An Introduction to Atmospheric Radiation

SECOND EDITION

K. N. LIOU

An Introduction to Atmospheric Radiation

Second Edition

An Introduction to Atmospheric Radiation
Second Edition

K. N. Liou
DEPARTMENT OF ATMOSPHERIC SCIENCES
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA

ACADEMIC PRESS

An imprint of Elsevier Science

Amsterdam Boston London New York Oxford Paris
San Diego San Francisco Singapore Sydney Tokyo

Contents

Chapter 3   Absorption and Scattering of Solar Radiation in the Atmosphere

Chapter 4   Thermal Infrared Radiation Transfer in the Atmosphere

Chapter 5   Light Scattering by Atmospheric Particulates

Chapter 6   Principles of Radiative Transfer
in Planetary Atmospheres

Chapter 7   Application of Radiative Transfer Principles to Remote Sensing

Chapter 8   Radiation and Climate

This Page Intentionally Left Blank

Preface

The preparation of the second edition of *An Introduction to Atmospheric Radiation* began when I moved to UCLA in September 1997. I was to teach a number of courses associated with atmospheric radiation and remote sensing, and I needed updated teaching material. Moreover, in the 20 years since the publication of the first edition, global climate research had yielded a wealth of information in the atmospheric radiation field, principally in conjunction with the global warming resulting from greenhouse gases, the effects of aerosols and clouds on climate and climate change, and the need for global observational data to support and perfect weather and climate modeling by means of remote sensing. It was also with a sense of duty and commitment to the atmospheric radiation field that I resolved to complete a volume that reflected the current state of the field, including the latest advances.

The second edition has been extensively revised to address the fundamental study, physical understanding, and quantitative analysis of the interactions of solar and terrestrial radiation with molecules, aerosols, and cloud particles in planetary atmospheres through the theory of radiative transfer and radiometric observations made from the ground, the air, and space.

The second edition contains about 70% new material. However, to the greatest extent possible, I have followed the structure of the chapters developed for the first edition, based on which substantial revisions and additions were made. New subjects include, but are not limited to, the correlated k-distribution method for infrared radiative transfer, light scattering by ice crystals and nonspherical aerosols, and a number of advanced topics in radiative transfer not covered by the plane-parallel assumption, presented in Chapters 4, 5, and 6, respectively.

The subject of remote sensing is a major discipline in itself and also relates to many other fields. However, in Chapter 7, I have confined the discussion of the application of fundamental radiative transfer principles to the development of remote sensing methodologies for the inference of atmospheric and surface parameters.

An understanding of the earth's climate and climate change must begin with a comprehensive understanding of the radiative processes in the earth–atmosphere system. In Chapter 8, I have based presentation of the role of radiation in climate on heat balance considerations, and present it in the context of simplified one-dimensional and global climate models.

I am indebted to Jennifer Kibbe, Melissa Licker, and Yoshihide Takano for assistance in the preparation of the manuscript in terms of typing, editing, and computer graphics. It would not have been possible to complete the second edition within the timetable set by myself and Academic Press without their dedicated contributions. This is particularly so in view of my other academic commitments, including my responsibilities as chair of the department. I have also had the great fortune of working with a number of bright and talented graduate students who complemented my strength in research and produced many of the results that are presented in the text.

A number of colleagues have kindly offered helpful comments and suggestions on various chapters. I am particularly grateful to Ute Böttger, Thomas Charlock, Annmarie Eldering, Frank Evans, Qiang Fu, Nori Fukuta, Michael Mishchenko, Steve S. C. Ou, Irina Sokolik, Ping Yang, and Charlie Zender. Some of them have used a draft of Chapters 1–4 in teaching atmospheric radiation classes.

I thank Richard Goody, who has written two texts on the subject of atmospheric radiation, and Wilfried Brutsaert, who is completing a text on fluid dynamics in parallel to this text, for their continuous encouragement of my pursuit of academic and research excellence. Finally, I thank James Holton, an editor of the International Geophysics Series, and Frank Cynar, Senior Publishing Editor of Physical Sciences at Academic Press, for their encouragement and support of this project. During the preparation of the second edition, my research programs have benefitted from the continuous support of the National Science Foundation, NASA, the Department of Energy, and the Air Force Office of Scientific Research.

K. N. LIOU

Chapter 1 | Fundamentals of Radiation for Atmospheric Applications

This text is intended for the study, understanding, and quantitative analysis of atmospheric radiation, a field in which the interactions of solar and terrestrial radiation with molecules, aerosols, and cloud particles in planetary atmospheres, as well as with the surface, are studied through the theory of radiative transfer and radiometric observations made from the ground, the air, and space. The field is closely associated with the investigation of atmospheric greenhouse effects resulting from external radiative perturbations in climate systems and the development of methodologies for inferring atmospheric and surface parameters by means of remote sensing. In the following, we begin with a discussion of various concepts, definitions, and units that are pertinent to the field of atmospheric radiation.

1.1 Concepts, Definitions, and Units

1.1.1 Electromagnetic Spectrum

The most important process responsible for energy transfer in the atmosphere is electromagnetic radiation. Electromagnetic radiation travels in wave form, and all electromagnetic waves travel at the same speed, the speed of light. This is $2.99793 \pm 1 \times 10^8$ m sec^{-1} in a vacuum and very nearly the same speed in air. Visible light, gamma rays, x-rays, ultraviolet light, infrared radiation, microwaves, television signals, and radio waves constitute the *electromagnetic spectrum*.

The retina of the human eye is sensitive to electromagnetic waves with frequencies between 4.3×10^{14} vibrations per second (usually written as cycles per second and abbreviated cps) and 7.5×10^{14} cps. Hence, this band of frequencies is called the *visible* region of the electromagnetic spectrum. The eye, however, does not respond to frequencies of electromagnetic waves higher than 7.5×10^{14} cps. Such waves, lying beyond the violet edge of the spectrum, are called *ultraviolet* light. The human eye also does not respond to electromagnetic waves with frequencies lower than 4.3×10^{14} cps. These waves, having frequencies lower than the lowest frequency of visible light at the red end of the spectrum and higher than about 3×10^{12} cps, are called *infrared light* or *infrared radiation*. Just beyond the infrared portion of the

spectrum are *microwaves,* which cover the frequencies from about 3×10^{10} cps to 3×10^{12} cps. The most significant spectral regions associated with radiative energy transfer in planetary atmospheres lie between ultraviolet light and microwaves.

The *x-ray* region of the electromagnetic spectrum consists of waves with frequencies ranging from about 3×10^{16} cps to 3×10^{18} cps, and is adjacent to the ultraviolet region in the spectrum. The *gamma-ray* region of the spectrum has the highest frequencies of all, ranging upward from about 3×10^{19} cps. *Radio* waves have the lowest frequencies in the spectrum, extending downward from about 3×10^{5} cps.

Electromagnetic waves are often described in terms of their wavelength rather than their frequency. The following general formula connects frequency $\tilde{\nu}$ and wavelength λ:

$$\lambda = c/\tilde{\nu}, \tag{1.1.1}$$

where c represents the speed of light in a vacuum. It is conventional to use micrometers (μm; 1 μm $= 10^{-4}$ cm) to denote the wavelengths of radiation from the sun. Other units, known as nanometers (nm; 1 nm $= 10^{-7}$ cm $= 10^{-3}$ μm) and angstroms (Å; 1 Å $= 10^{-4}$ μm), have also been frequently used, particularly for ultraviolet radiation. Equation (1.1.1) is valid for any type of wave and is not restricted to electromagnetic waves. It is customary to use wavenumber ν to describe the characteristics of infrared radiation. It is defined by

$$\nu = \tilde{\nu}/c = 1/\lambda. \tag{1.1.2}$$

Thus, a 10 micrometer (μm) wavelength is equal to a 1000 cm^{-1} wavenumber. In the microwave region, however, a frequency unit called gigahertz (GHz) is commonly used. One GHz is equal to 10^{9} cycles per second. It follows that 1 cm is equivalent to 30 GHz. Figure 1.1 shows the complete electromagnetic spectrum along with each region's corresponding frequency, wavenumber, and wavelength.

1.1.2   Solid Angle

The analysis of a radiation field often requires the consideration of the amount of radiant energy confined to an element of solid angle. A solid angle is defined as the ratio of the area σ of a spherical surface intercepted at the core to the square of the radius, r, as indicated in Fig. 1.2. It can be written as

$$\Omega = \sigma/r^{2}. \tag{1.1.3}$$

Units of solid angle are expressed in terms of the steradian (sr). For a sphere whose surface area is $4\pi r^{2}$, its solid angle is 4π sr.

To obtain a differential elemental solid angle, we construct a sphere whose central point is denoted as O. Assuming a line through point O moving in space and intersecting an arbitrary surface located at a distance r from point O, then as is evident from Fig. 1.3, the differential area in polar coordinates is given by

$$d\sigma = (rd\theta)(r\sin\theta \, d\phi). \tag{1.1.4}$$

Name of region	Wavelength (μm)	Frequency (GHz)	Wavenumber (cm⁻¹)
Gamma rays	10^{-5}	3×10^{10}	10^9
X rays	10^{-2}	3×10^7	10^6
Ultraviolet	3×10^{-1}	10^6	0.33×10^5
Visible			
Infrared	1	3×10^2	10^4
	10^3		10
Microwaves	10^4 (1cm)	3×10^1	1
Spacecraft	10^6	3×10^{-1}	10^{-2}
Television & FM	10^7	3×10^{-2}	10^{-3}
Shortwave	10^8	3×10^{-3}	10^{-4}
AM Radio waves	10^9	3×10^{-4}	10^{-5}

Violet 0.4μm
Purple
Blue
Green
Yellow
Orange
Red 0.7μm

Figure 1.1 The electromagnetic spectrum in terms of wavelength in μm, frequency in GHz, and wavenumber in cm^{-1}.
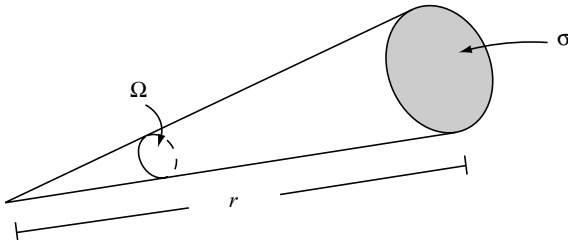
Figure 1.2 Definition of a solid angle Ω, where σ denotes the area and r is the distance.
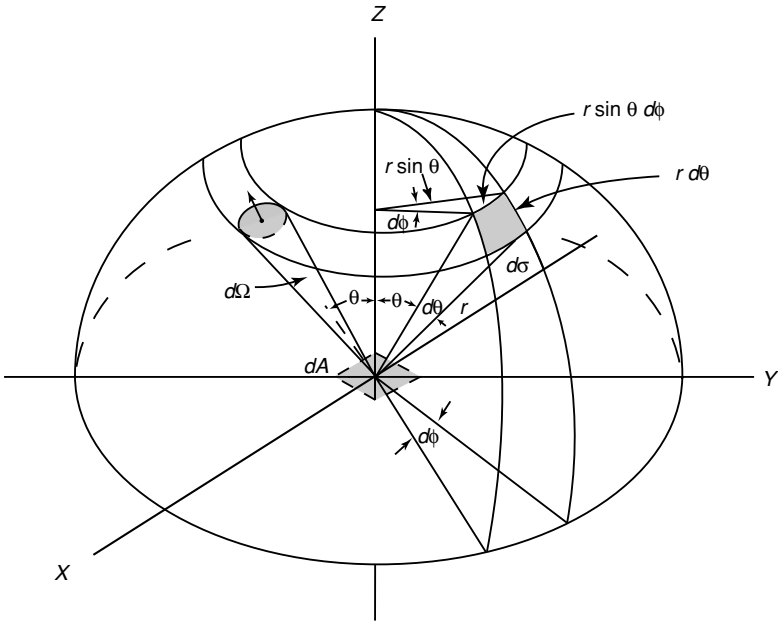
Figure 1.3 Illustration of a differential solid angle and its representation in polar coordinates. Also shown for demonstrative purposes is a pencil of radiation through an element of area dA in directions confined to an element of solid angle $d\Omega$. Other notations are defined in the text.

Hence, the differential solid angle is

$$d\Omega = d\sigma/r^2 = \sin\theta \, d\theta \, d\phi, \tag{1.1.5}$$

where θ and ϕ denote the zenith and azimuthal angles, respectively, in polar coordinates.

1.1.3 Basic Radiometric Quantities

Consider the differential amount of radiant energy dE_λ in a time interval dt and in a specified wavelength interval, λ to $\lambda + d\lambda$, which crosses an element of area dA depicted in Fig. 1.3, in directions confined to a differential solid angle, which is oriented at an angle θ to the normal of dA. This energy is expressed in terms of the specific intensity I_λ by

$$dE_\lambda = I_\lambda \cos\theta \, dA \, d\Omega \, d\lambda \, dt, \tag{1.1.6}$$

where $\cos\theta \, dA$ denotes the effective area at which the energy is being intercepted. Equation (1.1.6) defines the *monochromatic intensity* (or *radiance*) in a general way as follows:

$$I_\lambda = \frac{dE_\lambda}{\cos\theta \, d\Omega \, d\lambda \, dt \, dA}. \tag{1.1.7}$$

Thus, the intensity is in units of energy per area per time per wavelength and per steradian. It is evident that the intensity implies a directionality in the radiation stream. Commonly, the intensity is said to be confined in a pencil of radiation.

The *monochromatic flux density* or the *monochromatic irradiance* of radiant energy is defined by the normal component of I_λ integrated over the entire hemispheric solid angle and may be written as

$$F_\lambda = \int_\Omega I_\lambda \cos\theta \, d\Omega. \tag{1.1.8}$$

In polar coordinates, we write

$$F_\lambda = \int_0^{2\pi} \int_0^{\pi/2} I_\lambda(\theta, \phi) \cos\theta \, \sin\theta \, d\theta \, d\phi. \tag{1.1.9}$$

For isotropic radiation (i.e., if the intensity is independent of the direction), the monochromatic flux density is then

$$F_\lambda = \pi I_\lambda. \tag{1.1.10}$$

The total flux density of radiant energy, or irradiance, for all wavelengths (energy per area per time), can be obtained by integrating the monochromatic flux density over the entire electromagnetic spectrum:

$$F = \int_0^\infty F_\lambda \, d\lambda. \tag{1.1.11}$$

Moreover, the total flux f, or radiant power W (energy per time), is defined by

$$f = \int_A F \, dA. \tag{1.1.12}$$

The monochromatic flux density in the frequency domain may be written in the form

$$F_{\tilde{\nu}} = \frac{dF}{d\tilde{\nu}}. \tag{1.1.13}$$

From the relationship between wavelength and frequency denoted in Eq. (1.1.1), we have

$$F_{\tilde{\nu}} = -(\lambda^2/c)F_\lambda. \tag{1.1.14}$$

Likewise, the intensity in wavelength and frequency domains is connected by

$$I_{\tilde{\nu}} = -(\lambda^2/c)I_\lambda. \tag{1.1.15}$$

A similar relation between the monochromatic flux density, or intensity, in wavenumber and wavelength (or frequency) domains may be expressed by means of Eq. (1.1.2).

When the flux density or the irradiance is from an emitting surface, the quantity is called the *emittance*. When expressed in terms of wavelength, it is referred to as the *monochromatic emittance*. The intensity or the radiance is also called the *brightness* or *luminance* (photometric brightness). The total flux from an emitting surface is

Table 1.1

Symbols, Dimensions, and Units of Various Radiometric Quantities

Symbol	Quantity	Dimension[a]	Unit[b]
E	Energy	ML^2T^{-2}	Joule (J)
f	Flux (luminosity)	ML^2T^{-3}	Joule per second (J sec^{-1}, W)
F	Flux density (irradiance) Emittance	MT^{-3}	Joule per second per square meter (W m^{-2})
I	Intensity (radiance) Brightness (luminance)	MT^{-3}	Joule per second per square meter per steradian (W m^{-2} sr^{-1})

[a]M is mass, L is length, and T is time.
[b]1 watt (W) = 1 J sec^{-1}.

often called *luminosity*. The basic radiometric quantities are summarized in Table 1.1, along with their symbols, dimensions, and units.

1.1.4   Concepts of Scattering and Absorption

Most of the light that reaches our eyes comes not directly from its source but indirectly through the process of *scattering*. We see diffusely scattered sunlight when we look at clouds or at the sky. Land and water surfaces and the objects surrounding us are visible through the light that they scatter. An electric lamp does not send us light directly from the luminous filament but usually glows with the light that has been scattered by the glass bulb. Unless we look directly at a light source, such as the sun, a flame, or an incandescent filament with a clear bulb, we see light that has been scattered. In the atmosphere, we see many colorful examples of scattering generated by molecules, aerosols, and clouds containing water droplets and ice crystals. Blue sky, white clouds, and magnificent rainbows and halos, to name a few, are all optical phenomena produced by scattering. Scattering is a fundamental physical process associated with light and its interaction with matter. It occurs at all wavelengths throughout the entire electromagnetic spectrum.

Scattering is a physical process by which a particle in the path of an electromagnetic wave continuously abstracts energy from the incident wave and reradiates that energy in all directions. Therefore, the particle may be thought of as a point source of the scattered energy. In the atmosphere, the particles responsible for scattering range in size from gas molecules ($\sim 10^{-4}$ μm) to aerosols (~ 1 μm), water droplets (~ 10 μm), ice crystals (~ 100 μm), and large raindrops and hail particles (~ 1 cm). The effect of particle size on scattering is inferred by a physical term called the *size parameter*. For a spherical particle, it is defined as the ratio of the particle circumference to the incident wavelength, λ; i.e., $x = 2\pi a/\lambda$, where a is the particle radius. If $x \ll 1$, the scattering is called *Rayleigh scattering*. An excellent example of this case is the scattering of visible light (0.4–0.7 μm) by atmospheric molecules, leading to the explanation of blue sky and sky polarization to be discussed in Chapter 3. For
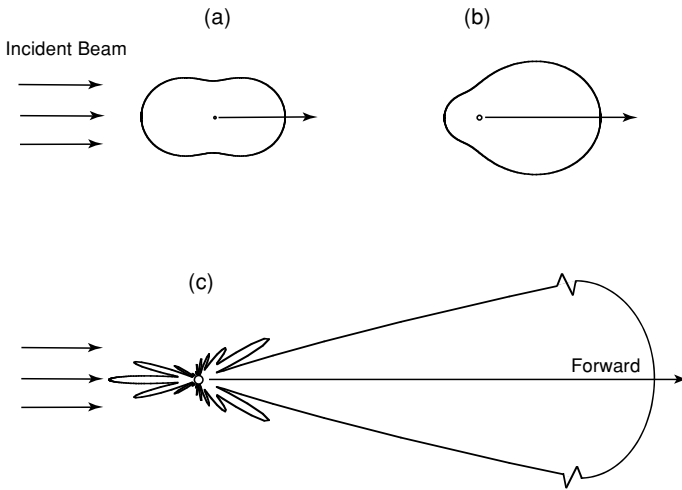
Figure 1.4   Demonstrative angular patterns of the scattered intensity from spherical aerosols of three sizes illuminated by the visible light of 0.5 μm: (a) 10^{-4} μm, (b) 0.1 μm, and (c) 1 μm. The forward scattering pattern for the 1 μm aerosol is extremely large and is scaled for presentation purposes.

particles whose sizes are comparable to or larger than the wavelength, i.e., $x \gtrsim 1$, the scattering is customarily referred to as *Lorenz–Mie scattering*. The mathematical theory of Lorenz–Mie scattering for spherical particles will be presented in Chapter 5. Figure 1.4 illustrates the scattering patterns of spherical aerosols of size $10^{-4}, 0.1$, and 1 μm illuminated by a visible light of 0.5 μm. A small particle tends to scatter light equally in the forward and backward directions. When the particle becomes larger, the scattered energy becomes increasingly concentrated in the forward direction with increasingly complex scattering features. Because of the spherical symmetry with respect to the incoming light beam, the scattering patterns for other planes are the same as the ones presented in Fig. 1.4. The scattering of sunlight by spherical cloud droplets and raindrops produces the magnificent rainbows and glory that we see in our daily life.

In situ observations and electronic microscopic photography have shown that aerosols in the atmosphere, such as minerals, soot, and even oceanic particles, exhibit a wide variety of shapes ranging from quasi-spherical to highly irregular geometric figures with internal structure. The shape and size of ice crystals are governed by temperature and supersaturation, but they generally have a basic hexagonal structure. In the atmosphere, if ice crystal growth involves collision and coalescence, the crystal's shape can be extremely complex. Recent observations based on aircraft optical probes and replicator techniques for widespread midlatitude, tropical, arctic, and contrail cirrus show that these clouds are largely composed of ice crystals in the shape of bullet rosettes, solid and hollow columns, plates, and aggregates, and ice crystals with irregular surfaces with sizes ranging from a few micrometers to thousands of micrometers. The scattering of sunlight by some of the defined ice crystals produces
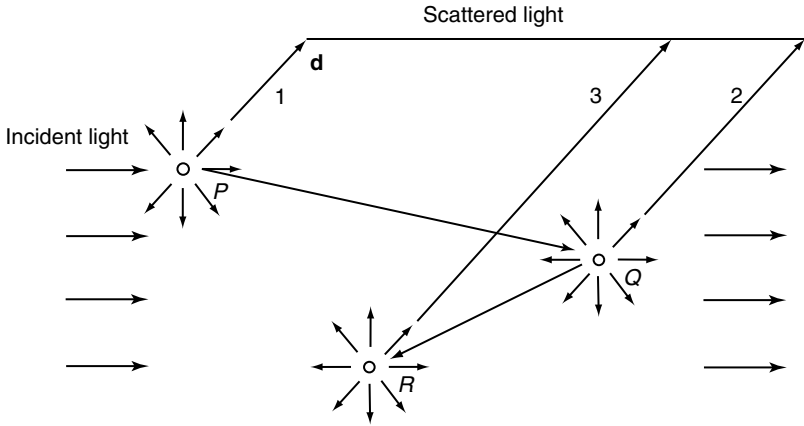
Figure 1.5   Multiple scattering process involving first (P), second (Q), and third (R) order scattering in the direction denoted by **d.**

fascinating optical phenomena including $22°$ and $46°$ halos, sundogs, and numerous arcs and bright spots. Light scattering by nonspherical and inhomogeneous particles, a contemporary research subject, will be covered in Chapter 5, which presents a combination of geometric ray-tracing and numerical solution approaches.

In atmospheric scattering, it is generally assumed that the light scattered by molecules and particulates has the same frequency (or wavelength) as the incident light. It is noted, however, that high-energy laser light can produce phenomena such as Raman scattering in shift frequencies, which can be employed for the remote sensing of water vapor. Atmospheric molecules and particulates are separated widely enough so that each particle scatters light in exactly the same way as if all other particles did not exist. This is referred to as *independent scattering*. The assumption of independent scattering greatly simplifies the problem of light scattering by a collection of particles, because it allows the use of energy quantity instead of electric field in the analysis of the propagation of electromagnetic waves in planetary atmospheres.

In a scattering volume, which contains many particles, each particle is exposed to, and also scatters, the light that has already been scattered by other particles. To demonstrate this concept we refer to Fig. 1.5. A particle at position P removes the incident light by scattering just once, i.e., single scattering, in all directions. Meanwhile, a portion of this scattered light reaches the particle at position Q, where it is scattered again in all directions. This is called *secondary scattering*. Likewise, a subsequent third-order scattering involving the particle at position R takes place. Scattering more than once is called *multiple scattering*. It is apparent from Fig. 1.5 that some of the incident light that has been first scattered away from direction **d** may reappear in this direction by means of multiple scattering. Multiple scattering is an important process for the transfer of radiant energy in the atmosphere, especially when aerosols and clouds are involved. Chapter 6 deals with the theory of multiple scattering in planetary atmospheres.

Scattering is often accompanied by absorption. Grass looks green because it scatters green light while it absorbs red and blue light. The absorbed energy is converted into some other form, and it is no longer present as red or blue light. In molecular atmospheres, there is very little absorption of energy in the visible spectrum. Clouds also absorb very little visible light. Both scattering and absorption remove energy from a beam of light traversing the medium. The beam of light is attenuated, and we call this attenuation *extinction*. Thus, extinction is a result of scattering plus absorption. In a nonabsorbing medium, scattering is the sole process of extinction.

In the field of light scattering and radiative transfer, it is customary to use a term called *cross section,* analogous to the geometrical area of a particle, to denote the amount of energy removed from the original beam by the particle. When the cross section is associated with a particle dimension, its units are denoted in terms of area (cm^2). Thus, the extinction cross section, in units of area, is the sum of the scattering and absorption cross sections. However, when the cross section is in reference to unit mass, its units are given in area per mass ($cm^2 \ g^{-1}$). In this case, the term mass extinction cross section is used in radiative transfer. The mass extinction cross section is, therefore, the sum of the mass absorption and mass scattering cross sections. Furthermore, when the extinction cross section is multiplied by the particle number density (cm^{-3}), or when the mass extinction cross section is multiplied by the density ($g \ cm^{-3}$), the quantity is referred to as the *extinction coefficient,* whose units are given in terms of length (cm^{-1}). In the field of infrared radiative transfer, the mass absorption cross section is simply referred to as the *absorption coefficient*.

The absorption of energy by particles and molecules leads to emission. The concept of emission is associated with blackbody radiation and will be discussed in the following section. In addition, a number of minor atmospheric constituents exhibit complicated absorption line structures in the infrared region. Section 1.3 and Chapter 4 will provide discussions of the fundamentals of line formation and the transfer of infrared radiation in the atmosphere. A fundamental understanding of the scattering and absorption processes in the atmosphere is imperative for the study of the radiation budget and climate of planetary atmospheres and for the exploration of remote sounding techniques to infer atmospheric composition and structure.

1.2   Blackbody Radiation Laws

The laws of blackbody radiation are basic to an understanding of the absorption and emission processes. A blackbody is a basic concept in physics and can be visualized by considering a cavity with a small entrance hole, as shown in Fig. 1.6. Most of the radiant flux entering this hole from the outside will be trapped within the cavity, regardless of the material and surface characteristics of the wall. Repeated internal reflections occur until all the fluxes are absorbed by the wall. The probability that any of the entering flux will escape back through the hole is so small that the interior appears dark. The term *blackbody* is used for a configuration of material where absorption is complete. Emission by a blackbody is the converse of absorption. The flux
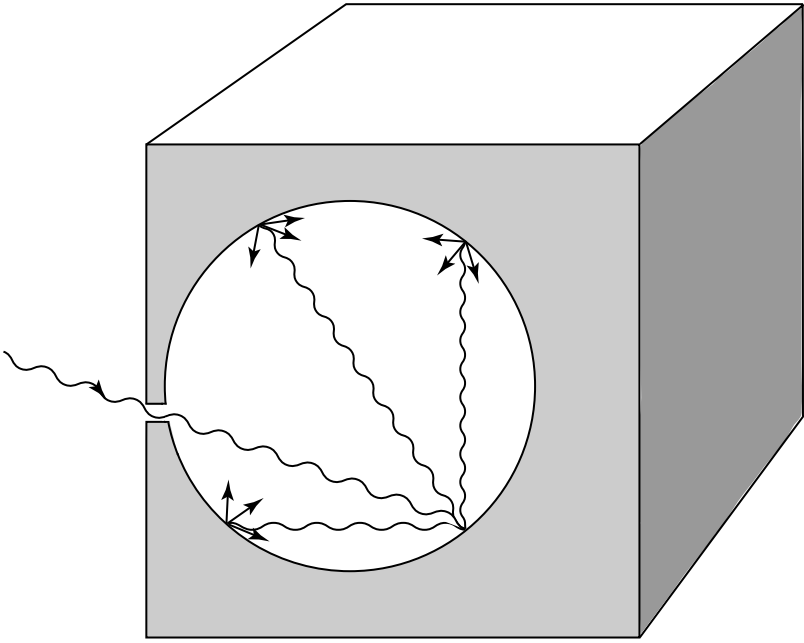
Figure 1.6   A blackbody radiation cavity to illustrate that absorption is complete.

emitted by any small area of the wall is repeatedly reflected and at each encounter with the wall, the flux is weakened by absorption and strengthened by new emission. After numerous encounters, emission and absorption reach an equilibrium condition with respect to the wall temperature. In the following, we present four fundamental laws that govern blackbody radiation, beginning with Planck's law.

1.2.1   Planck's Law

In his pursuit of a theoretical explanation for cavity radiation, Planck (1901) assumed that the atoms that make up the wall behave like tiny electromagnetic oscillators, each with a characteristic frequency of oscillation. The oscillators emit energy into the cavity and absorb energy from it. In his analysis, Planck was led to make two assumptions about the atomic oscillators. First, Planck postulated that an oscillator can only have energy given by

$$E = nh\tilde{\nu}, \tag{1.2.1}$$

where $\tilde{\nu}$ is the oscillator frequency, h is Planck's constant, and n is called the quantum number and can take on only integral values. Equation (1.2.1) asserts that the oscillator energy is quantized. Although later developments revealed that the correct formula for a harmonic oscillator is $E = (n + 1/2)h\tilde{\nu}$ [see Eq. (1.3.7)], the change introduces no difference to Planck's conclusions. Secondly, Planck postulated that the oscillators

do not radiate energy continuously, but only in jumps, or in quanta. These quanta of energy are emitted when an oscillator changes from one to another of its quantized energy states. Hence, if the quantum number changes by one unit, the amount of radiated energy is given by

$$\Delta E = \Delta n h \tilde{\nu} = h \tilde{\nu}. \tag{1.2.2}$$

Determination of the emitted energy requires knowing the total number of oscillators with frequency $\tilde{\nu}$ for all possible states in accord with Boltzmann statistics, as presented in Appendix A. Following the two preceding postulations and normalization of the average emitted energy per oscillator, the Planck function in units of energy/area/time/sr/frequency is given by

$$B_{\tilde{\nu}}(T) = \frac{2h\tilde{\nu}^3}{c^2(e^{h\tilde{\nu}/KT} - 1)}, \tag{1.2.3}$$

where K is Boltzmann's constant, c is the velocity of light, and T is the absolute temperature. The Planck and Boltzmann constants have been determined through experimentation and are $h = 6.626 \times 10^{-34}$ J sec and $K = 1.3806 \times 10^{-23}$ J deg^{-1}.

The Planck function relates the emitted monochromatic intensity to the frequency and the temperature of the emitting substance. By utilizing the relation between frequency and wavelength shown in Eq. (1.1.15), Eq. (1.2.3) can be rewritten as follows:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5(e^{hc/K\lambda T} - 1)} = \frac{C_1\lambda^{-5}}{\pi(e^{C_2/\lambda T} - 1)}, \tag{1.2.4}$$

where $C_1 = 2\pi hc^2$ and $C_2 = hc/K$ are known as the first and second radiation constants, respectively. Figure 1.7 shows curves of $B_\lambda(T)$ versus wavelength for a number of emitting temperatures. It is evident that the blackbody radiant intensity increases with temperature and that the wavelength of the maximum intensity decreases with increasing temperature. The Planck function behaves very differently when $\lambda \to \infty$, referred to as the *Rayleigh–Jeans distribution,* and when $\lambda \to 0$, referred to as the *Wien distribution*.

1.2.2   Stefan–Boltzmann Law

The total radiant intensity of a blackbody can be derived by integrating the Planck function over the entire wavelength domain from 0 to ∞. Hence,

$$B(T) = \int_0^\infty B_\lambda(T)\, d\lambda = \int_0^\infty \frac{2hc^2\lambda^{-5}}{(e^{hc/K\lambda T} - 1)}\, d\lambda. \tag{1.2.5}$$

On introducing a new variable $x = hc/k\lambda T$, Eq. (1.2.5) becomes

$$B(T) = \frac{2k^4T^4}{h^3c^2} \int_0^\infty \frac{x^3\, dx}{(e^x - 1)}. \tag{1.2.6}$$
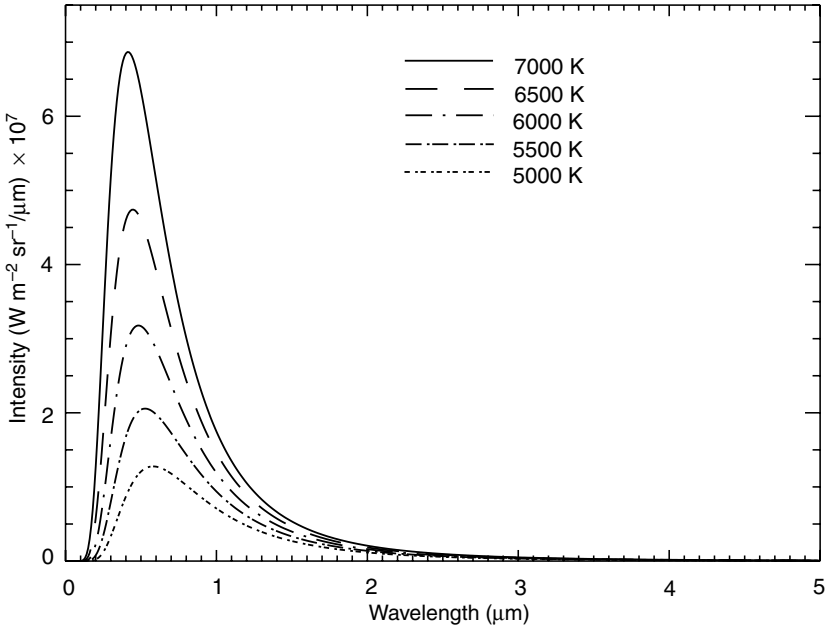
Figure 1.7   Blackbody intensity (Planck function) as a function of wavelength for a number of emitting temperatures.

The integral term in Eq. (1.2.6) is equal to $\pi^4/15$. Thus, defining

$$b = 2\pi^4 K^4/(15c^2 h^3), \tag{1.2.7}$$

we then have

$$B(T) = bT^4. \tag{1.2.8}$$

Since blackbody radiation is isotropic, the flux density emitted by a blackbody is therefore [see Eq. (1.1.10)]

$$F = \pi B(T) = \sigma T^4, \tag{1.2.9}$$

where σ is the Stefan–Boltzmann constant and is equal to 5.67×10^{-8} J m^{-2} sec^{-1} deg^{-4}. Equation (1.2.9) states that the flux density emitted by a blackbody is proportional to the fourth power of the *absolute* temperature. This is the Stefan–Boltzmann law, fundamental to the analysis of broadband infrared radiative transfer.

1.2.3   Wien's Displacement Law

Wien's displacement law states that the wavelength of the maximum intensity of blackbody radiation is inversely proportional to the temperature. By differentiating

the Planck function with respect to wavelength, and by setting the result equal to zero, i.e.,

$$\frac{\partial B_\lambda(T)}{\partial \lambda} = 0, \tag{1.2.10}$$

we obtain the wavelength of the maximum (Exercise 1.4)

$$\lambda_m = a/T, \tag{1.2.11}$$

where $a = 2.897 \times 10^{-3}$ m deg. From this relationship, we can determine the temperature of a blackbody from the measurement of the maximum monochromatic intensity. The dependence of the position of the maximum intensity on temperature is evident from the blackbody curves displayed in Fig. 1.7.

1.2.4 Kirchhoff's Law

The preceding three fundamental laws are concerned with the radiant intensity emitted by a blackbody, which is dependent on the emitting wavelength and the temperature of the medium. A medium may absorb radiation of a particular wavelength, and at the same time also emit radiation of the same wavelength. The rate at which emission takes place is a function of temperature and wavelength. This is the fundamental property of a medium under the condition of *thermodynamic equilibrium*. The physical statement regarding absorption and emission was first proposed by Kirchhoff (1860).

To understand the physical meaning of Kirchhoff's law, we consider a perfectly insulated enclosure having black walls. Assume that this system has reached the state of thermodynamic equilibrium characterized by uniform temperature and isotropic radiation. Because the walls are black, radiation emitted by the system to the walls is absorbed. Moreover, because there is an equilibrium, the same amount of radiation absorbed by the walls is also emitted. Since the blackbody absorbs the maximum possible radiation, it has to emit that same amount of radiation. If it emitted more, equilibrium would not be possible, and this would violate the second law of thermodynamics. Radiation within the system is referred to as blackbody radiation as noted earlier, and the amount of radiant intensity is a function of temperature and wavelength.

On the basis of the preceding discussion, the emissivity of a given wavelength, ε_λ (defined as the ratio of the emitting intensity to the Planck function), of a medium is equal to the absorptivity, A_λ (defined as the ratio of the absorbed intensity to the Planck function), of that medium under thermodynamic equilibrium. Hence, we may write

$$\varepsilon_\lambda = A_\lambda. \tag{1.2.12}$$

A medium with an absorptivity A_λ absorbs only A_λ times the blackbody radiant intensity $B_\lambda(T)$ and therefore emits ε_λ times the blackbody radiant intensity. For a blackbody, absorption is a maximum and so is emission. Thus, we have

$$A_\lambda = \varepsilon_\lambda = 1 \tag{1.2.13}$$

for all wavelengths. A *gray body* is characterized by incomplete absorption and emission and may be described by

$$A_\lambda = \varepsilon_\lambda < 1. \tag{1.2.14}$$

Kirchhoff's law requires the condition of thermodynamic equilibrium, such that uniform temperature and isotropic radiation are achieved. Obviously, the radiation field of the earth's atmosphere as a whole is not isotropic and its temperatures are not uniform. However, in a localized volume below about 60–70 km, to a good approximation, it may be considered to be isotropic with a uniform temperature in which energy transitions are governed by molecular collisions. It is in the context of this local thermodynamic equilibrium (LTE) that Kirchhoff's law is applicable to the atmosphere. Departure from the LTE conditions will be discussed in Section 1.3.3.

1.3   Absorption Line Formation and Line Shape

1.3.1   Line Formation

Inspection of high-resolution spectroscopy reveals that the emission spectra of certain gases are composed of a large number of individual and characteristic spectral lines. In the previous section, we indicated that Planck successfully explained the nature of radiation from heated solid objects of which the cavity radiator formed the prototype. Such radiation generates continuous spectra, as opposed to line spectra. Planck's quantization ideas, properly extended, however, lead to an understanding of line spectra as well. In the following, we use the simplest hydrogen model to discuss emission and absorption line formation.

1.3.1.1   BOHR'S MODEL

Investigation of the hydrogen spectrum led Bohr (1913) to postulate that the circular orbits of the electrons were quantized; that is, their angular momentum could have only integral multiples of a basic value. Bohr assumed that the hydrogen atom exists, like Planck's oscillators, in certain stationary states in which it does not radiate. Radiation occurs only when the atom makes a transition from one state with energy E_k to a state with lower energy E_j. Thus, we write

$$E_k - E_j = h\tilde{\nu}, \tag{1.3.1}$$

where $h\tilde{\nu}$ represents the quantum of energy carried away by the photon, which is emitted from the atom during the transition. The lowest energy state is called the *ground state* of the atom. When an electron of an atom absorbs energy due to a collision and jumps into a larger orbit, the atom is said to be in an *excited state*. Then, according to Eq. (1.3.1), a sudden transition will take place, and the atom emits a photon of energy and collapses to a lower energy state. This is illustrated in Fig. 1.8 for a hydrogen atom. Also shown in this figure is the absorption of a photon by a stationary hydrogen atom.

EMISSION
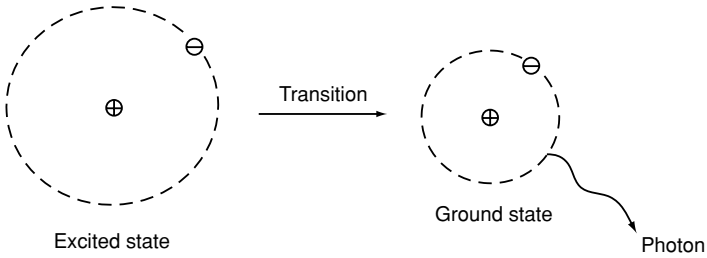
ABSORPTION
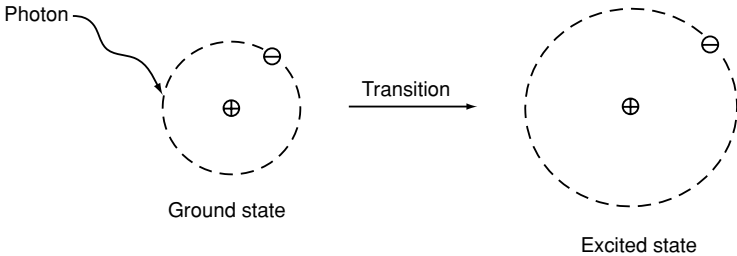
Figure 1.8 Illustration of emission and absorption for a hydrogen atom that is composed of one proton and one electron. The radius of the circular orbit r is given by $n^2 \times 0.53$ Å, where n is the quantum number, and 1 Å $= 10^{-8}$ cm.

Bohr further postulated that the angular momentum L can take on only discrete values by

$$L = n(h/2\pi), \quad n = 1, 2, 3 \ldots . \tag{1.3.2}$$

With this selection rule, Bohr showed from the equation of motion for an electron that the total energy state of the system is given by

$$E_n = -\frac{me^4}{8\varepsilon_0^2 h^2} \frac{1}{n^2} = -\frac{R_H hc}{n^2}, \quad n = 1, 2, 3 \ldots, \tag{1.3.3}$$

where m is the mass of the electron, e is the charge carried by the electron, ε_0 is the permittivity constant given by 8.85×10^{-12} coul/volt/m, with 1 volt $= 1$ joule/coul, and R_H is the *Rydberg constant* for hydrogen with a value of 1.097×10^5 cm^{-1}. It follows from Eq. (1.3.1) that the wavenumber of emission or absorption lines in the hydrogen spectrum is

$$\nu = R_H \left(\frac{1}{j^2} - \frac{1}{k^2} \right), \tag{1.3.4}$$
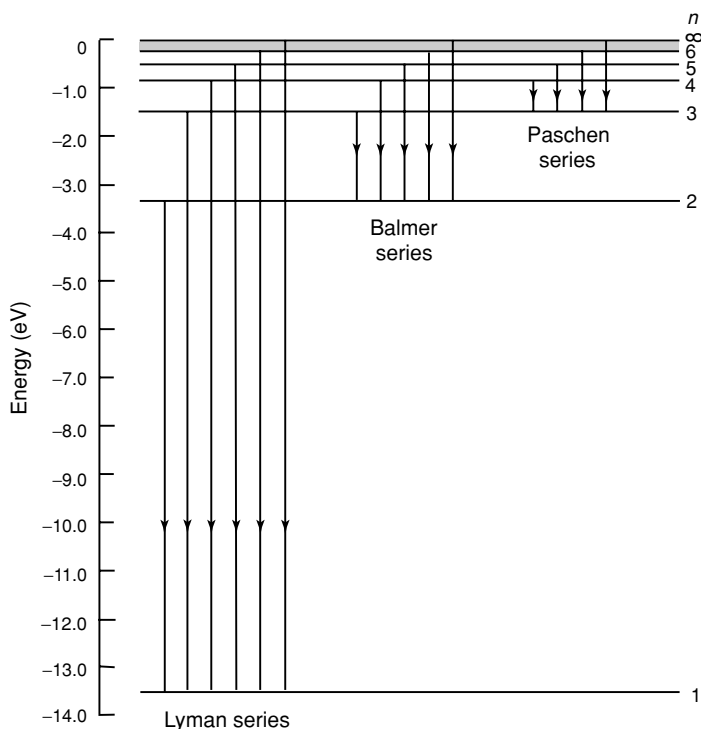
Figure 1.9   Energy level diagram for a hydrogen atom showing the quantum number n for each level and some of the transitions that appear in the spectrum. An infinite number of levels is crowded in between the levels marked $n = 6$ and $n = \infty$.

where j and k are integers defining, respectively, the lower and higher energy states. Figure 1.9 shows the energy diagram for hydrogen. In the field of spectroscopy, energy is usually given in units of electron volts (eV) or in units of wavenumber (cm^{-1}). An electron volt is the energy acquired by an electron accelerated through a potential difference of one volt, and is equivalent to 1.602×10^{-19} J. Exercise 1.11 requires the derivation of Eq. (1.3.3) based on the definitions of kinetic and potential energies of the system.

Each quantum jump between fixed energy levels results in the emission or absorption of a characteristic frequency or wavelength. These quanta appear in the spectrum as emission or absorption lines. For the simple hydrogen atom described previously the line spectrum is relatively simple, whereas the spectra of water vapor, carbon dioxide, and ozone molecules are considerably more complex.

1.3.1.2   VIBRATIONAL AND ROTATIONAL TRANSITIONS
In the preceding discussion, we used the electronic transitions of the hydrogen atom to illustrate emission and absorption. It is now helpful to introduce the ways in which a molecule can store various energies. Any moving particle has kinetic

energy as a result of its motion in space. This is known as *translational energy*. The average translational kinetic energy of a single molecule in the x, y, and z directions is found to be equal to $KT/2$, where K is the Boltzmann constant and T is the absolute temperature. A molecule, composed of atoms, can rotate, or revolve, about an axis through its center of gravity and, therefore, has *rotational energy*. The atoms of the molecule are bounded by certain forces like springs such that the individual atoms can vibrate about their equilibrium positions relative to one another. The molecule, therefore, will also have *vibrational energy*. These three molecular energy types are based on a rather mechanical model of the molecule that ignores the detailed structure of the molecule in terms of nuclei and electrons. It is possible, however, for the energy of a molecule to change as a result of a change in the energy state of the electrons of which it is composed, as demonstrated by Bohr's model. Thus, the molecule has *electronic energy*. The last three energy types are quantized and take only discrete values. The absorption and emission of radiation takes place when the atoms or molecules undergo transitions from one energy state to another. In general, these transitions are governed by selection rules.

In radiative transitions, the molecule must couple with an electromagnetic field so that energy exchanges can take place. This coupling is generally provided by the electric dipole moment of the molecule. If the effective centers of the positive and negative charges of the molecule have nonzero separation, then the dipole moment exists. Radiatively active gases in the infrared, such as H_2O and O_3, have permanent electric dipole moments due to their asymmetrical charge distributions. Linear molecules such as N_2 and O_2, however, are inactive in the infrared because of their symmetrical charge distributions. However, they have weak magnetic dipole moments that allow radiative activities to take place in the ultraviolet and, to a lesser extent, in the visible region.

Rotational energy changes are relatively small, with a minimum on the order of 1 cm^{-1} (see the conversion to energy in Exercise 1.12). For this reason, pure rotational lines occur in the microwave and far-infrared spectra. Many of the rotational energy levels above the lowest level are populated at terrestrial temperatures. Changes in vibrational energy are generally greater than 600 cm^{-1}, which is much larger than the minimum changes in rotational energy. Thus, vibrational transitions never occur alone but are coupled with simultaneous rotational transitions. This coupling gives rise to a group of lines known as the vibrational–rotational band in the intermediate infrared spectrum. An electronic transition typically involves a few electron volts ($\sim 10^4 \text{ cm}^{-1}$) of energy. Because a high-energy photon is required for the transition, absorption and emission usually occur in the ultraviolet or visible spectrum. Atoms can produce line spectra associated with electronic energy. Molecules, however, can have two additional types of energy, leading to complex band systems.

In Subsection 1.3.1.1, we discussed the physical meaning of stationary states for a hydrogen atom. Schrödinger (1926) first introduced the idea of stationary states corresponding to standing matter waves and used this idea as the foundation of *wave mechanics*. In quantum mechanics, to determine the energy states produced by vibrational and rotational transitions, a term referred to as the *Hamiltonian* operator, H,

was introduced as a convenient operator by replacing variables in the classical expression for the energy, E, of a system composed of the atomic nuclei and electrons that form a molecule. Schrödinger's equation can be written in terms of the first-order differential equation involving the wave function and the Hamiltonian operator, as shown in Appendix B. The Hamiltonian may be linearly separated into a time-dependent term and a time-independent term. The stationary states of the molecules can be deduced from the time-independent term, giving discrete eigenvalues (energy levels), E_n, and eigenfunctions, φ_n. Transitions between energy levels result in the absorption and emission of photons with frequency $\tilde{\nu}$ following Planck's relation. The time-dependent term may be treated as a perturbation from which the rate of change of the probability that a stationary state is occupied can be evaluated.

The Hamiltonian operator for the harmonic-oscillator rigid rotator is separable for vibrational and rotational motions so that energies may be added for a combined state. For the rotational states, the kinetic energy of a rigid rotating dipole is equal to one-half the product of angular momentum, L, and angular velocity, ω, i.e., $L\omega/2$, where $L = I\omega$ and I is the moment of inertia. From the solution of the time-independent Schrödinger equation, the quantum restrictions on angular momentum are given by

$$L = \frac{h}{2\pi}[J(J + 1)]^{1/2}, \tag{1.3.5}$$

where J is the rotational quantum number (an integer). Thus, the quantized rotational energy can be written as

$$E_J = Bhc J(J + 1), \tag{1.3.6}$$

where $B = h/8\pi^2 Ic$ is the rotational constant. This expression is valid for a rigid rotating dipole assuming spherical tops or linear molecules. For asymmetric tops, an additional term is required. The selection rule for radiation transition is governed by $\Delta J = \pm 1$, applicable to the harmonic-oscillator rigid-rotator model. From Planck's relation in Eq. (1.3.1), the spectral line location can be derived and is given by $\nu = 2BJ'$ (cm^{-1}), where J' can be any quantum number. Because of the selection rule, the separation in wavenumber of adjacent lines is simply $2B$ (cm^{-1}), as shown in Fig. 1.10a. As noted above, because of the small energy of a rotational transition, pure rotational spectra occur only in the far infrared and microwave regions.

For vibrational states, the quantized energy levels for a harmonic vibration are given by

$$E_\upsilon = h\tilde{\nu}_k(\upsilon_k + 1/2), \tag{1.3.7}$$

where υ_k is the vibrational quantum number (an integer) and subscript k denotes the normal modes. For triatomic molecules such as H_2O and O_3, there are three normal modes, referred to as *fundamentals*. For linear molecules such as CO_2 and NO_2, there are four fundamentals, but two orthogonal bending modes are degenerate and so only three fundamentals exist (see Fig. 3.3). The term *degenerate* is used to denote states with the same energy but with different sets of quantum numbers.
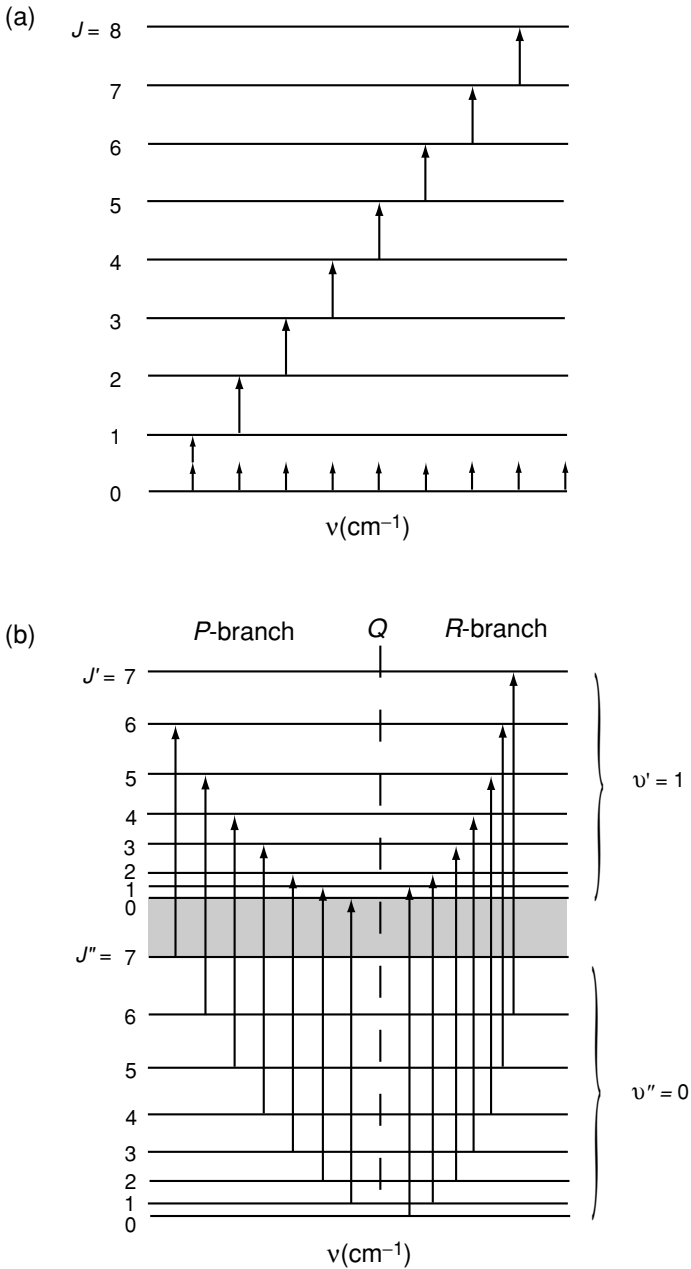
Figure 1.10 (a) Rotational transition following the selection rule $\Delta J = +1$ and equally spaced spectral lines in wavenumber. (b) Simultaneous vibrational and rotational transitions where $\Delta J = -1$ produces the P-branch and $\Delta J = +1$ generates the R-branch. $\Delta J = 0$ shows the Q-branch that overlaps with the vibrational wavenumber, but see text for discussion.

Molecular vibration produces an oscillating electric dipole moment that is sufficient for both vibrational and rotational transitions. Thus, both transitions occur simultaneously and the resulting energy level is the sum of the separate transition energies. Because the energy of the vibrational transition is much larger than that of the rotational transition as noted earlier, and since many rotational levels are active, the spectrum of the combined transitions is an array of rotational lines grouped around the vibrational wavenumber, as illustrated in Fig. 1.10b. From Eqs. (1.3.6) and (1.3.7), the sum of the rotational and vibrational energies is

$$E_{v,J} = BhcJ(J+1) + h\tilde{\nu}_k(\upsilon_k + 1/2). \tag{1.3.8}$$

In simple cases, the selection rule is $\Delta\upsilon_k = \pm 1$, except that $\Delta\upsilon_k = -1$ cannot be applied to $\upsilon_k = 0$. Consider the transitions $(\upsilon'_k, J') \leftarrow (\upsilon''_k, J'')$, and let $\upsilon''_k = 0$, the ground state, and $\upsilon'_k = 1$, the first excited state. J' and J'' denote the higher and lower rotational states, respectively, as shown in Fig. 1.10b. The selection rules in this case are $\Delta\upsilon_k = 1$ (fundamental) and $\Delta J = \pm 1$. Most of the molecules are in the ground state but are distributed over numerous levels of J''. Thus, any molecules advancing to the level υ' can go either to the next higher rotational level, for $\Delta J = +1$, or to the next lower level, for $\Delta J = -1$, with about equal probability. Using Planck's relation, we find that the spectral wavenumber of the line is given by

$$\nu = \nu_k \begin{cases} +2BJ', \ J' = 1, 2\ldots, \Delta J = +1, \\ -2B(J'+1), \ J' = 0, 1, \ldots, \Delta J = -1, \end{cases} \tag{1.3.9}$$

where J' is the rotational quantum number in the excited vibrational state υ'. Because many closely spaced rotational energy levels are involved, numerous transitions generate a band of rotational lines grouped on each side of the vibrational wavenumber, ν_k, with spacing of $2B$ cm^{-1}, as for a pure rotational spectrum. Several of the simultaneous transitions available to diatomic molecules and linear triatomic molecules (e.g., CO_2) in normal modes $k = 1(\upsilon_1)$ and $k = 3(\upsilon_3)$ fundamentals are shown schematically in Fig. 1.10b. The group with lower energy ($\Delta J = -1$) and hence the lower wavenumber portion of the band, is called the P-branch. The higher wavenumber part is referred to as the R-branch, corresponding to $\Delta J = +1$. The rotational level spacings in the υ' level are somewhat smaller than those in the υ'' level because of the increased moment of inertia in higher vibrational levels. The lengths of the arrows do not increase by a constant amount from the left to the right and the wavenumber spacing of the lines decreases slightly. The branches P and R are called *parallel branches* because the dipole moment oscillates parallel to the internuclear axis (see Fig. 3.3). For such vibrational modes the transition $\Delta J = 0$ is forbidden. In quantum mechanics, it is customary to refer to transitions as *forbidden* (or unfavorable) and *allowed*. For the vibrational mode $k = 2(\upsilon_2)$ of linear triatomic and the three modes of bent triatomic molecules (see Fig. 3.3), the change of dipole moment has a component perpendicular to an internuclear axis. The rotational selection rule is now $\Delta J = 0, \pm 1$, which produces a Q-branch that corresponds to $\Delta J = 0$, known as the *perpendicular branch*. This branch occurs at the vibrational frequency itself. In simple cases, it appears as a broad unresolved line. But if the moment of inertia differs

in the υ' and υ'' levels, the Q-branch may be seen as a group of very closely spaced lines.

The vibrational and rotational transitions discussed above are for the harmonic-oscillator rigid rotator in which the selection rules are given by $\Delta J = \pm 1$ and $\Delta \upsilon = \pm 1$. Because of the anharmonicity of the oscillator, the transition $\upsilon = 1 \leftarrow 0$ differs from the transition $\upsilon = 2 \leftarrow 1$. The upper-state band $2 \leftarrow 1$ does not have the same frequency as the ground-state band $1 \leftarrow 0$. Moreover, anharmonicity also changes the selection rules from those for a harmonic oscillator in which all integral changes of the quantum numbers are allowed. For example, $\Delta \upsilon = 2$ gives the first *overtone band* with twice the frequency of the normal (fundamental) mode. Simultaneous changes in two different vibrational quantum numbers give rise to *combination* and *difference bands* with frequencies that are the sum or difference of the normal-mode frequencies. They normally have smaller transition probabilities than fundamentals.

1.3.2   Line Broadening

Monochromatic emission is practically never observed. Energy levels during energy transitions are normally changed slightly due to both external influences on atoms and molecules, and the loss of energy in emission. As a consequence, radiation emitted during repeated energy transitions is nonmonochromatic, and spectral lines of finite widths are observed. The broadening of spectral lines is caused by: (1) the damping of oscillator vibrations resulting from the loss of energy in emission (the broadening of lines in this case is considered to be normal); (2) the perturbations due to reciprocal collisions between the absorbing molecules and between the absorbing and nonabsorbing molecules; and (3) the *Doppler effect* resulting from the difference in thermal velocities of atoms and molecules. The broadening of lines due to the loss of energy in emission (natural broadening) is practically negligible as compared to that caused by collisions and the Doppler effect. In the upper atmosphere, we find a combination of collision and Doppler broadenings, whereas in the lower atmosphere, below about 20 km, collision broadening prevails because of the pressure effect.

1.3.2.1   PRESSURE BROADENING

The shape of spectral lines due to collisions, referred to as *pressure broadening,* is given by the *Lorentz profile* (Lorentz, 1906). It is expressed by the formula

$$k_\nu = \frac{S}{\pi} \frac{\alpha}{(\nu - \nu_0)^2 + \alpha^2} = S f(\nu - \nu_0), \qquad (1.3.10)$$

where k_ν denotes the absorption coefficient, ν_0 is the wavenumber of an ideal, monochromatic line, α is the half-width of the line at the half-maximum and is a function of pressure and to a lesser degree of the temperature, $f(\nu - \nu_0)$ represents the shape factor of a spectral line, and the line strength or line intensity S is defined by

$$\int_{-\infty}^{\infty} k_\nu d\nu = S. \qquad (1.3.11)$$

In this case, we say the absorption coefficient is normalized. Figure 1.11 illustrates the Lorentz profile.
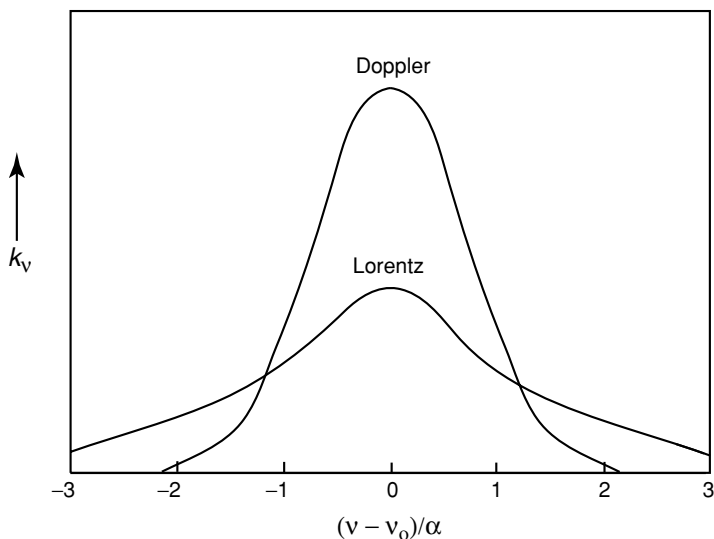
Figure 1.11　Demonstrative Lorentz and Doppler line shapes for the same intensities and line widths.

The Lorentz shape of absorption lines is fundamental to the theory of infrared radiative transfer in the atmosphere and thus, we should give a brief explanation of how the formula denoted in Eq. (1.3.10) is derived. An isolated molecule emits or absorbs an almost purely harmonic wave given by

$$f(t) = A \cos 2\pi \nu_0 ct, \qquad (1.3.12a)$$

where c is the velocity of light and A is an arbitrary amplitude. During the period $-t/2$ to $t/2$, the distribution of amplitude $g(\nu)$ of the wave in the discrete wavenumber domain may be obtained from the Fourier cosine transform as follows:

$$g(\nu) = \sqrt{\frac{2}{\pi}} \int_0^t (A \cos 2\pi \nu_0 ct') \cos 2\pi \nu ct' dt'$$

$$= \frac{A}{(2\pi)^{3/2}c} \left[\frac{\sin \pi(\nu_0 + \nu)ct}{\nu_0 + \nu} + \frac{\sin \pi(\nu_0 - \nu)ct}{\nu_0 - \nu} \right]. \qquad (1.3.12b)$$

Generally, the widths of absorption lines are much smaller than ν_0, so that the first term in Eq. (1.3.12b) may be neglected when it is compared to the second.

The only deviation from purely harmonic behavior would be produced by the damping due to the loss of energy in emission. In the infrared, the spectroscopic effect of this damping is extremely small. However, if a radiating molecule collides with another molecule, it alters the radiating harmonic wave train due to the intermolecular forces, and the frequency of the emitting molecules would be temporarily shifted by an appreciable amount. Since the collision may be considered to be instantaneous, we may assume that the principal effect of the collision is to destroy the phase coherence of the emitted wave train. That is to say, after the collision the molecule starts emitting

at another phase and these new phases are now randomly distributed. From general statistical principles, the time between collisions is distributed according to Poisson's law that the probability a collision occurs between t and $t + dt$ is e^{-t/t_0}, where t_0 is the mean time between collisions. All the initial phases of the wave trains must be averaged. Thus, the absorption coefficient will be given by

$$k_\nu = A' \int_0^\infty [g(\nu)]^2 e^{-t/t_0} \, dt, \qquad (1.3.13)$$

where $[g(\nu)]^2$ is the distribution of intensity, and A' is a certain constant. Further, by letting $1/t_0 = 2\pi \alpha c (\alpha$ in cm$^{-1})$ and using Eq. (1.3.11), we find that Eq. (1.3.13) becomes equivalent to Eq. (1.3.10). Here, $2\pi \alpha c$ is the number of collisions per molecule per unit time. [Exercise 1.14 requires the derivation of Eq. (1.3.10) from Eq. (1.3.13).] The Lorentz line shape also can be derived from the classical theory of absorption and dispersion as shown in Appendix D.

   From the kinetic theory of gases, the dependence of the half-width α on pressure and temperature is given by

$$\alpha = \alpha_0 (p/p_0)(T_0/T)^n, \qquad (1.3.14)$$

where α_0 is the width at the standard pressure, p_0(1013mb), and temperature, T_0(273K). The index n ranges from $^1/_2$ to 1, depending on the type of molecule. When $n = {}^1/_2$, it is known as the classical value. Under the reference condition, α_0 ranges from about 0.01 to 0.1 cm^{-1} for most radiatively active gases in the earth's atmosphere and depends on the spectral line. For the CO_2 molecule, it is fairly constant with a value of about 0.07 cm^{-1} (see Section 4.2.1 for further discussion).

1.3.2.2   DOPPLER BROADENING

   Assuming that there is no collision broadening in a highly rarefied gas, a molecule in a given quantum state radiates at wavenumber ν_0. If this molecule has a velocity component in the line of sight (the line joining the molecule and the observer), and if $v \ll c$, the velocity of light, the wavenumber

$$\nu = \nu_0(1 \pm v/c). \qquad (1.3.15)$$

Note that because of the conventional use of notation the wavenumber ν and the velocity v appear indistinguishable. Let the probability that the velocity component lies between v and $v + dv$ be $p(v) \, dv$. From the kinetic theory, if the translational states are in thermodynamic equilibrium, $p(v)$ is given by the Maxwell–Boltzmann distribution so that

$$p(v) \, dv = (m/2\pi KT)^{1/2} \exp(-mv^2/2KT) \, dv, \qquad (1.3.16)$$

where m is the mass of the molecule, K is the Boltzmann constant, and T is the absolute temperature.

   To obtain the Doppler distribution, we insert the expression of ν in Eq. (1.3.15) into Eq. (1.3.16), and perform normalization to an integrated line intensity S defined

in Eq. (1.3.11). After these operations, we find the absorption coefficient in the form

$$k_v = \frac{S}{\alpha_D \sqrt{\pi}} \exp\left[-\left(\frac{v - v_0}{\alpha_D}\right)^2\right], \qquad (1.3.17)$$

where

$$\alpha_D = v_0 (2KT/mc^2)^{1/2}, \qquad (1.3.18)$$

is a measure of the Doppler width of the line. The half-width at the half-maximum is $\alpha_D \sqrt{\ln 2}$. The Doppler half-width is proportional to the square root of the temperature.

A graphical representation of the Doppler line shape is also shown in Fig. 1.11. Since the absorption coefficient of a Doppler line is dependent on $\exp[-(v - v_0)^2]$, it is more intense at the line center and much weaker in the wings than the Lorentz shape. This implies that when a line is fully absorbed at the center, any addition of absorption will occur in the wings and will be caused by collision effects rather than Doppler effects.

1.3.2.3   VOIGT PROFILE

In the altitude region extending from about 20 to 50 km, effective line shapes are determined by both collision- and Doppler-broadening processes. We must add the Doppler shift component to the pressure-broadened lines at wavenumbers $v' - v_0$ in order to combine the two effects. The Doppler line redistributes the Lorentz line at wavenumber v' to v. The line shapes for pressure and Doppler broadening may then be expressed by $f(v' - v_0)$ and $f_D(v - v')$, respectively. To account for all possible thermal velocities, a convolution of the Lorentz and Doppler line shapes can be performed to obtain

$$f_v(v - v_0) = \int_{-\infty}^{\infty} f(v' - v_0) f_D(v - v') \, dv'$$

$$= \frac{1}{\pi^{3/2}} \frac{\alpha}{\alpha_D} \int_{-\infty}^{\infty} \frac{1}{(v' - v_0)^2 + \alpha^2} \exp\left[\frac{-(v - v')^2}{\alpha_D^2}\right] dv'. \quad (1.3.19a)$$

This line shape is referred to as the *Voigt profile*.

To simplify the representation of the Voigt profile, we let $t = (v - v')/\alpha_D$, $y = \alpha/\alpha_D$, and $x = (v - v_0)/\alpha_D$. Thus, we have

$$f_v(v - v_0) = \frac{1}{\alpha_D \sqrt{\pi}} K(x, y), \qquad (1.3.19b)$$

where the Voigt function is defined by

$$K(x, y) = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{1}{y^2 + (x - t)^2} e^{-t^2} \, dt. \qquad (1.3.20)$$

Many attempts have been made to simplify the computation of the Voigt function. Closed-form approximations can be developed. The Voigt profile satisfies the requirement of normalization such that

$$\int_{-\infty}^{\infty} f_v(v - v_0)d(v - v_0) = 1. \tag{1.3.21}$$

Exercise 1.17 requires the derivation of Eq. (1.3.19b) and the proofs that in the limits of $\alpha \to 0$ and $\alpha_D \to 0$, the Voigt profile reduces to the Doppler and Lorentz shapes, respectively.

One final note is in order. The line strength of a specific molecule that undergoes a transition from an upper energy state i to a lower energy state j is proportional to the square of the transition probability defined by

$$\mathbf{R}_{ij} = \int \psi_i^* \mathbf{M} \psi_j \, dV, \tag{1.3.22}$$

where $\mathbf{M}$ is the matrix of the dipole moment related to the time-dependent Hamiltonian, V is the volume, $\psi_{i,j}$ are wave functions of the upper and lower energy states that can be determined from Schrödinger's equation, and ψ^* is the conjugate of ψ. The line strength for absorption also depends on the ratio of the population, n_j, of the lower energy state of the transition to the total population of the absorbing gas, n. At thermodynamic equilibrium, this ratio is defined by the Boltzmann factor in the form

$$n_j/n = g_j e^{-E_j/KT} \Big/ \sum_i g_i e^{-E_i/KT}, \tag{1.3.23}$$

where the integer g_j is called degenerate or statistical weight, which is the number of distinct states having energy E_j. The denominator on the right side of Eq. (1.3.23) is the *partition function,* which can be determined for both vibrational and rotational states.

1.3.3   Breakdown of Thermodynamic Equilibrium

In Section 1.2.4, we pointed out that in thermodynamic equilibrium, the source function is given by the Planck function, which depends only on temperature, frequency, and the velocity of light, and that within a small constant-temperature enclosure in which nothing changes, an element of matter absorbs and emits according to Planck's and Kirchhoff's laws. However, as was first pointed out by Einstein, emission is also affected by the incident radiation field, referred to as *induced emission*. In the following, we wish to address the extent to which the source function and absorption coefficient can be changed from their equilibrium values by the action of incident radiation. Since the discussion now involves departure from the equilibrium state, the thermodynamic arguments cannot be followed. We must now use a microscopic statistical model to understand the condition under which Kirchhoff's law cannot be applied.

Thermodynamic equilibrium can be defined in terms of Boltzmann's law for the distribution of molecules between two states. Consider a simple case where emission
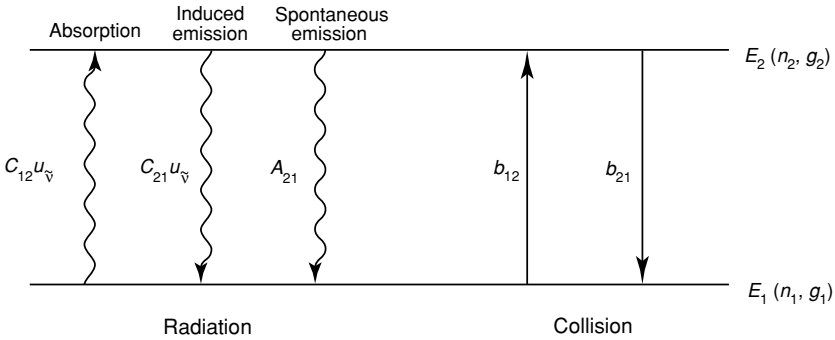
Figure 1.12 Radiative and collisional transitions connecting two energy states E_1 and E_2 where n_i are the numbers of molecules at the level E_i per unit volume; g_i are the statistical weights; b_{21} denotes the probability of transition by collisions per unit of time from the upper level E_2 to the lower level E_1 and b_{12} from E_1 to E_2; C_{12}, C_{21}, and A_{21} are the Einstein coefficients associated with emission and absorption; and $u_{\tilde{v}}$ denotes the radiant energy density.

(or absorption) occurs through transitions between the energy levels E_1 and E_2 of a two-level energy system, as shown in Fig. 1.12. Let n_1 and n_2 be the numbers of molecules at these levels per unit volume; and g_1 and g_2 denote the statistical weights, defined in Eq. (1.3.23). In accordance with Boltzmann's law, the distribution of molecules between two states is defined by

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \exp\left(-\frac{E_2 - E_1}{KT}\right) = \frac{g_2}{g_1} e^{-h\tilde{v}/KT}, \qquad (1.3.24)$$

where $\tilde{v}$ is the emitting frequency from Planck's relation denoted in Eq. (1.3.1). Under complete equilibrium, Eq. (1.3.24) is obeyed for all energy states throughout the medium. It is known that collisions acting alone will bring about a Boltzmann distribution and consequently a Planck source function. However, in a collisionless medium, radiation can bring about almost any population of energy levels through absorption and induced emission involving a radiating molecule and a photon (C coefficients), as shown in Fig. 1.12. Radiative transitions can also take place spontaneously without the presence of a photon or a colliding molecule (A coefficients). This occurs from the upper level to the lower level. Including both collision and radiation processes, one can show from the balance of transitions between two energy levels that the state population ratio is given by

$$\frac{n_2}{n_1} = \frac{g_2}{g_1} \frac{\eta + u_{\tilde{v}}}{\eta \exp(h\tilde{v}/KT) + 8\pi h\tilde{v}^3/c^3 + u_{\tilde{v}}}, \qquad (1.3.25)$$

where $u_{\tilde{v}}$ is the energy density and the coefficient that governs the relative importance of collision and radiation is defined by

$$\eta = \frac{b_{12} \text{ (collision)}}{C_{12} \text{ (radiation)}}. \qquad (1.3.26)$$

Exercise 1.18 requires the derivation of Eq. (1.3.25). The population of energy levels and the resulting source functions will be governed by the conflict between radiative and collisional effects. The rate of collisional adjustment of state populations is determined by a relaxation time proportional to the pressure. Radiative adjustment is determined by the natural lifetime of the excited states with respect to radiative transitions and is dependent on specific molecules but independent of the state of the medium. When the collision events are much larger than the radiative transitions, then $\eta \gg 1$ and Eq. (1.3.25) reduces to Eq. (1.3.24). In this case, LTE should occur and Planck's law will be valid. On the other hand, if $\eta \ll 1$, then a different source function would be required. In the earth's atmosphere, pressure varies rapidly with height and since collisional processes are dependent on pressure, there will be a sharply defined relaxation level below which Planck's law (LTE) is valid for transitions between energy levels but above which a different source function will be required (non-LTE). This level occurs at about 60–70 km in the earth's atmosphere.

1.4   Introduction to Radiative Transfer

1.4.1   The Equation of Radiative Transfer

A pencil of radiation traversing a medium will be weakened by its interaction with matter. If the intensity of radiation I_λ becomes $I_\lambda + dI_\lambda$ after traversing a thickness ds in the direction of its propagation, then

$$dI_\lambda = -k_\lambda \rho I_\lambda \, ds, \qquad (1.4.1)$$

where ρ is the density of the material, and k_λ denotes the mass extinction cross section (in units of area per mass) for radiation of wavelength λ. As discussed in Section 1.1.4, the mass extinction cross section is the sum of the mass absorption and scattering cross sections. Thus, the reduction in intensity is due to absorption by the material as well as to scattering by the material.

On the other hand, the radiation's intensity may be strengthened by emission from the material plus multiple scattering from all other directions into the pencil under consideration at the same wavelength (see Fig. 1.13). We define the source function coefficient j_λ such that the increase in intensity due to emission and multiple scattering is given by

$$dI_\lambda = j_\lambda \rho \, ds, \qquad (1.4.2)$$

where the source function coefficient j_λ has the same physical meaning as the mass extinction cross section. Upon combining Eqs. (1.4.1) and (1.4.2), we obtain

$$dI_\lambda = -k_\lambda \rho I_\lambda \, ds + j_\lambda \rho \, ds. \qquad (1.4.3)$$

Moreover, it is convenient to define the source function J_λ such that

$$J_\lambda \equiv j_\lambda / k_\lambda. \qquad (1.4.4)$$
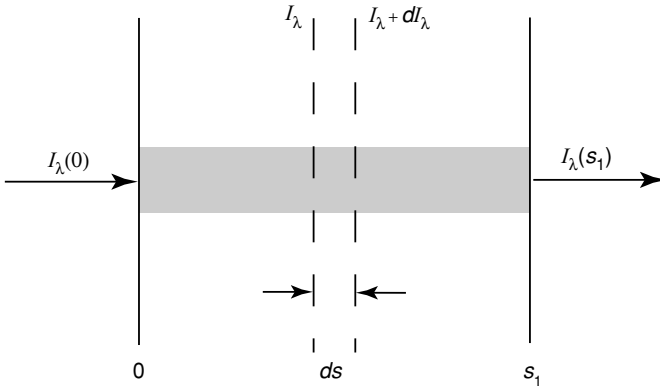
Figure 1.13   Depletion of the radiant intensity in traversing an extinction medium.

In this manner, the source function has units of radiant intensity. It follows that Eq. (1.4.3) may be rearranged to yield

$$\frac{dI_\lambda}{k_\lambda \rho \, ds} = -I_\lambda + J_\lambda. \tag{1.4.5}$$

This is the general radiative transfer equation without any coordinate system imposed, which is fundamental to the discussion of any radiative transfer process.

1.4.2   Beer–Bouguer–Lambert Law

Consider a direct light beam from the sun, which covers the wavelengths from about 0.2 to 5 μm. Emission contributions from the earth–atmosphere system can be generally neglected, as discussed in Section 1.2. Moreover, if the diffuse radiation produced by multiple scattering can be neglected, then Eq. (1.4.5) reduces to the following form:

$$\frac{dI_\lambda}{k_\lambda \rho \, ds} = -I_\lambda. \tag{1.4.6}$$

Let the incident intensity at $s = 0$ be $I_\lambda(0)$. Then the emergent intensity at a distance s away shown in Fig. 1.13 can be obtained by integrating Eq. (1.4.6) and is given by

$$I_\lambda(s_1) = I_\lambda(0) \exp\left(-\int_0^{s_1} k_\lambda \rho \, ds\right). \tag{1.4.7}$$

Assuming that the medium is homogeneous, so that k_λ is independent of the distance s, and defining the path length

$$u = \int_0^{s_1} \rho \, ds, \tag{1.4.8}$$

Eq. (1.4.7) can be expressed by

$$I_\lambda(s_1) = I_\lambda(0) e^{-k_\lambda u}. \tag{1.4.9}$$

This is known as Beer's law or Bouguer's law or Lambert's law, referred to here as the Beer–Bouguer–Lambert law, which states that the decrease in the radiant intensity traversing a homogeneous extinction medium is in accord with the simple exponential function whose argument is the product of the mass extinction cross section and the path length. Since this law involves no directional dependence, it is applicable not only to the intensity quantity but also to the flux density and the flux.

From Eq. (1.4.9), we can define the monochromatic transmissivity T_λ as follows:

$$T_\lambda = I_\lambda(s_1)/I_\lambda(0) = e^{-k_\lambda u}. \tag{1.4.10}$$

Moreover, for a nonscattering medium, the monochromatic absorptivity, representing the fractional part of the incident radiation that is absorbed by the medium, is given by

$$A_\lambda = 1 - T_\lambda = 1 - e^{-k_\lambda u}. \tag{1.4.11}$$

Equations (1.4.10) and (1.4.11) are normally expressed in the wavenumber domain in conjunction with the application of infrared radiation transfer. Finally, if there is a scattering contribution from the medium, certain portions of the incident radiation may reflect back to the incident direction. In this case, we may define the monochromatic reflectivity R_λ, which is the ratio of the reflected (backscattered) intensity to the incident intensity. On the basis of the conservation of energy, we must have

$$T_\lambda + A_\lambda + R_\lambda = 1 \tag{1.4.12}$$

for the transfer of radiation through a scattering and absorbing medium.

1.4.3   Schwarzschild's Equation and Its Solution

Consider a nonscattering medium that is in local thermodynamic equilibrium. A beam of intensity I_λ passing through it will undergo the absorption and emission processes simultaneously. This is the case for the transfer of thermal infrared radiation emitted from the earth and the atmosphere. The source function, as defined in Eq. (1.4.4), is given by the Planck function and can be expressed by

$$J_\lambda = B_\lambda(T). \tag{1.4.13}$$

Hence, the equation of radiative transfer can now be written as

$$\frac{dI_\lambda}{k_\lambda \rho \, ds} = -I_\lambda + B_\lambda(T), \tag{1.4.14}$$

where k_λ is now the absorption coefficient. The first term in the right-hand side of Eq. (1.4.14) denotes the reduction of the radiant intensity due to absorption, whereas the second term represents the increase in the radiant intensity arising from blackbody emission of the material. To seek a solution for Schwarzschild's equation, we define the monochromatic optical thickness of the medium between points s and s_1 as shown
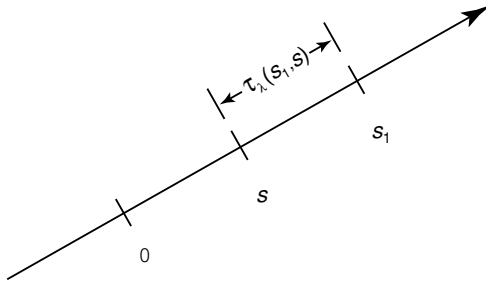
Figure 1.14   Configuration of the optical thickness τ_λ defined in Eq. (1.4.15).

in Fig. 1.14 in the form

$$\tau_\lambda(s_1, s) = \int_s^{s_1} k_\lambda \rho \, ds'. \tag{1.4.15}$$

By noting that

$$d\tau_\lambda(s_1, s) = -k_\lambda \rho \, ds, \tag{1.4.16}$$

Eq. (1.4.14) becomes

$$-\frac{dI_\lambda(s)}{d\tau_\lambda(s_1, s)} = -I_\lambda(s) + B_\lambda[T(s)]. \tag{1.4.17}$$

Upon multiplying Eq. (1.4.17) by a factor $e^{-\tau_\lambda(s_1,s)}$, and integrating the thickness ds from 0 to s_1, we obtain

$$-\int_0^{s_1} d[I_\lambda(s)e^{-\tau_\lambda(s_1,s)}] = \int_0^{s_1} B_\lambda[T(s)]e^{-\tau_\lambda(s_1,s)}d\tau_\lambda(s_1, s). \tag{1.4.18}$$

Consequently, we have

$$I_\lambda(s_1) = I_\lambda(0)e^{-\tau_\lambda(s_1,0)} + \int_0^{s_1} B_\lambda[T(s)]e^{-\tau_\lambda(s_1,s)}k_\lambda \rho \, ds. \tag{1.4.19}$$

The first term in Eq. (1.4.19) is essentially equivalent to Eq. (1.4.7), representing the absorption attentuation of the radiant intensity by the medium. The second term denotes the emission contribution from the medium along the path from 0 to s_1. If the temperature and density of the medium and the associated absorption coefficient along the path of the beam are known, Eq. (1.4.19) can be integrated numerically to yield the intensity at the point s_1. Applications of Eq. (1.4.19) to infrared radiative transfer and to the remote sounding of atmospheric temperature profiles and compositions from orbiting meteorological satellites will be discussed in Chapters 4 and 7, respectively.

In the discussion of the absorption and emission lines in the spectra of sun and stars, Schwarzschild (1914) presented Eq. (1.4.14) within the context of Kirchhoff's law and derived an integral solution for the condition without scattering. It is thus referred to as Schwarzschild's equation.
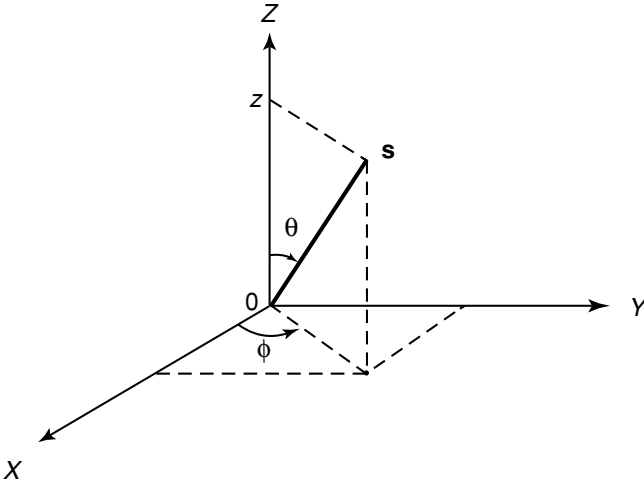
Figure 1.15   Geometry for plane-parallel atmospheres where θ and ϕ denote the zenith and azimuthal angles, respectively, and **s** represents the position vector.

1.4.4   The Equation of Radiative Transfer for Plane-Parallel Atmospheres

For many atmospheric radiative transfer applications, it is physically appropriate to consider that the atmosphere in localized portions is plane-parallel such that variations in the intensity and atmospheric parameters (temperature and gaseous profiles) are permitted only in the vertical direction (i.e., height and pressure). In this case, it is convenient to measure linear distances normal to the plane of stratification (see Fig. 1.15). If z denotes this distance, then the general equation of radiative transfer defined in Eq. (1.4.5) becomes

$$\cos\theta \frac{dI(z;\theta,\phi)}{k\rho\,dz} = -I(z;\theta,\phi) + J(z;\theta,\phi), \qquad (1.4.20)$$

where θ denotes the inclination to the upward normal, and ϕ the azimuthal angle in reference to the x axis. Here, we have omitted the subscript λ on various radiative quantities.

Introducing the normal optical thickness (or depth)

$$\tau = \int_{z}^{\infty} k\rho\,dz' \qquad (1.4.21)$$

measured downward from the outer boundary, we have

$$\mu \frac{dI(\tau;\mu,\phi)}{d\tau} = I(\tau;\mu,\phi) - J(\tau;\mu,\phi), \qquad (1.4.22)$$

where $\mu = \cos\theta$. This is the basic equation for the problem of multiple scattering in plane-parallel atmospheres.
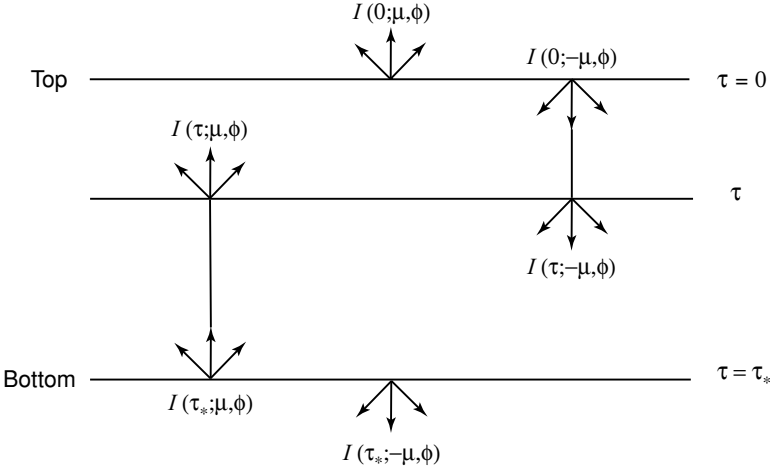
Figure 1.16   Upward (μ) and downward ($-\mu$) intensities at a given level τ and at top ($\tau = 0$) and bottom ($\tau = \tau_*$) levels in a finite, plane-parallel atmosphere.

Following the same procedure as that described in Section 1.4.3, Eq. (1.4.22) can be solved to give the upward and downward intensities for a finite atmosphere that is bounded on two sides at $\tau = 0$ and $\tau = \tau_*$ as illustrated in Fig. 1.16. To obtain the upward intensity ($\mu > 0$) at level τ, we multiply Eq. (1.4.22) by a factor $e^{-\tau/\mu}$ and perform integration from τ to $\tau = \tau_*$. This leads to

$$I(\tau; \mu, \phi) = I(\tau_*; \mu, \phi)e^{-(\tau_*-\tau)/\mu} + \int_\tau^{\tau_*} J(\tau'; \mu, \phi)e^{-(\tau'-\tau)/\mu}\frac{d\tau'}{\mu} \quad (1 \geq \mu > 0).$$

$$(1.4.23)$$

To derive the downward intensity ($\mu < 0$) at level τ, a factor $e^{\tau/\mu}$ is used and μ is replaced by $-\mu$. After carrying out integration from $\tau = 0$ to τ, we obtain

$$I(\tau; -\mu, \phi) = I(0; -\mu, \phi)e^{-\tau/\mu} + \int_0^\tau J(\tau'; -\mu, \phi)e^{-(\tau-\tau')/\mu}\frac{d\tau'}{\mu} \quad (1 \geq \mu > 0).$$

$$(1.4.24)$$

In Eqs. (1.4.23) and (1.4.24), $I(\tau_*; \mu, \phi)$ and $I(0; -\mu, \phi)$ represent the inward source intensities at the bottom and top surfaces, respectively, as shown in Fig. 1.16.

For applications to planetary atmospheres, it is desirable to measure the emergent outward intensities at the top and bottom of the atmosphere in conjunction with the remote sensing of atmospheric compositions and radiation balance studies. Upon setting $\tau = 0$ in Eq. (1.4.23), we have

$$I(0; \mu, \phi) = I(\tau_*; \mu, \phi)e^{-\tau_*/\mu} + \int_0^{\tau_*} J(\tau'; \mu, \phi)e^{-\tau'/\mu}\frac{d\tau'}{\mu}, \quad (1.4.25)$$

where the first and second terms represent, respectively, the bottom surface contribution (attenuated to the top) and the internal atmospheric contribution. On the other

hand, if we set $\tau = \tau_*$ in Eq. (1.4.24), we obtain

$$I(\tau_*; -\mu, \phi) = I(0; -\mu, \phi)e^{-\tau_*/\mu} + \int_0^{\tau_*} J(\tau'; -\mu, \phi)e^{-(\tau_* - \tau')/\mu}\frac{d\tau'}{\mu}, \quad (1.4.26)$$

where, again, the first and second terms represent the top surface contribution (attenuated to the bottom) and the internal atmospheric contribution, respectively. Detailed applications of the preceding equations associated with infrared radiation transfer and multiple scattering will be discussed in Chapters 4 and 6.

1.4.5   Radiative Transfer Equations for Three-Dimensional Inhomogeneous Media

In several atmospheric conditions, the plane-parallel assumption may not be valid. These include the transfer of radiation in the atmosphere where spherical geometry must be accounted for, and in clouds with finite dimension and/or inhomogeneity in the horizontal direction. The latter has been a subject of contemporary research and development in conjunction with studies of clouds in climate and remote sensing. Although this topic will be further elaborated upon in Chapter 6, here we provide some introductory notes consistent with the preceding presentation. We begin with the general equation of radiative transfer discussed in Section 1.4.1. Letting the extinction coefficient be $\beta_e = k_\lambda \rho$ and omitting the subscript λ for simplicity, we write

$$-\frac{dI}{\beta_e \, ds} = I - J. \quad (1.4.27)$$

The differential operator can be defined in time and space as follows:

$$\frac{d}{ds} = \frac{1}{c}\frac{\partial}{\partial t} + \mathbf{\Omega} \cdot \nabla, \quad (1.4.28)$$

where c is the velocity of light, $\mathbf{\Omega}$ is a unit vector specifying the direction of scattering through a position vector $\mathbf{s}$, and t is time. Under the condition that radiation is independent of time (steady state), such as the illumination from the sun, Eq. (1.4.27) can be expressed by

$$-\frac{1}{\beta_e(\mathbf{s})}(\mathbf{\Omega} \cdot \nabla)I(\mathbf{s}, \mathbf{\Omega}) = I(\mathbf{s}, \mathbf{\Omega}) - J(\mathbf{s}, \mathbf{\Omega}), \quad (1.4.29a)$$

where the source function, J, can be produced by the single scattering of the direct solar beam, multiple scattering of the diffuse intensity, and emission of the medium.

In Cartesian coordinates (x, y, z), we have

$$\mathbf{\Omega} \cdot \nabla = \Omega_x \frac{\partial}{\partial x} + \Omega_y \frac{\partial}{\partial y} + \Omega_z \frac{\partial}{\partial z}, \quad (1.4.29b)$$

where the directional cosines are given by

$$\Omega_x = \frac{\partial x}{\partial s} = \sin\theta \, \cos\phi = (1 - \mu^2)^{1/2} \cos\phi, \quad (1.4.30a)$$

$$\Omega_y = \frac{\partial y}{\partial s} = \sin\theta \ \sin\phi = (1-\mu^2)^{1/2} \sin\phi, \qquad (1.4.30b)$$

$$\Omega_z = \frac{\partial z}{\partial s} = \cos\theta = \mu, \qquad (1.4.30c)$$

where θ and ϕ are the zenith and azimuthal angles defined previously, and $|\mathbf{s}| = s = (x^2 + y^2 + z^2)^{1/2}$. In general, analytic solutions for Eq. (1.4.29a) do not exist and it must be solved numerically. In cases where the medium is homogeneous with respect to its single-scattering properties including the extinction coefficient, Eq. (1.4.29a) reduces to a first-order partial differential equation from which simplified solutions can be derived. Interested readers should refer to Chapter 6 for further details.

Exercises

1.1   What is the meaning of isotropic radiation? Show that for isotropic radiation, the monochromatic flux density is $F_\lambda = \pi I_\lambda$.

1.2   A meteorological satellite circles the earth at a height h above the earth's surface. Let the radius of the earth be a_e and show that the solid angle under which the earth is seen by the satellite sensor is $2\pi[1 - (2a_e h + h^2)^{1/2}/(a_e + h)]$.

1.3   Express the Planck function in the wavelength and wavenumber domains based on the Planck function in the frequency domain.

1.4   From Eq. (1.2.10), show that Eq. (1.2.11) is true.

1.5   Show that the maximum intensity of the Planck function is proportional to the fifth power of the temperature.

1.6   An infrared scanning radiometer aboard a meteorological satellite measures the outgoing radiation emitted from the earth's surface in the $10\ \mu$m window region. Assuming that the effect of the atmosphere between the satellite and the surface can be neglected, what would be the temperature of the surface if the observed radiance at $10\ \mu$m is $9.8\ \mathrm{W\,m^{-2}\ \mu m^{-1}\ sr^{-1}}$?

1.7   A black land surface with a temperature of $15°$C emits radiation at all frequencies. What would be the emitted radiances at $0.7\ \mu$m, $1000\ \mathrm{cm^{-1}}$, and 31.4 GHz? Use appropriate Planck functions in the calculations.

1.8   Assuming the average normal body temperature is $98°$F, what would be the emittance of the body? If it is not a blackbody but absorbs only 90% of the incoming radiation averaged over all wavelengths, what would be the emittance in this case? Also, at which wavelength does the body emit the maximum energy?

1.9   (a) The photosphere of the sun has a temperature of about 5800 K. Assuming it is a blackbody, compute the percentage of its emitting intensity at wavelengths longer than $5\ \mu$m. (b) The earth–atmosphere system has an equilibrium temperature of about 255 K. Assume it can be considered a blackbody and compute the percentage of its emitting intensity at wavelengths shorter than $5\ \mu$m.

1.10 Show that when $\lambda \to \infty$, the Planck intensity is directly proportional to the temperature, referred to as the Rayleigh–Jeans distribution. When $\lambda \to 0$, derive the expression referred to as the Wien distribution. Plot the Planck intensity for the temperatures of the sun and the earth–atmosphere system and compare these two approximations with the exact values.

1.11 (a) From Newton's second law of motion and Coulomb's law, find the kinetic energy of an electron in a hydrogen atom moving with a velocity v in a circular orbit of radius r centered on its nucleus. Express r in terms of the quantum number n using the selection rule for the angular momentum mvr. Then find the potential energy of the proton–electron system. By combining the kinetic and potential energy, derive Eq. (1.3.3). (b) Considering only the transitions between the ground state ($n = 1$) and the excited states and letting the highest quantum number be 6, compute the wavelengths of hydrogen emission lines.

1.12 In spectroscopy, the wavenumber v in cm^{-1} is not only used to specify a spectral location, but also as a measure of energy itself. From the energy equation, compute 1 cm^{-1} of energy of terms of 1 joule per molecule.

1.13 (a) Derive Eq. (1.3.9) from Eq. (1.3.8). (b) For the Q-branch, show that the line spacing is proportional to the difference of the rotational constants in the ground and excited states.

1.14 Derive Eq. (1.3.10) from Eq. (1.3.13).

1.15 Prove that the line intensity $S = \int_{-\infty}^{\infty} k_v \, dv$ for Lorentz, Doppler, and Voigt absorption lines.

1.16 Calculate and plot the shape factor of the Lorentz and Doppler profiles for ozone whose half-width is assumed to be 0.1 cm^{-1} in the wavenumber domain at the standard temperature and pressure.

1.17 Derive Eq. (1.3.19b) from Eq. (1.3.19a). In the limits of $\alpha \to 0$ and $\alpha_D \to 0$, show that the Voigt profile reduces to the Doppler and Lorentz shapes, respectively.

1.18 (a) From Eq. (1.3.24) for collision, derive an expression for b_{12}/b_{21} from n_2/n_1. (b) For radiation, show that $C_{21}/C_{12} = g_1/g_2$ and $A_{21}/C_{12} = g_1/g_2 \cdot 8\pi h\tilde{v}^3/c^3$, where $u_{\tilde{v}}$ is the energy density defined in Appendix A. (c) Then, based on the balance of transitions between two energy levels (Fig. 1.12), show that the state population ratio for the general case is given by Eq. (1.3.25).

1.19 A He–Ne laser beam at 0.6328 μm with an output power of 5 mW (10^{-3} W) is passing through an artificial cloud layer 10 m in thickness and is directed at $30°$ from the normal to the layer. Neglecting the effect of multiple scattering, calculate the extinction coefficients (per length) if the measured powers are 1.57576 and 0.01554 mW. Also calculate the normal optical depths in these cases.

1.20 The contrast of an object against its surroundings is defined by

$$C \equiv (B - B_0)/B_0,$$

where B represents the brightness of the object. $B = B_0$ when $x \to \infty$, and $B = 0$ when $x = 0$, where x denotes the distance between the object and the observer. For the normal eye, the threshold contrast has a value of ± 0.02. Assuming that the extinction coefficient β_e is independent of the wavelength in the visible, show that the visual range or visibility is given by $x = 3.912/\beta_e$. If extinction coefficients for a clear atmosphere containing a background aerosol and for a dense fog are 0.08 and 100 km^{-1}, respectively, what would be the visibilities in these cases?

1.21 In reference to Fig. 1.13, if the reflectivity at both $s = 0$ and $s = s_1$ is R_λ, what would be the value of $I_\lambda(s_1)$? Consider sequentially the transmission at $s = 0$, the absorption in the medium, and the transmission at $s = s_1$.

1.22 By differentiation with respect to the optical thickness τ, show that Eqs. (1.4.23) and (1.4.24) reduce to Eq. (1.4.22), the equation of transfer for plane-parallel atmospheres. Use the following Leibnitz's rule for differentiation of integrals:

$$\frac{d}{dy} \int_{\phi_1(y)}^{\phi_2(y)} F(x, y)\,dx = \int_{\phi_1(y)}^{\phi_2(y)} \frac{\partial F}{\partial y}\,dx - F(\phi_1, y)\frac{d\phi_1}{dy} + F(\phi_2, y)\frac{d\phi_2}{dy}.$$

1.23 Consider an isothermal nonscattering atmosphere with a temperature T and let the surface temperature of such an atmosphere be T_s. Derive an expression for the emergent flux density at the top of an atmosphere whose optical depth is τ_* by using Eq. (1.4.25) and show that it can be expressed by the *exponential integral of third order* given by

$$E_3(\tau_*) = \int_0^1 e^{-\tau_*/\mu} \mu\,d\mu.$$

Suggested Reading

Goody, R. M., and Yung, Y. L. (1989). *Atmospheric Radiation. Theoretical Basis*, 2nd ed. Oxford University Press, New York. Chapter 3 presents an authoritative and advanced discussion of the vibration–rotation spectra of molecules.

Jastrow, R., and Thompson, M. H. (1984). *Astronomy: Fundamentals and Frontiers*, 2nd ed. Wiley, New York. Chapter 4 contains an excellent description in lay terms of absorption line formation.

McCartney, E. J. (1983). *Absorption and Emission by Atmospheric Gases*. Wiley, New York. Chapters 4 and 5 provide enlightened discussions of the fundamentals of rotational, vibrational, and electronic transitions.

Penner, S. S. (1950). *Quantitative Molecular Spectroscopy and Gas Emissivities*. Addison-Wesley, Reading, MA. Chapters 1–3 provide helpful explanations of blackbody radiation and absorption line profiles.

Chapter 2 | Solar Radiation at the Top of the Atmosphere

2.1   The Sun as an Energy Source

Once upon a time about 4.6 billion years ago, the sun condensed out of the center of a thin, hot, spinning disk of interstellar material, according to a theory proposed by Laplace (1796). The sun is a typical second-generation G2 star in the cosmic hierarchy. Among the billions of stars in the universe, the sun is about average in mass but below average in size. The sun has one unique feature in that it is 300,000 times closer to the earth than the next nearest star. With a mean distance of about 1.5×10^8 km between the earth and the sun, virtually all of the energy that the earth receives and that sets the earth's atmosphere and oceans in motion comes from the sun.

The sun is a gaseous sphere with a visible radius of about 6.96×10^5 km and a mass of approximately 1.99×10^{30} kg. Its main ingredients are primordial hydrogen (H) and helium (He), plus a small amount of heavier elements including oxygen (O), carbon (C), nitrogen (N), neon (Ne), iron (Fe), silicon (Si), magnesium (Mg), sulfur (S), and calcium (Ca). Hydrogen makes up roughly 90% of the mass, and the remaining 10% or so is helium. The temperature of the sun decreases from a central value of about 5×10^6 K to about 5800 K at the surface. The density within the sun falls off very rapidly with increasing distance from the center. The central density is about 150 g cm^{-3}, and at the surface, it is about 10^{-7} g cm^{-3}. The average density is about 1.4 g cm^{-3}. Approximately 90% of the sun's mass is contained in the inner half of its radius.

Solar energy is believed to be generated by the steady conversion of four hydrogen atoms to one helium atom in fusion reactions, which take place in the deep interior of the sun with temperatures up to many millions of degrees, as shown in Fig. 2.1. The amount of energy released in nuclear fusion causes a reduction of the sun's mass. According to Einstein's law relating mass and energy, $E = mc^2$, and converting the energy radiated by the sun, we find that almost 5 million tons of mass per second are radiated by the sun in the form of electromagnetic energy. In a billion years, it is believed that the sun will radiate into space about 10^{26} kg, which is less than one part in 10^4 of its total mass. Thus, only an insignificant fraction of the sun's substance has been lost by electromagnetic radiation. It is estimated that only 5% of the sun's total mass has been converted from hydrogen to helium in its lifetime thus far.
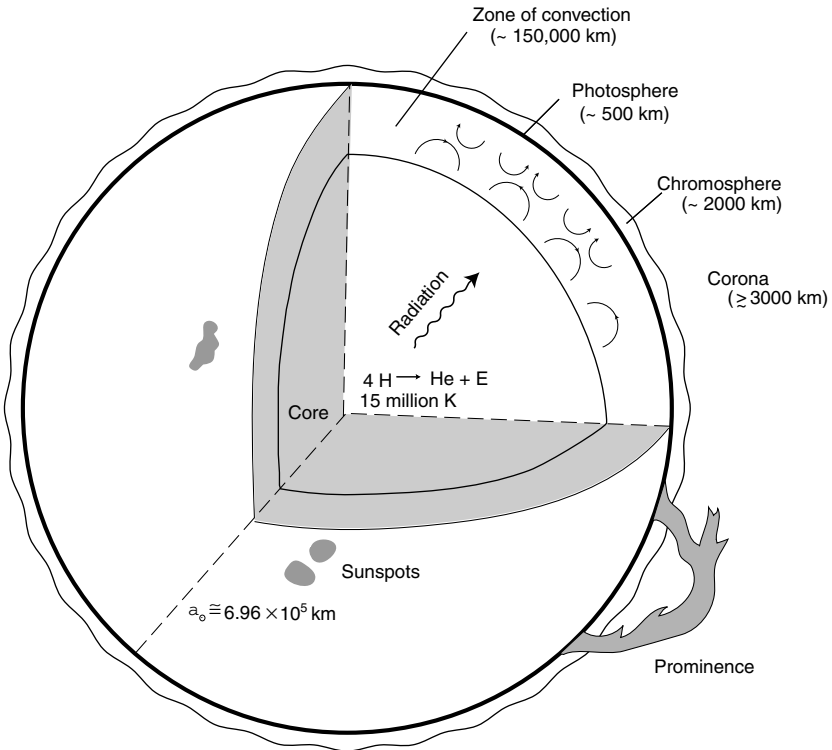
Figure 2.1   A cross section of the sun illustrating the solar interior and atmosphere. The solar interior includes the core with a temperature of about 1.5×10^7 K, the radiation zone, and the convective zone. The solar atmosphere includes the photosphere, the chromosphere, and the corona. The former two layers are exaggerated for illustration purposes.

As a result of the extremely high temperatures in the deep interior of the sun, collisions between atoms are sufficiently violent to eject many electrons from their orbits. Only the tightly bound inner electrons of heavy atoms will be retained. The energy emitted by nuclear fusion in the form of photons can pass through the inner part of the sun without being absorbed by the electrons. However, closer to the sun's surface, the temperature decreases and the heavier atoms such as iron begin to recapture their outer electrons. These outer electrons are bound to the nucleus by relatively small forces and can be easily separated from the nucleus by the absorption of photons. It follows that the flow of photons coming from the interior is blocked by the appearance of the absorbing atoms. The blocking of these photons will cause the temperature to drop sharply at some depth below the surface. Thus, the outer region of the sun consists of a layer of relatively cool gas resting on top of a hotter interior. As a consequence, the gas at the bottom of the cool outer layer is heated by the hot gas in the interior. It undergoes expansion and rises toward the surface. Once it reaches the surface, the hot gas loses its heat to space, cools, and descends into the interior. The entire outer

Table 2.1

Magnitude and Variability of the Solar Sources of Terrestrial Energy[a]

Source	Energy (W m^{-2})	Solar cycle change (W m^{-2})	Terrestrial deposition altitude
Solar radiation			
Total irradiance	1366	1.3	Surface, troposphere
UV 2000–3000 Å	15.4	0.16	0–50 km
UV 0–2000 Å	0.1	0.02	50–500 km
Particles			
Solar protons	0.002		30–90 km
Galactic cosmic rays	0.000007		0–90 km
Solar wind	0.0003		About 500 km

[a]Data taken from National Research Council (1994).

layer breaks up into ascending columns of heated gas and descending columns of cooler gas. The region in which this large-scale upward and downward movement of gases occurs is called the *zone of convection* (Fig. 2.1), which extends from a depth of about 150,000 km to the surface of the sun. Below this depth, it is believed that energy is transported within the sun by means of electromagnetic radiation, i.e., by the flow of photons. Near the surface, however, because of the substantial blocking of radiant energy by the absorption of heavier elements, energy is transferred partly by convection and partly by electromagnetic radiation. Above the surface, energy transport is again by means of electromagnetic radiation. The sun provides electromagnetic, particle, and plasma energy to the earth as summarized in Table 2.1. It is clear that electromagnetic radiation characterized by wavelengths from gamma rays to radio waves (see Fig. 1.1) is by far the largest solar energy source for the earth and the most important for its weather and climate processes.

2.1.1 The Structure of the Sun

The visible region of the sun is called the *photosphere,* where most of the electromagnetic energy reaching the earth originates. Although the sun is a gaseous body, the photosphere is referred to as the *surface* of the sun. The photosphere is marked by relatively bright *granules* about 1500 km in diameter, which are separated by dark regions and variable features called *sunspots*. The bright granules are fairly uniformly distributed over the solar disk and are believed to be associated with ascending hot gases in the uppermost layer of the zone of convection discussed previously.

The photosphere is a comparatively thin layer about 500 km thick that constitutes the source of the sun's visible radiation, as illustrated in Fig. 2.1. The temperature in this layer varies from 8000 K in the lower layer to 4000 K in the upper layer. Matching the theoretical Planck curve versus wavelength (see Fig. 2.9) with the measured spectral radiant energy emitted by the sun, the best agreement was found for a temperature of approximately 5800 K. This temperature is an average over the

temperature range of the photosphere. The *effective temperature* of the photosphere may also be obtained by measuring the luminosity of the sun. On the basis of the Stefan–Boltzmann Law, and with the knowledge that the sun's radius assumes a sharply defined surface, the effective temperature also yields a value close to 5800 K. Radiation emitted from the photosphere is essentially continuous and will be discussed further in Section 2.3.1.

The region above the photosphere is called the *solar atmosphere*. It is characterized by tenuous and transparent solar gases. The solar atmosphere is divided into two regions called the *chromosphere* and the *corona* (see Fig. 2.1). The chromosphere lies above the photosphere to a height of approximately 2000 km. Its temperature increases from a minimum of about 4000 K and stays between 4000 and 6000 K up to about 2000 km. Above this height, the temperature rises drastically, reaching about 10^5–10^6 K. The temperature structure of the solar atmosphere is shown in Fig. 2.2, in
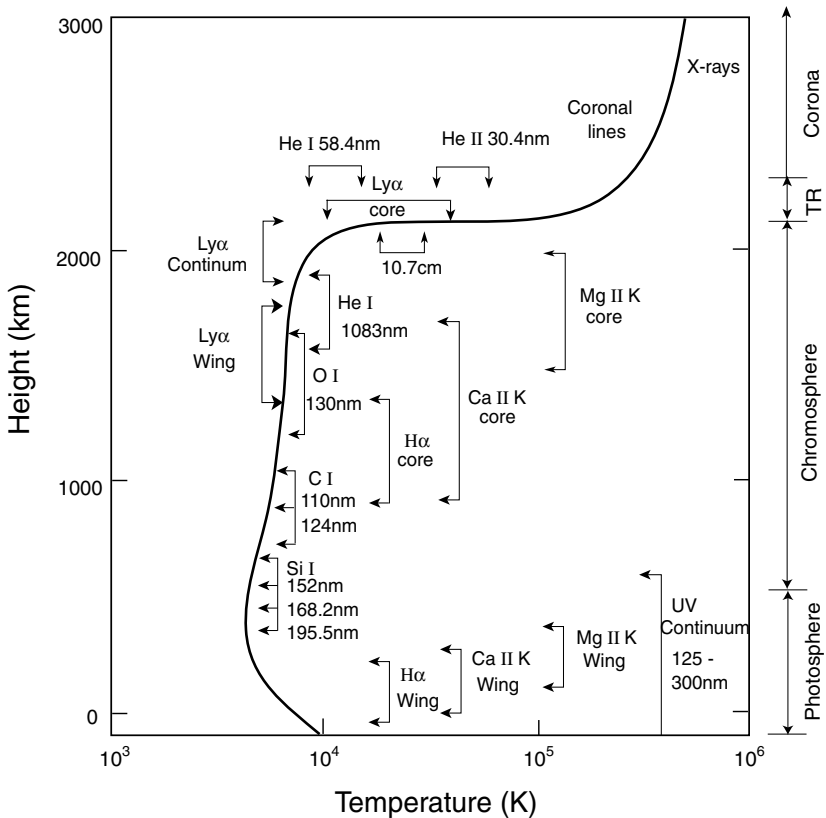
Figure 2.2   The temperature structure of the solar atmosphere and the regions where absorption lines are formed. Data constructed from the results published in Lean (1987) and Livingston *et al.* (1988). The symbols I and II denote a neutral atom and a singly ionized atom, respectively, and TR refers to the transition region.

which the regions of formation of different absorption lines discussed below are also displayed. The layer with a minimum temperature of 4000 K extends to a few thousand kilometers, consisting of relatively cool gases lying over the hotter gases. These cool gases absorb continuous radiation emitted from the photosphere at wavelengths characteristic of the atoms in the sun, and generate the solar absorption spectrum. As was discussed in Section 1.3.1, when an atom absorbs radiant energy, it is excited to a new energy level. The excited atom then makes a transition to a lower excited state or to the ground state, during which a quantum of energy is emitted. Consequently, the emission spectrum of the chromosphere is formed. Since the absorption spectrum is produced by the initial transition of atoms from a low-energy to a high-energy state, while the emission spectrum results from the subsequent transition of the same atoms in the reverse direction, it is clear that the lines in the sun's emission spectrum are the same as those in its absorption spectrum. When the photosphere is eclipsed by the moon or by instrument, line emissions, mostly from hydrogen, helium, and calcium, can be observed. Because a bright line emission spectrum flashes into view briefly at the beginning and the end of a total eclipse, it is called the *flash spectrum*. The Hα line at 6563 Å is one of the strongest absorption lines in the solar spectrum. Because of the large amount of energy emitted in this line, the chromosphere becomes visible and has a characteristic reddish appearance during an eclipse.

Above the chromosphere lies the region of the solar atmosphere called the *corona*. The corona layer extends out from the edge of the solar disk many millions of kilometers. It is visible as a faint white halo during total eclipses. Figure 2.3 illustrates the solar corona during the total eclipses of March 1970 and July 1991. It is generally believed that the corona has no outer boundary. A stream of gas called *solar wind* (see Table 2.1) flows out of the corona and into the solar system continuously. An instrument called a *coronagraph* has been used in the past to study both the chromosphere and the corona in the absence of a natural eclipse. Strong emission lines of hydrogen and helium originating within the chromosphere disappear with increasing altitude and are replaced by the continuous spectrum of white light characteristic of the corona. The spectrum of the corona contains a number of weak emission lines, of which the most intense is the green line of ionized iron. The generation of this emission line requires an enormous amount of energy, and it is believed that the temperature in large regions of the corona is close to 10^6 K.

2.1.2 Solar Surface Activity: Sunspots

Several observable features of the sun are particularly interesting and important because of their transient nature. The best known and most frequently observed of these variable features are *sunspots,* which are relatively dark regions on the photosphere—the surface of the sun. Sunspots have an average size of about 10,000 km but range from barely visible to areas that cover more than 150,000 km on the sun's surface. The spots usually occur in pairs, or in complex groups, which follow a leader spot in the direction of the sun's rotation. Small sunspots persist for several days or a week, while the largest spots may last for several weeks, long enough for these spots to
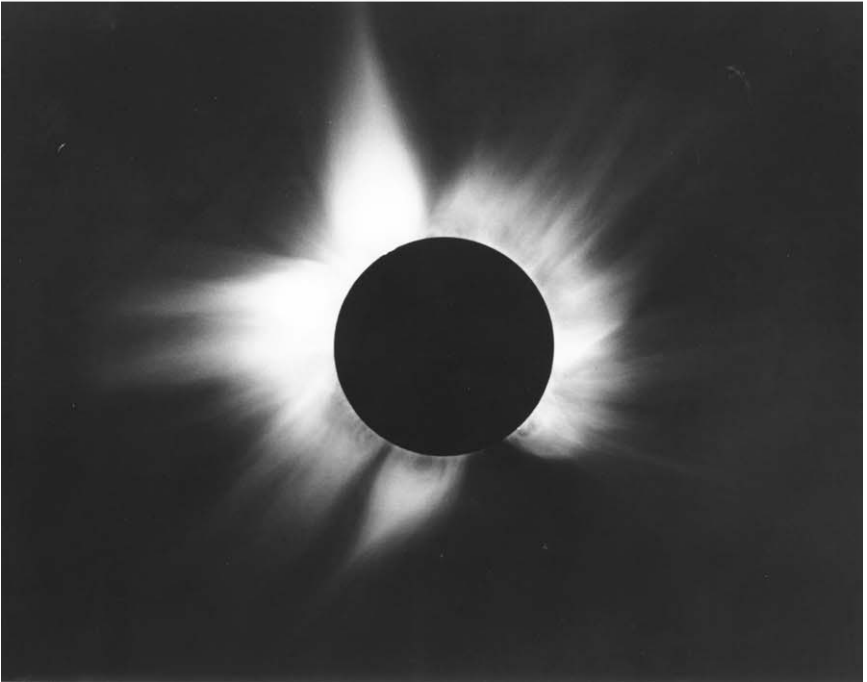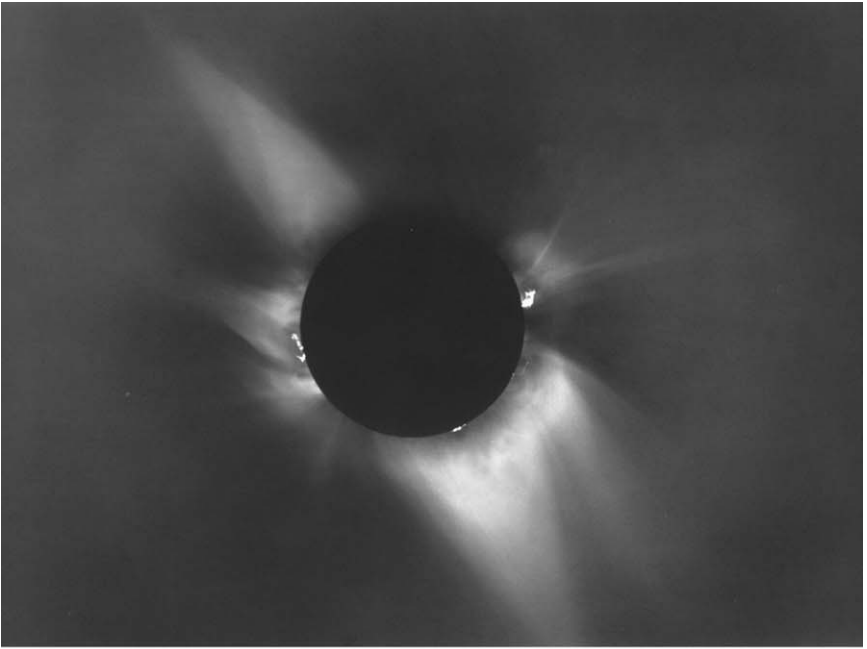
Figure 2.3 The top picture is the total solar eclipse of July 11, 1991, photographed from Mauna Loa, Hawaii. The bottom picture is the solar corona during the total eclipse of March 7, 1970. Features are visible at a distance of about 4.5 solar radii or 3 million kilometers (courtesy of Rhodes College and High Altitude Observatory, Boulder, Colorado).

reappear during the course of the sun's 27-day rotation. Sunspots are almost entirely confined to the zone of latitude between 40° and the equator, and they never appear near the poles. Just after their minimum occurrence, spots first appear near latitude 27° in both hemispheres. As the cycle proceeds, they drift toward the equator and disappear close to latitude 8°.

Sunspots are cooler regions with an average temperature of about 4000 K, compared to an average temperature of 6000 K for the photosphere. Because of their relatively low temperature, sunspots appear black. Sunspot activity has been observed with the aid of high-power telescopes. The number of sunspots that appear on the solar disk averaged over a period of time is highly variable. There are periods of time when spots are relatively numerous, whereas a few years later spots occur hardly at all. These periods are called *sunspot maxima* and *sunspot minima,* respectively. The periodic change in the sunspot number is referred to as the *sunspot cycle*. For more than 200 years, the number of spots appearing every day and the position of these spots on the face of the sun have been recorded continuously. The average length of time between sunspot maxima is about 11 years; the so-called *11-year cycle*. Figure 2.4
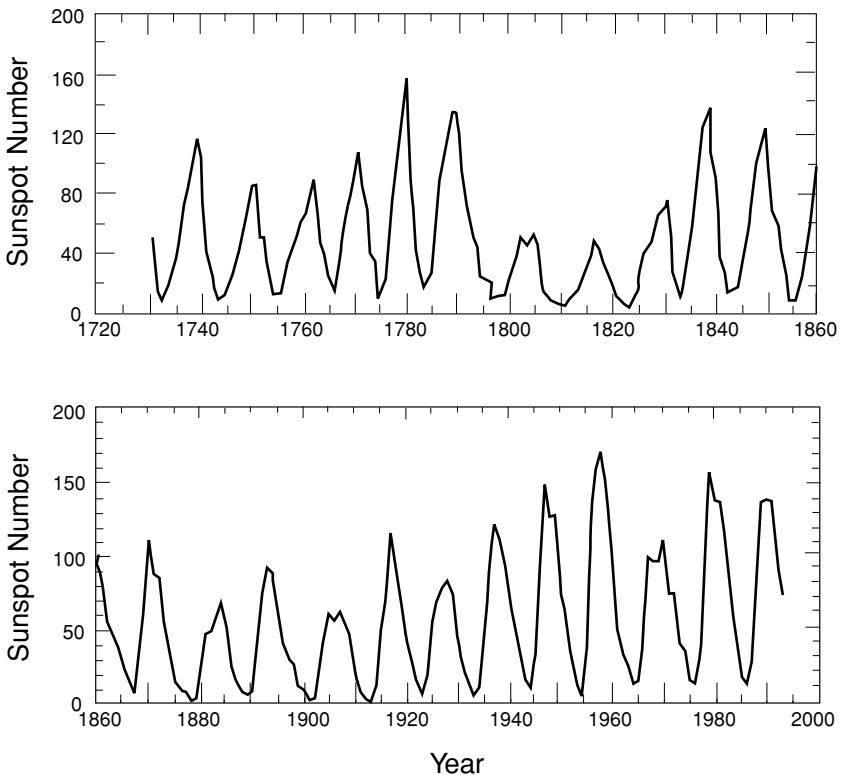
Figure 2.4   Variability of the sunspot number as function of year. Results from 1730 to 1870 are taken from Eddy (1977); those after 1860 are from Lean and Rind (1998).

depicts the variation in the number of sunspots since about 1730. The number ranges from a few to a maximum of about 150. In the years of sunspot maxima, the sun's surface is violently disturbed, and outbursts of particles and radiation are commonly observed. During sunspot minima periods, however, outbursts are much less frequent.

It is believed that sunspots are associated with the very strong magnetic fields that exist in their interiors. Magnetic field measurements show that pairs of sunspots often have opposite magnetic polarities. For a given sunspot cycle, the polarity of the leader spot is always the same for a given hemisphere. With each new sunspot cycle, the polarities reverse. The cycle of sunspot maxima having the same polarity is referred to as the *22-year cycle*. Sunspot activity has been found to have a profound influence on many geophysical phenomena and atmospheric processes.

Sunspots are not the only source of solar variability. When a sunspot is near to the limb, it can be seen to be surrounded by networks of enhanced emission, called *faculae,* which appear before and disappear after the sunspots. Although faculae have less magnetic flux than sunspots, they extend over considerably more of the sun's disk and persist longer. When the sun is viewed by monochromatic light from a single element such as the hydrogen Hα line, the sunspots are visible but are surrounded by bright areas, known as *plages*. These prominent outburst features also occur at high latitudes, but are usually observed in the vicinity of large, complex sunspot groups. They are known as *solar flares* and are associated with great increases of hydrogen Lyman α at 1216 Å and other ultraviolet radiation. The burst of radiation and energy particles from a large flare may produce interference with radio communications and cause substantial variations in the earth's magnetic field.

The sun's other transient features are the *prominences* produced by photospheric eruptions. They extend into the chromosphere and can be observed on the limb of the sun. A typical prominence may be 30,000 km high and 200,000 km long with a temperature of about 5000 K. Because prominences are cooler than the photosphere, they may be seen in Hα light as dark filaments on a bright background. Disturbances in the corona are closely related to the sunspot cycle and changes in sunspot number, based on solar radio emission observations. Outbursts are accompanied by large increases in the far-ultraviolet and x-ray emission from the sun.

All of the preceding variabilities are associated with magnetic activity. Variations in the magnetic field are produced by the interactions among the convective motion, the solar rotation, and the general magnetic field of the sun. Sunspots contain the strongest magnetic fields. As noted earlier, the polarity in sunspot pairs reverses in successive sunspot cycles.

2.2   The Earth's Orbit about the Sun and Solar Insolation

2.2.1   Orbital Geometry

The earth is one of the nine planets in the solar system. The four planets closest to the sun (i.e., Mercury, Venus, Earth, and Mars), are referred to as the terrestrial

Figure 2.5   The earth–sun geometry. *P* denotes the perihelion, *A* the aphelion, *AE* the autumnal equinox, *VE* the vernal equinox, *WS* the winter solstice, and *SS* the summer solstice, **n** is normal to the ecliptic plane, **a** is parallel to the earth's axis, δ is the declination of the sun, ϵ the oblique angle of the earth's axis, ω the longitude of the perihelion relative to the vernal equinox, ν the true anomaly of the earth at a given time, λ the true longitude of the earth, *O* the center of the ellipse, *OA* (or *OP* $= a$) the semimajor axis, *OB* ($= b$) the semiminor axis, *S* the position of the sun, *E* the position of the earth, and *ES* ($= r$) the distance between the earth and the sun.

planets, and the remaining planets (i.e., Jupiter, Saturn, Uranus, Neptune, and Pluto), are called the major planets. All of the planets revolve around the sun in the same direction, and with the exception of Uranus, they also rotate in the same direction about their axes. In addition, all the planetary orbits except Mercury and Pluto lie in almost the same plane.

Once every 24 hours the earth makes a steady rotation eastward around the axis of its poles. This rotation is the cause of the most obvious of all time periods, involving the alternation of day and night, which comes about as the sun shines on the different parts of the earth exposed to it. At the same time, the orbital motion of the earth, with a mass of 6×10^{24} kg, in an ellipse with the sun at one focus, takes approximately 365 days to complete. The distance between the earth and the sun varies. The earth's orbit about the sun and the earth's rotation about its axis, which is tilted as shown in Fig. 2.5, are the most important factors determining the amount of solar radiant energy reaching the earth, and hence, the climate and climatic changes of the earth–atmosphere system. Because of the rotation of the earth about its axis, the earth assumes the shape of an oblate spheroid, having equatorial and polar radii of 6378.17 and 6356.79 km, respectively.

The position of the sun is defined by the *solar zenith angle,* which is determined from other known angles. In reference to Fig. 2.6 let *P* be the point of observation and *OZ* the zenith through this point. Assume that the sun is in the direction of *OS* or *PS* and let *D* be the point directly under the sun. Then the plane of *OZ* and *OS* will intersect the surface of the earth in a great circle. The angle *ZOS,* measured by the arc *PD* of this circle, is equal to the sun's zenith distance θ_0. In the spherical triangle *NPD*, the arc *ND* is equal to 90° minus the solar inclination δ, which is the angular distance

Figure 2.6   Relationship of the solar zenith angle θ_0 to the latitude φ, the solar inclination angle δ, and the hour angle h. P and D are the point of observation and the point directly under the sun, respectively (see text for further explanation).

of the sun north (positive) or south (negative) of the equator. The arc NP is equal to $90°$ minus the latitude ϕ of the observation point. The angle h is the hour angle, or the angle through which the earth must turn to bring the meridian of P directly under the sun, and is given by $h = 2\pi t/\Delta t_\odot$, where $\Delta t_\odot = 86{,}400$ sec. From spherical trigonometry, the cosine of the solar zenith angle as given in Fig. 2.6 and Appendix C is

$$\cos \theta_0 = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos h. \tag{2.2.1}$$

At the poles, $\varphi = \pm\pi/2$, so that $\cos \theta_0 = \sin \delta$ or $\theta_0 = \pi/2 - \delta$. The solar elevation angle $(\pi/2 - \theta_0)$ is equal to the declination of the sun. During the 6 months of daylight, the sun simply circles around the horizon and never rises more than about $23.5°$. At solar noon at any latitude, the hour angle $h = 0$. Thus, $\cos \theta_0 = \cos(\varphi - \delta)$ or $\theta_0 = \varphi - \delta$. Except at the poles, $\theta_0 = \pi/2$ at sunrise and sunset. From Eq. (2.2.1), the half-day, H (i.e., from sunrise to noon or from noon to sunset), is defined by

$$\cos H = -\tan \varphi \tan \delta. \tag{2.2.2}$$

If $\varphi = 0$ (equator) or $\delta = 0$ (equinoxes), then $\cos H = 0$. The length of the solar day is 12 hours. The latitude of the polar night $(H = 0)$ is determined by $\varphi = 90° - |\delta|$ in the winter hemisphere. The solar inclination is a function of the day of the year only and is independent of the location of the observation point. It varies from $23°27'$ on June 21 to $-23°27'$ on December 22. The hour angle is zero at solar noon and increases by $15°$ for every hour before or after solar noon.

The plane of the earth's orbit is called the *plane of the ecliptic,* as shown in Fig. 2.5. The position of the sun is located at one of the foci of an ellipse. Let a and b denote the semimajor and semiminor axes of the ellipse. The eccentricity of the earth's orbit

is defined as the ratio of the distance between the two foci to the major axis of the ellipse and is given by

$$e = (a^2 - b^2)^{1/2}/a. \tag{2.2.3}$$

The tilt of the earth's axis with respect to normal to the ecliptic plane is defined by the oblique angle ε. The longitude of the perihelion, i.e., the closest point of the earth to the sun, relative to the vernal equinox, is defined by the angle ω. For a given time, the position of the earth is defined by the true anomaly v in reference to the perihelion.

From the geometry shown in Fig. 2.5, the declination of the sun can be expressed in terms of the oblique angle of the earth's axis, the longitude of the perihelion relative to the vernal equinox, and the true anomaly of the earth at a given time. Also, letting the true longitude of the earth, counted counterclockwise from the vernal equinox, be λ, from three-dimensional geometry one can derive the following relationships:

$$\sin \delta = \sin \varepsilon \sin(v + \omega) = \sin \varepsilon \sin \lambda. \tag{2.2.4}$$

Having defined the relevant geometric parameters, we shall now introduce the three basic laws governing the motion of the planet. The planet's orbital ellipse can be described by Kepler's first law (the law of orbits), in which the distance is related to the true anomaly and eccentricity in the form

$$r = \frac{a(1 - e^2)}{1 + e \cos v}. \tag{2.2.5}$$

This equation can be derived from the geometry of an ellipse (Exercise 2.2). To compute the solar flux over a certain time period, Kepler's second law (the law of areas) is required. This law is a statement of the conservation of angular momentum that the radius vector, drawn from the sun to the planet, sweeps out equal areas in equal times (Exercise 2.3). Letting $\widetilde{T}$ denote the tropical year (i.e., the time between successive arrivals of the sun at the vernal equinox), which is equal to 365.2422 mean solar days, and noting that the area of an ellipse is πab, we have

$$r^2 \frac{dv}{dt} = \frac{2\pi a^2}{\widetilde{T}}(1 - e^2)^{1/2}. \tag{2.2.6}$$

In order to compute solar insolation, which will be defined in Subsection 2.2.3, the mean distance between the earth and the sun must be defined. From Kepler's second law, we may define a mean distance based on the conservation of angular momentum such that

$$r_0^2 = \frac{1}{2\pi} \int_0^{2\pi} r^2 dv = a^2(1 - e^2)^{1/2} \cong a^2. \tag{2.2.7}$$

Kepler's third law (the law of periods) states that

$$a^3/\widetilde{T}^2 = k, \tag{2.2.8}$$

where $\widetilde{T}$ is the planet's period of revolution around the sun, and k has the same value for all planets. This law is a consequence of the balance between the gravitational

and centrifugal forces governing a planet in orbit. The semimajor axis of the earth's orbit is invariant. Because the factor $(1 - e^2)^{1/2}$ is very close to 1, the mean distance between the earth and the sun may be set as the invariant semimajor axis (i.e., $r_0 = a$).

Solar insolation as a function of latitude and the time of year, which will be defined in Eq. (2.2.21), requires the values of $(a/r)^2$ and δ, which in turn are computed from the eccentricity e, the oblique angle ε, and the longitude of the perihelion relative to the vernal equinox ω. Based on celestial mechanics, the secular variations in these three parameters are associated with the perturbations that other principal planets exert on the earth's orbit. Milankovitch (1941) has provided mathematical expressions for the computation of solar insolation including orbital parameters. Berger (1978) developed simplified trigonometric expansions for the efficient computation of the aforementioned three parameters. Figure 2.7 shows their values for the past 200,000 years. The eccentricity varies from about 0.01 to 0.04 with a mean value of about 0.017 and has a characteristic period of about 100,000 years. The oblique angle varies
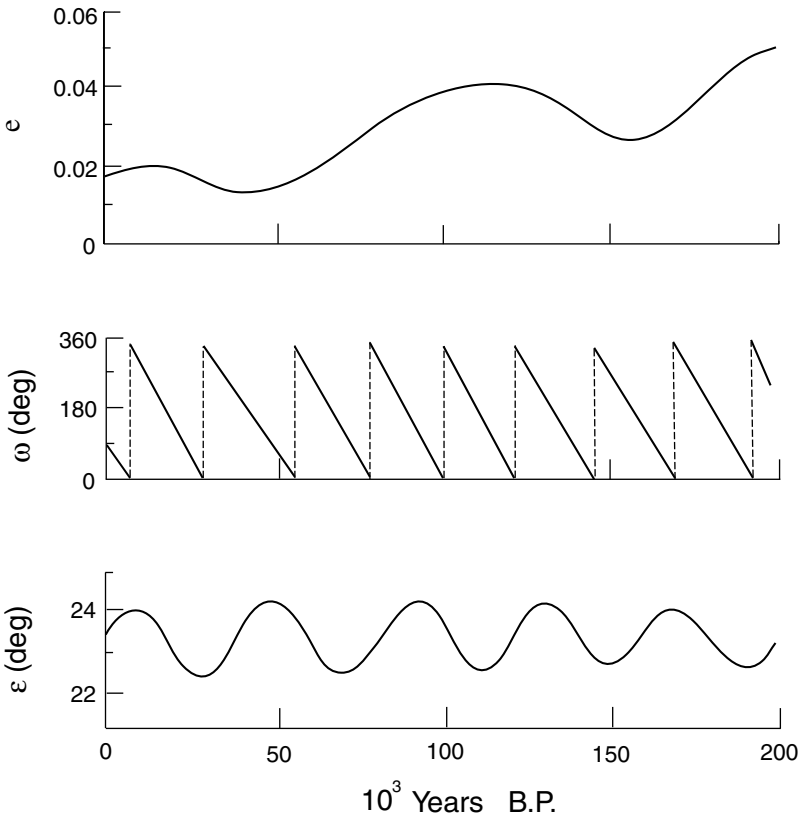

Figure 2.7   The eccentricity e, the obliquity of the ecliptic ε, and the longitude of the perihelion ω of the earth as functions of year before the present.

Table 2.2

Coefficients for the Calculation of the Sun–Earth Distance and the Declination Angle

n	a_n	b_n	c_n	d_n
0	1.000110	0	0.006918	0
1	0.034221	0.001280	−0.399912	0.070257
2	0.000719	0.000077	−0.006758	0.000907
3	—	—	−0.002697	0.000148

from about 22° to 24.5° with a dominant period of about 41,000 years. The longitude of the perihelion has a periodicity of about 21,000 years due to the advance of the perihelion by about 25 minutes each year, referred to as the periodic precession index.

The sun–earth distance can be approximated with an accuracy of about 10^{-4} as follows:

$$\left(\frac{a}{r}\right)^2 = \sum_{n=0}^{2}(a_n \cos nt + b_n \sin nt), \qquad (2.2.9)$$

where $t = 2\pi d/365$, with $d = 0$ for January 1 and $d = 364$ for December 31. Also, the declination angle can be evaluated from

$$\delta = \sum_{n=0}^{3}(c_n \cos nt + d_n \sin nt), \qquad (2.2.10)$$

with an accuracy of 0.0006 radians. The coefficients a_n, b_n, c_n, and d_n are listed in Table 2.2.

The most distinguishable feature of climatic change is the seasons. The revolution of the earth about the sun and the tilt of the earth's axis cause this seasonal variation. At the time of the summer solstice, which occurs about June 22, the sun appears directly overhead at noon at latitude 23.5°N, called the *Tropic of Cancer*. The elevation of the sun above the horizon and the length of the day reach their maximum values in the northern hemisphere at the summer solstice, and everywhere north of the Arctic Circle (latitude 66.5°N), the sun remains above the horizon all day. In the southern hemisphere on the June solstice, the sun's elevation is at a minimum, the days are shortest, and everywhere south of the Antarctic Circle (latitude 66.5°S), the sun does not rise above the horizon. This is the beginning of the northern hemisphere summer, whereas the southern hemisphere summer begins with the winter solstice on about December 22. Having reached the southernmost point in its annual migration, the sun then stands directly overhead at noon at latitude 23.5°S, called the *Tropic of Capricorn*. Both the elevation of the sun above the horizon and the length of the day are then at their minimum values in the northern hemisphere and their maximum values in the southern hemisphere, and the sun does not rise within the Arctic Circle or set within the Antarctic Circle. At the vernal (spring) and autumnal equinoxes, the days and nights everywhere are equal (12 hours), and the sun appears directly overhead on the equator at noon. The sun crosses the equator from north to south at

the autumnal equinox, and from south to north at the vernal equinox. The distances between the centers of the sun and the earth vary between the extreme values of 147×10^6 km at about winter solstice, and 153×10^6 km at about summer solstice. The mean distance is about 150×10^6 km, as noted earlier in Section 2.1.

2.2.2   Definition of the Solar Constant

The distribution of electromagnetic radiation emitted by the sun as a function of the wavelength incident on the top of the earth's atmosphere is called the *solar spectrum*. The *solar constant S* is a quantity denoting the amount of total solar energy (i.e., covering the entire solar spectrum) reaching the top of the atmosphere. It is defined as the flux of solar energy (energy per unit time) across a surface of unit area normal to the solar beam at the mean distance between the sun and the earth.

The sun emits energy at the rate of 6.2×10^7 W m^{-2}. On the basis of the energy conservation principle and if there is no intervening medium present, the energy emitted from the sun must remain the same at some distance away. Thus,

$$F_\odot^* 4\pi a_\odot^2 = S 4\pi r_0^2, \qquad (2.2.11)$$

where $F_\odot^*$ denotes the solar emittance, $a_\odot$ the radius of the sun, and r_0 the mean distance between the sun and the earth. Hence, the solar constant may be expressed by

$$S = F_\odot^* (a_\odot/r_0)^2. \qquad (2.2.12a)$$

Since the sun may be considered as an isotropic emitter, the solar intensity (or brightness) I is then given by $F_\odot^*/\pi$ [see Eq. (1.1.10)]. Thus, Eq. (2.2.12a) can be rewritten in the form

$$S = I \cdot \pi a_\odot^2/r_0^2 = I \cdot \Omega, \qquad (2.2.12b)$$

where the solid angle Ω is the angle from which the earth sees the sun. The intensity, i.e., the energy contained within the solid angle, is invariant: it is the same at the position of the sun as it is at the position of the earth. If the intensity varies within the solid angle, then we must use Eq. (1.1.9) to obtain the flux density. A number of exercises at the end of the chapter require the use of the intensity concept.

The total energy intercepted by the earth whose radius is a_e is given by $S\pi a_e^2$. If this energy is spread uniformly over the full surface of the earth, then the amount received per unit area and unit time at the top of the atmosphere is given by

$$\bar{Q}_s = S\pi a_e^2/(4\pi a_e^2) = S/4. \qquad (2.2.13)$$

To estimate the equilibrium temperature $T_\odot$ of the sun, we use the blackbody assumption. From the Stefan–Boltzmann law, i.e., $F_\odot^* = \sigma T_\odot^4$, we find

$$T_\odot^4 = (r_0/a_\odot)^2(S/\sigma). \qquad (2.2.14)$$

Inserting values of S, σ, r_0, and $a_\odot$ into Eq. (2.2.14), we obtain an equilibrium temperature of about 5800 K for the sun. Thus, once the solar constant is measured, the effective temperature of the sun can be computed.

2.2.3   Distribution of Solar Insolation

Solar insolation is defined as the flux of solar radiation per unit of horizontal area for a given locality. It depends primarily on the solar zenith angle and to some extent on the variable distance of the earth from the sun. The flux density at the top of the atmosphere may be expressed by

$$F = F_\odot \cos \theta_0, \tag{2.2.15}$$

where $F_\odot$ represents the solar flux density at the top of the atmosphere when the instantaneous distance between the earth and sun is r, and θ_0 denotes the solar zenith angle. The definition of the solar constant is $S = F_\odot (r/r_0)^2$. Thus, we have

$$F(t) = S(r_0/r)^2 \cos \theta_0, \tag{2.2.16}$$

where S is the solar constant corresponding to the mean earth–sun distance r_0 defined previously. If we define the solar heating received at the top of the atmosphere per unit area as Q, then the solar flux density may be written as

$$F = \frac{dQ}{dt}. \tag{2.2.17}$$

Thus, the insolation for a given period of time is

$$Q = \int_t F(t) \, dt. \tag{2.2.18}$$

The total solar energy received by a unit of area per one day may be calculated by integrating total insolation over the daylight hours. Upon substituting Eq. (2.2.16) into (2.2.18) and noting that variation of the distance r in one day can be neglected, we can define the daily insolation as follows:

$$Q \cong S \left(\frac{r_0}{r}\right)^2 \int_{sunrise}^{sunset} \cos \theta_0(t) \, dt. \tag{2.2.19}$$

Inserting Eq. (2.2.1) into Eq. (2.2.19) and denoting the angular velocity of the earth ω by $dh/dt \, (= 2\pi \text{ rad/day})$, we obtain

$$Q = S \left(\frac{r_0}{r}\right)^2 \int_{-H}^{H} (\sin \varphi \sin \delta + \cos \varphi \cos \delta \cos h) \frac{dh}{\omega}, \tag{2.2.20}$$

where H represents a half-day, i.e., from sunrise or sunset to solar noon. After performing this simple integration, daily solar insolation is given by

$$Q \cong \frac{S}{\pi} \left(\frac{a}{r}\right)^2 (\sin \varphi \sin \delta H + \cos \varphi \cos \delta \sin H), \tag{2.2.21}$$

where we have set $r_0 = a$ as shown in Eq. (2.2.7), and H in the first term on the right-hand side is expressed in units of radians ($180° = \pi$ rad). Note that the factor $(a/r)^2$ never departs from unity by more than 3.5%. It ranges from 1.0344 on January 3 to 0.9674 on July 5. The computations of (a/r) and δ have been discussed previously.
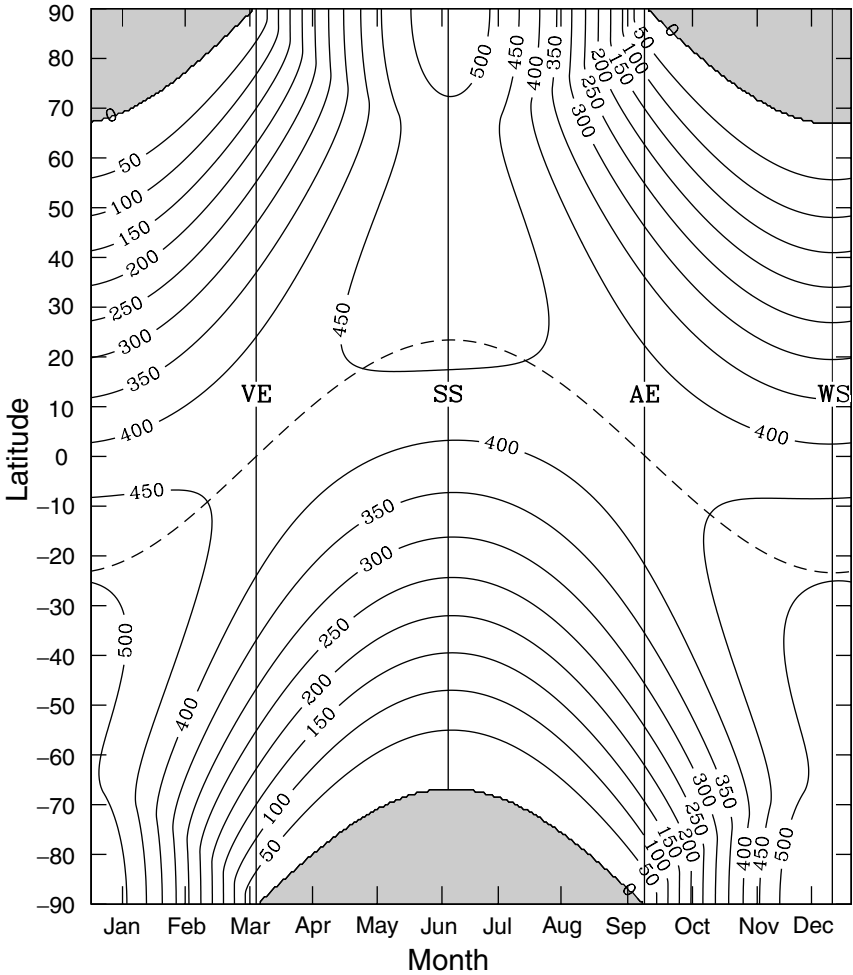
Figure 2.8   Daily mean solar insolation ($Q/24$ hr) as a function of latitude and day of year in units of W m^{-2} based on a solar constant of 1366 W m^{-2}. The shaded areas denote zero insolation. The position of vernal equinox (VE), summer solstice (SS), autumnal equinox (AE), and winter solstice (WS) are indicated with solid vertical lines. Solar declination is shown with a dashed line.

Daily solar insolation is shown in Fig. 2.8. The distribution of solar insolation is independent of longitude and is slightly asymmetric between the northern and southern hemispheres. The sun is closest to the earth in January (winter in the northern hemisphere), so that the maximum solar insolation received in the southern hemisphere is greater than that received in the northern hemisphere. At the equinoxes, solar insolation is at a maximum at the equator and is zero at the poles. At the summer solstice of the northern hemisphere, daily insolation reaches a maximum at the North Pole because of the 24-hour-long solar day. At the winter solstice, the

sun does not rise above the horizon north of about 66.5°, where solar insolation is zero.

The calculation of seasonal and annual insolations is rather involved and will not be detailed here. However, in the domain of the true longitude of the earth as shown in Fig. 2.5, the year can be divided into astronomical spring $(0 - \pi/2)$, summer $(\pi/2 - \pi)$, autumn $(\pi - 3\pi/2)$, and winter $(3\pi/2 - 2\pi)$. Seasonal solar insolation can be evaluated using these divisions. Because of the cosine property of the true longitude of the earth, λ, solar insolation is the same for spring and summer, and for autumn and winter. Consequently, it suffices to consider the summer half-year (spring plus summer) and winter half-year (autumn plus winter) for the calculation of seasonal solar insolation. Considering the total solar insolation for the winter $(\pi, 2\pi)$ half-year and expressing this insolation in terms of that for the summer $(0, \pi)$ half-year, and performing a lengthy algebraic analysis, we find

$$Q_{s,w} = \frac{S\widetilde{T}}{2\pi(1-e^2)^{1/2}}[\widetilde{S}(\varphi, \varepsilon) \pm \sin \varphi \sin \varepsilon], \qquad (2.2.22)$$

where the insolation function is defined by

$$\widetilde{S}(\varphi, \varepsilon) = \frac{\sin \varphi \sin \varepsilon}{2\pi} \int_0^{2\pi} (H - \tan H) \sin \lambda \, d\lambda, \qquad (2.2.23)$$

and the positive and negative signs are for the summer and winter solar insolations, respectively (see Exercise 2.15 for the reason for this difference). The half-day defined in Eq. (2.2.2) is given by

$$\cos H = -\frac{\tan \varphi \sin \varepsilon \sin \lambda}{(1 - \sin^2 \varepsilon \sin^2 \lambda)^{1/2}}, \qquad (2.2.24)$$

where the definition of the declination angle defined in Eq. (2.2.4) is used. The annual insolation for a given latitude is the sum of the summer and winter insolations and is given by

$$Q_a = \frac{S\widetilde{T}\widetilde{S}(\varphi, \varepsilon)}{\pi(1-e^2)^{1/2}}. \qquad (2.2.25)$$

Because the insolation function is the same for the northern and southern hemispheres, i.e., $\widetilde{S}(\varphi, \varepsilon) = \widetilde{S}(-\varphi, -\varepsilon)$, annual solar insolation is the same for corresponding latitudes in each hemisphere.

Finally, annual global solar insolation can be evaluated by using the instantaneous solar insolation for the entire earth, which is given by $S(a/r)^2\pi a_e^2$. Distributing this energy over the surface area of the earth, $4\pi a_e^2$, the mean solar insolation for one day is given by $\Delta t_\odot(a/r)^2 S/4$. We can perform an integration over a year via Kepler's second law to obtain

$$Q_{at} = \int_0^{\widetilde{T}} \frac{S\Delta t_\odot}{4}\left(\frac{a}{r}\right)^2 \frac{dt}{\Delta t_\odot} = \frac{S}{4}\widetilde{T}(1-e^2)^{-1/2} \cong \frac{S}{4}\widetilde{T}(1 + e^2/2). \qquad (2.2.26)$$

The annual global insolation is proportional to $(1 + e^2/2)$, but is independent of the declination of the sun δ and the true anomaly ν.

2.3   Solar Spectrum and Solar Constant Determination

2.3.1   Solar Spectrum

The solar spectrum covers wavelengths ranging from gamma rays to radio waves, as shown in Fig. 1.1. Because of the nonquantized electronic transitions, most solar energy is carried by the continuum, i.e., radiation is continuous rather than selective. The single most important contributor is hydrogen, both in its neutral state and as negative ions. A radiation transition from one level to another is characterized by an absorption or an emission line whose frequency is governed by Planck's relation. However, in the ionization process the atom (or molecules) may absorb more than the minimum energy required to remove the electron. This additional energy may be thought of as supplying kinetic energy to the freed electron and is not quantized. As a consequence, absorption is not selective but rather continuous. The ionization continuum occurs on the high-frequency (shorter wavelength) side of the ionization frequency. Neutral hydrogen has ionization continua associated with lines, some of which were defined in Fig. 1.9. Metallic atoms also contribute to the continuum in the ultraviolet spectrum. The continuum absorption in the visible and infrared spectrum, however, is produced by negative hydrogen ions.

Electromagnetic radiation emerging from within the sun is continuously emitted and absorbed by atoms. As shown in Fig. 2.2, the radiative temperature first drops off to a minimum value of about 4500 K just above the photosphere, and then levels off and slowly rises in the chromosphere, followed by a rapid rise in the transition region to several million degrees in the corona. At each temperature, probabilities of the electronic transition exist that any atom will achieve a particular excited state, leading to the formation of absorption lines at different levels in the solar atmosphere. The core of a line forms at the temperature where the maximum transition probabilities of an electron moving from one orbital level to another occur (see Fig. 1.8). The wings of a line form at different temperature levels because of the required transition probabilities. Each absorption line has a preferred formation region in the solar atmosphere. Those lines that absorb very little radiation are known as weak lines, which can form in narrow layers of the solar atmosphere. Some of the absorption lines in the solar atmosphere were displayed in Fig. 2.2.

In view of the preceding discussion, the solar spectrum consists of a continuous emission with a superimposed line structure. The visible and infrared spectrum of the photosphere shows absorption lines, known as the *Fraunhofer spectrum*. The strongest of these lines are produced by H, Mg, Fe, Ca, and Si, as well as singly ionized Ca and Mg. Most of the lines shorter than 1850 Å produced from the photosphere exhibit in emission. Light from the chromosphere and the corona has emission lines at all observed wavelengths.
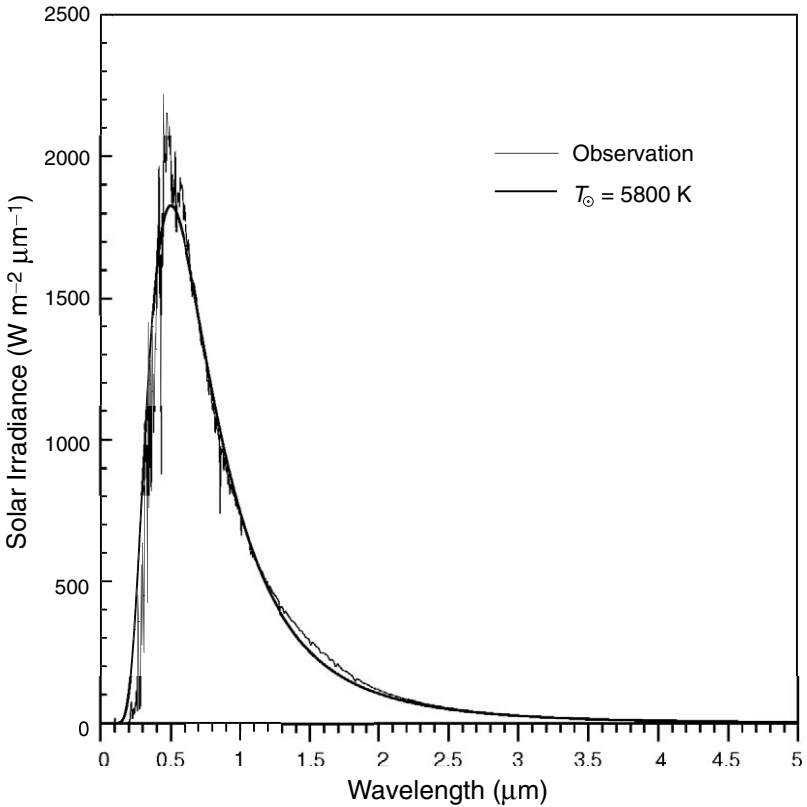
Figure 2.9   Solar irradiance for a 50 cm^{-1} spectral interval at the top of the atmosphere based on the results presented in the MODTRAN 3.7 program. Also shown is the Planck flux with a temperature of 5800 K accounting for the mean distance between the earth and the sun.

Figure 2.9 shows the spectral solar irradiance observation at the top of the atmosphere averaged over a 50 cm^{-1} spectral interval as a function of wavelength up to 5 μm, based on the results presented in the MODTRAN 3.7 program (Anderson *et al.,* 1995). Although the total solar irradiance derived from this program is 1373 W m^{-2}, the spectral solar irradiance curve presented here is scaled with respect to the recently proposed solar constant of 1366 W m^{-2} (see Section 2.3.3 for further discussion). A 50 cm^{-1} spectral average has been performed to smooth out the rapid fluctuations produced by the absorption/emission line structure. However, some variabilities can still be seen, particularly in the ultraviolet spectrum. Also shown is the Planck curve with an emitting temperature of 5800 K, taking into account the mean distance between the sun and the earth. This temperature appears to fit closely with the visible and infrared spectrum characteristic of radiation emitted from the photosphere. For atmospheric applications, it is critically important to have reliable spectral solar irradiances for use in radiative transfer models. Table 2.3 gives tabulated data from 0.2 to 5 μm with

Table 2.3

Distribution of Solar Spectral Irradiance S_λ from 0.2 to 100 μm in Terms of the Accumulated Energy and Percentage Based on the Values Listed in the MODTRAN 3.7 Program[a]

λ (μm)	S_λ (W m^{-2} μm^{-1})	$S_{0-\lambda}$ (W m^{-2})	$S_{0-\lambda}$ (%)	λ (μm)	S_λ (W m^{-2} μm^{-1})	$S_{0-\lambda}$ (W m^{-2})	$S_{0-\lambda}$ (%)
0.20	2.0832E+01	2.08317E+00	0.15250	3.8	1.0564E+01	1.35239E+03	99.00390
0.30	5.4765E+02	5.68479E+01	4.16163	3.9	9.6162E+00	1.35335E+03	99.07430
0.40	1.4042E+03	1.97272E+02	14.44155	4.0	8.6980E+00	1.35422E+03	99.13797
0.50	1.9619E+03	3.93464E+02	28.80410	4.1	7.9180E+00	1.35502E+03	99.19593
0.60	1.7632E+03	5.69780E+02	41.71153	4.2	7.2072E+00	1.35574E+03	99.24870
0.70	1.4300E+03	7.12778E+02	52.17994	4.3	6.5062E+00	1.35639E+03	99.29633
0.80	1.1257E+03	8.25347E+02	60.42075	4.4	5.7954E+00	1.35697E+03	99.33875
0.90	8.8835E+02	9.14182E+02	66.92404	4.5	5.2622E+00	1.35749E+03	99.37727
1.00	7.2943E+02	9.87125E+02	72.26392	4.6	4.8180E+00	1.35798E+03	99.41255
1.10	5.8743E+02	1.04587E+03	76.56425	4.7	4.4724E+00	1.35842E+03	99.44529
1.20	4.8921E+02	1.09479E+03	80.14558	4.8	4.1565E+00	1.35884E+03	99.47573
1.30	4.0851E+02	1.13564E+03	83.13614	4.9	3.8504E+00	1.35922E+03	99.50391
1.40	3.4450E+02	1.17009E+03	85.65813	5.0	3.5740E+00	1.35958E+03	99.53008
1.50	2.9066E+02	1.19916E+03	87.78592	6.0	1.8385E+00	1.36303E+03	99.78240
1.60	2.4644E+02	1.22380E+03	89.58999	7.0	1.0108E+00	1.36404E+03	99.85639
1.70	2.0453E+02	1.24425E+03	91.08726	8.0	5.9672E−01	1.36464E+03	99.90007
1.80	1.6829E+02	1.26108E+03	92.31927	9.0	3.7458E−01	1.36501E+03	99.92751
1.90	1.3725E+02	1.27481E+03	93.32404	10.0	2.4702E−01	1.36526E+03	99.94559
2.00	1.1624E+02	1.28643E+03	94.17501	11.0	1.6932E−01	1.36543E+03	99.95798
2.10	9.7416E+01	1.29617E+03	94.88816	12.0	1.2005E−01	1.36555E+03	99.96677
2.20	8.2132E+01	1.30439E+03	95.48942	13.0	8.7276E−02	1.36563E+03	99.97315
2.30	6.9594E+01	1.31134E+03	95.99889	14.0	6.5062E−02	1.36570E+03	99.97792
2.40	5.9198E+01	1.31726E+03	96.43226	15.0	4.9463E−02	1.36575E+03	99.98154
2.50	5.1023E+01	1.32237E+03	96.80577	16.0	3.8307E−02	1.36579E+03	99.98434
2.60	4.4280E+01	1.32679E+03	97.12994	17.0	3.0112E−02	1.36582E+03	99.98655
2.70	3.8672E+01	1.33066E+03	97.41305	18.0	2.3991E−02	1.36584E+03	99.98831
2.80	3.3815E+01	1.33404E+03	97.66058	19.0	1.9351E−02	1.36586E+03	99.98973
2.90	2.9589E+01	1.33700E+03	97.87720	20.0	1.5797E−02	1.36588E+03	99.99088
3.00	2.6133E+01	1.33962E+03	98.06850	30.0	3.4388E−03	1.36598E+03	99.99860
3.10	2.3093E+01	1.34193E+03	98.23756	40.0	1.0465E−03	1.36599E+03	99.99937
3.20	2.0476E+01	1.34397E+03	98.38746	50.0	4.2098E−04	1.36600E+03	99.99968
3.30	1.8186E+01	1.34579E+03	98.52059	60.0	2.0151E−04	1.36600E+03	99.99983
3.40	1.6191E+01	1.34741E+03	98.63913	70.0	1.0860E−04	1.36600E+03	99.99991
3.50	1.4562E+01	1.34887E+03	98.74574	80.0	6.3779E−05	1.36600E+03	99.99995
3.60	1.3032E+01	1.35017E+03	98.84114	90.0	3.9985E−05	1.36600E+03	99.99998
3.70	1.1670E+01	1.35134E+03	98.92657	100.0	2.6459E−05	1.36600E+03	100.0000

[a] The solar constant is taken to be 1366 W m^{-2}.

a 0.1-μm spectral interval. From 5 to 100 μm, solar irradiance accounts for about 6 W m^{-2}. Based on these values, about 50% of the total solar irradiance lies in wavelengths longer than the visible, about 40% in the visible region, and about 10% in wavelengths shorter than the visible. Note that from 3.5 to 5 μm, the emitted thermal infrared radiation from the earth and the atmosphere system becomes significant.

According to solar flux observations, the ultraviolet region ($<0.4\,\mu$m) of the solar spectrum deviates greatly from the visible and infrared regions in terms of
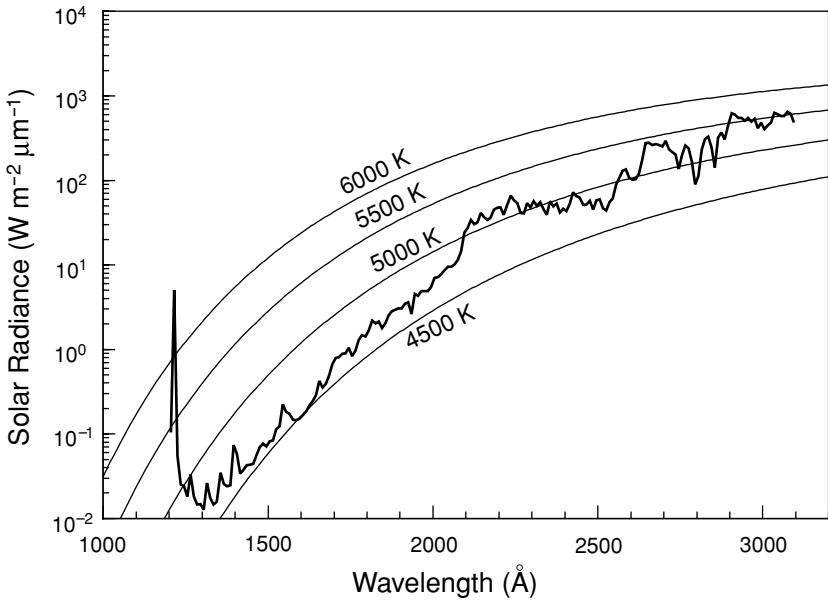
Figure 2.10 Observed irradiance outside the earth's atmosphere in the ultraviolet region (data taken from Brasseur and Simon, 1981) and comparison with the Planck curves for temperatures ranging from 4500 K to 6000 K.

the equivalent blackbody temperature of the sun. Figure 2.10 illustrates a detailed observed solar spectrum from about 1000 to 3000 Å, along with blackbody temperatures of 4500, 5000, 5500, and 6000 K. In the interval 2100–3000 Å, the equivalent blackbody temperature of the sun lies somewhat above 5000 K. It falls gradually to a minimum level of about 4700 K at about 1400 Å. From there toward shorter wavelengths, a larger amount of energy flux is observed at the Lyman α emission line of 1216 Å associated with the transition of the first excited and ground states of hydrogen atoms. The ultraviolet portion of the solar spectrum below 3000 Å contains a relatively small amount of energy. However, because the ozone and the molecular and atomic oxygen and nitrogen in the upper atmosphere absorb all this energy, it represents the prime source of the energy in the atmosphere above 10 km.

2.3.2 Determination of the Solar Constant: Ground-Based Method

For historical reasons, we shall first introduce the ground-based method for the determination of the solar constant. Ground-based observations of solar irradiance for the purpose of determining the solar constant require three primary instruments. These are the *pyrheliometer,* the *pyranometer,* and the *spectrobolometer*. The pyrheliometer is used to measure the direct, plus some diffuse, solar radiation, while the pyranometer, utilizing a suitable shield to block the direct solar radiation from striking the instrument, measures only the diffuse solar radiation for arriving at a pyrheliometer

correction. The amount of direct sunlight can then be calculated by subtracting the flux density measured by the pyranometer from that measured by the pyrheliometer. The spectrobolometer is a combination of a spectrograph and a coelostat. A coelostat is a mirror that follows the sun and focuses its rays continuously on the entrance slit of the spectrograph, which disperses the solar radiation into different wavelengths by means of a prism or diffraction grating. In the Smithsonian solar constant measurements, about 40 standard wavelengths between 0.34 and 2.5 μm are measured nearly simultaneously from the record of the spectrograph. The instrument for these measurements is called a *bologram*. There are two techniques of measuring the solar constant from the ground-based radiometer, called the *long* and *short* methods of the Smithsonian Institution. The long method is more fundamental and establishes the basis for the short method. The long method uses the Beer–Bouguer–Lambert law and is introduced in the following.

Consider an atmosphere consisting of plane-parallel layers. At a given position of the sun, which is denoted by the solar zenith angle θ_0, the effective path length of the air mass is $u \sec \theta_0$, where

$$u = \int_{z_1}^{z_\infty} \rho dz. \tag{2.3.1}$$

In this equation z_1 is the height of the station and z_∞ denotes the top of the atmosphere. On the basis of the Beer–Bouguer–Lambert law, the irradiance F of the direct solar radiation of wavelength λ observed at the surface level is given by

$$F_\lambda = F_{\lambda 0} \exp(-k_\lambda u \sec \theta_0) = F_{\lambda 0} T_\lambda^m, \tag{2.3.2}$$

where $F_{\lambda 0}$ is the monochromatic solar irradiance at the top of the atmosphere, k_λ denotes the monochromatic mass extinction cross section, T_λ is the monochromatic transmissivity defined in Eq. (1.4.10), and m $(= \sec \theta_0)$ represents the ratio of the air mass between the sun and the observer to the air mass with respect to the local zenith distance. Upon taking the logarithm, we find

$$\ln F_\lambda = \ln F_{\lambda 0} + m \ln T_\lambda. \tag{2.3.3}$$

Observations of F_λ may be made for several zenith angles during a single day. If the atmospheric properties do not change during the observation period, then the transmissivity T_λ is constant. A plot of $\ln F_\lambda$ versus m shown in Fig. 2.11 may be extrapolated to the zero point, which represents the top of the atmosphere $(m = 0)$. This is referred to as the *Langley plot*. If observations of the monochromatic irradiance are carried out for wavelengths covering the entire solar spectrum, then we have

$$F_\odot = \int_0^\infty F_{\lambda 0} d\lambda \approx \sum_{i=1}^{N} F_{\lambda_i 0} \Delta \lambda_i, \tag{2.3.4}$$

where N is the total number of the monochromatic irradiances measured. The irradiance $F_\odot$ corresponds to the actual distance between the earth and the sun, r. By using
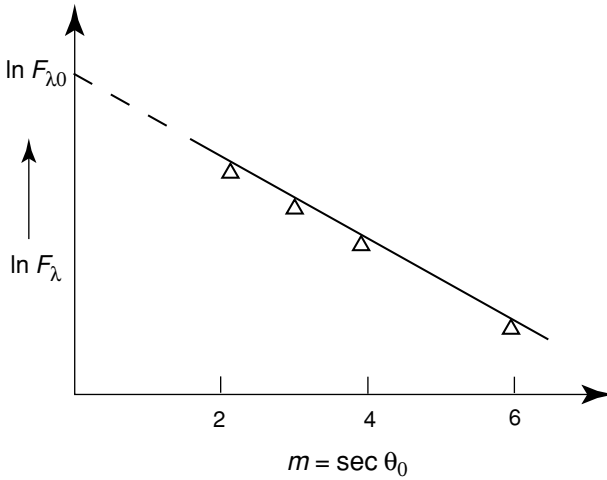
Figure 2.11   Hypothetical observed monochromatic solar irradiances F_λ as a function of the effective path length from which the solar irradiance at the top of the atmosphere can be graphically determined, referred to as the Langley plot.

the mean distance $r_0 = a$, the solar constant is defined by

$$S = F_\odot (r/a)^2. \tag{2.3.5}$$

The foregoing outlines the theoretical procedures of the Smithsonian long method for the determination of the solar constant. However, the atmosphere is essentially opaque for wavelengths shorter and longer than about 0.34 μm and 2.5 μm, respectively. Consequently, flux density observations cannot be made in these regions. Therefore, empirical corrections are needed for the omitted ranges, which account for about 8% of the solar flux.

There are other sources of error inherent in the Smithsonian long method caused by (1) empirical corrections for the absorption of ultraviolet by ozone, and the absorption of infrared by water vapor and carbon dioxide in the wings of the solar spectrum; (2) an unknown amount of diffuse radiation entering the aperture of the observing instrument; (3) variations of k_λ and the possible effects of aerosols during a series of measurements; and (4) measurement errors. Therefore, in spite of careful evaluation and observation, a certain amount of error is inevitable.

Employing the Smithsonian long method, each determination requires about 2 to 3 hours of observation time, plus twice that much time for the data reduction. In addition, there is no assurance that atmospheric properties and solar conditions will remain unchanged during the observation period. Because of this uncertainty and the burdensome, time-consuming work involved, a short method was devised to determine the solar constant.

In the short method, the diffuse component of solar radiation (the sky brightness) is measured for a given locality over a long period of time, so that a mean diffuse intensity

can be determined. Thus, a pyranometer reading of diffuse solar radiation will differ from the mean by an amount ε, called the *pyranometer excess*. In reference to Section 1.1.4 and Fig. 3.9, the attenuation of solar radiation on a clear day is due to scattering by molecules and aerosol particles, and absorption by various gases, primarily water vapor. If total precipitable water is given by w, an empirical relationship between the attenuation of direct solar irradiance and scattering and absorption effects may be expressed in the form $F_\lambda = w + q_\lambda \varepsilon$, where q_λ is a constant empirically determined for each wavelength for a given locality. With q_λ known, the spectral value of the solar irradiance can be found from the observed precipitable water and a pyranometer reading.

On the basis of a long series of previous observations of F_λ, m, and T_λ at a given location where the solar constant measurement has been made, a graph of F_λ versus air mass m can be constructed for a set value of T_λ. Thus, for a particular measurement of F_λ with a known air mass m, the corresponding transmissivity T_λ can be found from the graph. Once T_λ has been determined, solar irradiance at the top of the atmosphere $F_{0\lambda}$ can be evaluated through Eq. (2.3.2). After this point, evaluation of the solar constant proceeds in the same manner as in the long method. In the short method, the required measurements include a bologram of the sun, an observation of sky brightness by the pyranometer, and air mass determined by the position of the sun from a theodolite. These three measurements take only about 10 to 15 minutes. From the thousands of observations at various locations around the world during a period of more than half a century, the best value of the solar constant determined by the Smithsonian methods is 1353 W m^{-2}.

The presence of aerosols in the atmosphere imposes limitations on the accuracy of ground-based radiometric measurements of the solar constant (see, e.g., Reagan *et al.,* 1986). To minimize atmospheric effects, a number of measurements have also been made in the upper atmosphere and outer space. These have included observations made from balloons floating in the 27- to 35-km altitude range, jet aircraft at about 12 km, the X-15 rocket aircraft at 82 km, and the Mars Mariner VI and VII spacecrafts entirely outside the atmosphere. The solar constant derived from these experiments varies. Based on a series of measurements from high-altitude platforms, a standard solar constant of 1353 ($\pm$21) W m^{-2} was issued in 1976 by the National Aeronautics and Space Administration (Thekaekara, 1976).

2.3.3   Satellite Measurements of the Solar Constant

Measurements of incoming solar irradiance have been routinely made from satellite platforms since the mid-1970s. However, the high-accuracy, high-stability satellite-borne radiometer was only developed and incorporated in the Nimbus 7 satellite in 1978. This radiometer was an electrically calibrated cavity radiometer. The basic concept of blackbody cavity radiation was shown in Fig. 1.6. Radiometers of this design for use in satellites had a black painted cavity that absorbed nearly all the solar radiation impinging on it. The absorbed radiation raised the temperature of the cavity so that a radiant power could be measured corresponding to the increase in temperature. Such a cavity can also be heated by an electrical element in a manner equivalent to the incident sunlight. Because the input electrical power can be measured

accurately, the temperature response of the radiometer can be calibrated, and it is thus referred to as the *self-calibrating radiometer*.

Solar constant data have been derived from total solar irradiance measurements made by self-calibrating radiometers aboard a number of satellites since 1978. These include the Nimbus 7 Earth Radiation Budget (ERB) mission in 1978; the Solar Maximum Mission (SMM) Active Cavity Radiometer Irradiance Monitor 1 (ACRIM I) in 1980; the Earth Radiation Budget Experiment (ERBE) on board the NASA Earth Radiation Budget Satellite (ERBS, 1984), NOAA 9 (1984), and NOAA 10 (1986); and the ACRIM II measurements on board the Upper Atmosphere Research Satellite (UARS, 1991). Figure 2.12 shows the daily measurements of the solar constant from these satellites from 1979 to 1996. The solid lines are 81-day running means of the daily data. The absolute radiance scale of the ACRIM II data has been adjusted to
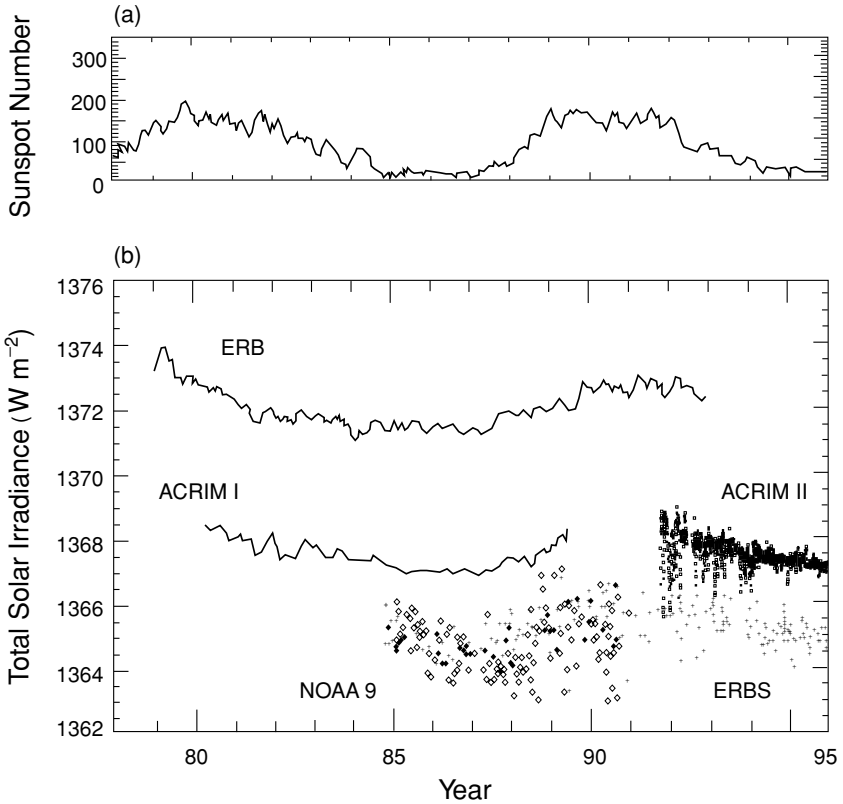

Figure 2.12 Solar activity variations from 1978 to 1996 illustrated by (a) the sunspot number and (b) changes in total solar irradiance. The results were obtained from the ERB radiometer on the Nimbus-7 satellite, ACRIM I on the Solar Maximum Mission (SMM) satellite, ACRIM II on the UARS, and the ERBE program (NOAA-9 and ERBS). The solid lines are 81-day running means of the daily data. Total solar irradiance increases during times of maximum solar activity relative to its levels in the intervening activity minimum. The differences in absolute irradiance levels among various measurements are of instrumental origin (data taken from Lean and Rind, 1998).

match the results of ACRIM I. The differences in absolute irradiance levels among various measurements depicted in this figure are attributed to the instrument sensitivity changes related to temperature or aspect drifts. In particular, ERB and ERBS data differed by about 10 watts per square meter. The top panel shows the sunspot numbers over the same period. It is quite clear that the data displayed in Fig. 2.12 provide irrefutable evidence of the 11-year solar constant cycle. When solar activity is high, as indicated by the sunspot number, the total and UV radiative outputs from the sun increase. Dark sunspots on the solar disk reduce total radiative output because their emission is less than that of the surrounding disk. However, after sunspots develop, magnetic regions involving faculae and plages where emission is enhanced also increase. These regions are evident as complexes of bright emission. The sun's irradiance fluctuates because radiation sources are not homogeneously distributed on its disk. Magnetic fields erupting from the solar convection zone (Section 2.1) into the overlying solar atmosphere generate active regions and complexes in which the local radiation is altered relative to the background solar disk. Magnetic activity erupts, evolves, and decays at different rates throughout the 11-year cycle, generating sunspots, plages, and faculae that modulate total and spectral solar radiative outputs. Finally, it should be noted that our knowledge of the 11-year irradiance cycle is imperfect because of uncertainties arising from the limited duration of space-borne solar monitoring that barely exceeds one 11-year cycle, as well as instrumental uncertainties that cause variable signals in individual satellite solar radiometers.

A number of analyses of the mean total solar irradiance have been reported. Based on the analysis of the solar irradiance measurements taken by the cavity sensor in a number of the satellites depicted in Fig. 2.12, a mean value for the solar constant of 1366 W m^{-2} with a measurement uncertainty of $\pm 3 \text{ W m}^{-2}$ has been suggested (Lean and Rind, 1998). The solar constant value is critical in the interpretation of measured solar absorption and heating rates in the atmosphere.

Exercises

2.1 Compute the solar elevation angle at solar noon at the poles, $60°$ N(S), $30°$ N(S), and the equator. Also compute the length of the day (in terms of hours) at the equator and at $45°$ N at the equinox and solstice.

2.2 From the geometry of an ellipse and the equation defining it, derive Kepler's first law denoted in Eq. (2.2.5).

2.3 Based on the conservation of angular momentum that the radius vector drawn from the sun to the planet sweeps out equal areas in equal time, derive Kepler's second law denoted in Eq. (2.2.6).

2.4 (a) Derive Kepler's third law by equating Newton's law of universal gravitation and the centrifugal force required to keep the planet in a circular orbit. (b) Given that the NOAA polar satellites orbit at about 850 km above the earth's surface, what would be the period of these satellites? (c) Geostationary satellites have the same angular velocity as the earth. What would be the required height for these satellites?

2.5 Given the solar constant of 1366 W m^{-2}, the mean earth–sun distance of 150×10^6 km, and the sun's radius of 0.70×10^6 km, calculate the equilibrium temperature of the sun.

2.6 If the average output of the sun is 6.2×10^7 W m^{-2}, and the radius of the earth is 6.37×10^3 km, what is the total amount of energy intercepted by the earth in one day?

2.7 Compute the fraction of the sun's emittance intercepted by the earth.

2.8 Consider a circular cloud whose diameter is 2 km and assume that it is an infinitely thin blackbody with a temperature of $10°C$. How much energy does it emit toward the earth? How much energy from this cloud is detected on a square centimeter of the earth's surface when the center of the cloud is 1 km directly over the receiving surface?

2.9 Assume that $\bar{r}$ is the mean albedo of the earth (*albedo* is defined as the ratio of the amount of flux reflected to space to the incoming solar flux), and that the earth–atmosphere system is in equilibrium. Show that the equilibrium temperature of the earth–atmosphere system $T_e = [(1 - \bar{r})S/4\sigma]^{1/4}$.

2.10 The following table gives the distances of various planets from the sun and their albedos. Employing the result in Exercise 2.9, compute the equilibrium temperatures of these planets.

Planet	Distance from sun (relative to earth)	Albedo (%)
Mercury	0.39	6
Venus	0.72	78
Earth	1.00	30
Mars	1.52	17
Jupiter	5.20	45

2.11 The height of earth-synchronous (geostationary) orbiting satellites, such as GOES satellites, is about 35,000 km. Using the solid angle derived in Exercise 1.2, calculate the equilibrium temperature of such a satellite in the earth–satellite system, assuming an effective equilibrium temperature of 255 K for the earth and assuming that the satellite is a blackbody.

2.12 Show that the change in the earth's equilibrium temperature T_e in terms of the earth–sun distance r is given by $\delta T_e/T_e = \delta r/2r$. The distance between the earth and the sun varies by about 3.3% with a maximum and minimum on January 3 and July 5, respectively. Compute the seasonal change in the earth's equilibrium temperature.

2.13 Calculate the daily insolation at the top of the atmosphere at (a) the south pole at the winter solstice; and (b) the equator at the vernal equinox. Use the mean earth–sun distance in your calculations and check your values with those shown in Fig. 2.8.

2.14 Prove that annual insolation is the same for corresponding latitudes in the two hemispheres [you may use the results in Eqs. (2.2.23) and (2.2.24) for analysis].

2.15 Show that the difference between the length of summer and that of winter is given by $\widetilde{T}4e\sin\omega/\pi$. In carrying out this exercise, first define the length using the astronomical season definition and then utilize Kepler's expressions by approximation.

2.16 Reproduce the daily solar insolation graph presented in Fig. 2.8 using Eqs. (2.2.21), (2.2.9), and (2.2.10).

2.17 Compute and plot the solar irradiance at the top of the earth's atmosphere emitted from temperatures of 5000, 5500, and 6000 K. Compare your results with those presented in Figs. 2.9 and 2.10.

2.18 On a clear day, measurements of the direct solar flux density F at the earth's surface in the 1.5- to 1.6-μm wavelength interval give the following values:

Zenith angle (degree):	$40°$	$50°$	$60°$	$70°$
F (W m^{-2}):	13.95	12.55	10.46	7.67

Find the solar flux density at the top of the atmosphere and the transmissivity of the atmosphere for normal incidence [see Eq. (1.4.10)] in this wavelength interval.

Suggested Reading

Berger, A. L. (1988). Milankovich theory and climate. *Rev. Geophys*. **26,** 624–657. This paper offers an authoritative overview of contemporary theory of the earth's orbit about the sun and its impact on climate.

Coulson, K. L. (1975). *Solar and Terrestrial Radiation*. Academic Press, New York. Chapters 3 and 4 give a comprehensive illustration of various kinds of pyrheliometers and pyranometers.

Hoyt, D. V., and Schatten, K. H. (1997). *The Role of the Sun in Climate Change*. Oxford University Press, New York. Chapters 2 and 3 contain a readable discussion of the composition of the sun and the solar constant from a historical perspective.

Jastrow, R., and Thompson, M. H. (1984). *Astronomy: Fundamentals and Frontiers,* 2nd ed. Wiley, New York. Chapter 12 provides an in-depth discussion and delightful photos of the structure and composition of the sun.

Lean, J., and Rind, D. (1998). Climate forcing by changing solar radiation. *J. Climate* **11,** 3069–3094. This paper includes a comprehensive and authoritative review of the solar constant and solar spectrum measurements and variabilities.

Chapter 3 | Absorption and Scattering of Solar Radiation in the Atmosphere

3.1 Composition and Structure of the Earth's Atmosphere

It is now generally accepted that the terrestrial planets were formed by the accretion of solid materials that condensed from the solar nebula about 4.5 billion years ago (BY). The earth's present atmosphere is believed to be a secondary atmosphere that was generated from volatile compounds contained within the solid planetesimals from which the earth formed. Any primary atmosphere that was captured must have been lost because the cosmic abundances, the composition of most stars including the sun, which contains about 90% H and 10% He by mass, had been significantly depleted. The heavy bombardment of the earth ended about 3.8 BY ago and life was probably extant by 3.5 BY, at which time the atmosphere might have contained CH_4 and NH_3. The post-heavy bombardment atmosphere was probably dominated by CO_2 and N_2 with traces of CO and H_2O, but lacking free O_2, referred to as a weakly reduced atmosphere, associated with volcanic activities. Liquid water is believed to have existed on the earth's surface. A large amount of CO_2 or other greenhouse gases is likely to have evolved to compensate for the faint young sun at about 3.5–3.8 BY. It has been suggested that the biota played an integral role in controlling atmospheric CO_2. As a consequence of photosynthesis and organic carbon burial, atmospheric O_2 levels rose naturally. The major increase appears to have occurred between about 1.9 and 2.2 BY. The level of free O_2 is also believed to be associated with the formation of the ozone layer that provided an effective screen for ultraviolet solar radiation. In the following, we define the region of the earth's atmosphere and discuss its present composition.

3.1.1 Thermal Structure

To describe the interaction of the earth's atmosphere with solar radiation, we must first understand the atmosphere's composition and structure. For the purpose of defining the region of the atmosphere associated with the absorption and scattering of sunlight, we first present the vertical temperature profile for the standard atmosphere, which is shown in Fig. 3.1. This profile represents typical conditions in the mid-latitude regions. According to the standard nomenclature defined by the International Union of Geodesy and Geophysics (IUGG) in 1960, the vertical profile is divided into
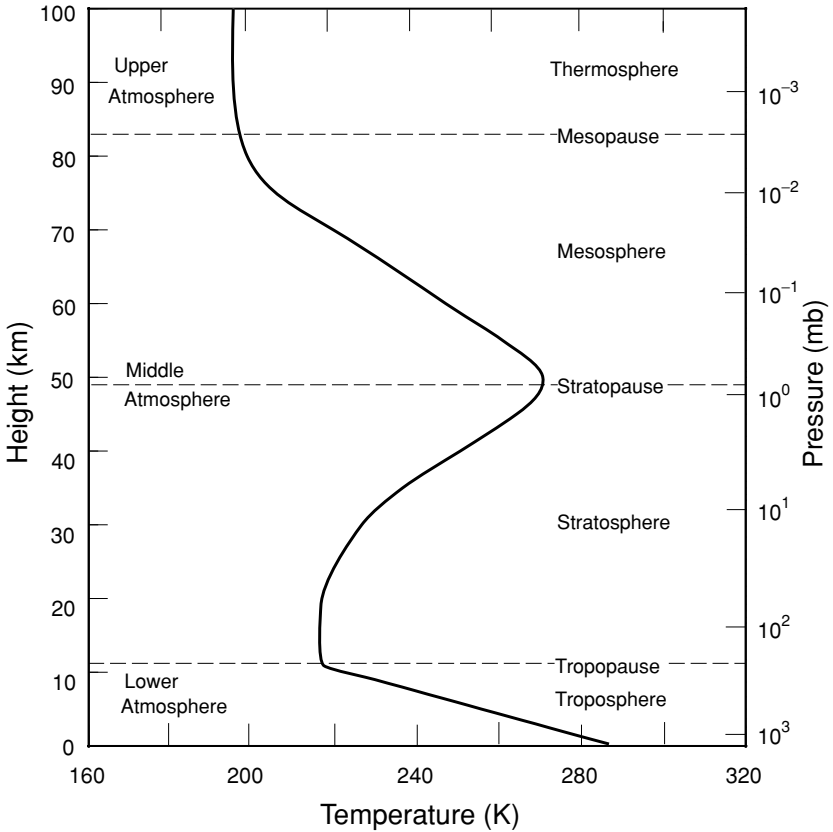
Figure 3.1   Vertical temperature profile after the U.S. Standard Atmosphere and definitions of atmospheric nomenclature.

four distinct layers as displayed in Fig. 3.1. These are the *troposphere, stratosphere, mesophere,* and *thermosphere*. The tops of these layers are called the tropopause, stratopause, mesopause, and thermopause, respectively.

The troposphere is characterized by a decrease in temperature with respect to height from a mean surface temperature of about 288 K to a temperature of about 220 K with a typical lapse rate of 6.5 K km^{-1}. The temperature structure in this layer is a consequence of the radiative balance and the convective transport of energy from the surface to the atmosphere. Virtually all water vapor, clouds, and precipitation are confined in this layer. The stratosphere is characterized by an isothermal layer from the tropopause to about 20 km, above which the temperature increases to the stratopause with a temperature of about 270 K. Ozone occurs chiefly in the stratosphere. In addition, thin layers of aerosol are observed to persist for a long period of time within certain altitude ranges of the stratosphere. The state of the stratosphere is primarily determined by the absorption of solar fluxes by ozone and through the emission

of infrared fluxes by carbon dioxide. Like the troposphere, the temperatures in the mesosphere decrease with height from about 50 to 85 km. Above this height and extending upward to an altitude of several hundred kilometers lies the thermosphere where temperatures range from 500 K to as high as 2000 K, depending on the level of solar activity. The outermost region of the atmosphere above the thermosphere is called the *exosphere*.

Atmospheric nomenclature is customarily derived from its thermal state, as just defined. However, there is practically no physical distinction between the stratosphere and the mesosphere. In recent years, the atmosphere has been defined by the following terms: *lower atmosphere* (troposphere), *middle atmosphere* (stratosphere and mesosphere), and *upper atmosphere* (above 80 km). Moreover, the lowest 1 km or so of the atmosphere differs from the remaining troposphere in that interactions with the surface are strong and significant. This is referred to as the *planetary boundary layer*.

3.1.2 Chemical Composition

The earth's atmosphere is presently composed of two groups of gases, one with nearly permanent concentrations and another with variable concentrations. The atmosphere also contains various kinds of aerosols, clouds and precipitation, which are highly variable in space and time. Table 3.1 lists the chemical formulas and volume ratio for the concentrations of permanent and variable gases. Nitrogen, oxygen, and argon account for more than 99.96% of the atmosphere by volume. The permanent gases have virtually constant volume ratios up to an altitude of about 60 km.

Although listed as a permanent constituent, the carbon dioxide concentration has been increasing by about 0.4% per year as a result of the combustion of fossil fuels,

Table 3.1

The Composition of the Atmosphere[a]

Permanent constituents		Variable constituents	
Constituent	% by volume	Constituent	% by volume
Nitrogen (N_2)	78.084	Water vapor (H_2O)	0–0.04
Oxygen (O_2)	20.948	Ozone (O_3)	$0–12 \times 10^{-4}$
Argon (Ar)	0.934	Sulfur dioxide (SO_2)[b]	0.001×10^{-4}
Carbon dioxide (CO_2)	0.036	Nitrogen dioxide (NO_2)[b]	0.001×10^{-4}
Neon (Ne)	18.18×10^{-4}	Ammonia (NH_3)[b]	0.004×10^{-4}
Helium (He)	5.24×10^{-4}	Nitric oxide (NO)[b]	0.0005×10^{-4}
Krypton (Kr)	1.14×10^{-4}	Hydrogen sulfide (H_2S)[b]	0.00005×10^{-4}
Xenon (Xe)	0.089×10^{-4}	Nitric acid vapor (HNO_3)	Trace
Hydrogen (H_2)	0.5×10^{-4}	Chlorofluorocarbons	Trace
Methane (CH_4)	1.7×10^{-4}	($CFCl_3$, CF_2Cl_2	
Nitrous oxide (N_2O)[b]	0.3×10^{-4}	CH_3CCl_3, CCl_4, etc.)	
Carbon monoxide (CO)[b]	0.08×10^{-4}		

[a] After the U.S. Standard Atmosphere (1976) with modifications.
[b] Concentration near the earth's surface.

absorption and release by the oceans, and photosynthesis. In addition, a number of measurement series indicate that the atmospheric methane concentration, with a present value of ~1.7 parts per million by volume (ppmv), has increased by 1–2% per year and that it may have been increasing for a long period of time. The most likely cause of the increase in the CH_4 concentration is the greater biogenic emissions associated with a rising human population. Rice paddies seem to be another prime source of CH_4. There is no direct evidence of an increase in carbon monoxide concentration. However, deforestation, biomass burning, and modification of CH_4 sources could lead to changes in the atmospheric CO concentration. There is also some evidence of an increase in nitrous oxide. A possible global increase of ~0.2% per year in N_2O has been suggested. This increase is attributed to the combustion of fossil fuels and, in part, to fertilizer denitrification.

The amounts of variable gases listed in Table 3.1 are small, but they are extremely important in the radiation budget of the atmosphere. Water vapor is the major radiative and dynamic element in the earth's atmosphere. The H_2O concentration varies significantly with both space and time. The spatial distribution of tropospheric H_2O is determined by the local hydrological cycle via evaporation, condensation, and precipitation, and by large-scale transport processes. Specific humidity decreases rapidly with pressure, almost following an exponential function. Specific humidity also decreases with latitude. More than 50% of water vapor is concentrated below ~850 mb, while more than 90% is confined to the layers below ~500 mb. The variability of the H_2O concentration shows a bimodal distribution with a maximum in the subtropics of both hemispheres below ~700 mb. The variability is very small in the equatorial region and poleward of ~60°. The stratospheric H_2O concentration is relatively small, with a value of ~3–4 ppmv in the lower stratosphere. It has been suggested that H_2O in the lower stratosphere is controlled by the temperature of the tropical tropopause, and by the formation and dissipation of cirrus anvils due to outflow from cumulonimbus.

The ozone concentration also varies significantly with space and time, but ozone occurs principally at altitudes from ~15 to 30 km, an area referred to as the *ozone layer*. The maximum ozone concentration occurs at ~20–25 km, depending on latitude and season. Atmospheric ozone is continually created and destroyed by photochemical processes associated with solar ultraviolet radiation. The absorption of deadly solar ultraviolet radiation by the ozone layer is essential to life on earth. Many photochemical reactions associated with O_3 involve H_2O, CH_4, and CO. Total ozone varies significantly in terms of latitude and season, with the maximum occurring during the polar night.

Nitrogen oxides ($NO_x = NO$, NO_2) appear to be important in the determination of both tropospheric and stratospheric O_3 concentrations. Atmospheric NO_x are emitted by transportation and combustion processes at the surface and by high-flying aircraft in the upper troposphere and lower stratosphere. In the stratosphere, the major source of NO_x is the dissociation of N_2O by excited oxygen atoms. In the lower atmosphere, the major source of NO_x appears to be the anthropogenic combustion of fossil fuels and biomass burning.

Chlorofluorocarbons are also recognized as presenting a potential threat to the ozone layer. Large amounts of these chemicals are produced by industry and are used in solvents, refrigerants, and spray-can propellants. Chlorofluorocarbons include $CFCl_3$ (CFC-11), CF_2Cl_2 (CFC-12), CF_3Cl (CFC-13), CF_3CCl_3 (CFC-113), CF_4CCl_2 (CFC-114), CF_2ClCF_3 (CFC-115), CHF_2Cl (CFC-22), CH_3Cl, and CCl_4.

Sulfur dioxide in the stratosphere is largely produced by volcanic eruptions. SO_2 and other sulfur-based gases are believed to be the primary precursors of stratospheric aerosols. Emissions of SO_2 from the surface may be important in the formation of tropospheric aerosols as well and, hence, related to the production of acid rain through cloud and precipitation processes. Figure 3.2 shows representative vertical profiles of the gases listed in Table 3.1 for mean midlatitude conditions.

The atmosphere continuously contains aerosol particles ranging in size from $\sim 10^{-3}$ to $\sim 20~\mu m$. These aerosols are known to be produced by natural processes as well as by human activity. Natural aerosols include volcanic dust, smoke from forest fires,
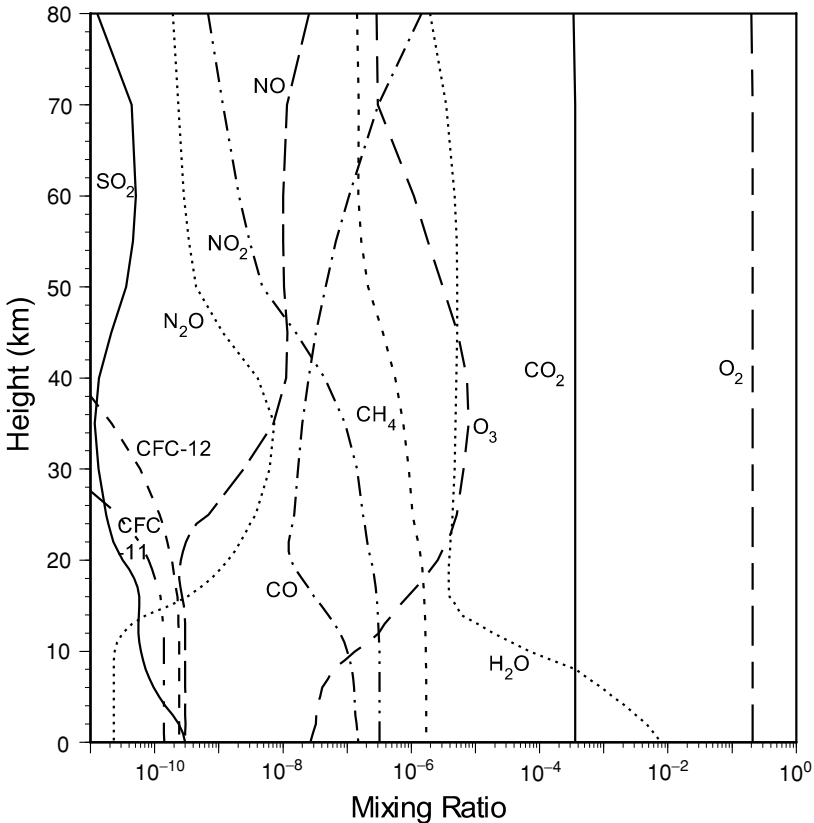
Figure 3.2 Representative vertical profiles of mixing ratios of selected species for midlatitude conditions.

particles from sea spray, windblown dust, and small particles produced by the chemical reactions of natural gases. Primary man-made aerosols include particles directly emitted during combustion processes and particles formed from gases emitted during combustion. The atmospheric aerosol concentration varies with locality; the largest concentrations generally occur in urban and desert areas. In normal conditions, the background aerosol concentration has a visibility of ∼20–50 km. Aerosol concentrations generally decrease rapidly with height in the troposphere. Some aerosols are effective condensation and ice nuclei upon which cloud particles may form. For the hygroscopic type, the size of the aerosol depends on relative humidity. Thin layers of aerosols are observed to persist for a long period of time in some altitudes of the stratosphere, as noted previously.

Clouds are global in nature and regularly cover more than 50% of the sky. There are various types of clouds. Cirrus in the tropics and stratus in the Arctic and near the coastal areas are climatologically persistent. The microphysical composition of clouds in terms of particle size distribution and cloud thickness varies significantly with cloud type. Clouds can also generate precipitation, a discrete event generally associated with midlatitude weather disturbances and tropical cumulus convection.

3.2   Atmospheric Absorption

Understanding the atmospheric absorption produced by the various molecules listed in Table 3.1 requires an understanding of their molecular structure. Consider first the diatomic structure (N_2, O_2, CO). Its two nuclei can only move toward and away from each other during vibration. Thus, diatomic molecules have but one vibrational mode, known as symmetric stretch and denoted by v_1. Because of symmetrical charge distributions, N_2 and O_2 molecules lack a permanent dipole moment that could acquire oscillating momentum during vibration. As a result, they have little radiative activity in the visible and infrared regions (see also Subsection 1.3.1.2).

For triatomic molecules with a linear symmetrical configuration (CO_2, N_2O), there are three vibrational modes: v_1 for symmetric stretch, v_2 for bending motion, and v_3 for antisymmetric stretch, as shown in Fig. 3.3. As a result of their linear symmetry, the CO_2 molecules have no permanent electric dipole moment. Because of vibrational symmetry, the symmetric stretch mode, v_1, is radiatively inactive at its fundamental, although it has been identified in the Raman spectrum near 7.5 μm. The bending mode, v_2, consists of v_{2a} and v_{2b} vibrations at the same frequency, referred to as degenerate, as noted in Subsection 1.3.1.2. The triatomic structure for H_2O and O_3 molecules forms an isosceles triangle that is obtuse, referred to as the asymmetric top (bent triatomic) configuration. This molecular shape has three fundamental vibration modes, as shown in Fig. 3.3 (Herzberg, 1945). Not shown in the figure is the CH_4 molecule, which has a spherical top configuration, no permanent electric dipole moment, and four fundamental vibration modes.

The rotation of a rigid body is defined by its center of gravity with respect to three orthogonal axes. The axes of rotational freedom of linear and asymmetric top
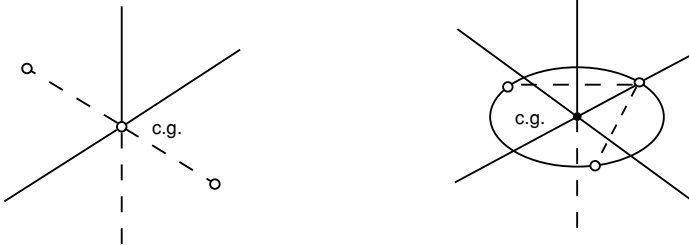
Vibration Modes
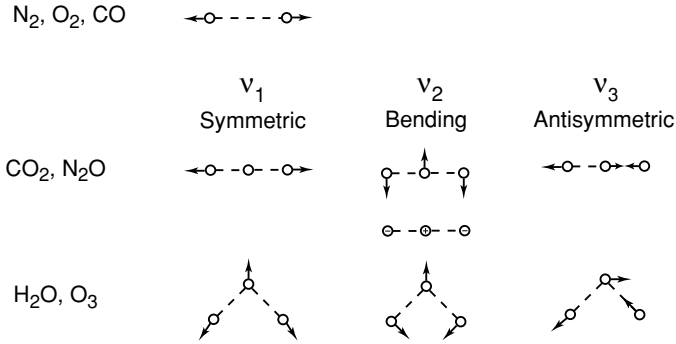
Rotation

Figure 3.3   Vibrational modes of diatomic and triatomic atmospheric molecules and the axes of rotational freedom for linear and asymmetric top molecules.

molecules are displayed in Fig. 3.3. A diatomic and a linear triatomic molecule have two equal moments of inertia and two degrees of rotational freedom. Asymmetric top molecules have three unequal moments and three degrees of rotational freedom. Molecular structures of greater complexity have additional degrees of freedom.

As discussed in Section 2.3.1, electrons can be removed from an orbit if a characteristic amount of energy is available. This amount with respect to the ground level is called the *ionization potential*. For atoms and molecules with more than one electron, the ionization potential usually refers to the most loosely bound electron, the one that requires the least energy for removal. Likewise, a certain amount of energy can dissociate or separate the atoms and is referred to as the *dissociation potential*.

Electronic energy is closely related to vibrational energy because both derive from the elastic valence bonds that bind the atoms into a molecular entity. The sign and magnitude of the force between two atoms in a molecule depend on two factors: the
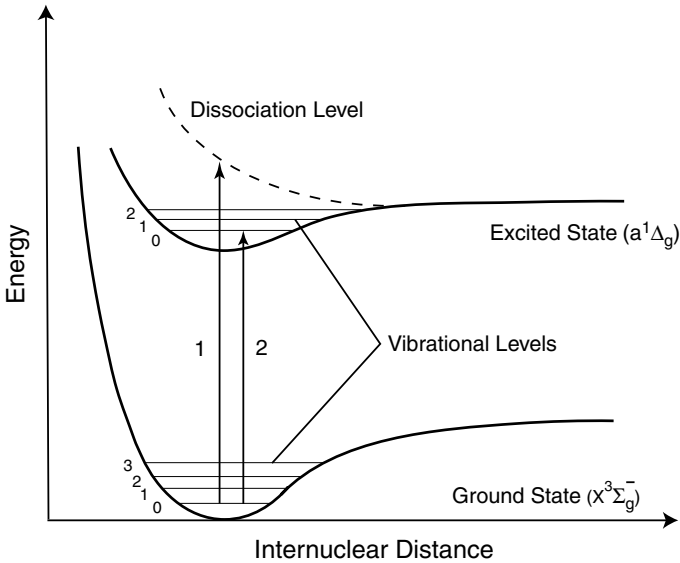
Figure 3.4 Illustrative potential energy curves for two electronic states of a diatomic molecule. The horizontal lines in the potential well represent vibrational energy levels.

distance between the two nuclei and their electronic configuration. It is customary to represent the dependence of this force on the internuclear distance as a *potential curve,* with potential energy as the ordinate and distance as the abscissa, as shown in Fig. 3.4 in which two possible potential curves are displayed schematically. These two curves are representative of the ground and excited states of a diatomic molecule. At large distances the two atoms of a molecule exert no significant attractive or repulsive force on one another. As the atoms approach one another in the ground state of the molecule, they begin to exert attractive force, as shown by the decrease in potential energy. The ground electronic state of the molecule exhibits maximum stability at the minimum in the potential curve that represents a stable bound configuration of the molecule. As the distance becomes smaller, the atoms begin to repel one another so that potential energy increases.

When a high-energy photon is absorbed, the electron configuration changes to one that has potential energy even when the nuclei are at equilibrium distance. This is a transition to an excited state whose potential energy is represented by the upper curve. The energy gain due to the transition is given by the difference between the two minima. Also shown are the vibrational levels denoted by the horizontal lines in the potential well.

As noted earlier and as shown in Fig. 3.4, the absorption of a photon of appropriate energy can lead to the excitation of a molecule from one electronic state to another. If the excited state is unstable, two possibilities may take place: the molecule may decompose into its constituent atoms and photoionization may also occur. At

atmospheric temperatures, most molecules are in the ground vibrational state. Possible transitions are denoted by the vertical lines labeled 1 and 2 in Fig. 3.4. Transition 1 leads from the ground state to a state that is not quantized, denoted by the dissociation level, which has more energy than the quantized level. Thus, a continuum of wavelengths near the energy associated with transition 1 are possible. The spectrum should then consist of a smooth continuum at short wavelengths above the dissociation limit. Transition 2, on the other hand, requires absorption into a specific vibrational level of the upper electronic state. It is quantized at the energy level defined by transition 2. The spectra corresponding to these transitions are then shown at discrete wavelengths. Many absorption processes associated with diatomic molecules are largely a combination of electronic and vibrational transitions that occur in the ultraviolet (UV), visible, and near infrared (IR) regions. As noted in Subsection 1.3.1.2, it is customary to refer to transitions as forbidden and allowed. However, even highly forbidden transitions occur under certain conditions.

As discussed in Section 1.1.4, the strength of absorption is customarily represented by cross section σ_a in units of cm^2. Also frequently used is the absorption coefficient k in units of (cm atm)$^{-1}$. The two are related through the Loschmidt's number, $N_0(= 2.687 \times 10^{19}$ particles cm^{-3} at standard temperature of 273 K and standard pressure of 1013 mb), and the fact that $k = \sigma_a \cdot N_0$.

3.2.1 Absorption in the Ultraviolet

3.2.1.1 MOLECULAR NITROGEN

The UV absorption spectrum of the most abundant gas in the atmosphere consists of a band system from about 1450 to 1120 Å, referred to as the *Lyman–Birge–Hopfield bands,* associated with the forbidden ground-state transition. Dissociation and ionization continua have been observed below 800 Å. From about 1000 to 850 Å, the nitrogen spectrum consists of various sharp bands whose absorption coefficients are highly variable. Absorption of N_2 in the solar spectrum is generally considered to be insignificant. The photodissociation of N_2 in the atmosphere plays a minor role in atmospheric chemistry below 100 km.

3.2.1.2 MOLECULAR OXYGEN

The UV absorption spectrum of O_2 begins with the weak *Herzberg band* between 2600 and 2000 Å, due primarily to the forbidden ground-state transition and dissociation continuum, which lead to the formation of two oxygen atoms in the ground ^{3}P state. Absorption by this band system is weak and of little importance in the absorption of solar radiation because of overlap with the much stronger O_3 bands in this spectral region. It is, however, considered to be of significance in the formation of ozone. Adjacent to the Herzberg continuum are the *Schumann–Runge bands* produced by ground-state transitions that occupy the spectral region from 2000 to 1750 Å. At 1750 Å, the bands converge to a stronger dissociation continuum in which one of the oxygen atoms is formed in the excited ^{1}D state. Referred to as the *Schumann–Runge continuum,* this extends to about 1300 Å and represents the most important
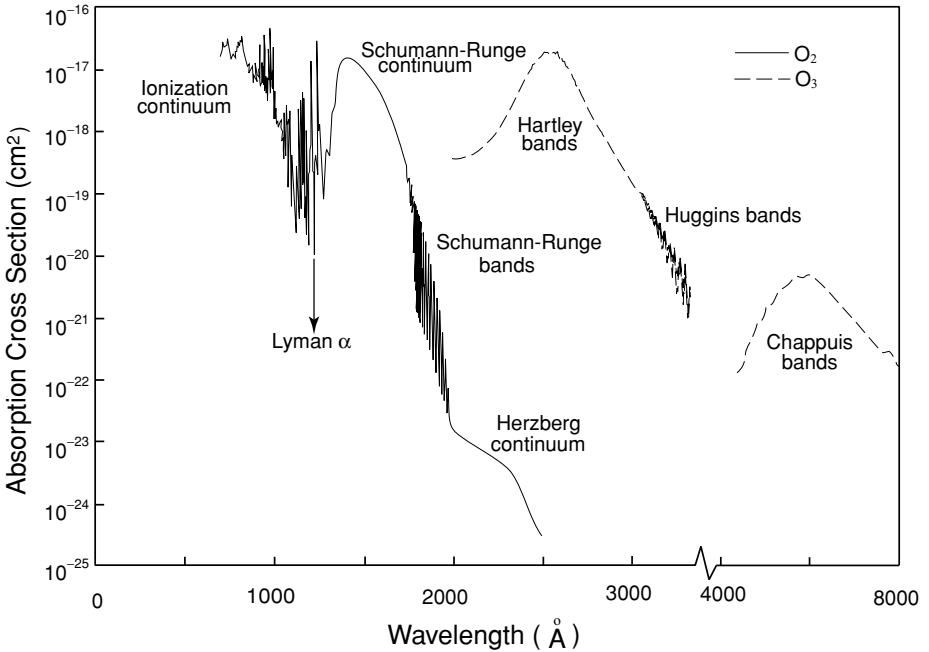
Figure 3.5 Absorption cross section of ozone and molecular oxygen in the ultraviolet spectral region. Data taken from Brasseur and Solomon (1986), Vigroux (1953), and Griggs (1968).

absorption spectrum of O_2. At shorter wavelengths some diffuse bands occur, which have not yet been identified. These are interspersed with a series of windows allowing some wavelengths to penetrate deeply into the atmosphere. Of particular interest is the Lyman α line located at 1216 Å (see the solar spectrum in Fig. 2.9), which happens to lie in one of the windows. Between 850 and 1100 Å are a series of distinct Rydberg bands, associated with transitions between excited states, known as the *Hopfield bands*. Below about 1026 Å, O_2 absorption is in the form of an ionization continuum. The spectral distribution of the absorption cross section of O_2 is shown in Fig. 3.5.

Because of the absorption of solar UV radiation, some of the oxygen and nitrogen molecules in the upper atmosphere undergo photochemical dissociation into atomic oxygen and nitrogen. Atomic nitrogen exhibits an absorption spectrum from about 10 to about 1000 Å. Although atomic nitrogen probably is not abundant enough to be a significant absorber in the upper atmosphere, it may play an important role in the absorption of UV radiation in the thermosphere. Atomic oxygen also shows an absorption continuum in the region of 10 to 1000 Å. Because of the absorption of solar UV radiation, a portion of molecular and atomic oxygen and nitrogen becomes ionized. The ionized layers in the upper atmosphere are formed mainly as a result of these processes.

3.2.1.3 OZONE

The absorption of ozone in the solar spectral region is due to electronic transitions. The strongest ozone bands are the *Hartley bands,* which cover the region from 2000 to 3000 Å and are centered at 2553 Å. The absorption of solar flux in these ozone bands takes place primarily in the upper stratosphere and in the mesosphere. The weak bands between 3000 and 3600 Å have more structure and are called the *Huggins bands*. Ozone also shows weak absorption bands in the visible and near-IR regions from about 4400 to 11,800 Å. These bands are referred to as *Chappuis bands*. The absorption coefficients in these bands are slightly dependent upon temperature. They have been measured by a number of earlier researchers (Inn and Tanaka, 1953; Vigroux, 1953) and more recently by Molina and Molina (1986) and Anderson and Mauersberger (1992). Absorption cross sections of the preceding three O_3 bands are also displayed in Fig. 3.5.

3.2.1.4 OTHER MINOR GASES

The NO_2 molecule absorbs solar flux in the UV as well as in the visible section of the wavelength range between 0.2 and 0.7 μm. Accurate absorption cross sections are required for the retrieval of aerosols and ozone using ground-based sunphotometer measurements (see Section 7.2.1). The UV absorption cross sections of NO, N_2O, H_2O, CO_2, and other trace gases listed in Table 3.1 have been measured by numerous researchers and they are important in the discussion of atmospheric chemistry and ionization in the middle and upper atmospheres. However, either because they occur in too small quantities or because they are dissociated at high levels, these gases absorb relatively little energy in the UV and are overshadowed by the absorption of O_2 and O_3. The most important absorption bands in the UV particularly associated with photochemical processes are listed in Table 3.2.

3.2.1.5 ABSORPTION OF SOLAR RADIATION

In reference to Section 1.4.2, consider an atmosphere that is plane-parallel and nonscattering. We may define a normal absorption optical depth for a monochromatic

Table 3.2

Important Absorption Spectral Regions Associated with Photochemistry in the Atmosphere

Wavelength range (Å)	Absorber	Principal location
1000–1750	O_2 Schumann–Runge continuum	Thermosphere
	O_2 1216 Lyman α line	Mesosphere
1750–2000	O_2 Schumann–Runge bands	Mesosphere
2000–2420	O_2 Herzberg continum; O_3 Hartley band	Stratosphere
2420–3100	O_3 Hartley band; $O(^1D)$ formation	Stratosphere
3100–4000	O_3 Huggins bands; $O(^3P)$ formation	Stratosphere/ troposphere
4000–8500	O_3 Chappuis bands	Troposphere

wavelength and for a given absorber as follows:

$$\tau(z) = \int_z^\infty n(z)\sigma_a dz, \tag{3.2.1}$$

where σ_a is the absorption cross section (cm^2), and n is the number density (cm^{-3}) for a specific absorber. The attenuation of solar flux that enters the atmosphere at a solar zenith angle of θ_0 is given by

$$F_{\text{act}}(z) = F_\odot e^{-\tau(z)/\mu_0}, \tag{3.2.2a}$$

where $\mu_0 = \cos\theta_0$ and $F_\odot$ is the solar irradiance at the top of the atmosphere. The flux here is in reference to the direct solar beam without accounting for the cosine dependence [see Eq. (1.1.8)] and is referred to as *actinic flux* associated with photo-dissociation, defined by

$$F_{\text{act}}(z) = \int_0^{2\pi} \int_{-1}^{1} I(z; \mu, \phi)\, d\mu\, d\phi, \tag{3.2.2b}$$

with units of photons cm^{-2} sec^{-1}. This definition can be applied to both direct and diffuse beams.

The solar energy absorbed per time and volume along the direction of the solar beam, referred to as the *volume absorption rate* (photons cm^{-3} sec^{-1}), can be expressed by

$$q(z) = \frac{d F_{\text{act}}(z)}{dz/\mu_0} = F_\odot n(z)\sigma_a e^{-\tau(z)/\mu_0}$$

$$= J(z)n(z), \tag{3.2.3a}$$

where the monochromatic *photodissociation coefficient, $J(z)$*(sec^{-1}), is defined by

$$J(z) = \sigma_a F_{\text{act}}(z). \tag{3.2.3b}$$

To gain a physical insight into solar attenuation, consider an atmosphere containing a concentration of an absorber that varies exponentially with altitude according to a scale height, H (Exercise 3.1), in the form

$$n(z) = n_0 e^{-z/H}, \tag{3.2.4}$$

where n_0 is the value of n at some arbitrary level at which z is taken to be zero. On substituting Eq. (3.2.4) into Eqs. (3.2.3a) and (3.2.1), we obtain

$$q(z) = F_\odot \sigma_a n_0 \exp\left(-\frac{z}{H} - \frac{1}{\mu_0}\sigma_a n_0 H e^{-z/H}\right). \tag{3.2.5a}$$

We may define a level, z_0, at which the normal absorption optical depth is equal to 1 such that

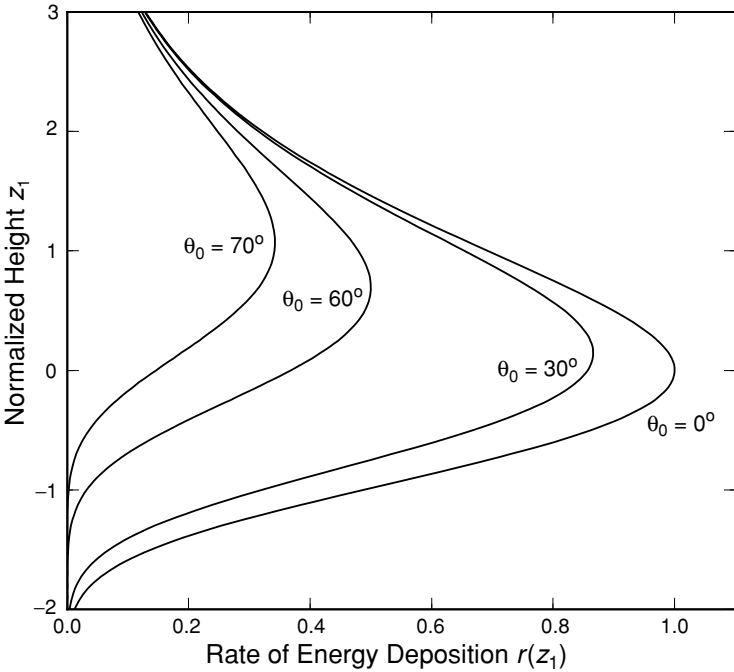$$\tau(z_0) = 1, \text{ i.e., } \sigma_a H n_0 = e^{z_0/H}. \tag{3.2.5b}$$

Figure 3.6 Normalized rate of energy deposition as a function of normalized height for a number of solar zenith angles (see text for the definition of r and z_1).

By carrying out some straightforward analyses, we find

$$r(z_1) = \frac{q(z)}{q[z_0(\mu_0 = 1)]} = \exp\left(1 - z_1 - \frac{1}{\mu_0}e^{-z_1}\right), \qquad (3.2.6)$$

where $z_1 = (z - z_0)/H$ (Exercise 3.2).

The function $r(z_1)$ is illustrated in Fig. 3.6 for a number of solar zenith angles θ_0. This function displays a well-defined layered structure, referred to as the *Chapman layer* (Chapman, 1931). The small value of $q(z)$ at the top level is associated with the extremely small value of $n(z)$ due to exponential falloff and a constant $J(z)$ value at the top of the atmosphere. Deep down in the atmosphere $n(z)$ increases but $J(z)$ declines more rapidly, again leading to a small value of $q(z)$, as shown in Eq. (3.2.3a). The rate of energy deposition is also related to the solar zenith angle. When the sun is low, the maximum deposition level moves higher because of the increase in the effective optical depth. As demonstrated in Eq. (3.2.5b), the optical depth of unity is the level at which maximum energy deposition occurs.

The preceding discussion of the Chapman layer is limited to one particular concentration. In practice, however, the combined effects of several absorbing gases must

be accounted for. Thus, the total absorption optical depth for a given wavelength is

$$\tau(z, \lambda) = \sum_i \int_z^\infty n_i(z)\sigma_a(n_i, \lambda)\, dz, \qquad (3.2.7a)$$

and the photodissociation coefficient defined in Eq. (3.2.3b) for a spectral interval can be written as

$$J_i(z) = \int_{\Delta\lambda} \sigma_a(n_i, \lambda) F_\odot(\lambda) e^{-\tau(z,\lambda)/\mu_0}\, d\lambda. \qquad (3.2.7b)$$

When the solar zenith angle is larger than about $75°$, as in the case of sunrise or sunset, the effect of the earth's curvature must be accounted for in the calculation of the effective optical depth. In this case a more complicated function should replace $1/\mu_0$ (Exercise 3.3).

To illustrate the relative absorption effects of oxygen and nitrogen molecules, Fig. 3.7 shows the reduction of solar flux when it penetrates the atmosphere. The curve represents the altitude where the optical depth is unity. Alternately, we can express $F_{act} = F_\odot e^{-1}$, referred to as the *e-folding transfer*. The EUV (extreme ultraviolet) fluxes are absorbed at high altitudes, resulting in the dissociation and ionization of the major constituents in the thermosphere and leading to the formation of the layers of the ionosphere. At longer wavelengths, from 1750 to about 2400 Å, the solar flux penetrates deeper into the atmosphere and is chiefly absorbed by O_2 in the Schumann–Runge band and Herzberg continuum (Table 3.2), leading to the production of O and
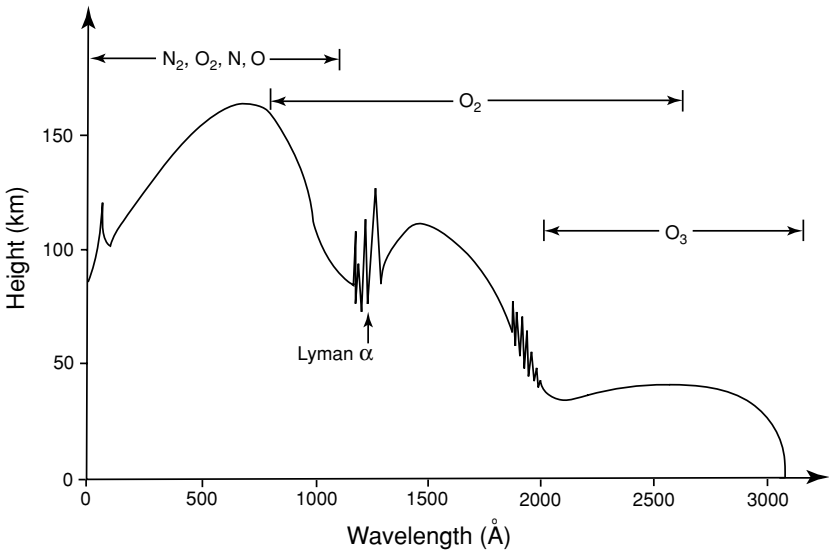
Figure 3.7   Depth of penetration of solar UV radiation in the earth's atmosphere for $\mu_0 = 1$ and an averaged ozone profile. The line indicates the height where the optical depth is equal to 1 (data taken from Herzberg, 1965).

O_3. Maximum ozone absorption occurs at about 50 km, associated with absorption in the Hartley band in the wavelength range 2400–3100 Å. Above 3100 Å, the atmosphere is relatively transparent except for Rayleigh scattering and scattering produced by aerosols and clouds.

3.2.2 Photochemical Processes and the Formation of Ozone Layers

Because of the absorption spectrum of various molecules and atoms in the solar UV region, a great variety of photochemical processes take place in the upper and middle atmospheres. Those involving various forms of oxygen are critical in determining the amount of ozone in the stratosphere. The classical photochemistry of the middle atmosphere concerning the ozone problem was first postulated by Chapman (1930), who described the following five basic reactions.

Ozone is basically formed by the three-body collision

$$O + O_2 + M \xrightarrow{K_{12}} O_3 + M, \tag{3.2.8a}$$
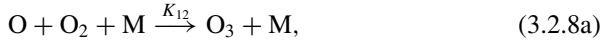
where M is any third atom or molecule, and K_{12} is the rate coefficient involving O and O_2. Atomic oxygen is produced when the oxygen molecule is dissociated by a quantum of solar energy

$$O_2 + h\tilde{\nu}(\lambda < 2420\,\text{Å}) \xrightarrow{J_2} O + O, \tag{3.2.8b}$$
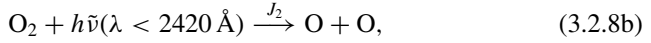
where J_2 is the dissociating quanta per molecule absorbed by O_2. Ozone is destroyed by photodissociation,

$$O_3 + h\tilde{\nu}(\lambda < 11,000\,\text{Å}) \xrightarrow{J_3} O + O_2, \tag{3.2.8c}$$
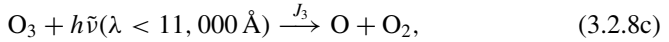
as well as by collision with oxygen atoms,

$$O_3 + O \xrightarrow{K_{13}} 2O_2, \tag{3.2.8d}$$
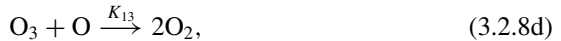
where J_3 is the dissociating quanta per molecule absorbed by O_3, and K_{13} denotes the rate coefficient involving O_3 and O. At the same time, oxygen atoms generated by reactions (3.2.8b) and (3.2.8c) may undergo a three-body collision,

$$O + O + M \xrightarrow{K_{11}} O_2 + M, \tag{3.2.8e}$$
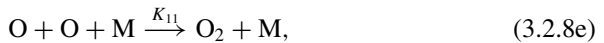
with K_{11} denoting the rate coefficient involving O and O. Normally, the reaction denoted in (3.2.8e) may be neglected below 50 to 60 km.

The preceding five reactions take place simultaneously. If an equilibrium state is reached, the number of ozone molecules formed exactly equals the number destroyed in unit volume and time. To evaluate the equilibrium amount of ozone, let [O], [O_2], [O_3], and [M] be the number densities, respectively, for O, O_2, and O_3, and air molecules. Based on the observed concentrations of oxygen species, almost all oxygen below 70 km is in the form of O_2, which is fairly constant with a value of about 21%. Thus, below 70 km, [O_2] may be treated as an independent variable. It follows that the

photochemical processes given by Eqs. (3.2.8a)–(3.2.8c) may be expressed in terms of the rate of change of the number density of O and O_3 in the forms:

$$\frac{\partial[O]}{\partial t} = -K_{12}[O][O_2][M] + 2J_2[O_2] - K_{13}[O][O_3]$$

$$+J_3[O_3] - 2K_{11}[O][O][M], \tag{3.2.9a}$$

$$\frac{\partial[O_3]}{\partial t} = K_{12}[O][O_2][M] - K_{13}[O][O_3] - J_3[O_3], \tag{3.2.9b}$$

with the photodissociation coefficients given by

$$J_2(z) = \int_0^{\lambda_1} \sigma_a(O_2, \lambda) F_\odot(\lambda) e^{-\tau(z,\lambda)/\mu_0} d\lambda, \tag{3.2.10a}$$

$$J_3(z) = \int_0^{\lambda_2} \sigma_a(O_3, \lambda) F_\odot(\lambda) e^{-\tau(z,\lambda)/\mu_0} d\lambda, \tag{3.2.10b}$$

where $\lambda_1 = 0.2420 \, \mu m$ and $\lambda_2 = 1.1 \, \mu m$.

On adding Eqs. (3.2.9a) and (3.2.9b), we find

$$\frac{\partial}{\partial t}([O] + [O_3]) = 2J_2[O_2] - 2K_{13}[O][O_3]. \tag{3.2.11a}$$

Below 75 km, $K_{13}[O][O_3] \ll K_{12}[O][O_2][M]$, so that Eq. (3.2.9b) may be simplified to give

$$\frac{\partial[O_3]}{\partial t} \cong -J_3[O_3] + K_{12}[O][O_2][M]. \tag{3.2.11b}$$

Under the assumption of photochemical equilibrium, $\partial[O]/\partial t = \partial[O_3]/\partial t = 0$. Thus, from Eqs. (3.2.11a) and (3.2.11b), the solution for ozone is given by

$$[O_3]_{eqb} \cong \left(\frac{J_2 K_{12}}{J_3 K_{13}}[M] \right)^{1/2} [O_2]. \tag{3.2.12}$$

At low levels, $[O_3]$ is small since J_2 is small. At high levels, $[O_3]$ is also small because $[O_2]$ and $[M]$ are both small. $[O_3]$ has a maximum in the stratosphere. Exercise 3.4 requires the calculation of $[O_3]$ using Eq. (3.2.12).

Figure 3.8 depicts the equilibrium ozone concentration determined from the classical theory. In the same diagram, the observational range of ozone number densities (shaded area) is also shown. It is evident that the classical theory overestimates the ozone number densities at almost all heights. The total ozone in an atmospheric column based on theoretical calculations exceeds the observed values by as much as a factor of three or four. Obviously, additional loss mechanisms are required to explain the observed data.

In the early 1970s, two independent theoretical analyses suggested that ozone can be destroyed by minute concentrations of chemical species associated with human activities. These species are chlorine atoms in the chlorine molecule and nitrogen atoms in the nitrogen molecule. The importance of chlorine was established by Molina
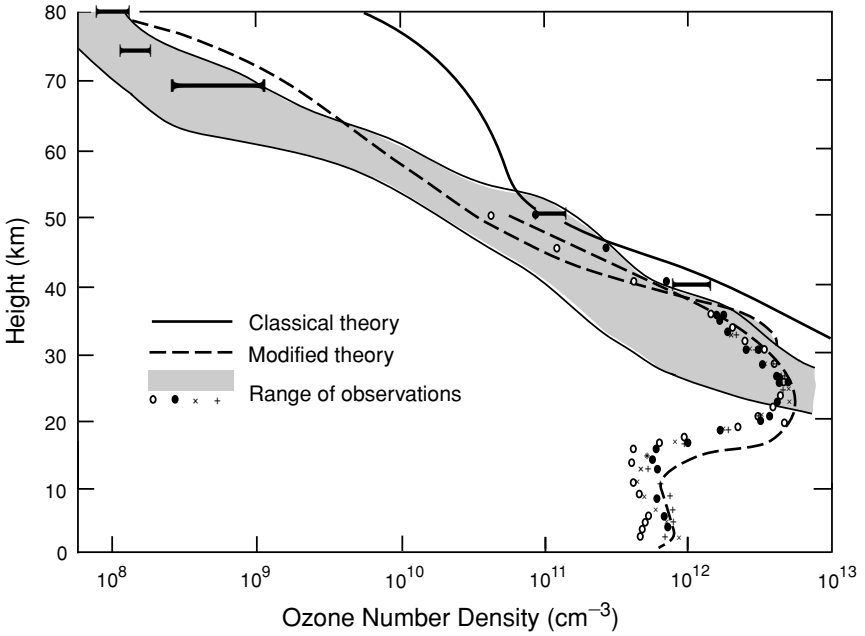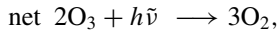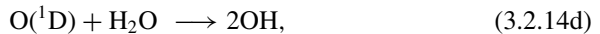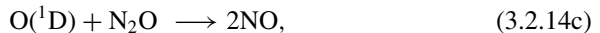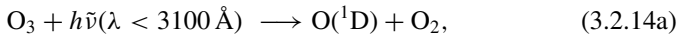
Figure 3.8   Observational range in ozone number densities and theoretical calculations for equilibrium ozone number densities (data from Leovy, 1969, 20–80 km; Logan *et al.,* 1978, 0–50 km).

and Rowland (1974) from their study of the chlorofluorocarbons (CFCs) produced by refrigerants and spray-can propellants. Ozone reduction from the effect of nitrogen oxides was first reported by Crutzen (1970) and that from the exhaust of a fleet of supersonic transports was documented by Johnston (1971).

The catalytic destruction reactions of ozone have been found to be

$$O_3 + h\tilde{\nu} \longrightarrow O + O_2, \tag{3.2.13a}$$

$$O + XO \longrightarrow X + O_2, \tag{3.2.13b}$$

$$X + O_3 \longrightarrow XO + O_2, \tag{3.2.13c}$$

$$\text{net } 2O_3 + h\tilde{\nu} \longrightarrow 3O_2,$$

where X may be nitric oxide (NO), chlorine (Cl), hydroxyl radical (OH), or atomic hydrogen (H). The possible sources of NO and OH are the following reactions:

$$O_3 + h\tilde{\nu}(\lambda < 3100\,\text{Å}) \longrightarrow O(^1D) + O_2, \tag{3.2.14a}$$

$$O(^1D) + M \longrightarrow O + M, \tag{3.2.14b}$$

$$O(^1D) + N_2O \longrightarrow 2NO, \tag{3.2.14c}$$

$$O(^1D) + H_2O \longrightarrow 2OH, \tag{3.2.14d}$$

where O(^{1}D) denotes the excited atomic oxygen in the ^{1}D state, which is essential to these reactions. It is clear that the high concentration of ozone from Chapman's theoretical prediction is due to the neglect of these additional loss mechanisms and possibly others. If we introduce reactions involving (3.2.13a)–(3.2.14d), the calculated equilibrium ozone concentration labeled as the modified theory appears to closely match the observed values (Logan *et al.,* 1978).

Ozone is a natural trace ingredient of the atmosphere that occurs at an average concentration of about 3 parts per million by volume. Its concentration varies with season and latitude and is modulated by radiative and dynamic processes in the middle atmosphere. High intensities of UV radiation shorter than 3200 Å, which are harmful to nearly all forms of life, are largely ($\sim$99%) screened out by ozone. It has been postulated that surface life on earth did not evolve until after the ozone layer was formed. The effect of small increases in the intensity of ultraviolet radiation due to the reduction of ozone by human activities has been a subject of extensive scientific research.

More recently, there has been continuous concern with the question of whether the ozone layer in the stratosphere is actually thinning globally because of the increased use of CFCs. A decrease in Antarctic ozone during the austral spring, just after the end of the Antarctic polar night, has been discovered. Data from the British Antarctic Survey collected at the Halley Bay station indicate a 40% total ozone decrease during the month of October between 1977 and 1984 (Farman *et al.,* 1985). This decrease has been confirmed by data derived from the Total Ozone Mapping Spectrometer and the Solar Backscatter Ultraviolet Spectrometer on board Nimbus 7. The largest percentage decrease corresponds roughly to the polar vortex, covering approximately the entire area of the Antarctic continent. It has been suggested that the so-called *ozone hole* is the precursor of a major decrease in the ozone layer worldwide caused by the addition of CFCs to the atmosphere. It has also been noted that the ozone hole appears to have a dynamic origin and is related to the rather special meteorological conditions prevalent over Antarctica each October. The cold polar vortex traps the air parcels for weeks, during which time *polar stratospheric clouds* are formed. These clouds may facilitate certain chemical reactions that favor the destruction of ozone (Toon and Turco, 1991).

3.2.3   Absorption in the Visible and Near Infrared

3.2.3.1   MOLECULAR OXYGEN AND OZONE

Molecular oxygen has absorption bands in the visible and near-IR. The ground electronic state of O_2 is designated by X and two excited states by a and b. The a$\leftarrow$X and b$\leftarrow$X transitions coupled with vibrational–rotational transitions produce weak absorption lines in the near-IR and visible, respectively. These are referred to as *infrared bands* and *red bands*. The most important red bands are the A band centered at 0.762 μm (0$\leftarrow$0), the B band at 0.688 μm (1$\leftarrow$0), and the γ band at 0.628 μm (2$\leftarrow$0). See Fig. 3.4 for the meaning of b$\leftarrow$X (ground state electronic transition)

and $0 \leftarrow 0$ (vibrational transition). Because the positions of these three bands are near the peak of the solar spectrum, absorption of the solar flux due to O_2 in the visible is important in the middle and upper atmospheres and could affect the solar flux available at the surface and in the troposphere. Line-by-line information for O_2 is available and a precise computation of the band absorption can be performed. The oxygen A band has been utilized for the remote sensing of the level of cloud pressure from space. Absorption of O_3 in the visible and near-IR, the Chappuis band, was discussed in Subsection 3.2.1.3.

3.2.3.2 WATER VAPOR

The water vapor molecule has an asymmetric top (bent triatomic) configuration with the oxygen atom in the middle and a bond angle of $104.45°$, as shown in Fig. 3.3. The distance between the oxygen and hydrogen atoms is 0.958 Å. It has three rotational constants. The hydrogen atom has two isotopic forms, ^{1}H and ^{2}D, with relative abundances of 99.9851 and 0.0149%, respectively. The isotopic forms of water vapor that are important in infrared radiative transfer are $HH^{16}O$, $HH^{18}O$, $HD^{16}O$, and $HD^{18}O$. Each of these molecules has a different vapor pressure, and their abundances are a function of the hydrological cycle. The H_2O molecule has three fundamental vibration modes (Fig. 3.3). The bending vibration, v_2, has the lowest wavenumber; both v_1 and v_3 have wavenumbers about twice the wavenumber for v_2. The v_2 fundamental band of H_2O is centered at 6.25 μm and is important in thermal infrared radiation transfer and remote sensing applications (Section 4.2.1). The v_1 and v_3 fundamentals of H_2O produce bands centered at 3657.05 cm^{-1} (2.74 μm, $100 \leftarrow 000$) and 3755.93 cm^{-1} (2.66 μm, $001 \leftarrow 000$), respectively. These two bands are close to one another and combine to form a strong band in the solar spectrum, referred to as the 2.7 μm band. The three integers represent the three vibrational quantum numbers $v_1 v_2 v_3$. The ground state is designated by 000. As described in Subsection 1.3.1.2, the transition from the ground state to the first excited state 001 or 100 is referred to as the fundamental. The $2v_2$ band is centered at 3161.60 cm^{-1} (3.2 μm band, $020 \leftarrow 000$) in the tail of the solar spectrum. In addition, the solar spectrum contains a large number of overtone and combination bands, which arise from ground state transitions. In the near-IR region, these bands absorb a significant amount of solar flux in the lower atmosphere. They are centered at 0.94, 1.1, 1.38, and 1.87 μm, and are commonly identified in groups by the Greek letters (ρ, σ, τ), ϕ, Ψ, and ω, respectively. Although the overtone and combination bands centered at 0.72 and 0.82 μm are relatively weak, their contribution to the solar heating of the atmosphere appears to be not insignificant. There are also a number of H_2O lines in the visible region associated with overtone and combination transitions. However, their contributions to absorption and the heating rate appear to be quite small.

3.2.3.3 CARBON DIOXIDE

The CO_2 molecule has a linear symmetrical configuration, with the carbon atom in the middle and an oxygen atom on each side, as shown in Fig. 3.3. The length of the C–O bond in the fundamental vibration is 1.1632 Å and it has one rotational constant.

The natural isotopes of the carbon atom are ^{12}C and ^{13}C with relative abundances of 98.892 and 1.108%, respectively. For the oxygen atom, the isotopes are ^{16}O (99.758%), ^{17}O (0.0373%), and ^{18}O (0.2039%). Thus, several CO_2 isotopes are present in the atmosphere. The most significant of these for the radiation problem are $^{12}C^{16}O^{16}O$, $^{13}C^{16}O^{16}O$, and $^{12}C^{16}O^{18}O$.

As a result of its symmetrical arrangement, the CO_2 molecule has no permanent dipole moment and no permitted rotation band. The symmetrical stretch mode, ν_1, is radiatively inactive at its fundamental. The bending mode, ν_2, produces the most important band in the thermal infrared: the 15 μm CO_2 band. This subject will be discussed further in Section 4.2.1. Because of asymmetric stretching vibration, the ν_3 fundamental transition has a wavenumber centered at 2349.16 cm^{-1} (the 4.3 μm band) for $^{12}C^{16}O^{16}O$. In addition, the combination band $\nu_1 + \nu_3 - 2\nu_2$ is centered at 2429.37 cm^{-1} and occurs when the ν_3 transition originates at the vibrational level $\nu = 2$ or higher. Because of parallel vibration, the Q branch corresponding to $\Delta J = 0$ does not appear. There are also a number of bands for the isotopes. The 4.3 μm CO_2 band is in the tail of the solar spectrum and its effect on solar absorption is insignificant. This band has been used in connection with the 15 μm band for the nighttime detection of atmospheric temperatures (see Section 7.4.3). Carbon dioxide exhibits a number of rather weak overtone and combination bands in the solar region: 2.0, 1.6, and 1.4 μm. The stronger 2.7 μm band of CO_2 overlaps with the 2.7 μm band of H_2O and contributes to the absorption of solar flux in the lower stratosphere.

3.2.3.4   OTHER MINOR GASES

The N_2O molecule is linear and asymmetric (NNO). The three fundamentals are ν_1 (1284.91 cm^{-1}), ν_2 (558.77 cm^{-1}), and ν_3 (2223.76 cm^{-1}). The ν_3 band (4.5 μm) is in the solar region, but is insignificant in solar absorption. There are also four overtone and combination bands: 4.06, 3.90, 2.97, and 2.87 μm. The CO molecule has fundamental (2.34 μm) and first-overtone (4.67 μm) bands in the solar spectrum. The CH_4 molecule is a spherical top with four independent (ν_1, ν_2, ν_3, ν_4) fundamentals. The ν_3 fundamental occurs at 3018.92 cm^{-1} (3.31 μm). Other overtone and combination bands are located at 3.83, 3.53, 3.26, 2.37, 2.30, 2.20, and 1.66 μm. The NO_2 molecule absorbs solar flux in the wavelength range 0.2–0.7 μm, as pointed out in Subsection 3.2.1.4. Absorption of solar radiation by the transitional molecules, $O_2 \cdot O_2$ and $O_2 \cdot N_2$, has also been reported (Solomon *et al.,* 1998). A summary of the preceding absorption bands is given in Table 3.3.

3.2.3.5   TRANSFER OF DIRECT SOLAR FLUX IN THE ATMOSPHERE

The monochromatic downward solar flux density, following the discussion in Section 1.4.2, may be written as

$$F_\lambda(z) = \mu_0 F_\odot(\lambda) \exp\left(-\frac{k_\lambda u(z)}{\mu_0}\right), \tag{3.2.15}$$

Table 3.3

Solar Absorption Bands of Atmospheric Gases

Species	Band (μm)	$\Delta\nu$ (cm^{-1})	Major transitions
	3.2, 2.7	2500–4500	020, 001, 100
	1.87	4800–6200	110, 011
	1.38	6400–7600	021, 200, 101
	1.1	8200–9400	111
H_2O	0.94	10,100–11,300	121, 201, 003
	0.82	11,700–12,700	211
	0.72	13,400–14,600	221, 202, 301
	Visible	15,000–22,600	500, 203
	4.3	2000–2400	00^01
	2.7	3400–3850	10^01
CO_2	2.0	4700–5200	20^01
	1.6	6100–6450	30^01
	1.4	6850–7000	00^03
	4.74	2000–2300	101
	3.3	3000–3100	003
O_3	Visible	10,600–22,600	—
	UV	29,000–50,000	—
	1.58	6300–6350	$a \leftarrow X\ (0 \leftarrow 1)$
	1.27	7700–8050	$a \leftarrow X\ (0 \leftarrow 0)$
	1.06	9350–9400	$a \leftarrow X\ (1 \leftarrow 0)$
O_2	0.76(A)	12,850–13,200	$b \leftarrow X\ (0 \leftarrow 0)$
	0.69(B)	14,300–14,600	$b \leftarrow X\ (1 \leftarrow 0)$
	0.63(γ)	14,750–15,900	$b \leftarrow X\ (2 \leftarrow 0)$
$O_2 \cdot O_2$	Visible	7600–30,000	—
$O_2 \cdot N_2$	1.26	7600–8300	—
	4.5	2100–2300	00^01
N_2O	4.06, 3.9	2100–2800	$12^00, 20^00$
	2.97, 2.87	3300–3500	$02^01, 10^01$
	3.83, 3.53		
	3.31, 3.26	2500–3200	0002, 0101, 0200
CH_4	2.37, 2.30		
	2.20	4000–4600	1001, 0011, 0110
	1.66	5850–6100	0020
CO	4.67	2000–2300	1
	2.34	4150–4350	2
NO_2	Visible	14,400–50,000	—

where $k_\lambda u$ represents the optical depth with k_λ the absorption coefficient (cm^2 g^{-1}). It is customary to use the path length (g cm^{-2}) to denote the amount of absorber, particularly for water vapor, in the form

$$u(z) = \int_z^{z_\infty} \rho_a(z')\, dz', \qquad (3.2.16)$$

where ρ_a denotes the density of the absorbing gas. The flux density here is a hemispheric

quantity weighted by the cosine of the solar zenith angle. It is this flux density that is related to the solar heating rate defined in Section 3.5.

Because of the structure of absorption lines, it is important to define the spectral absorptance, also referred to as absorptivity or absorption, in the form

$$A_{\bar{\lambda}}(u/\mu_0) = \int_{\Delta\lambda} \left(1 - e^{-k_\lambda u/\mu_0}\right) \frac{d\lambda}{\Delta\lambda}, \tag{3.2.17}$$

where $\Delta\lambda$ denotes the spectral interval. If the solar flux variation is small in this interval, the downward spectral flux density may be expressed by

$$F_{\bar{\lambda}}(z) \cong \mu_0 F_\odot(\bar{\lambda})[1 - A_{\bar{\lambda}}(u/\mu_0)]. \tag{3.2.18}$$

Computation of the spectral absorptance is the key to the evaluation of the downward solar flux in the atmosphere, which can be done by line-by-line techniques or simplified methodologies such as the correlated k-distribution method to be introduced
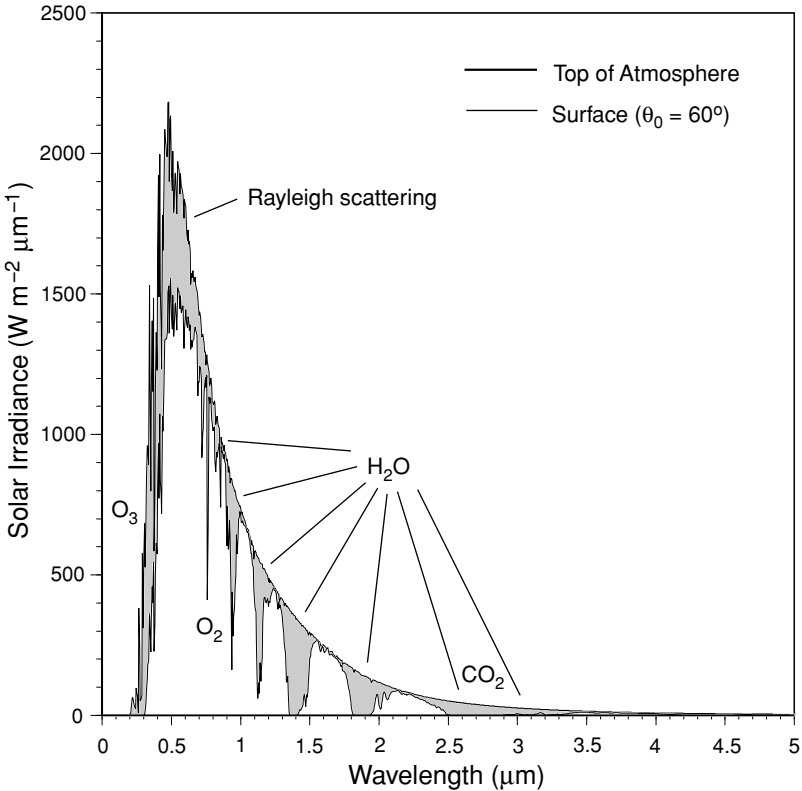
Figure 3.9   Solar irradiance curve for a 50 cm^{-1} spectral interval at the top of the atmosphere (see Fig. 2.9) and at the surface for a solar zenith angle of 60° in an atmosphere without aerosols or clouds. Absorption and scattering regions are indicated. See also Table 3.3 for the absorption of N_2O, CH_4, CO, and NO_2.

in Section 4.5. Exercise 3.6 requires the evaluation of spectral absorptance in the limit of strong line approximation using the Lorentz line shape. The total amount of water vapor can be estimated by this approximation using the 0.94 μm band, which is the foundation for the determination of precipitable water from sunphotometers (see Section 7.2).

Figure 3.9 shows the depletion of solar flux in a clear atmosphere. The top curve is an observed solar irradiance with a spectral resolution of 50 cm^{-1} at the top of the earth's atmosphere as depicted in Fig 2.9. The depletion of solar irradiance in the UV region (<0.4 μm) is chiefly due to the absorption of molecular oxygen and ozone discussed previously. In the visible, the depletion of solar flux is caused by the absorption produced by oxygen red bands, the ozone Chappuis band, and some water vapor weak bands; but the chief attenuation is associated with Rayleigh scattering to be discussed in Section 3.3. In the near-IR, the prime absorber is water vapor with contributions from carbon dioxide in the 2.7 μm band. Other minor absorbers such as N_2O, CO, and CH_4 also contribute to the depletion of solar flux but are less significant. It is evident that water vapor is the most important absorber in the solar near-IR spectrum, which contains about 50% of solar energy.

3.3   Atmospheric Scattering

3.3.1   Rayleigh Scattering

The simplest and in some ways the most important example of a physical law of light scattering with various applications is that discovered by Rayleigh (1871). His findings led to the explanation of the blue color of the sky. In this section we formulate the scattering of unpolarized sunlight by air molecules and describe its important application to the atmosphere.

3.3.1.1   THEORETICAL DEVELOPMENT
Consider a small homogeneous, isotropic, spherical particle whose radius is much smaller than the wavelength of the incident radiation. The incident radiation produces a homogeneous electric field $\mathbf{E}_0$, called the *applied field*. Since the particle is very small, the applied field generates a dipole configuration on it. The electric field of the particle, caused by the electric dipole, modifies the applied field inside and near the particle. Let $\mathbf{E}$ be the combined field, i.e., the applied field plus the particle's own field, and further, let $\mathbf{p}_0$ be the induced dipole moment. Then we apply the electrostatic formula to give

$$\mathbf{p}_0 = \alpha \mathbf{E}_0. \tag{3.3.1}$$

This equation defines the polarizability α of a small particle. The dimensions of $\mathbf{E}_0$ and $\mathbf{p}_0$ are in units of charge per area and charge times length, respectively, and α has the dimension of volume. In general, α is a tensor, because the vectors $\mathbf{p}_0$ and $\mathbf{E}_0$ may not align along the three mutually perpendicular directions. In the very common case where these two vectors coincide, α is a scalar.

The applied field $\mathbf{E}_0$ generates oscillation of an electric dipole in a fixed direction. The oscillating dipole, in turn, produces a plane-polarized electromagnetic wave, the scattered wave. To evaluate the scattered electric field in regions that are far away from the dipole, we let r denote the distance between the dipole and the observation point, γ the angle between the scattered dipole moment $\mathbf{p}$ and the direction of observation, and c the velocity of light. According to the classical electromagnetic solution given by Hertz (1889), the scattered electric field is proportional to the acceleration of the scattered dipole moment and $\sin \gamma$, but is inversely proportional to the distance r. Thus, we have

$$\mathbf{E} = \frac{1}{c^2} \frac{1}{r} \frac{\partial^2 \mathbf{p}}{\partial t^2} \sin \gamma. \tag{3.3.2}$$

In an oscillating periodic field, the scattered dipole moment may be written in terms of the induced dipole moment as

$$\mathbf{p} = \mathbf{p}_0 e^{-ik(r-ct)}. \tag{3.3.3}$$

Note that k is the wavenumber, and $kc = \omega$ is the circular frequency. By combining Eqs. (3.3.1) and (3.3.3), Eq. (3.3.2) yields

$$\mathbf{E} = -\mathbf{E}_0 \frac{e^{-ik(r-ct)}}{r} k^2 \alpha \sin \gamma. \tag{3.3.4}$$

Now we consider the scattering of sunlight by air molecules. Let the plane defined by the directions of incident and scattered waves be the reference plane (plane of scattering). Since any electric vector may be arbitrarily decomposed into orthogonal components, we may choose the two components perpendicular (E_r) and parallel (E_l) to the plane of scattering. The sunlight is characterized by the same electric field in the r and l directions and by a random phase relation between these two components, and is referred to as *natural* or *unpolarized* light (see Section 6.6 for a more advanced discussion of the representation of polarized light). In this case, we may consider separately the scattering of the two electric field components E_{0r} and E_{0l} by molecules assumed to be homogeneous, isotropic, spherical particles. Based on Eq. (3.3.4), we have

$$E_r = -E_{0r} \frac{e^{-ik(r-ct)}}{r} k^2 \alpha \sin \gamma_1, \tag{3.3.5a}$$

$$E_l = -E_{0l} \frac{e^{-ik(r-ct)}}{r} k^2 \alpha \sin \gamma_2. \tag{3.3.5b}$$

Referring to Fig. 3.10, we see that $\gamma_1 = \pi/2$ and $\gamma_2 = \pi/2 - \Theta$, where Θ is defined as the scattering angle, which is the angle between the incident and scattered waves. Note that γ_1 is always equal to $90°$ because the scattered dipole moment (or the scattered electric field) in the r direction is normal to the scattering plane defined previously.

In matrix form, we may write

$$\begin{bmatrix} E_r \\ E_l \end{bmatrix} = -\frac{e^{-ik(r-ct)}}{r} k^2 \alpha \begin{bmatrix} 1 & 0 \\ 0 & \cos \Theta \end{bmatrix} \begin{bmatrix} E_{0r} \\ E_{0l} \end{bmatrix}. \tag{3.3.6}$$
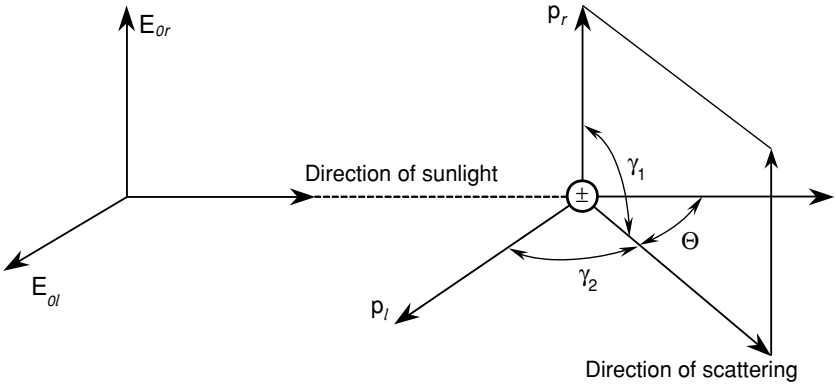
Figure 3.10 Scattering by a dipole. The incident electric field, a vector, can be arbitrarily decomposed into a parallel (l) and a perpendicular (r) component, where each undergoes the scattering by the dipole. We may select the component that is always perpendicular to the scattering plane that is defined by the incident and scattering beams (i.e., $\gamma_1 = 90°$). All the notations are defined in the text.

A complete description of the intensity of a light beam and its polarized state will be given in Section 6.6 in which the Stokes parameters are introduced. For the sake of the continuity of the present discussion, however, we may define the intensity components (per solid angle) of the incident and scattered radiation in the forms $I_0 = C|E_0|^2$ and $I = C|E|^2$, where C is a certain proportionality factor such that C/r^2 implies a solid angle. It follows that Eqs. (3.3.5) and (3.3.6) can be expressed in the form of intensities as

$$I_r = I_{0r} k^4 \alpha^2 / r^2, \tag{3.3.7a}$$

$$I_l = I_{0l} k^4 \alpha^2 \cos^2 \Theta / r^2, \tag{3.3.7b}$$

where I_r and I_l are polarized intensity components perpendicular and parallel to the plane containing the incident and scattered waves, i.e., the plane of scattering. The total scattered intensity of the unpolarized sunlight incident on a molecule in the direction of Θ is then

$$I = I_r + I_l = (I_{0r} + I_{0l} \cos^2 \Theta) k^4 \alpha^2 / r^2. \tag{3.3.8}$$

But for unpolarized sunlight, $I_{0r} = I_{0l} = I_0/2$, and by noting that $k = 2\pi/\lambda$, we obtain

$$I = \frac{I_0}{r^2} \alpha^2 \left(\frac{2\pi}{\lambda} \right)^4 \frac{1 + \cos^2 \Theta}{2}. \tag{3.3.9}$$

This is the original formula derived by Rayleigh, and we call the scattering of sunlight by molecules *Rayleigh scattering*. By this formula, the intensity of unpolarized sunlight scattered by a molecule is proportional to the incident intensity I_0 and is inversely proportional to the square of the distance between the molecule and the

point of observation. In addition to these two factors, the scattered intensity also de-
pends on the polarizability, the wavelength of the incident wave, and the scattering
angle. The dependence of these three parameters on the scattering of sunlight by
molecules introduces a number of significant physical features.

3.3.1.2 PHASE FUNCTION, SCATTERING CROSS SECTION, AND POLARIZABILITY

On the basis of Eqs. (3.3.7) and (3.3.9), the intensity scattered by a molecule
depends on the polarization characteristics of the incident light. For vertically (r)
polarized incident light, the scattered intensity is independent of the direction of the
scattering plane. In this case then, the scattering is isotropic. On the other hand,
for horizontally (l) polarized incident light, the scattered intensity is a function of
$\cos^2 \Theta$. When the incident light is unpolarized, such as sunlight, the scattered intensity
depends on $(1 + \cos^2 \Theta)$. The angular scattering patterns in space for the three types
of incident polarization are illustrated in Fig. 3.11. We see that the scattering of
unpolarized sunlight by molecules (Rayleigh scattering) has maxima in the forward
($0°$) and backward ($180°$) directions, whereas it shows minima in the side directions
($90°$ and $270°$). Light scattered by particles or molecules is not confined only to
the plane of incidence, but is visible in all azimuthal directions. Because of the
spherical symmetry assumed for molecules, scattering patterns are symmetrical in
three-dimensional space, as demonstrated in Fig. 3.11.

To describe the angular distribution of scattered energy in conjunction with mul-
tiple scattering and radiative transfer analyses and applications for planetary atmo-
spheres, we find it necessary to define a nondimensional parameter called the *phase
function, $P(\cos \Theta)$*, such that

$$\int_0^{2\pi} \int_0^{\pi} \frac{P(\cos \Theta)}{4\pi} \sin \Theta \, d\Theta \, d\phi = 1. \tag{3.3.10}$$

By this definition, the phase function is said to be normalized to unity. Upon per-
forming simple integrations, the phase function of Rayleigh scattering for incident
unpolarized sunlight is given by

$$P(\cos \Theta) = \frac{3}{4}(1 + \cos^2 \Theta). \tag{3.3.11}$$

Employing the definition of the phase function, Eq. (3.3.9) may be rewritten in the
form

$$I(\Theta) = \frac{I_0}{r^2} \alpha^2 \frac{128\pi^5}{3\lambda^4} \frac{P(\Theta)}{4\pi}. \tag{3.3.12}$$

It follows that the angular distribution of the scattered intensity is directly proportional
to the phase function.

The scattered flux f (or power, in units of energy per time) can be evaluated by
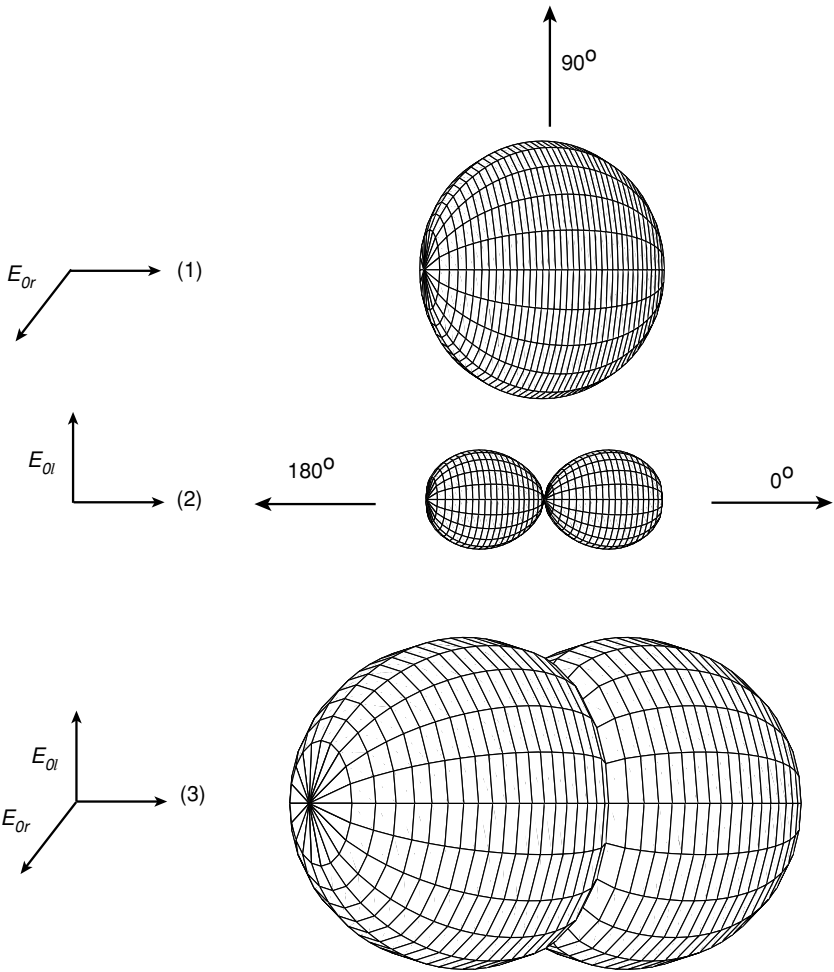integrating the scattered flux density ($I\Delta\Omega$) over the appropriate area a distance r

Figure 3.11 Polar diagram of the scattered intensity for Rayleigh molecules: (1) polarized incident light with the electric vector perpendicular to the scattering plane, (2) polarized incident light with the electric vector on the scattering plane, and (3) unpolarized incident light.

away from the scatterer. Thus,

$$f = \int_{\Omega} (I \Delta \Omega) r^2 \, d\Omega, \tag{3.3.13a}$$

where $r^2 d\Omega$ represents the area according to the definition of the solid angle. Inserting the expressions for scattered intensity and the differential solid angle defined in Eqs. (3.3.12) and (1.1.5), respectively, into Eq. (3.3.13a) and carrying out integrations over the solid angle of a sphere, we obtain the equivalent isotropically scattered flux

in the form

$$f = F_0 \alpha^2 128\pi^5/(3\lambda^4), \tag{3.3.13b}$$

where the incident flux density F_0 is equal to $I_0 \Delta\Omega$. Moreover, we may define the scattering cross section per one molecule as

$$\sigma_s = f/F_0 = \alpha^2 128\pi^5/(3\lambda^4). \tag{3.3.14}$$

The scattering cross section (in units of area) represents the amount of incident energy that is removed from the original direction because of a single scattering event such that the energy is redistributed isotropically on the area of a sphere whose center is the scatterer and whose radius is r. In terms of the scattering cross section, the scattered intensity can be expressed by

$$I(\Theta) = I_0 \frac{\sigma_s}{r^2} \frac{P(\Theta)}{4\pi}. \tag{3.3.15}$$

This is the general expression for scattered intensity, which is valid not only for molecules but also for particles whose size is larger than the incident wavelength, as will be discussed in Section 5.2.

The *polarizability* α, which was used in the preceding equations, can be derived from the principle of the dispersion of electromagnetic waves and is given by

$$\alpha = \frac{3}{4\pi N_s} \left(\frac{m^2 - 1}{m^2 + 2}\right), \tag{3.3.16}$$

where N_s is the total number of molecules per unit volume and m is the nondimensional refractive index of molecules. This equation is called the *Lorentz–Lorenz formula,* and its derivation is given in Appendix D. The refractive index is an optical parameter associated with the velocity change of electromagnetic waves in a medium with respect to a vacuum. Its definition and physical meanings are also given in Appendix D. Normally, the refractive indices of atmospheric particles and molecules are composed of a real part m_r and an imaginary part m_i corresponding, respectively, to the scattering and absorption properties of particles and molecules. In the solar visible spectrum, the imaginary parts of the refractive indices of air molecules are so insignificantly small that absorption of solar radiation by air molecules may be neglected in the scattering discussion. The real parts of the refractive indices of air molecules in the solar spectrum are very close to 1, but they depend on the wavelength (or frequency) of the incident radiation as illustrated in Appendix D. Because of this dependence, white light may be *dispersed* into component colors by molecules that function like prisms. The real part of the refractive index derived in Appendix D [(Eq. D.17)] may be approximately fitted by

$$(m_r - 1) \times 10^8 = 6432.8 + \frac{2{,}949{,}810}{146 - \lambda^{-2}} + \frac{25{,}540}{41 - \lambda^{-2}}, \tag{3.3.17}$$

where λ is in units of micrometers. Since m_r is close to 1, for all practical purposes, Eq. (3.3.16) may be approximated by

$$\alpha \approx \frac{1}{4\pi N_s}(m_r^2 - 1). \tag{3.3.18}$$

Thus, the scattering cross section defined in Eq. (3.3.14) becomes

$$\sigma_s = \frac{8\pi^3(m_r^2 - 1)^2}{3\lambda^4 N_s^2} f(\delta). \tag{3.3.19}$$

A correction factor $f(\delta)$ is added in Eq. (3.3.19) to take into consideration the anisotropic property of molecules, where $f(\delta) = (6 + 3\delta)/(6 - 7\delta)$ with the anisotropic factor δ of 0.035. Anisotropy implies that the refractive index of molecules varies along the x, y, and z directions, and thus is a vector, not a scalar. Hence, the polarizability α is a tensor, as noted previously.

The optical depth of the entire molecular atmosphere at a given wavelength may be calculated from the scattering cross section in the form

$$\tau(\lambda) = \sigma_s(\lambda) \int_0^{z_\infty} N(z)\, dz, \tag{3.3.20}$$

where $N(z)$ denotes the number density of molecules as a function of height, and z_∞ is the top of the atmosphere. The optical depth represents the attenuation power of molecules with respect to a specific wavelength of the incident light. Exercises 3.7–3.11 require the calculation of a number of parameters based on Rayleigh scattering results.

3.3.1.3 BLUE SKY AND SKY POLARIZATION

Returning to Eq. (3.3.12), we see that the scattered intensity depends on the wavelength of incident light and the index of refraction of air molecules contained in the polarizability term. According to the analyses given in Appendix D and Eq. (3.3.17), the index of refraction also depends slightly on the wavelength. However, the dependence of the refractive index on the wavelength is relatively insignificant in calculating the scattered intensity as compared to the explicit wavelength term. Thus, the intensity scattered by air molecules in a specific direction may be symbolically expressed in the form

$$I_\lambda \sim 1/\lambda^4. \tag{3.3.21}$$

The inverse dependence of the scattered intensity on the wavelength to the fourth power is a direct consequence of the theory of Rayleigh scattering and is the foundation for the explanation of blue sky.

In reference to the observed solar energy spectrum displayed in Fig. 3.9, a large portion of solar energy is contained between the blue and red regions of the visible spectrum. Blue light ($\lambda \approx 0.425\ \mu$m) has a shorter wavelength than red light ($\lambda \approx 0.650\ \mu$m). Consequently, according to Eq. (3.3.21) blue light scatters about 5.5 times

more intensity than red light. It is apparent that the λ^{-4} law causes more blue light to be scattered than red, green, and yellow, and so the sky, when viewed away from the sun's disk, appears blue. Moreover, since molecular density decreases drastically with height, it is anticipated that the sky should gradually darken to become completely black in outer space in directions away from the sun. And the sun itself should appear whiter and brighter with increasing height. As the sun approaches the horizon (at sunset or sunrise), sunlight travels through more air molecules, and therefore more and more blue light and light with shorter wavelengths are scattered out of the beam of light, and the luminous sun shows a deeper red color than at its zenith. However, since violet light ($\sim 0.405\ \mu m$) has a shorter wavelength than blue, a reasonable question is, why doesn't the sky appear violet? This is because the energy contained in the violet spectrum is much less than that contained in the blue spectrum, and also because the human eye has a much lower response to the violet color.

Another important phenomenon explained by the Rayleigh scattering theory is sky polarization. For many atmospheric remote sensing applications utilizing polarization, a parameter called the *degree of linear polarization* has been used (Subsection 7.3.5.2). In the case of Rayleigh scattering it is given by

$$LP(\Theta) = -\frac{I_l - I_r}{I_l + I_r} = -\frac{\cos^2 \Theta - 1}{\cos^2 \Theta + 1} = \frac{\sin^2 \Theta}{\cos^2 \Theta + 1}. \qquad (3.3.22)$$

In the forward and backward directions the scattered light remains completely unpolarized, whereas at the $90°$ scattering angle, the scattered light becomes completely polarized. In other directions, the scattered light is partially polarized with the percentage of polarization ranging from 0 to 100%. Interested readers may wish to refer to Section 6.6 for further details on this subject.

The theory of Rayleigh scattering developed in Section 3.3.1 is based on the assumption that molecules are homogeneous and isotropic spheres. However, molecules are in general anisotropic, whereby their polarizability, as defined in Eq. (3.3.16), varies along three axes and, hence, is a tensor instead of a scalar. The anisotropic effect of molecules reduces the degree of linear polarization defined in Eq. (3.3.22) by only a small percentage. At the $90°$ scattering angle, the degree of linear polarization for dry air is about 0.94. Further, the theory of Rayleigh scattering developed previously considers only single (or primary) scattering, i.e., where scattering occurs only once. But in the earth's atmosphere, which contains a large number of molecules and aerosol particles, light may undergo an infinite number of scattering events. In addition, the earth's surface also reflects light that reaches it. Multiple scattering processes involving the atmosphere and the surface become complicated and require a more advanced treatment of radiative transfer theory, which will be discussed in Chapter 6.

The theory of Rayleigh scattering predicts *neutral points,* i.e., points of zero polarization, only at the exact forward and backward directions. However, owing to multiple scattering of molecules and particulates, and reflection of the surface, there normally exist a number of neutral points in cloudless atmospheres. The first observations of neutral points and partially polarized sky light were made by Arago in

1809. He discovered the existence of a neutral point at a position in the sky about $25°$ above the antisolar direction (the direction exactly opposite that of the sun). The other two neutral points, which normally occur in the sunlit sky $25°$ above and $20°$ below the sun, were discovered by Babinet in 1840 and by Brewster in 1842, respectively. These three neutral points were named to honor these three discoverers. The neutral points in the sky vary and depend on the turbidity (an indication of the amount of aerosol loadings in the atmosphere), the sun's elevation angle, and the reflection characteristics of the surface at which observations are made.

Figure 3.12 illustrates the distribution of the degree of polarization and neutral points for a pristine, clear atmosphere (January 20, 1977) and for an atmospheric condition under the El Chichon volcanic cloud (July 27, 1982) observed at the Mauna Loa Observatory from a polarimeter developed by Coulson (1983). The observations
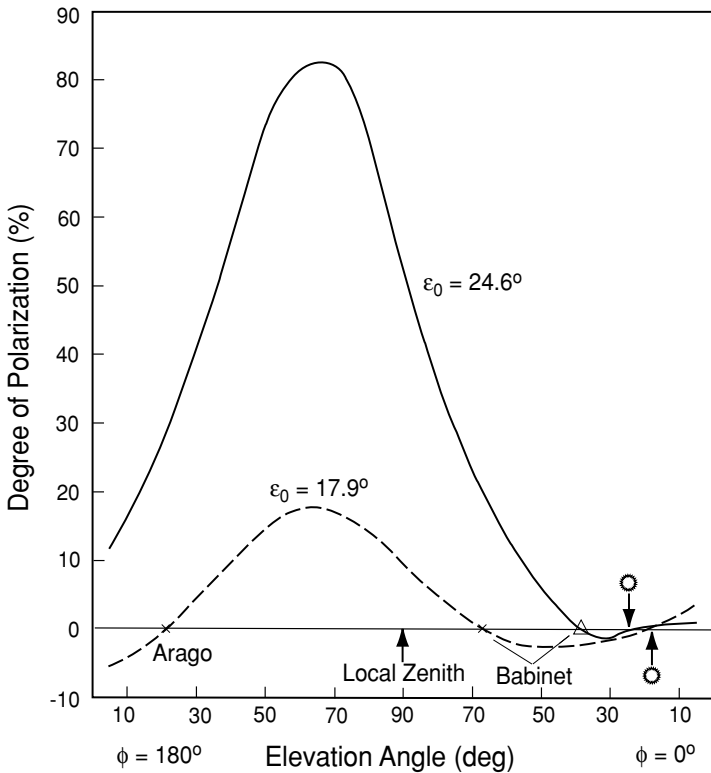
Figure 3.12 Illustration of neutral points in the distribution of the degree of polarization through the plane of the sun's vertical at a wavelength of 0.7 μm observed at the Mauna Loa Observatory for a clear atmospheric condition on January 20, 1977 (solid line), and for an atmosphere under the volcanic cloud on July 27, 1982 (dashed line). The azimuthal angles $\phi = 0°$ and $\phi = 180°$ are on the sun's vertical plane. The sun's elevation angles ε_0 for these two cases are indicated in the graph, as are the positions of Arago and Babinet (data taken from Coulson, 1983).

were made on the sun's vertical plane, referred to as the principal plane in radiative transfer, using a wavelength of 0.7 μm. The solar elevation angle, ε_0 (90° − solar zenith angle θ_0), differed slightly on these two dates, but the observed polarization patterns suffice to demonstrate their substantial variabilities in clear and turbid atmospheres. The clear Rayleigh atmosphere produced a maximum polarization of about 80%, 60% more than that generated in the volcanic cloud condition. The neutral points in the Rayleigh scattering atmosphere occurred at the positions close to the sun (forward direction) and about 20° above the sun, the Babinet point, which was about 50° above the sun when a significant aerosol loading was present. In this case, the Arago point was also shown at about 20° above the horizon at the opposite position of the sun. Because of the sun's position, the Brewster point was not observed. The neutral points' positions are dependent on the aerosol optical depth and composition. Thus, a systematic observation of these points could be a valuable approach for inferring aerosol optical properties and perhaps composition information.

3.3.2 Light Scattering by Particulates: Approximations

In Section 1.1.4, we defined the size parameter, $x = 2\pi a/\lambda$, where a is the particle radius. Rayleigh scattering is concerned with scattering events when $x \ll 1$. When $x \gtrsim 1$, scattering events are often called *Lorenz–Mie scattering*. Lorenz (1890) and Mie (1908) independently derived the solution for the interaction of a plane wave with an isotropic homogenous sphere. The mathematical theory of Lorenz–Mie scattering begins with Maxwell's equations and will be detailed in Chapter 5, along with some new developments in research on light scattering by nonspherical ice crystals and aerosols. In this section, however, we shall present a brief discussion of Lorenz–Mie scattering and two elementary approximations: geometric optics and anomalous diffraction.

3.3.2.1 LORENZ–MIE SCATTERING
The intensity scattered by a particle as a function of direction, as presented in Eq. (3.3.15), is given by

$$I(\Theta) = I_0 \Omega_{\text{eff}} \frac{P(\Theta)}{4\pi} = I_0 \left(\frac{\sigma_s}{r^2} \right) \frac{P(\Theta)}{4\pi}, \qquad (3.3.23)$$

where I_0 is the incident intensity, P is the phase function normalized according to Eq. (3.3.10), Ω_{eff} is the effective solid angle upon which scattering occurs, r is the distance between the particle and the observer, σ_s is the scattering cross section, and 4π is the solid angle for the entire spherical space. The scattering cross section can be derived from the Lorenz–Mie theory of light scattering by spheres and is given by the following expansion:

$$\sigma_s/\pi a^2 = Q_s = c_1 x^4 \left(1 + c_2 x^2 + c_3 x^4 + \cdots\right), \qquad (3.3.24)$$

where a is the radius, $x = 2\pi a/\lambda$, Q_s is referred to as the *scattering efficiency,* and the coefficients in the case of nonabsorbing particles are given by

$$c_1 = \frac{8}{3}\left(\frac{m^2 - 1}{m^2 + 2}\right)^2, \quad c_2 = \frac{6}{5}\left(\frac{m^2 - 1}{m^2 + 2}\right),$$

$$c_3 = \frac{3}{175}\frac{m^6 + 41m^4 - 28m^2 + 284}{(m^2 + 2)^2} + \frac{1}{900}\left(\frac{m^2 + 2}{2m^2 + 2}\right)^2\left[15 + (2m^2 + 3)^2\right].$$

The leading term is the dipole mode contribution associated with Rayleigh scattering. Note that for light scattering by spheres, we may replace the total number of molecules per volume N_s by $1/V$ where $V = 4\pi a^3/3$. For molecules, $a \sim 10^{-4}$ μm, so that $x \sim 10^{-3}$ in the visible. Thus, the higher order terms can be neglected and the scattered intensity is proportional to λ^{-4}. For aerosols and cloud particles, $a \gtrsim 10^{-1}$ μm, and $x \gtrsim 1$ in the visible. In this case, the scattered intensity is less wavelength dependent and is primarily dependent on particle size. As a result, clouds and nonabsorbing aerosols in the atmosphere generally appear white. In a cloudy atmosphere, the sky appears blue diluted with white scattered light, resulting in a less pure blue sky than would have been expected from pure Rayleigh scattering.

On the basis of Eq. (3.3.23), the scattered intensity is dependent on the phase function, which can be computed from the Lorenz–Mie theory for spheres. Figure 3.13 shows typical examples of the phase function for polydispersed cloud droplets (~ 10 μm) and aerosols (~ 1 μm) illuminated by a visible light. Also shown is the phase function for Rayleigh scattering. The mean size parameters in these cases are about 100, 10, and 10^{-3}, respectively. The scattering by cloud droplets is characterized by a strong forward diffraction; a minimum at $\sim 100°$ scattering angle; a peak at $\sim 138°$ scattering angle, the well-known rainbow feature; and a peak in the backscattering direction associated with the glory pattern. The diffraction pattern and the rainbow feature will be discussed further later; the explanation of the glory pattern requires more advanced discussion and will be presented in Chapter 5. The scattering of typical aerosols also displays a forward diffraction maximum and a maximum pattern in the 150°–170° scattering region (see also Fig. 1.4).

3.3.2.2   GEOMETRIC OPTICS

The principles of geometric optics are the asymptotic approximations of the fundamental electromagnetic theory and are valid for light-scattering computations involving a particle whose dimension is much larger than the wavelength, i.e., $x \gg 1$. In this case, a light beam can be thought of as consisting of a bundle of separate parallel rays that hit the particles, which is referred to as the *localization principle.* Each ray will then undergo reflection and refraction and will pursue its own path along a straight line outside and inside the scatterer with propagation directions determined by the *Snell law,* as shown in Fig. 3.14a. In the context of geometric optics, the total electric field is assumed to consist of the diffracted rays and the reflected and refracted rays, as illustrated in Fig. 3.14b, using a sphere as an example. The diffracted rays
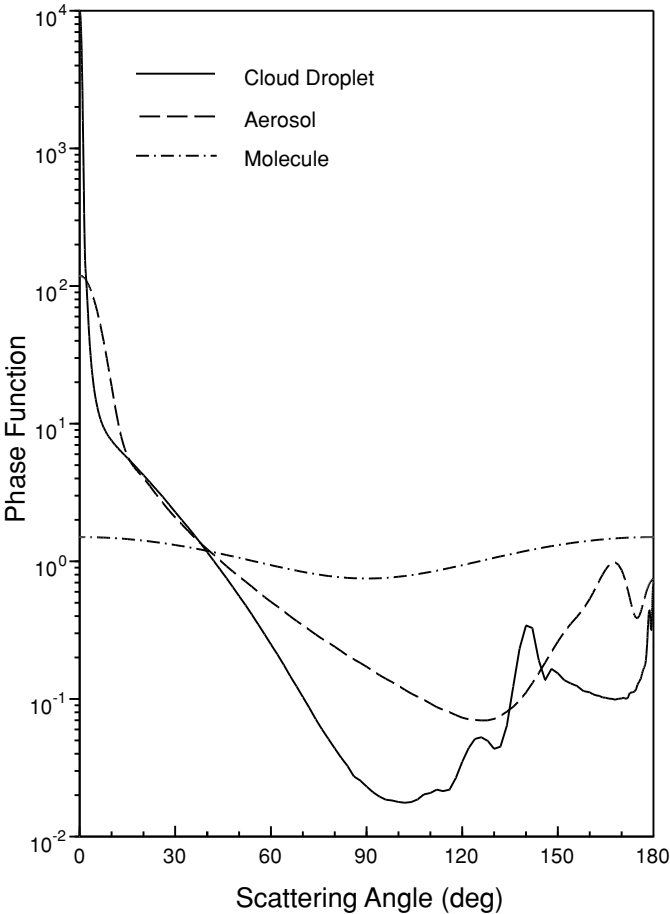
Figure 3.13   Normalized phase functions for cloud droplets ($\sim$10 μm), aerosols ($\sim$1 μm), and molecules ($\sim$10^{-4} μm) illuminated by a visible wavelength of 0.5 μm, computed from the Lorenz–Mie theory.

pass around the scatterer. The rays impinging on the scatterer undergo local reflection and refraction, referred to as *Fresnelian interaction*. The energy that is carried by the diffracted and the Fresnelian rays is assumed to be the same as the energy that is intercepted by the particle cross section projected along the incident direction.

In reference to Fig. 3.14a, let v_1 and v_2 be the velocities of propagation of plane waves in the two media such that $v_1 > v_2$. Also, let θ_i and θ_t be the angles corresponding to the incident and refracted waves. Thus, we have

$$\sin \theta_i / \sin \theta_t = v_1/v_2 = m, \tag{3.3.25}$$

where m is the index of refraction for the second medium with respect to the first.
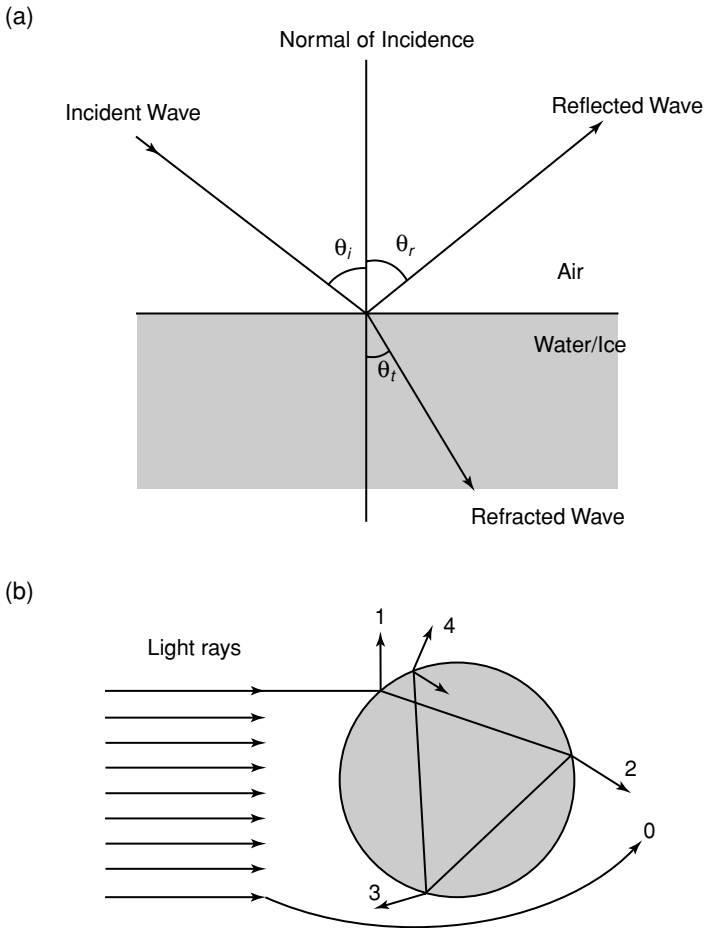
Figure 3.14   (a) Reflection and refraction of a plane wave from air to water/ice surface. (b) Representation of light rays scattered by a sphere based on the geometric optics principle: 0, exterior diffraction; 1, external reflection; 2, two refractions; 3, one internal reflection; and 4, two internal reflections.

For the purpose of this discussion, we shall assume that there is no absorption in the medium. This is the *Snell law* relating the incident and refracted angles through the index of refraction. Exercises 3.12 and 3.13 require the derivation of the minimum deviations of light rays that produce *rainbows* from spherical water droplets and *halos* from hexagonal ice crystals. Moreover, white sunlight is decomposed into component colors after the rays undergo geometric reflection and refraction through water droplets and ice crystals.

The diffraction component in geometric optics can be determined from *Babinet's principle*. This principle states that the diffraction pattern in the far field, referred to as *Fraunhofer diffraction,* from a circular aperture is the same as that from an opaque

disk or sphere of the same radius. Based on this principle and geometric consideration, the scattered intensity is proportional to

$$I_p = \frac{x^4}{4} \left[\frac{2J_1(x \sin \Theta)}{x \sin \Theta} \right]^2, \tag{3.3.26}$$

where J_1 is the first-order *Bessel function* and Θ is the scattering angle. Exercise 3.14 requires the calculation of maxima and minima of the diffraction pattern that can be used to explain an optical phenomenon known as the *corona*.

One final note is in order here. If a particle of any shape is much larger than the incident wavelength, the total energy removed is based on geometric reflection and refraction, giving an effective cross-section area equal to the geometric area A. In addition, according to Babinet's principle, diffraction takes place through a hole in this area, giving a cross-section area also equal to A. The total removal of incident energy is therefore twice the geometric area. Thus, the extinction cross section is given by

$$\sigma_e = 2A, \text{ or } Q_e = \sigma_e/A = 2, \tag{3.3.27}$$

where Q_e is called the *extinction efficiency*. This is referred to as the *optical theorem of extinction*. If a particle is nonabsorbing, then we have $Q_e = Q_s$, where the extinction and scattering efficiencies are the same.

3.3.2.3 ANOMALOUS DIFFRACTION THEORY

Consider large optically soft particles such that $x \gg 1$ and $|m - 1| \ll 1$. The second condition implies that rays are negligibly deviated as they cross the soft particle boundary and are negligibly reflected because the refractive indices inside and outside the particle are similar. In this case, the extinction is largely caused by absorption of the light beam passing through the particle, as well as by the interference of light passing through the particle and passing around the particle. This is the physical foundation for the anomalous diffraction theory originally developed by van de Hulst (1957). In reference to Fig. 3.15, let the plane wave be incident on a spherical particle with a radius a and a refractive index $m \to 1$. The wave front on the forward side of the particle can be divided into two types: one within the geometric shadow area denoted by $A = \pi a^2$, and one outside this area denoted by B. The incident rays can undergo diffraction and pass around the particle. The rays can also hit the particle and undergo reflection and refraction. Since $m \to 1$, we may assume that the rays enter into the particle and pass through it, as illustrated in Fig. 3.15. However, these rays will have phase lags due to the presence of the particle. The phase lag for the ray indicated in the figure is $2a \sin \alpha (m - 1) \cdot 2\pi/\lambda$. If we define the phase shift parameter

$$\rho = 2x(m - 1), \tag{3.3.28}$$

the phase lag can then be expressed by $\rho \sin \alpha$.

Consider a screen that collects the field. The resultant wave on the screen is the sum of the incident and scattered fields. If the incident field is assumed to be unity,
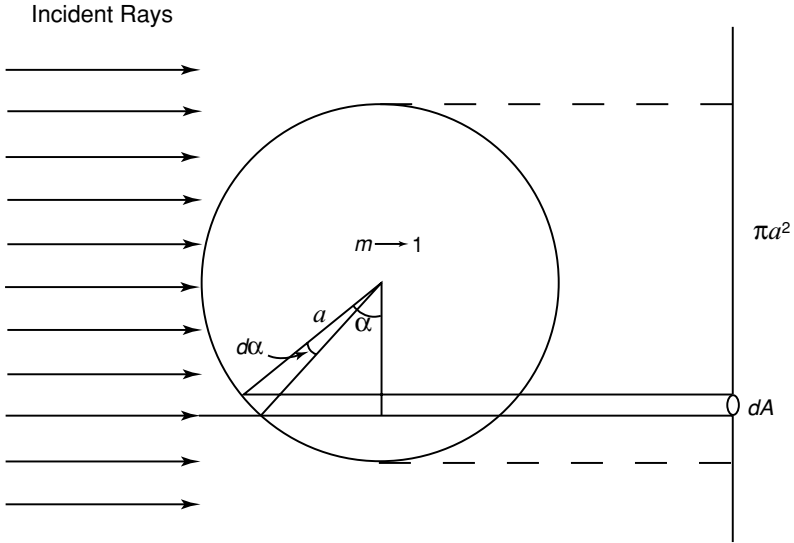
Figure 3.15 Geometry of anomalous diffraction through a sphere with a radius a and an index of refraction $m \to 1$. πa^2 denotes the geometric cross-section area of the sphere and dA denotes the differential cross-section area.

then in the forward direction ($\Theta = 0$), the change in the electric field is proportional to

$$A = \int\int \left(1 - e^{-i\rho \sin \alpha}\right) dx \, dy. \qquad (3.3.29a)$$

The differential area can be replaced by an area in the polar coordinate such that $dx \, dy = a \cos \alpha \, d(a \cos \alpha) \, d\phi$. Thus, we have

$$A = \int_0^{2\pi} \int_0^{\pi/2} \left(1 - e^{-i\rho \sin \alpha}\right) a^2 \sin \alpha \, d \sin \alpha \, d\phi = 2\pi a^2 K(i\rho), \qquad (3.3.29b)$$

where

$$K(i\rho) = \frac{1}{2} + \frac{e^{-i\rho}}{i\rho} + \frac{e^{-i\rho} - 1}{(i\rho)^2}. \qquad (3.3.30)$$

The extinction cross section σ_e is proportional to the differential change in the scattered intensity I. Since $I \sim |E|^2$, as shown in Eq. (3.3.6), $dI \sim 2d|E|$. Thus, we have $\sigma_e = 2\mathrm{Re}(A)$. It follows that the extinction efficiency is given by

$$Q_e = \sigma_e/\pi a^2 = 4\mathrm{Re}[K(i\rho)] = 2 - \frac{4}{\rho} \sin \rho + \frac{4}{\rho^2}(1 - \cos \rho), \qquad (3.3.31)$$

where Re denotes the real part of the function. Exercise 3.15 requires calculations of Q_e.

We may also determine the absorption efficiency by the following procedure. The ray path as shown in Fig. 3.15 is $l = 2a \sin \alpha$. The absorption coefficient $k_i = m_i 2\pi/\lambda$,

where m_i is the imaginary part of the refractive index. Thus, the absorption path length associated with the electric field is lk_i. The attenuation of the intensity of the ray is then $\exp(-2lk_i)$ and the absorption cross section for all possible rays is

$$\sigma_a = \int \int \left(1 - e^{-2lk_i}\right) dx\, dy. \qquad (3.3.32)$$

Following the procedure just illustrated, the absorption efficiency is given by

$$Q_a = \sigma_a/\pi a^2 = 1 + \frac{2}{b}e^{-b} + \frac{2}{b^2}(e^{-b} - 1), \qquad (3.3.33)$$

where $b = 4xm_i$ and $x = 2\pi a/\lambda$. The approximation based on the anomalous diffraction theory (ADT) is useful for the calculation of the extinction and absorption coefficients when $m \to 1$. It can also be applied to nonspherical particles such as spheroids and hexagons. Since refractions and reflections of rays are neglected in this approximation, its accuracy must be examined carefully when applied to the scattering of ice crystals ($m \sim 1.31$) and aerosols ($m \sim 1.5$). Finally, it should be noted that the ADT approximation cannot produce the phase function pattern.

3.4   Multiple Scattering and Absorption in Planetary Atmospheres

3.4.1   Fundamentals of Radiative Transfer

In Section 1.1.4, we pointed out that scattering is often coupled with absorption. In the following we formulate the fundamental equation governing the transfer of diffuse solar radiation in plane-parallel atmospheres. The term *diffuse* is associated with multiple scattering processes and is differentiated from *direct* solar radiation. In reference to Fig. 3.16 and considering a differential thickness Δz, the differential change of diffuse intensity emergent from below the layer is due to the following processes: (1) reduction from the extinction attenuation; (2) increase from the single scattering of the unscattered direct solar flux from the direction $(-\mu_0, \phi_0)$ to (μ, ϕ); (3) increase from multiple scattering of the diffuse intensity from directions (μ', ϕ') to (μ, ϕ); and (4) increase from emission within the layer in the direction (μ, ϕ). Consider a small volume containing a spectrum of molecules and/or particulates and denote the extinction, scattering, and absorption coefficients (in units of per length) as β_e, β_s, and β_a, respectively, defined by

$$\beta_{e,s,a} = \int_{\Delta z} \sigma_{e,s,a}(z)n(z)\,dz/\Delta z, \qquad (3.4.1)$$

where the symbol σ denotes the cross section and n is the number density. Moreover, let the phase function corresponding to a volume of particulates be P. Thus, $P(\mu, \phi; \mu', \phi')$ denotes the redirection of the incoming intensity defined by (μ', ϕ') to the outgoing intensity defined by (μ, ϕ). Also note that the differential length
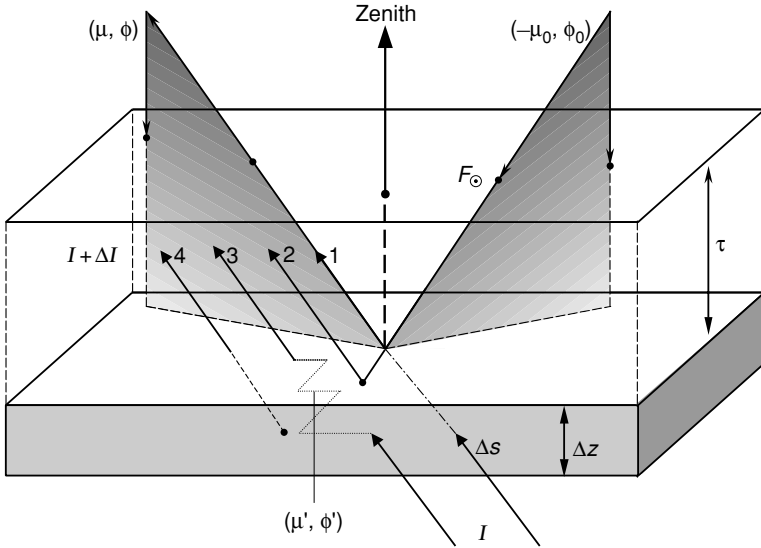
Figure 3.16 Transfer of diffuse solar intensity from below in plane-parallel layers: (1) attenuation by extinction; (2) single scattering of the unscattered solar flux; (3) multiple scattering; and (4) emission from the layer. All the radiative parameters are defined with reference to a small volume containing a spectrum of molecules and/or particulates. The notations are defined in the text.

$\Delta s = \Delta z/\mu$. Based on the preceding definitions, we may write (neglecting the wavelength index)

$$\frac{\Delta I(z; \mu, \phi)}{\Delta z/\mu} = -\beta_e I(z; \mu, \phi) + \beta_s F_\odot e^{-\tau/\mu} \cdot P(\mu, \phi; -\mu_0, \phi_0)/4\pi$$

$$+ \beta_s \int_0^{2\pi} \int_{-1}^1 I(z, \mu', \phi') \cdot P(\mu, \phi; \mu', \phi')/4\pi \, d\mu' d\phi' + \beta_a B[T(z)]. \tag{3.4.2}$$

All the terms are self-explanatory. However, it is noted that integration of the multiple scattering term is performed for diffuse intensity over the 4π solid angle and that radiative equilibrium is assumed such that emission is equal to absorption based on Kirchhoff's and Planck's laws (Section 1.2).

Further, we may define the single-scattering albedo as the ratio of the scattering coefficient to the extinction coefficient in the form

$$\tilde{\omega} = \frac{\beta_s}{\beta_e} \quad \text{or} \quad 1 - \tilde{\omega} = \frac{\beta_a}{\beta_e}. \tag{3.4.3}$$

The optical depth is defined by

$$\tau = \int_z^\infty \beta_e \, dz'. \tag{3.4.4}$$

Using the optical depth, Eq. (3.4.2) can be rewritten as follows:

$$\mu \frac{dI(\tau;\mu,\phi)}{d\tau} = I(\tau;\mu,\phi) - J(\tau;\mu,\phi), \qquad (3.4.5)$$

where the source function is given by [see also Eq. (1.4.22)]

$$J(\tau;\mu,\phi) = \frac{\tilde{\omega}}{4\pi} \int_0^{2\pi} \int_{-1}^1 I(\tau;\mu',\phi')P(\mu,\phi;\mu',\phi')d\mu'd\phi'$$

$$+ \frac{\tilde{\omega}}{4\pi} F_\odot P(\mu,\phi;-\mu_0,\phi_0)e^{-\tau/\mu_0} + (1-\tilde{\omega})B[T(\tau)]. \qquad (3.4.6)$$

The fundamental parameters that drive the transfer of diffuse intensity are the extinction coefficient (or the optical depth), the single-scattering albedo, and the phase function. It suffices to assume that these parameters are independent of τ (or z) in the present discussion. Interested readers should refer to Chapter 6 for an in-depth explanation of radiative transfer processes. For discussion of solar radiative transfer, the flux emitted from the earth and the atmosphere with an equilibrium temperature of $\sim$255 K is negligible in comparison to that emitted from the sun for $\lambda \leq 3.5$ μm. For some solar radiative transfer problems, we may then omit the last term in the source function.

The phase function represents the angular distribution of the scattered energy as a function of the scattering angle and has been presented for molecules, aerosols, and cloud particles in Fig. 3.13. From spherical geometry, the scattering angle is related to the incoming and outgoing directions in the form

$$\cos\Theta = \mu\mu' + (1-\mu^2)^{1/2}(1-\mu'^2)^{1/2}\cos(\phi'-\phi). \qquad (3.4.7)$$

We may express the phase function in terms of a known mathematical function for the purpose of solving Eq. (3.4.5), the first-order differential integral equation. The Legendre polynomials (Appendix E), by virtue of their unique mathematical properties, have been used extensively in the analysis of radiative transfer problems. In terms of Legendre polynomials P_ℓ, the phase function may be written in the form

$$P(\cos\Theta) = \sum_{\ell=0}^N \tilde{\omega}_\ell P_\ell(\cos\Theta), \qquad (3.4.8)$$

where the expansion coefficient, based on the orthogonal property, is given by

$$\tilde{\omega}_\ell = \frac{2\ell+1}{2} \int_{-1}^1 P(\cos\Theta)P_\ell(\cos\Theta)\,d\cos\Theta, \quad \ell = 0, 1\ldots, N. \qquad (3.4.9a)$$

When $\ell = 0$, $\tilde{\omega}_0 = 1$, representing the normalization of the phase function denoted in Eq. (3.3.10). When $\ell = 1$, we have

$$g = \frac{\tilde{\omega}_1}{3} = \frac{1}{2} \int_{-1}^1 P(\cos\Theta)\cos\Theta\,d\cos\Theta. \qquad (3.4.9b)$$

This term is referred to as the *asymmetry factor,* which is the first moment of the phase function and an important parameter in radiative transfer. For isotropic scattering, g is

zero, as it is for Rayleigh scattering (Exercise 3.16). The asymmetry factor increases as the diffraction peak of the phase function sharpens and can be negative if the phase function peaks in backward directions (90–180°). For Lorenz–Mie type particles, whose phase function has a generally sharp peak at the 0° scattering angle (Fig. 3.13), the asymmetry factor denotes the relative strength of forward scattering.

3.4.2   Approximations of Radiative Transfer

We shall present two useful approximations: one for remote sensing applications, and the other for radiation parameterization for use in climate studies.

3.4.2.1   SINGLE-SCATTERING APPROXIMATION

In a domain where the optical depth is small (e.g., $\tau < 0.1$), a large portion of scattering events is dominated by single scattering of the direct solar beam. This occurs in optically thin cirrus and aerosol atmospheres. In this case, the most important term in the source function is

$$J(\tau; \mu, \phi) \cong \frac{\tilde{\omega}}{4\pi} F_{\odot} P(\mu, \phi; -\mu_0, \phi_0) e^{-\tau/\mu_0}. \tag{3.4.10}$$

Consider a black surface such that the reflected upward intensity $I(\tau_*; \mu, \phi) = 0$, where τ_* is the total atmospheric optical depth. From Eq. (1.4.23), the upward intensity at the top of the atmosphere is

$$I(0; \mu, \phi) = \int_0^{\tau_*} J(\tau'; \mu, \phi) e^{-\tau'/\mu} \frac{d\tau'}{\mu}$$

$$= \frac{\mu_0 F_{\odot}}{\pi} \frac{\tilde{\omega}}{4(\mu + \mu_0)} P(\mu, \phi; -\mu_0, \phi_0) \left\{ 1 - \exp\left[-\tau_* \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) \right] \right\}. \tag{3.4.11a}$$

Moreover, for a small τ_*, we have

$$R(\mu, \phi; \mu_0, \phi_0) = \frac{\pi I(0; \mu, \phi)}{\mu_0 F_{\odot}} = \tau_* \frac{\tilde{\omega}}{4\mu\mu_0} P(\mu, \phi; -\mu_0, \phi_0). \tag{3.4.11b}$$

The term R is a nondimensional quantity, referred to as the *bidirectional reflectance*. This equation establishes the foundation for the retrieval of the optical depth of aerosols from satellites. It is clear that under the condition of optically thin atmosphere, the optical depth is directly proportional to the bidirectional reflectance that can be determined from satellite radiometric measurements, but is inversely proportional to the phase function. The latter dependence becomes an important issue in satellite remote sensing using reflected sunlight, a subject that will be discussed further in Section 7.3.1.

3.4.2.2   DIFFUSION APPROXIMATION

Consider a diffusion domain where the directional dependence of multiple scattering events is largely lost. In this case, it is appropriate to consider the transfer of hemispheric upward and downward flux densities defined by [see also Eq. (1.1.9)]

$$F^{\uparrow\downarrow}(\tau) = \int_0^{2\pi} \int_0^{\pm 1} I(\tau; \mu, \phi)\mu \, d\mu \, d\phi, \tag{3.4.12}$$

where the notations $\uparrow$ and $\downarrow$ correspond to $+$ and $-$, respectively. We may formulate the transfer problem based on the physical reasoning that the differential changes of the upward and downward flux densities must be related to these fluxes as well as to the direct downward flux from the sun. Thus, we write,

$$\frac{dF^{\uparrow}}{d\tau} = \gamma_1 F^{\uparrow} - \gamma_2 F^{\downarrow} - \gamma_3 \tilde{\omega} F_{\odot} e^{-\tau/\mu_0}, \tag{3.4.13a}$$

$$\frac{dF^{\downarrow}}{d\tau} = \gamma_2 F^{\uparrow} - \gamma_1 F^{\downarrow} + (1 - \gamma_3) \tilde{\omega} F_{\odot} e^{-\tau/\mu_0}, \tag{3.4.13b}$$

where γ_1, γ_2, and γ_3 are appropriate weighting coefficients related to multiple scattering events. The two flux equations were first formulated by Schuster (1905), although in a slightly different format. These equations can be derived from the well-known two-stream and Eddington approximations in which the three coefficients can be determined (see Chapter 6). Solutions for the upward and downward fluxes can be derived by setting $F_{\mathrm{dif}} = F^{\downarrow} - F^{\uparrow}$, and $F_{\mathrm{sum}} = F^{\downarrow} + F^{\uparrow}$. In this manner we can show that (Exercises 3.17 and 3.18)

$$\frac{d^2 F_{\mathrm{dif}}}{d\tau^2} = k^2 F_{\mathrm{dif}} + \chi e^{-\tau/\mu_0}, \tag{3.4.14}$$

where $k^2 = \gamma_1^2 - \gamma_2^2$ are the eigenvalues and χ is a certain coefficient. Equation (3.4.14) is referred to as the *diffusion equation for radiative transfer*. The general solution for this second-order nonhomogeneous differential equation is given by

$$F_{\mathrm{dif}} = c_1 e^{-k\tau} + c_2 e^{+k\tau} + \chi(1/\mu_0^2 - k^2)e^{-\tau/\mu_0}, \tag{3.4.15}$$

where $c_{1,2}$ are certain coefficients. Likewise, we can also derive a solution for F_{sum} which, together with F_{dif}, can be used to determine the analytic solutions for upward and downward flux densities. Interested readers should consult Section 6.5.2 for an advanced discussion of this topic.

Many general circulation and climate models utilize the two-stream or Eddington's approximation in the parameterization of radiative transfer because analytic solutions can be derived to achieve efficient computation that is critical for model simulations. In the following, the subject of atmospheric absorption in multiple scattering atmospheres that leads to the production of solar heating rates is further discussed.

3.5   Atmospheric Solar Heating Rates

The absorption of solar radiation by various gases is important because of its generation of heating in the atmosphere, which is also affected by multiple scattering processes. Consider a plane-parallel absorbing and scattering atmosphere illuminated by the solar spectral irradiance $F_\odot$ so that the downward flux density normal to the top of the atmosphere is given by $\mu_0 F_\odot$. Let the differential thickness within the atmosphere be Δz, and let the spectral downward and upward flux densities centered at wavelength λ be denoted by $F^\downarrow$ and $F^\uparrow$, respectively. We have omitted the wavelength dependence for simplicity of presentation. The net flux density (downward) at a given height z is then defined by

$$F(z) = F^\downarrow(z) - F^\uparrow(z). \tag{3.5.1a}$$

In reference to Fig. 3.17, because of absorption, the net flux density decreases from the upper levels to the progressively lower levels. The loss of net flux density, i.e., the
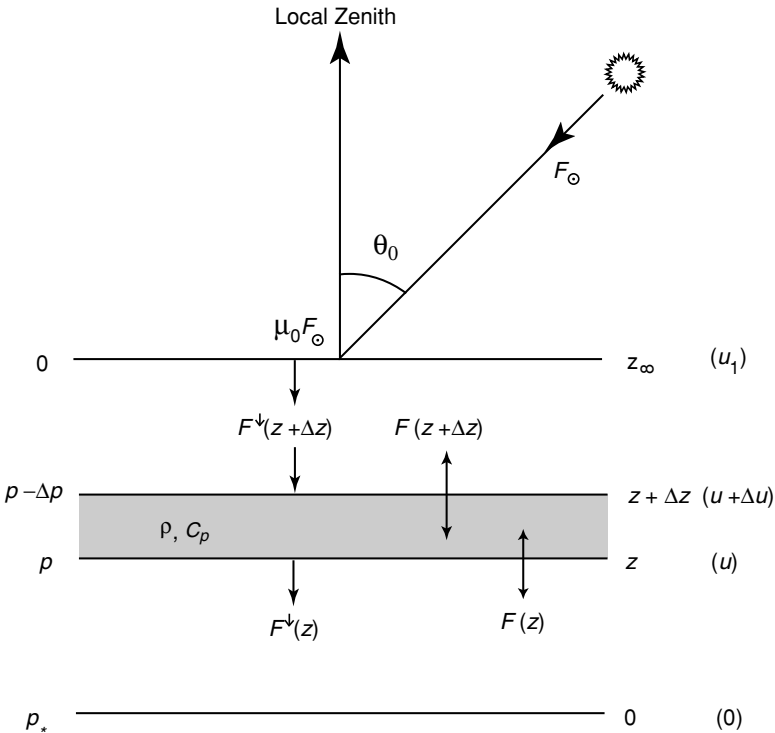
Figure 3.17   Divergence of the net flux density in z, p, and u coordinates. All the notations are defined in the text.

net flux density divergence for the differential layer is, therefore,

$$\Delta F(z) = F(z) - F(z + \Delta z). \tag{3.5.1b}$$

On the basis of the energy conservation principle, the absorbed radiant energy must be used to heat the layer. Thus, the heating experienced by a layer of air due to radiation transfer may be expressed in terms of the rate of temperature change. It is conventionally given by

$$\Delta F(z) = -\rho C_p \Delta z \frac{\partial T}{\partial t}, \tag{3.5.2}$$

where ρ is the air density in the layer, C_p is the specific heat at constant pressure, and t is the time. The heating rate for a differential layer Δz is, therefore,

$$\frac{\partial T}{\partial t} = -\frac{1}{\rho C_p} \frac{\Delta F(z)}{\Delta z} = \frac{g}{C_p} \frac{\Delta F(p)}{\Delta p} = -\frac{q}{C_p} \frac{\Delta F(u)}{\Delta u}, \tag{3.5.3}$$

where we have also expressed the heating rate in terms of pressure and path-length coordinates using the hydrostatic equation $dp = -\rho g \, dz$, and the definition of path length for a specific gas where q is the mixing ratio, g is the gravitational acceleration, and g/C_p is the well-known dry adiabatic lapse rate. If we divide the solar spectrum into N intervals and carry out the heating rate calculations for each spectral interval i, then the total heating rate due to solar radiation may be written in the form

$$\left(\frac{\partial T}{\partial t}\right)_s = \sum_{i=1}^{N} \left(\frac{\partial T}{\partial t}\right)_i. \tag{3.5.4}$$

Computation of the solar flux and heating rate in the atmosphere covering the entire solar spectrum is quite involved. In a clear atmosphere, we must include both absorption by various absorbing gases, chiefly H_2O, O_3, O_2, and CO_2, and scattering by molecules and aerosols, as well as reflection from the surface. The solar spectrum must be divided into a number of suitably grouped subspectral intervals in which a monochromatic radiative transfer program, such as the adding or discrete-ordinates method for inhomogeneous atmospheres discussed in Chapter 6, can be employed for the calculation of spectral fluxes and heating rates. The single-scattering properties of each subdivided interval for input into the radiative transfer model must include simultaneous contributions from the scattering and absorption of aerosols (and cloud particles), Rayleigh scattering, and gaseous absorption. For efficient spectral integration, we may employ the correlated k-distribution method introduced in Section 4.3.

In Fig. 3.18, we show typical solar heating rates and net flux profiles as functions of the cosine of the solar zenith angle μ_0 using the standard atmospheric profiles for H_2O, O_3, and other trace gases (see Fig. 3.2), along with a surface albedo of 0.1, as inputs of a radiative transfer model. The instantaneous solar heating rate profile is divided into two different levels to highlight the contributions from H_2O and O_3. The solar heating rate decreases as μ_0 decreases because the incoming solar irradiance available to the atmosphere is directly proportional to μ_0. Below about 10 km, the solar heating rate is primarily produced by water vapor with the heating rate ranging
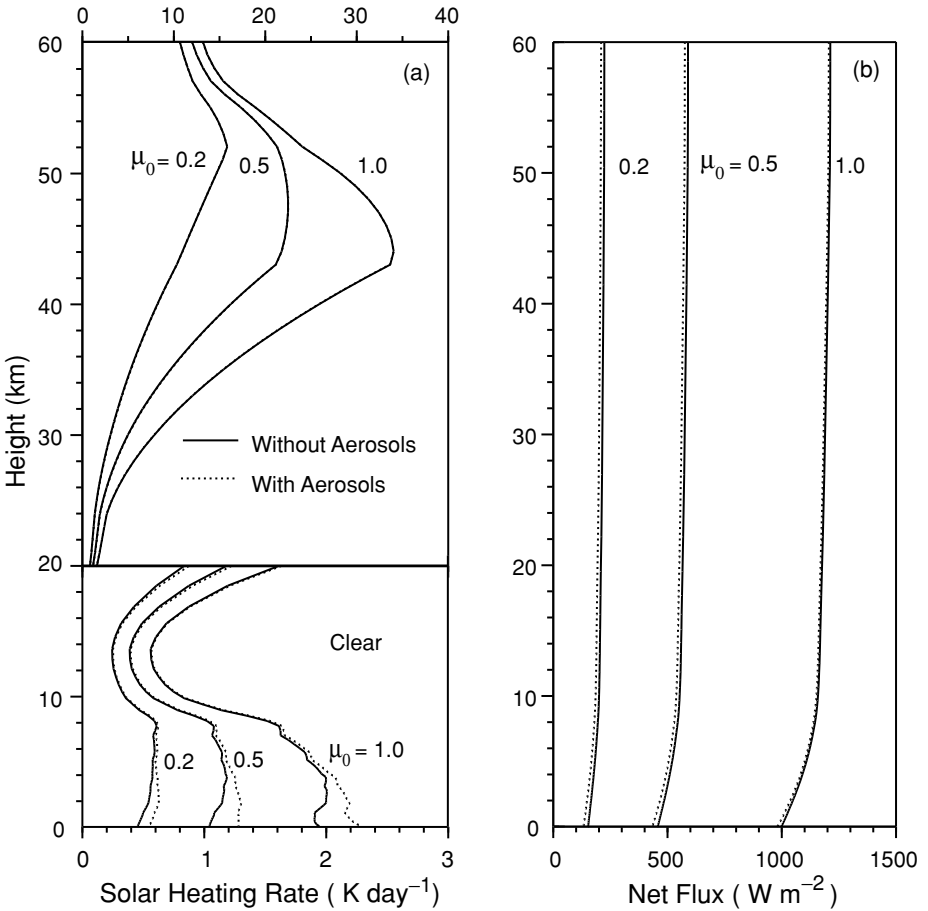
Figure 3.18   Solar heating rates and net fluxes as functions of height with and without the contribution of aerosols for a number of the cosines of solar zenith angles. The solar heating rates are instantaneous values and are separated in two regions to highlight the contributions from water vapor and ozone in the troposphere and stratosphere, respectively. A typical background aerosol profile with a visible optical depth of 0.15 is used to illustrate the effect of aerosols on the solar flux and heating rate. These results and those presented in Fig. 3.19 are computed from a line-by-line equivalent radiative transfer model that includes the contributions of gaseous absorption, multiple scattering, and the absorption of aerosol and cloud particles (Liou *et al.*, 1998).

from 0.5 to 2 K day^{-1} near the surface when the contribution from aerosols is not accounted for. The solar heating rate decreases rapidly with increasing altitude in phase with the exponential decrease of water vapor and reaches a minimum at about 15 km. Above 20 km, increased solar heating is produced primarily by the absorption of ozone. Solar net flux decreases significantly below about 10 km. When a standard aerosol profile with an optical depth of 0.15 at the 0.5 μm wavelength is added, the solar heating rate increases in the lower atmosphere because of the absorption of aerosols in the visible and near infrared. The effect of aerosols on the absorption of
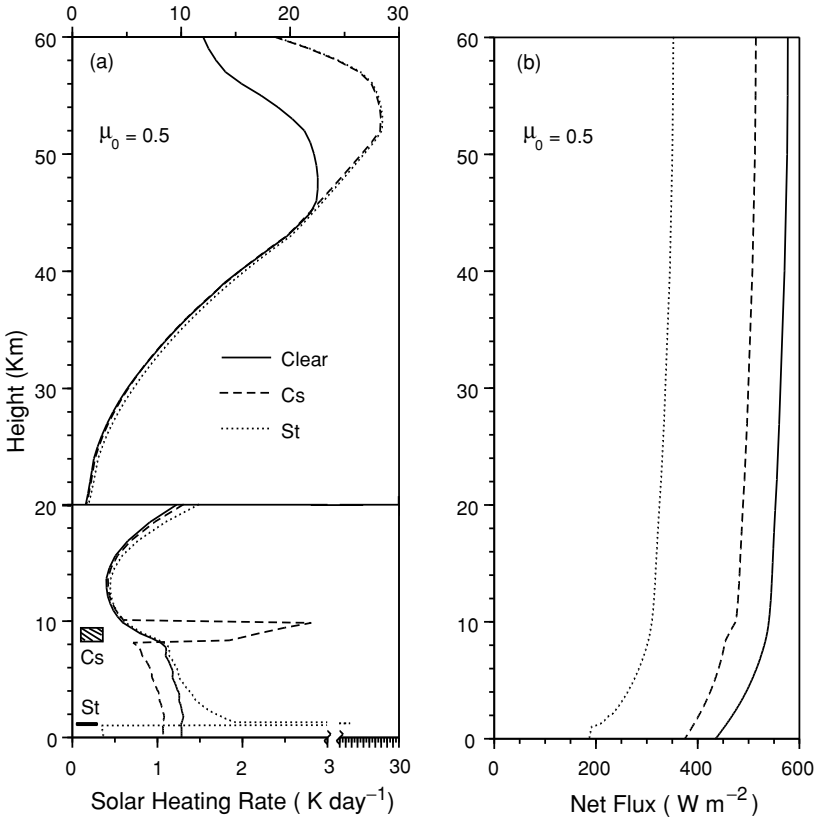
Figure 3.19 Comparison of (a) solar heating rates and (b) net fluxes in clear, cirrus (Cs), and stratus (St) cloudy conditions for a μ_0 of 0.5. The positions of these clouds are indicated in the diagram. The optical depths for Cs and St are 0.7 and 10, respectively, while the mean particle sizes are 42 and 8 μm, respectively.

solar fluxes depends on their chemical composition, particle size distribution, and vertical profile and is a subject of ongoing research.

The effects of clouds on solar heating and net flux profiles are investigated using typical single-layer cirrostratus (Cs) and stratus (St) clouds whose locations are shown in Fig. 3.19. We use a cosine of the solar zenith angle of 0.5 in this demonstration. The visible optical depths for Cs and St are 0.7 and 10, respectively, while the mean ice crystal maximum dimension and water droplet radius are 42 and 8 μm, respectively. In the case of low stratus, substantial instantaneous heating occurs at the cloud top with a value of about 22 K day^{-1}. Because of the reflection from clouds, ozone heating also increases. This increase appears to depend on the factors associated with cloud position and optical depth. In the overcast low stratus condition, net solar flux available at the surface is only about 187 W m^{-2}, in comparison to about 435 and 376 W m^{-2} in clear and cirrus cloud conditions, respectively.

Exercises

3.1 The scale height H is defined by $dp/p = -dz/H$. From the hydrostatic equation and the equation of state, show that $H = KT/Mg$, where K is the Boltzmann constant, M is the molecular weight of air, and g is gravity. Since the molecular translational energy is $\frac{1}{2} KT$, the scale height is then twice the distance through which atoms/molecules that have the equipartition of translational energy can rise in the vertical direction against the force of gravity.

3.2 Compute and display graphically $r(z_1)$ as a function of z_1, as defined in Eq. (3.2.6) for $\mu_0 = 1, 0.5$, and 0.2. Compare your results with those presented in Fig. 3.6 and explain the meaning of the Chapman layer.

3.3 In reference to the spherical atmosphere depicted in Fig. 3.20, derive the Chapman function $\text{Ch}(x, \theta_0)$. Compare this function with $1/\mu_0$ and determine
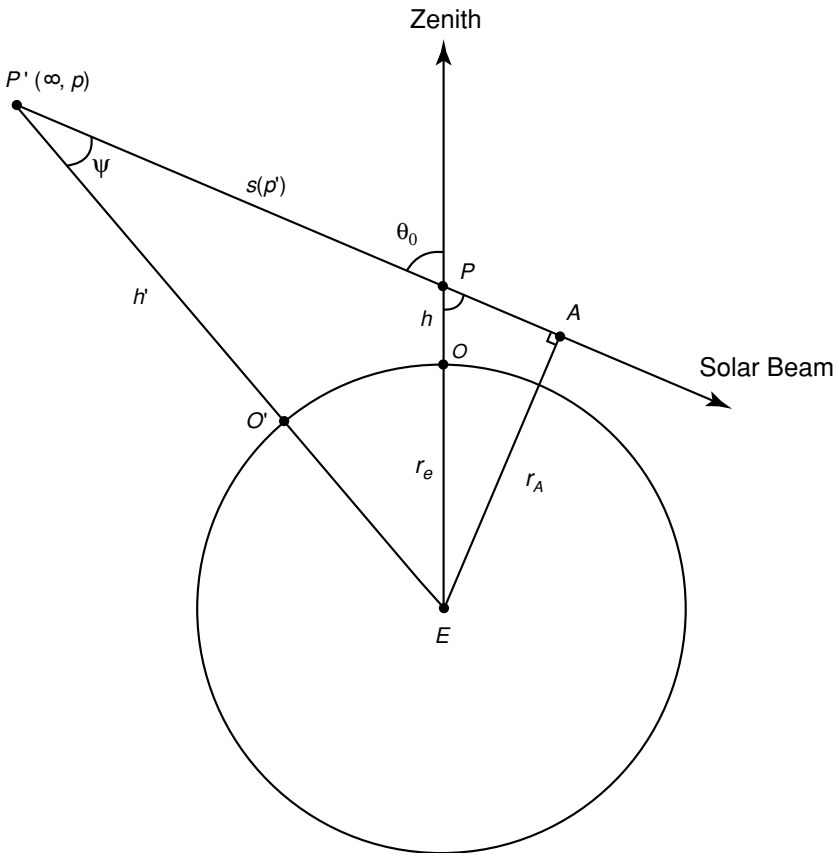
Figure 3.20 Spherical geometry for the evaluation of the Chapman function: $\theta_0 =$ solar zenith angle, $r_e =$ earth's radius, $r_A = AE$, $h = PO$, $h' = P'O'$, $s(p') =$ the path length from point p' to A. The objective is to compute the absorption at point P, which is at a height h above the earth's surface, based on the actual path length $s(p')$.

the limit of the solar zenith angle under which $1/\mu_0$ is a good approximation of the exponential attenuation calculation.

3.4   Given the values of $K_{12} = 5.6 \times 10^{-46}(300/T)^{2.36}\text{m}^6 \text{ sec}^{-1}$ and $K_{13} = 2.0 \times 10^{-17}\exp(-2280/T)$ m^3 sec^{-1}, and the J_2 and J_3 values in the following table, compute the equilibrium ozone concentration as a function of height and compare your results with those presented in Fig. 3.8. Use the standard atmospheric temperature and molecular number density profiles (Appendix G) in your calculations.

Height (km)	30	35	40	45	50
$J_2(\text{sec}^{-1}) \times 10^{-10}$	0.61	2.13	4.56	7.93	11.30
$J_3(\text{sec}^{-1}) \times 10^{-3}$	0.62	1.09	2.03	4.31	6.29

3.5   The principal photochemical reactions involving oxygen in the thermosphere are found to be

$$O_2 + h\tilde{\nu}(\lambda < 1751 \text{ Å}) \xrightarrow{J_2} O + O,$$

$$O + O + M \xrightarrow{K_{11}} O_2 + M,$$

$$O + O \xrightarrow{K'_{11}} O_2 + h\tilde{\nu}.$$
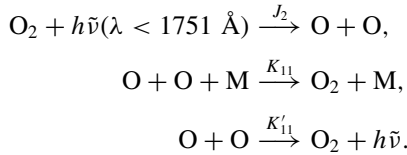
Express these photochemical processes in terms of the rate of change of the number density of O and O_2. Derive the number density of O under the photochemical equilibrium condition.

3.6   For very strong Lorentz lines (see Section 1.3.2), the half-width is much smaller than the spread of the line such that $\alpha \ll (\nu - \nu_0)$. Under this condition and using a single line, show that the spectral absorptivity is proportional to the square root of the path length. In your analysis, define the relevant parameters in the wavenumber domain and use the following integration:

$$\int_0^\infty \left(e^{-a^2/x^2} - e^{-b^2/x^2} \right) dx = \sqrt{\pi}(b - a).$$

3.7   The number of molecules per cubic centimeter of air at sea level in standard atmospheric conditions is about 2.55×10^{19} cm^{-3}. Calculate the scattering cross section of molecules at the 0.3, 0.5, and 0.7 μm wavelengths.

3.8   The number density profile as a function of height is given by the following table:

Height (km)	0	2	4	6	8	10	12	14	16
$N(\times 10^{18}$ cm$^{-3})$	25.5	20.9	17.0	13.7	10.9	8.60	6.49	4.74	3.46

Calculate the optical depth of a clear atmosphere at the wavelengths shown in Exercise 3.7.

3.9 For all practical purposes, we find that the refractive index m_r and the molecular density ρ are related by

$$(m_r - 1)_{\text{gas}} = \text{const} \times \rho.$$

At sea level, the refractive index of air is about 1.000292 for a wavelength of 0.3 μm. Find the refractive indices at the heights given in Exercise 3.8. Note that the density (g cm^{-3}) is related to the number density N(cm^{-3}) by $\rho = (M/N_0)N$, where M is the molecular weight of air (28.97 g mol^{-1}), and N_0 is Avogadro's number (6.02295 $\times 10^{23}$ mol^{-1}). Because the refractive index varies with the density of the atmosphere, light rays bend according to the atmospheric density profile and produce a number of atmospheric optical phenomena known as looming, sinking, and superior and inferior mirages.

3.10 An unpolarized ruby laser operated at 0.7 μm is projected vertically into a clear sky to investigate the density of the atmosphere. A detector located 10 km from the base of the laser is used to receive the flux density scattered from the laser beam by air molecules. Assuming that the laser output has a uniform distribution of flux density F_0 across the beam (i.e., $I_0 = F_0/\pi$ sr), and neglecting the effects of multiple scattering, find the scattered flux density at 6 and 10 km received by a detector whose field of view in a plane is 0.05 rad. Use the scattering cross section and molecular density profile obtained from Exercises 3.7 and 3.8.

3.11 (a) The radar backscattering coefficient (in units of per length) for a volume of identical cloud droplets is defined as

$$\beta_\pi = N_c \sigma_\pi = N_c \sigma_s P(\pi),$$

where N_c is the droplet number density, σ_π the backscattering cross section, and $P(\pi)$ the phase function at backscatter. Employing the Rayleigh scattering cross section and phase function, and noting that $N_c = 1/V$, where the volume of a spherical drop with a radius a is $V = 4\pi a^3/3$, show that

$$\beta_\pi = \frac{64\pi^5}{\lambda^4} N_c a^6 \left| \frac{m^2 - 1}{m^2 + 2} \right|^2.$$

(b) Assuming that the number density and the radius of cloud droplets are 100 cm^{-3} and 20 μm, respectively, calculate β_π for the following two radar wavelengths with the corresponding refractive indices for water:

λ(cm)	10	3.21
m	3.99–1.47i	7.14–2.89i

where $i = \sqrt{-1}$. Compute β_π again using only the real part of the refractive indices, and show the differences between the two computations.

3.12 From the geometry of a sphere with respect to the incident ray, show that the incident angle θ_i at which the minimum deviation occurs is given by

$$\cos^2 \theta_i = (m^2 - 1)/(p^2 - 1), \quad p \geq 2,$$

where $(p-1)$ denotes the number of reflection. The refractive index m for water is 1.33 in the visible. Compute the positions defined as the scattering angles for the first and second rainbows.

3.13 From the geometry of a hexagonal plate with respect to the incident ray, show that the angle of refraction at minimum deviation can be determined from

$$\sin\left[\frac{1}{2}(\theta' + A)\right] = m \sin\frac{A}{2},$$

where A denotes the prism angle and the refractive index of ice in the visible is 1.31. For $A = 60°$, $90°$, and $120°$, compute the positions of halos.

3.14 From Eq. (3.3.26), compute and plot the diffraction pattern as a function of $y = x \sin\theta$. What would be the position of the strongest corona produced by uniform-sized aerosols with a radius of 1 μm? Use a wavelength of 0.5 μm in your calculation.

3.15 (a) Compute the extinction coefficient as a function of the phase shift parameter, defined in Eqs. (3.3.31) and (3.3.28). (b) Estimate the aerosol particle size under which more blue light is available to an observer than red light based on the first maximum and minimum in the extinction curve assuming a refractive index of 1.5 for aerosols. This is related to an optical phenomenon referred to as *once in a blue moon.* Why is it so rare? Interested readers may also wish to refer to Figs. 5.1 and 5.7 for additional information on aerosol size distribution and extinction.

3.16 Show that for isotropic and Rayleigh scattering cases, the asymmetry factor is zero.

3.17 Consider the cases of pure scattering, referred to as *conservative scattering,* such that $\tilde{\omega} = 1$. Define the net flux associated with the diffuse beam as follows:

$$F(\tau) = \int_0^{2\pi} \int_{-1}^1 I(\tau, \mu, \phi)\mu \, d\mu \, d\phi.$$

Show from Eq. (3.4.13) that

$$\frac{dF(\tau)}{d\tau} = F_\odot e^{-\tau/\mu_0},$$

and that

$$F(\tau) + \mu_0 F_\odot e^{-\tau/\mu_0} = \text{constant}.$$

This is the so-called *flux integral.* In a pure scattering atmosphere, the total flux (direct plus diffuse solar beam) is conserved.

3.18 From the flux equations given in Eqs. (3.4.13a) and (3.4.13b), derive Eq. (3.4.14).

Suggested Reading

Brasseur, G., and Solomon, S. (1986). *Aeronomy of the Middle Atmosphere,* 2nd ed. D. Reidel, Dordrecht. Chapter 4 presents a comprehensive discussion of the

absorption of solar ultraviolet radiation pertaining to photochemical processes in the stratosphere and mesosphere.

Goody, R. M. (1995). *Principles of Atmospheric Physics and Chemistry*. Oxford University Press, New York. Chapters 4 and 5 contain concise discussions of the transfer of solar radiation and ozone formation.

Goody, R. M., and Yung, Y. L. (1989). *Atmospheric Radiation. Theoretical Basis,* 2nd ed. Oxford University Press, New York. Chapter 5 contains useful data on the absorption of solar radiation by various gases in the solar spectrum.

Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history, *Rev. Geophys*. **37,** 275–316. An authoritative and updated review of photochemical processes involving ozone formation and catalytic depletion in the earth's atmosphere.

van de Hulst, H. C. (1957). *Light Scattering by Small Particles*. Wiley, New York. A classic discussion of the principles of light scattering by molecules and small particles. See Chapters 3 and 12 in conjunction with the discussion in this chapter.

Chapter 4 | Thermal Infrared Radiation Transfer in the Atmosphere

4.1 The Thermal Infrared Spectrum and the Greenhouse Effect

Based on data gathered from satellite radiation budget experiments, the earth–atmosphere system reflects about 30% of the incoming solar radiation at the top of the atmosphere and absorbs the remaining part (see Section 8.2 for further discussion). Absorption and scattering of solar radiation take place in the atmosphere, processes that were discussed in Chapter 3. A large portion of the incoming solar radiation is absorbed by the earth's surface, consisting of approximately 70% ocean and 30% land. Over a climatological period of time, say, over a year or longer, the global equilibrium temperature of the earth–atmosphere system remains relatively constant. Consequently, radiant energy emitted from the sun that is absorbed in the earth–atmosphere system must be re-emitted to space so that an equilibrium energy state can be maintained. Just as the sun emits electromagnetic radiation covering all frequencies, so do the earth and the atmosphere, and this emitted radiation is referred to as *thermal infrared radiation*.

On the basis of the conservation of absorbed solar and emitted thermal infrared energies, and denoting the global albedo of the earth–atmosphere system by $\bar{r}$, the Stefan-Boltzmann law for emission gives us the following balanced equation:

$$S \cdot \pi a_e^2 (1 - \bar{r}) = \sigma T_e^4 \cdot 4\pi a_e^2, \qquad (4.1.1a)$$

where a_e is the earth's radius, S is the solar constant, representing the energy available at the top of the atmosphere, and T_e denotes the equilibrium temperature of the earth–atmosphere system. The factor of 4 accounts for the difference between the absorption and emission areas. Thus, we obtain

$$T_e = [S(1 - \bar{r})/4\sigma]^{1/4}. \qquad (4.1.1b)$$

As discussed in Section 2.3.3, the solar constant derived from recent satellite observations is 1366 W m^{-2}. Inserting the solar constant and global albedo values into the balanced equation, we find $T_e \sim 255$ K. The temperature profile as shown in Fig. 3.1 ranges from about 200 to 300 K.

From Planck's and Wien's displacement laws discussed in Chapter 1, the emitted Planck intensity (or radiance) from the earth and the atmosphere is smaller than that
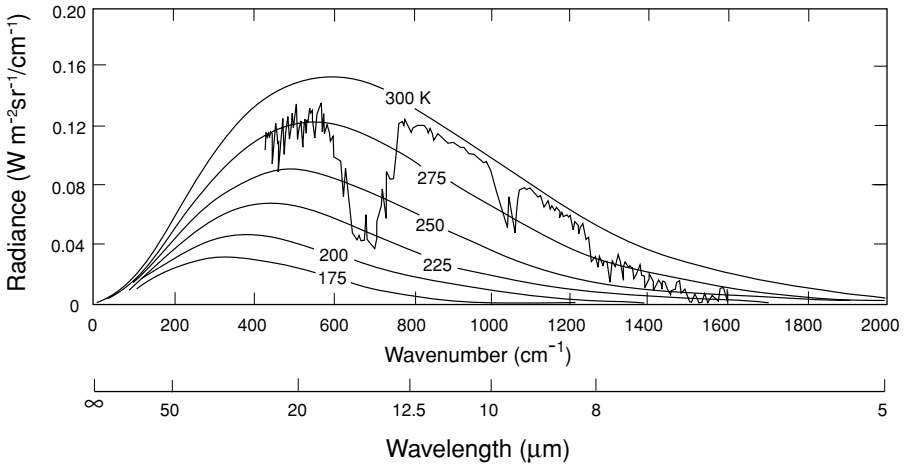
Figure 4.1 Theoretical Planck radiance curves for a number of the earth's atmospheric temperatures as a function of wavenumber and wavelength. Also shown is a thermal infrared emission spectrum observed from the Nimbus 4 satellite based on an infrared interferometer spectrometer.

of solar radiation, whereas the wavelength for the intensity peak of the earth's radiation field is longer than that of solar radiation. The energy emitted from the earth–atmosphere system is also referred to as *thermal IR* or *terrestrial radiation*. The spectral distribution of radiance emitted by a blackbody source at various temperatures in the earth's atmosphere in terms of wavenumber as well as wavelength is displayed in Fig. 4.1, where a measured atmospheric emission spectrum obtained from the infrared interferometer spectrometer (IRIS) instrument on board the Nimbus 4 satellite is also shown. The envelope of the emission spectrum is very close to the spectrum emitted from a blackbody with a temperature of about 290 K, which is about the temperature of the surface. It is evident that a large portion of thermal infrared energy is trapped by various gases in the atmosphere.

Among these gases, carbon dioxide, water vapor, and ozone are the most important absorbers. Some minor constituents, such as carbon monoxide, nitrous oxide, methane, and nitric oxide, which are not identified in Fig. 4.1, are relatively minor absorbers insofar as the heat budget of the earth–atmosphere system is concerned. Carbon dioxide absorbs a significant amount of thermal infrared radiation in the 15 μm band from about 600 to 800 cm^{-1}. This spectral region also corresponds to the maximum intensity of the Planck function in the wavenumber domain. Water vapor absorbs thermal infrared in the 6.3 μm band from about 1200 to 2000 cm^{-1} and in the rotational band (<500 cm^{-1}). Except for ozone, which has an absorption band in the 9.6 μm region, the atmosphere is relatively transparent from 800 to 1200 cm^{-1}. This region is referred to as the *atmospheric window*. As discussed in Section 3.1.2, the distribution of carbon dioxide is fairly uniform over the global space, although there has been observational evidence indicating a continuous global increase over the past century as a result of the increase in combustion of fossil fuels. This

leads to the question of possible changes in the earth's climate due to the increasing concentration of carbon dioxide (Section 8.4.1). Unlike carbon dioxide, however, water vapor and ozone are highly variable with respect to both time and geographical location.

In a clear atmosphere without clouds or aerosols, a large portion (about 50%) of solar energy transmits through the atmosphere and is absorbed by the earth's surface (see Fig. 3.9). In contrast, energy emitted from the earth is largely absorbed by carbon dioxide, water vapor, ozone, and other trace gases in the atmosphere as shown in Fig. 4.1. The trapping of thermal infrared radiation by atmospheric gases is typical of the atmosphere and is therefore called the *atmospheric effect*. It is also referred to as the *greenhouse effect* because of the similar way in which the glass covering of a greenhouse transmits solar radiation but absorbs emitted thermal infrared radiation. The climatological surface temperature is about 288 K. We may express the surface temperature in terms of the equilibrium temperature in the form

$$T_s = T_e + \gamma H, \tag{4.1.2}$$

where γ is the lapse rate and H is defined as the effective height of the greenhouse effect. Based on the standard lapse rate of 6.5 K km^{-1}, H is about 5 km.

A final note is in order. Solar radiation is also called *shortwave radiation* because solar energy is concentrated in shorter wavelengths with its peak at about 0.5 μm. Thermal infrared radiation emitted from the earth and the atmosphere is also referred to as *outgoing longwave radiation* (OLR) because its maximum energy is in the longer wavelengths at about 10 μm. The solar and infrared spectra are separated into two spectral ranges above and below about 5 μm, and the overlap between them is relatively small. This distinction makes it possible to treat the two types of radiative transfer and source functions separately and thereby simplify the complexity of the radiative transfer problem.

Following a discussion of the general characteristics of the vibrational–rotational spectra of water vapor, carbon dioxide, ozone and other minor gases, we present the fundamental theory of infrared radiative transfer and the line-by-line method for the numerical calculations involved. Next, we introduce approximate methods for infrared radiative transfer calculation, including the correlated k-distribution, band model, and broadband emissivity techniques. The subject of infrared radiative transfer in cloudy atmospheres is also further discussed. Finally, typical results of infrared heating and cooling rate profiles in both clear and cloudy atmospheres are presented.

4.2   Absorption and Emission in the Atmosphere

4.2.1   Absorption in the Thermal Infrared

4.2.1.1   WATER VAPOR

Following the discussion in the subsection on water vapor in Section 3.2.3, the pure H$_2$O rotational band ranges from 0 to 1000 cm^{-1}. This band is important in

the generation of tropospheric cooling. The ν_2 fundamental band at 1594.78 cm^{-1} (6.25 μm) is the most important vibrational–rotational band of water vapor. The two other fundamental bands, ν_1 and ν_3, are found to be close to one another and are centered at 2.7 μm. The H_2O isotopes $HH^{18}O$, $HH^{17}O$, and $HD^{16}O$ have been identified in both the rotational and ν_2 bands.

The region from 800 to 1200 cm^{-1}, the thermal infrared window, contains the moderately strong 9.6 μm band of ozone discussed later. Apart from the ozone band, absorption is continuous and is primarily due to the water vapor species. The attenuation due to the water vapor continuum in the 10 μm window remains a theoretical mystery. It has been suggested that the continuum results from the accumulated absorption of the distant wings of water vapor lines, principally in the far-infrared part of the spectrum. This absorption is caused by the collision broadening between absorbing molecules (H_2O–H_2O) and between absorbing and nonabsorbing molecules (H_2O–N_2). There is some evidence that contributions to continuous absorption may be caused by water dimer [($H_2O{\cdot}H_2O$)]. Absorption by water dimer depends significantly on water vapor pressure and temperature. Although accurate and well-controlled measurements are required in order to account for the water vapor continuum in real atmospheric situations, limited experimental measurements have been used to develop empirical parameterizations.

4.2.1.2   CARBON DIOXIDE

As discussed in Subsection 3.2.3.3, the bending mode, ν_2, is degenerate and consists of ν_{2a} and ν_{2b} vibrations at the same frequency. The CO_2 15 μm band represents this particular vibration. Owing to perpendicular vibration, the ν_2 fundamental transition is coupled with rotational transitions corresponding to changes in the quantum number $\Delta J = -1, 0, +1$. The spectral lines produced by these changes are referred to as the P, Q, and R branches, respectively, as noted in Section 1.3.1. For the P and R branches, the lines have wavenumbers lower and greater, respectively, than those of the line center. In the case of the Q branch, the lines are clustered near the center of the 15 μm band.

Because of the two vibrational modes, ν_{2a} and ν_{2b}, a $\pi/2$ phase difference is produced between them. The central carbon atom will perform a rotary motion about the symmetry axis and generate a component of angular momentum along this axis. This momentum is quantized with a quantum number l that is less than or equal to ν_2. This quantum number is added as a superscript to ν_2 in the designation of the vibrational level, e.g., 01^10, 02^00, 02^20. The $2\nu_2$ level of the carbon dioxide molecule lies very close to the ν_1 level, causing a resonance referred to as *Fermi resonance*. Because of the selection rules on the vibrational angular momentum, only the level 02^00 can combine with ν_1.

Besides the ν_2 fundamental band, numerous combination bands have been detected in the 15 μm region. Simultaneous transitions in two of the vibration modes are possible, resulting in weak combination (or difference) frequencies. There are also numerous hot bands in the 15 μm CO_2 band. These bands are produced by transitions between excited levels and are significant in cooling-rate calculations in the middle
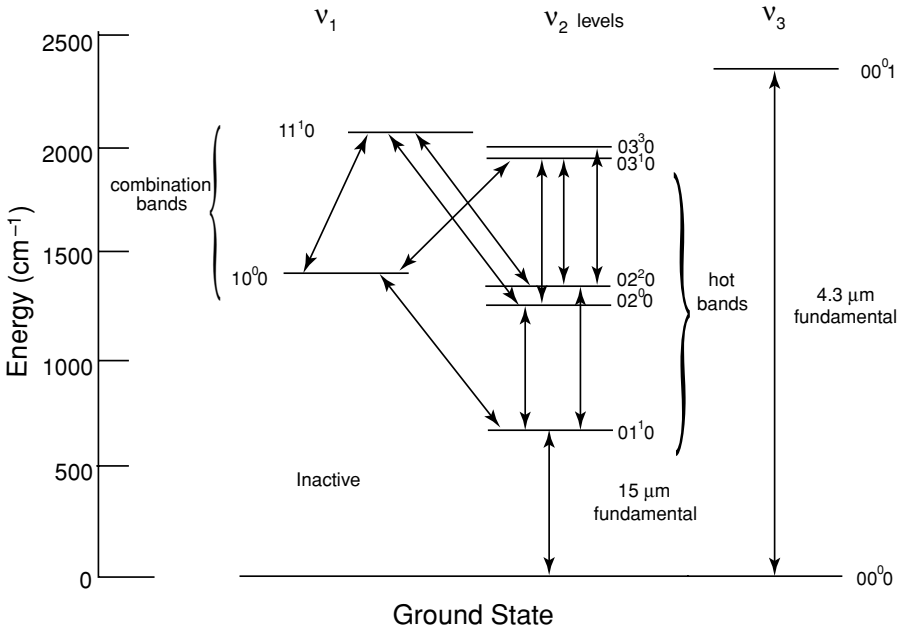
Figure 4.2   The most important vibrational transitions that produce the 15 μm CO_2 band, including the ν_2 fundamental, the hot bands, and the combination bands (data taken from Lopez-Puertas *et al.,* 1986). The rotational bands in this vibrational mode are the $P(\Delta J = -1)$, $Q(\Delta J = 0)$, and $R(\Delta J = +1)$ branches. The isotopes include $^{16}O^{12}C^{16}O$, $^{16}O^{13}C^{16}O$, and $^{16}O^{12}C^{18}O$.

atmosphere. A summary diagram of the 15 μm CO_2 energy transitions is presented in Fig. 4.2.

There are several overtone and combination bands of carbon dioxide in the IR. At atmospheric temperatures, most of the molecular population is in the vibrational level with quantum number $\upsilon = 0$. When transitions take place between nonadjacent levels ($\Delta \upsilon = 2, 3, 4$), weaker overtone frequencies are produced. Two moderately strong bands appear in the solar spectrum and are centered at 1063.8 and 961.0 cm^{-1}. Both are parallel bands and have been used for the development of the CO_2 laser emission at about 10.6 μm. The bands near 5 μm consist of the $3\nu_2$ band at 5.2 μm and several combination bands at 4.8 μm.

4.2.1.3   OZONE

The ozone molecule has an asymmetric top configuration similar to the water vapor molecule, but with a different apical angle, giving it a relatively strong rotational spectrum. The three isotopes, $^{16}O_3$, $^{16}O^{18}O^{16}O$, and $^{16}O^{16}O^{18}O$, have fundamental bands in the 9.6 and 14.27 μm regions. The ν_1 and ν_3 fundamental vibration modes are centered at 1110 and 1043 cm^{-1} and constitute the well-known 9.6 μm ozone band. The ν_2 fundamental band, centered at 705 cm^{-1} (14.27 μm), is well-masked

by the strong CO_2 15 μm band and appears to be less significant in atmospheric radiative transfer. There is also a relatively strong band of ozone at 4.75 μm that is produced by overtone and combination transitions. The electronic bands of ozone were discussed in Section 3.2.1.

4.2.1.4 METHANE

The methane molecule has a spherical top configuration. It has no permanent electric dipole moment and, hence, no pure rotational spectrum. There are four fundamental vibration modes. Of these, only ν_3 and ν_4, centered at 3020.3 and 1306.2 cm^{-1}, are active in the infrared spectrum. The ν_4 fundamental band of CH_4 is important in the climatic greenhouse effect. The inactive ν_1 and ν_2 fundamental bands are centered at 2914.2 and 1526 cm^{-1}. Methane also possesses a rich spectrum of overtone and combination bands that have been identified in the solar spectrum.

4.2.1.5 NITROUS OXIDE

The nitrous oxide molecule has a linear and asymmetric structure, with the configuration NNO. Similar to carbon dioxide, it has a single rotational constant and a detectable rotational spectrum. Numerous bands produced by the fundamental, overtone, and combination frequencies exist in the infrared. The three fundamental frequencies are centered at 1285.6 cm^{-1}(ν_1), 588.8 cm^{-1} (ν_2), and 2223.5 cm^{-1}(ν_3). The ν_1 fundamental band of nitrous oxide overlaps the ν_4 fundamental band of methane.

4.2.1.6 CHLOROFLUOROCARBONS

The methyl chloride (CH_3Cl) molecule has two bands of interest to atmospheric infrared radiative transfer: the ν_3 band at 732 cm^{-1} and the ν_2 band at 1350 cm^{-1}. In the ν_8 region, which is centered at 1161 cm^{-1} and has a band at 1095 cm^{-1}, Q branch features have been found for dichlorodifluoromethane (CF_2Cl_2). For the trichlorofluoromethane ($CFCl_3$) molecule, the ν_1 and ν_4 fundamental transitions are active and centered at 848 and 1085 cm^{-1}, respectively. The methylchloroform (CH_3CCl_3) molecule has a narrow Q branch associated with the ν_2 fundamental band at 1348.5 cm^{-1}. The carbon tetrachloride (CCl_4) molecule has an active band in the ν_3 region near 796 cm^{-1}. Absorption of these anthropogenic trace gases is primarily located in the window region. Thus, their potential increase can make the atmospheric window "dirty" and may lead to significant greenhouse effects.

All the preceding absorbing gases are identified in the infrared spectrum presented in Fig. 4.3, obtained from a scanning high-resolution interferometer sounder (S-HIS; see also Fig. 4.12), an instrument that measures the emitted thermal radiation between 3.3 and 18 μm from the NASA high-flying ER-2 aircraft at about 20 km. The spectral resolution of this interferometer varies but is on the order of 0.01 cm^{-1}. The spectrum is presented in terms of the brightness temperature (see Section 7.5.1 for the conversion of radiance to brightness temperature). Note that in addition to the absorption of the 4.3 μm CO_2 band (Section 3.2.3) in the solar spectral region of 3.3 to 5.0 μm, the absorption of nitrogen species is also identified.
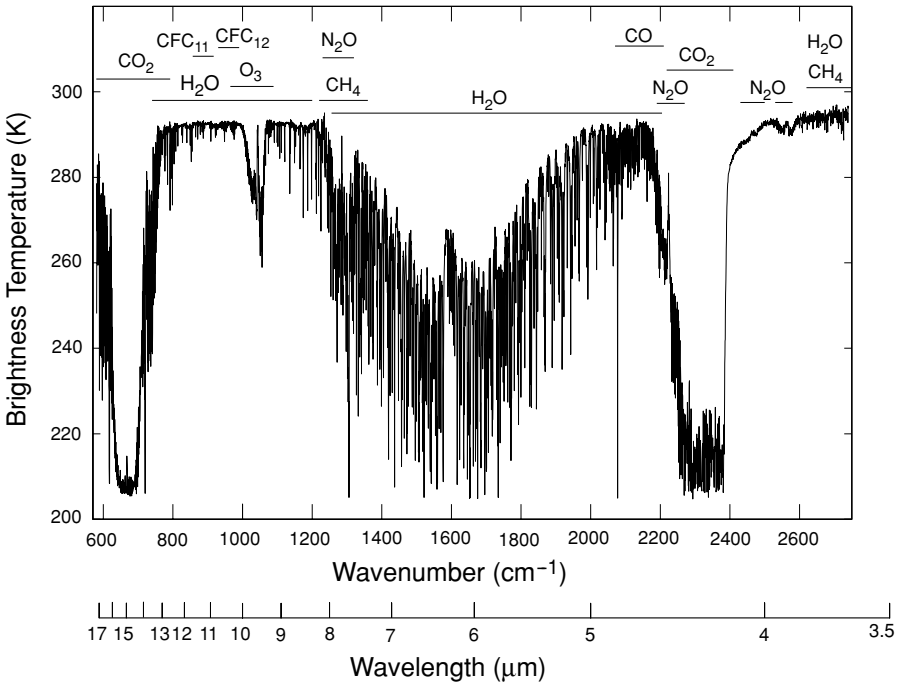
Figure 4.3 Observed infrared spectrum displaying all the absorption gases and their spectral location. This spectrum was obtained from the scanning high-resolution interferometer sounder (S-HIS), which measured the emitted thermal radiation between 3.3 and 18 μm, onboard the NASA ER-2 aircraft over the Gulf of Mexico southeast of Louisiana on April 1, 2001 (courtesy of Allen Huang and Dave Tobin of the University of Wisconsin).

4.2.2   Fundamentals of Thermal Infrared Radiative Transfer

Consider an absorbing and emitting medium. A pencil of radiation traversing this medium will be weakened by the interaction with matter through absorption. At the same time, this radiation may be strengthened by thermal emission from the medium (see Section 1.4.3). This pencil of radiation is usually represented by its intensity (or radiance), I_ν, in the field of radiative transfer. The general equation for radiative transfer in an absorbing and emitting medium can be written in terms of the differential change in the intensity in the form

$$-\frac{1}{k_\nu \rho_a} \frac{dI_\nu}{ds} = I_\nu - J_\nu, \tag{4.2.1}$$

where k_ν denotes the absorption coefficient, ρ_a is the density of absorbing gases, s is the slant path, and J_ν is the source function. For applications concerning the radiation budget of the planet, it suffices to consider the intensity as being independent of time. Moreover, it is commonly assumed that, in localized portions, the atmosphere is in thermodynamic equilibrium, as well as being plane-parallel. The first

assumption allows us to use the Planck intensity for the source function by virtue of Kirchhoff's law. The plane-parallel assumption implies that variations in the intensity and atmospheric parameters (temperature and gaseous profiles) are permitted only in the vertical direction (e.g., height or pressure). Under this assumption, absorption and emission processes would be symmetrical with respect to the azimuthal angle. It follows that the intensity is a function of the vertical position and zenith angle. Under these conditions, the basic equation that governs thermal IR radiation in the height coordinate may be written in the form

$$-\mu \frac{dI_\nu(z, \mu)}{k_\nu \rho_a dz} = I_\nu(z, \mu) - B_\nu(z), \tag{4.2.2}$$

where the Planck intensity $B_\nu(z) = B_\nu(T(z))$.

Because of the height dependence of both the gaseous density and absorption coefficient, it is convenient to define the *normal optical depth,* or simply *optical depth,* in the form

$$\tau = \int_z^{z_\infty} k_\nu(z')\rho_a(z')dz' = \int_0^p k_\nu(p')q(p')\frac{dp'}{g}, \tag{4.2.3}$$

where z_∞ denotes the height at the top of the atmosphere (TOA), $q = \rho_a/\rho$, the gaseous mixing ratio, and ρ is the air density. We have introduced the pressure coordinate p, using the hydrostatic equation. Figure 4.4 shows the coordinate systems in optical depth, height, and pressure. The differential optical depth can be readily obtained from Eq. (4.2.3) with the form $d\tau = -k_\nu(z)\rho_a(z)dz = k_\nu(p)q(p)dp/g$. In

$0, z_\infty$									$T_\infty, u_1, 0$
τ', z'									T', u', p'
τ, z									T, u, p
τ', z'									T', u', p'
$\tau_*, 0$									$T_s, 0, p_s$
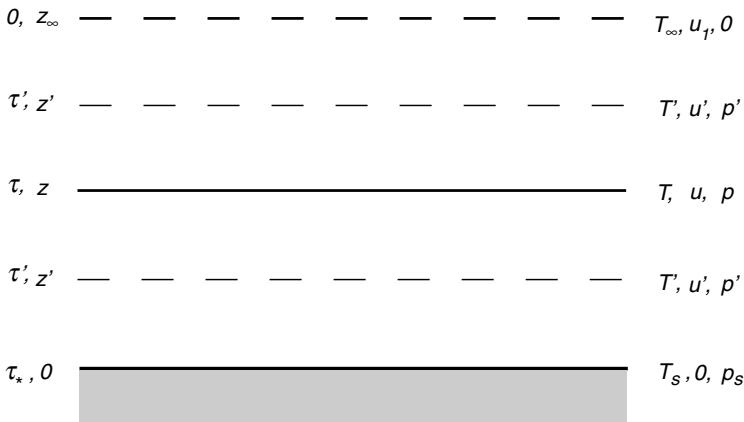
Figure 4.4 Coordinate systems in τ, z, u, T, and p for IR radiative transfer in plane-parallel atmospheres. u is the path length for absorbing gases defined from the surface upward. The total path length is denoted by u_1. T_∞ and z_∞ are temperature and height, respectively, at TOA. The surface temperature $T_s = T(\tau_*)$. The surface pressure is denoted by p_s.

terms of the τ coordinate, Eq. (4.2.2) may be rewritten as follows:

$$\mu \frac{dI_\nu(\tau, \mu)}{d\tau} = I_\nu(\tau, \mu) - B_\nu(\tau). \tag{4.2.4}$$

For the upward intensity, the zenith angle $0 \le \theta \le \pi/2$; that is, $0 \le \mu \le 1$. However, for the downward intensity, $\pi/2 \le \theta \le \pi$. In this case, we may set $\mu = -\mu$ for convenience in radiative transfer analyses.

Equation (4.2.4) represents a first-order differential equation. In order to solve both upward and downward components for an atmosphere with a total optical depth of τ_*, two boundary conditions are required. Under the plane-parallel assumption, these conditions are isotropic emissions from both the surface and TOA. In general, the earth's surface may be considered as a blackbody in the infrared, so that $I_\nu(\tau_*, \mu) = B_\nu(T(\tau_*))$. In addition, we may allow for a possible source of downward emission at TOA and write $I_\nu(0, -\mu) = B_\nu(\text{TOA})$. Normally, however, $B_\nu(\text{TOA}) \cong 0$. Subject to the preceding boundary conditions, the formal solutions for upward and downward intensities are given by

$$I_\nu^\uparrow(\tau, \mu) = B_\nu(\tau_*)e^{-(\tau_*-\tau)/\mu} + \int_\tau^{\tau_*} B_\nu(\tau')e^{-(\tau'-\tau)/\mu} \frac{d\tau'}{\mu}, \tag{4.2.5a}$$

$$I_\nu^\downarrow(\tau, -\mu) = \int_0^\tau B_\nu(\tau')e^{-(\tau-\tau')/\mu} \frac{d\tau'}{\mu}. \tag{4.2.5b}$$

We shall now define the *monochromatic transmittance* (also referred to as the *transmission function*) so that the exponential attenuation may be expressed by

$$T_\nu(\tau/\mu) = e^{-\tau/\mu}. \tag{4.2.6a}$$

The differential form is

$$\frac{dT_\nu(\tau/\mu)}{d\tau} = \frac{-1}{\mu}e^{-\tau/\mu}. \tag{4.2.6b}$$

The formal solutions for the intensities can then be expressed by

$$I_\nu^\uparrow(\tau, \mu) = B_\nu(\tau_*)T_\nu[(\tau_* - \tau)/\mu] - \int_\tau^{\tau_*} B_\nu(\tau')\frac{d}{d\tau'}T_\nu[(\tau' - \tau)/\mu]\,d\tau', \tag{4.2.7a}$$

$$I_\nu^\downarrow(\tau, -\mu) = \int_0^\tau B_\nu(\tau')\frac{d}{d\tau'}T_\nu[(\tau - \tau')/\mu]\,d\tau'. \tag{4.2.7b}$$

Equation (4.2.7a) can be applied to remote sensing from space if we set $\tau(\text{TOA}) = 0$.

For atmospheric heating rate calculations, the required quantities are the upward and downward flux densities (simply referred to as fluxes hereafter), which are the sum of the directional intensities from the upper and lower hemispheres, respectively. In accordance with the plane-parallel assumption, we have

$$F_\nu^{\uparrow\downarrow}(\tau) = 2\pi \int_0^1 I_\nu^{\uparrow\downarrow}(\tau, \pm\mu)\mu\,d\mu. \tag{4.2.8}$$

On noting the angular integration and in order to obtain the fluxes, we may define a physical parameter, referred to as *slab* or *diffuse transmittance,* in the form

$$T_\nu^f(\tau) = 2 \int_0^1 T_\nu(\tau/\mu)\mu \, d\mu. \tag{4.2.9}$$

With the aid of this definition, the expressions for fluxes may now be written in the form

$$F_\nu^\uparrow(\tau) = \pi B_\nu(\tau_*)T_\nu^f(\tau_* - \tau) - \int_\tau^{\tau_*} \pi B_\nu(\tau')\frac{d}{d\tau'}T_\nu^f(\tau' - \tau)\, d\tau', \tag{4.2.10a}$$

$$F_\nu^\downarrow(\tau) = \int_0^\tau \pi B_\nu(\tau')\frac{d}{d\tau'}T_\nu^f(\tau - \tau')\, d\tau'. \tag{4.2.10b}$$

The upward flux at a given level comes from two sources: the surface emission that is attenuated to that level and the emission contributions from the atmospheric layers characterized by Planck fluxes multiplied by the weighting function, $dT_\nu^f/d\tau$. Likewise, the downward flux at a given level is produced by the contributions from the atmospheric layers.

Finally, to account for the contributions from all wavenumbers in the thermal IR spectrum, an integration of the monochromatic flux with respect to wavenumber must be performed. Since τ is a function of the wavenumber, we use the height coordinate and write

$$F^{\uparrow\downarrow}(z) = \int_0^\infty F_\nu^{\uparrow\downarrow}(z)\, d\nu. \tag{4.2.11}$$

At this point, the transfer of thermal IR radiation in plane-parallel atmospheres (without clouds) is formally solved. Conditional to the definition of diffuse transmittance, the computation of atmospheric fluxes involves solving the integrations over the wavenumber and along the optical depth.

4.2.3 Line-By-Line (LBL) Integration

Absorption line parameters for various gases can be computed from fundamental quantum mechanics theory. Laboratory data are also available for a limited number of spectral intervals. Based on theory and measurements, line parameters have been compiled over the range 0 to 17,900 cm^{-1} (Rothman *et al.,* 1998). Data for more than 1 million lines have been presented. The absorption line parameters are listed in data form in terms of line position (in cm^{-1}), line intensity [in cm^{-1}/(mol cm^{-2}) at 296 K], air-broadened half-width (in cm^{-1}/atm at 296 K), and lower-state energy (in cm^{-1}). The molecular species include H_2O, CO_2, O_3, N_2O, CO, CH_4, and O_2.

For a given wavenumber and species, contributions to transmittance arise from the absorption coefficients for N lines. The optical depth is then

$$\tau = \sum_{j=1}^N \tau_j = \int_u \sum_{j=1}^N k_{\nu,j}(u)\, du, \tag{4.2.12}$$

where $j(= 1, 2, \ldots, N)$ is the index of the absorption coefficient for the jth line. Thus, the absorption coefficient can be expressed in terms of line strength and line shape in the form

$$k_\nu(p, T) = \sum_{j=1}^{N} S_j(T) f_{\nu, j}(p, T). \qquad (4.2.13)$$

In order to resolve individual lines, the absorption coefficient must be computed at wavenumber intervals that are smaller than the line half-width. In the upper stratosphere, absorption and emission processes are dominated by CO_2 and O_3. Broadening of the absorption lines is primarily due to the Doppler effect. The Doppler half-width in the 15 μm CO_2 and 9.6 μm O_3 bands is $\sim$0.0005–0.001 cm^{-1}. The spectral interval in these two bands covers about 400 cm^{-1}. Thus, the absorption must be calculated at more than half a million points if individual lines are to be resolved.

In the troposphere, absorption due to H_2O predominates. H_2O lines cover essentially the entire infrared spectrum with a spectral region of about 15,000 cm^{-1}. These lines are broadened by collisions, and their half-widths are >0.01 cm^{-1}. Computations must be performed at about 1 million points to resolve H_2O lines. For each point, there are numerous lines and atmospheric conditions that must be considered for applications to atmospheric radiative transfer. The computer time required for line-by-line calculations, even with the availability of a supercomputer, is formidable. This is especially true for flux calculations in which an integration over all absorption bands is necessary.

In the case of Lorentz line shapes, it is important to cut off the contribution of significant lines at computational points in a line-by-line program. The far wings of a pressure-broadened absorption line in the infrared have been suggested to be sub-Lorentzian. Two approaches have been used to compute the far wing contribution. The first approach is to multiply the Lorentz profile by an empirical function χ, such that $\chi = 1$ at the line center and $\chi = 0$ at some distance from the center. The second is to use the Lorentz profile for all wavenumbers but to cut the lines off at some distance from the line center. The lines can be cut off at a constant distance from the center or they can be cut off at a distance varying with the half-width; that is, the cutoff wavenumber $\nu_c = \beta\alpha$, with β a constant. Numerical computations show that absorption is affected only slightly by cutting a line off at a wavenumber of about 200 times the Lorentz half-width from the center.

In infrared radiative transfer calculations, it is advantageous to define the radiative parameters in a small spectral interval such that variation of the Planck function can be neglected. In terms of the basic parameters in the intensity and flux equations, we may define the spectral transmittance as

$$T_{\bar\nu}(u) = \int_{\Delta\nu} e^{-\tau} \frac{d\nu}{\Delta\nu} = \int_{\Delta\nu} \exp\left(-\int_u \sum_j k_{\nu, j}(u) du\right) \frac{d\nu}{\Delta\nu}. \qquad (4.2.14)$$

Thus, in order to calculate the spectral transmittance $T_{\bar\nu}(u)$ exactly, adequate and reliable summations must be performed to cover the absorption lines, the spectral

interval, and the nonhomogeneous path. Moreover, the absorption coefficient is a function of the line strength, which is in turn a function of temperature and the half-width, which is itself a function of pressure and temperature. Exact line-by-line calculations for the spectral transmittance are very tedious and require a significant amount of computer time. All the infrared radiative transfer theories that have been developed are essentially intended to simplify and economize the computation of the spectral transmittance by circumventing the integration over the spectral interval and nonhomogeneous path length.

For flux calculations, the diffuse transmittance defined in Eq. (4.2.9) is required. In general, a four-point Gaussian quadrature will give accurate results for integration over the cosine of the zenith angle μ. For many atmospheric applications, it suffices to use

$$T_{\bar{\nu}}^{f}(u) \cong T_{\bar{\nu}}(u/\bar{\mu}), \tag{4.2.15}$$

where $1/\bar{\mu}$, the inverse of the mean emergent angle, is referred to as the *diffusivity factor*. Based on numerical computations for thermal IR radiative transfer, a good approximation for $1/\bar{\mu}$ ranges from 1.66 to 2. The diffusivity of 1.66 was originally proposed by Elsasser (1942) and has been found to be a reliable approximation for the computation of radiative fluxes and heating rates in clear atmospheres. Finally, it should be noted that the notation T is being used in this text as transmittance defined by a wavenumber subscript in infrared radiative transfer discussion, and as the transmission function defined by incoming and outgoing directions in the presentation of multiple scattering processes. It is also being used as temperature throughout the text.

4.3   Correlated K-Distribution Method for Infrared Radiative Transfer

4.3.1   Fundamentals

The k-distribution method for the computation of infrared radiative transfer is based on the grouping of gaseous spectral transmittances according to the absorption coefficient k_{ν}. In a homogeneous atmosphere, spectral transmittance is independent of the ordering of k for a given spectral interval. Hence, wavenumber integration may be replaced by an integration over the k space. If the normalized probability distribution function for k_{ν} in the interval $\Delta\nu$ is given by $f(k)$ and its minimum and maximum values are $k_{\min}$ and $k_{\max}$, respectively, then the spectral transmittance may be expressed by

$$T_{\bar{\nu}}(u) = \int_{\Delta\nu} e^{-k_{\nu}u}\, \frac{d\nu}{\Delta\nu} = \int_{0}^{\infty} e^{-ku} f(k)\, dk, \tag{4.3.1a}$$

where we have set $k_{\min} \to 0$ and $k_{\max} \to \infty$, for mathematical convenience, and

$$\int_{0}^{\infty} f(k)\, dk = 1. \tag{4.3.1b}$$

From Eq. (4.3.1a), the probability distribution function is the inverse Laplace transform, L^{-1}, of the spectral transmittance such that

$$f(k) = L^{-1}(T_{\bar{\nu}}(u)). \tag{4.3.2}$$

If the spectral transmittance can be expressed in terms of an analytic exponential function and if the inverse Laplace transform can be performed, then an analytic expression can be derived for the probability distribution function (Exercises 4.3 and 4.4).

Moreover, a cumulative probability function may be defined in the form

$$g(k) = \int_0^k f(k)\,dk, \tag{4.3.3}$$

where $g(0) = 0$, $g(k \to \infty) = 1$, and $dg(k) = f(k)\,dk$. By definition, $g(k)$ is a monotonically increasing and smooth function in k space. By using the g function, the spectral transmittance can be written

$$T_{\bar{\nu}}(u) = \int_0^1 e^{-k(g)u}dg \cong \sum_{j=1}^M e^{-k(g_j)u}\Delta g_j. \tag{4.3.4}$$

From Eq. (4.3.3), since $g(k)$ is a smooth function in k space, the inverse will also be true here; that is, $k(g)$ is a smooth function in g space. Consequently, the integration in g space, which replaces the tedious wavenumber integration, can be evaluated by a finite sum of exponential terms, as shown in Eq. (4.3.4).

Figure 4.5a shows k_ν as a function of ν in a portion of the H_2O rotational band at a pressure of 600 mb and a temperature of 260 K. Figure 4.5b shows the probability distribution $f(k)$ as a function of k for this band [see the following for an evaluation of $f(k)$]. In Fig. 4.5c, the cumulative probability function $g(k)$ is shown as a function of k. We may then compute $k(g)$ as a function of g from Eq. (4.3.3). This curve is illustrated in Fig. 4.5d. Since g is a smooth monotonic function, a few quadrature points will suffice to achieve a high degree of accuracy in the transmittance computation. The physical foundation for the k distribution is quite simple, but it offers an advantage in the computation of infrared flux transfer. The idea of scrambling and ranking absorption lines was described in Ambartzumian's (1936) work on the equilibrium temperature of stellar atmospheres. Arking and Grossman (1972) used this method to discuss the line shape effect on the temperature of planetary atmospheres.

4.3.2   Application to Nonhomogeneous Atmospheres

The preceding theory of the k-distribution method assumes that the absorption coefficient is constant so that ν-integration can be replaced by g-integration. However, as pointed out in Section 4.2.3, the absorption coefficient varies greatly with pressure and temperature in terms of its half-width and line strength. Thus, in order to apply the k-distribution method to realistic atmospheres, variation in the absorption coefficient in the vertical must be accounted for. Basically, we must determine whether the
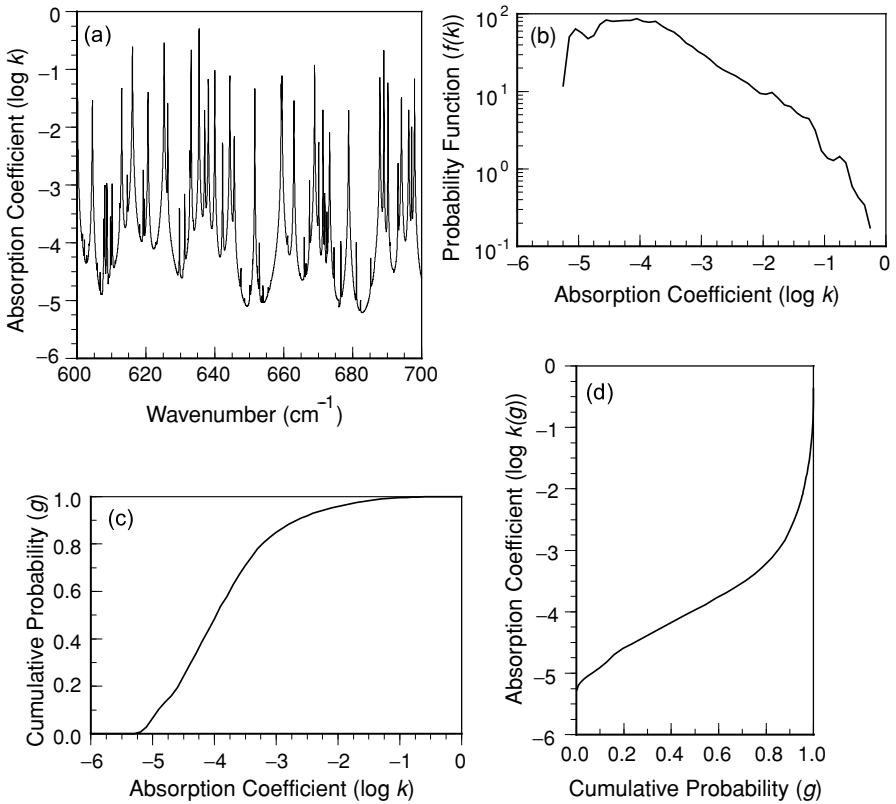
Figure 4.5   (a) Absorption coefficient k_ν in units of (cm atm)$^{-1}$ as a function of wavenumber with a resolution of 0.01 cm^{-1} in the H_2O rotational band with $p = 600$ mb and $T = 260$ K. (b) The probability function $f(k)$ of the absorption coefficient. (c) The cumulative probability function for $f(k)$ shown in (b), plotted as a function of k. (d) Same as (c), except that values of the absorption coefficient are expressed as a function of g.

following two integrations are equivalent:

$$T_{\bar{\nu}}(u) = \int_{\Delta\nu} \exp\left(-\int_u k_\nu \, du\right) \frac{d\nu}{\Delta\nu} \stackrel{?}{=} \int_0^1 \exp\left[-\int_u k(g) \, du\right] dg. \qquad (4.3.5)$$

Because the line profile varies significantly from one pressure (temperature) level to another, the rearrangement of all the lines so that the ν- and g-integrations are the same is not obvious. The method that assumes the two are equivalent is referred to as the correlated k-distribution (CKD) method. In the following, we examine the conditions under which the CKD method may be considered to be exact.

Consider first a single line of any shape in a spectral interval $\Delta\nu$. Let this interval be from $-\Delta\nu/2$ to $+\Delta\nu/2$. At the line center, $\nu(k_{\max}) = 0$ and at the end of the line $|\nu(k_{\min})| = \Delta\nu/2$, as shown in Fig. 4.6a. The cumulative probability function defined
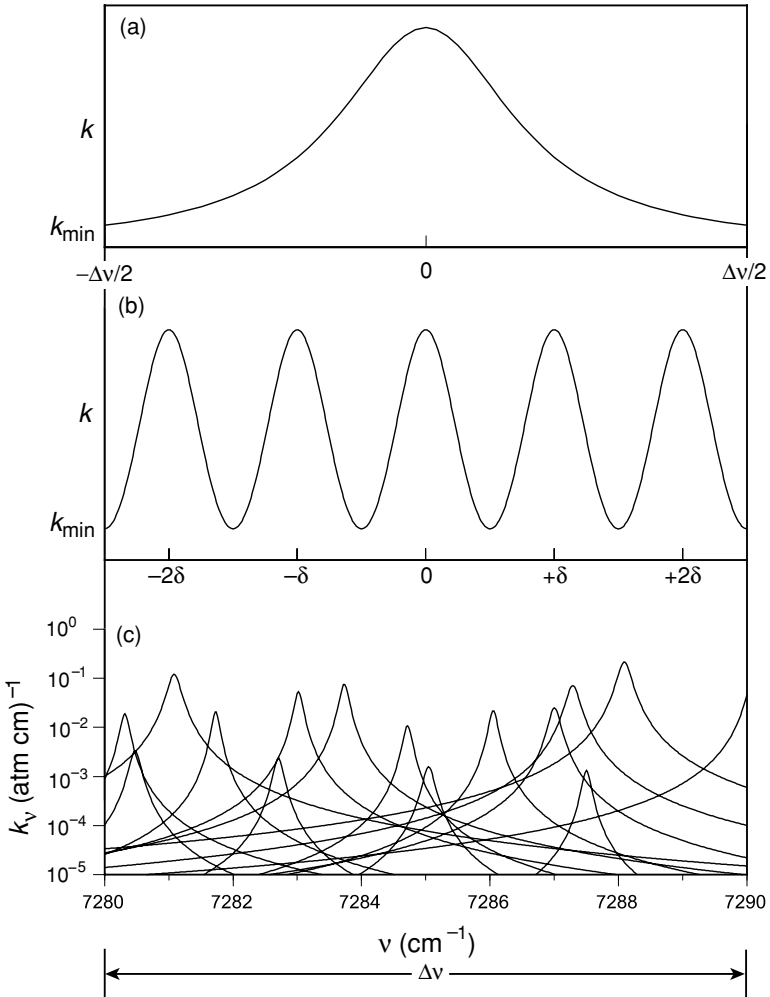
Figure 4.6   (a) A single line, (b) periodic (regular) lines where δ is the line spacing, and (c) Lorentz lines in a 10 cm^{-1} interval in the 1.38 μm H$_2$O band. Fifteen lines are shown.

in Eq. (4.3.3) is given by

$$g(k) = \frac{2}{\Delta v} \int_{k_{min}}^{k} \left| \frac{dv}{dk} \right|_{k=k'} dk' = \frac{2}{\Delta v} v(k) - 1. \tag{4.3.6a}$$

It follows that $(2/\Delta v)\, dv = dg$. That is, for a single line, we can replace the integration over the v-domain with an integration over the g-domain in Eq. (4.3.5). Consider now a group of n lines that occur periodically (or regularly) in a spectral interval Δv

(Fig. 4.6b). Let the line spacing be δ. Then we have

$$g(k) = \frac{2}{\delta} \int_{k_{\min}}^{k} \left| \frac{dv}{dk} \right|_{k=k'} dk' = \frac{2}{\delta} v(k) - 1.$$ (4.3.6b)

Thus, $(2/\delta) \, dv = dg$. Again, the integration over the v-domain can be replaced by an integration over the g-domain. Calculations of the spectral transmittance would be the same in the two approaches. For reference purposes, Fig. 4.6c shows a group of absorption lines assuming the Lorentz shape in a 10 cm^{-1} spectral interval within the $1.38 \ \mu\text{m} \ H_2O$ band.

Next, if the absorption coefficient and/or the path length is small, referred to as the *weak-line limit,* we have from Eq. (4.3.5), regardless of the line shape, the following:

$$T_{\bar{v}}(\text{weak}) \cong \int_{\Delta v} \left[1 - \int_u k_v \, du \right] \frac{dv}{\Delta v} = 1 - \int_u k_* \, du,$$ (4.3.7a)

where $k_* = \sum S_j / \Delta v$ and is independent of wavenumber. The last expression is obtained by employing the definition of the absorption coefficient given in Eq. (4.2.13). Since k_* is small, we may rewrite Eq. (4.3.7a) in the form

$$T_{\bar{v}}(\text{weak}) = \exp \left(- \int_u k_* \, du \right).$$ (4.3.7b)

Under the weak-line limit, the spectral transmittance may be expressed explicitly by an exponential function without integration over v. This is referred to as the *gray approximation*. In this case, replacement of $dv/\Delta v$ by dg can readily be made, since k_* is independent of v.

Now consider a group of lines. The absorption coefficient at a given wavenumber is contributed from all the lines and using the Lorentz profile, we have

$$k_v(p, T) = \sum_j \frac{S_j(T)}{\pi} \frac{\alpha_j(p, T)}{(v - v_0)^2 + \alpha_j^2(p, T)},$$ (4.3.8)

where the line strength is a function of temperature and the half-width is a function of both pressure and temperature [see Eq. (1.3.14)]. At the line center, $v = v_0$, $k_v \sim 1/\alpha_j(p, T) \sim 1/p$. However, in the far wing, $|v - v_0| \gg \alpha$, and $k_v \sim \alpha_j(p, T) \sim p$. Under normal atmospheric conditions, the half-widths of Lorentzian lines are small in comparison to the mean line spacing. This is particularly true for most water vapor lines. Near the line centers, absorption becomes saturated so that the transfer of radiation through the wing regions becomes predominant. This is the condition for the *strong-line limit*. In this case, we may write

$$k_v(p, T) \cong \sum_j \frac{S_j(T)}{\pi} \frac{\alpha_j(p, T)}{(v - v_0)^2} = k_v^* f^*(p, T),$$ (4.3.9)

where

$$k_v^* = 1/\pi(v - v_0)^2, \quad f^*(p, T) = \sum_j S_j(T)\alpha_j(p, T).$$

The condition in which the dependent variable ν may be separated from the absorption coefficient is referred to as the *scaling approximation*. Substituting Eq. (4.3.9) into Eq. (4.3.5) leads to

$$T_{\bar{\nu}}(\text{strong}) \cong \int_{\Delta\nu} e^{-k_\nu^* \tilde{u}} \frac{d\nu}{\Delta\nu} \equiv \int_0^1 e^{-k_\nu^*(g)\tilde{u}} dg, \tag{4.3.10}$$

as in the homogeneous case, where

$$\tilde{u} = \int_u f^*(p, T) \, du.$$

In essence, the scaling approximation allows the transformation of a nonhomogeneous path into a homogeneous path.

On the basis of the preceding discussion, the CKD method is exact for a single line and periodic lines, as well as in the limits of weak-line and strong-line approximations. The CKD method is an extension of the k-distribution to nonhomogeneous paths, the idea of which appeared in a technical report (Lacis *et al.,* 1979). In this method, the vertical nonhomogeneity of the atmosphere is accounted for by assuming a simple correlation of k-distributions at different temperatures and pressures such that the spectral transmittance can be expressed by

$$T_{\bar{\nu}}(u) \cong \int_0^1 \exp\left[-\sum_i k_i(g)\Delta u_i \right] dg. \tag{4.3.10}$$

Because the CKD approach allows the use of k-distributions at each altitude, an appropriate Voigt profile can also be accounted for throughout the atmosphere. Moreover, it can also be used for absorption bands in both solar and thermal infrared spectra and, at the same time, the results from this method can be directly incorporated into multiple-scattering processes associated with cloud and aerosol particles. Goody *et al.* (1989), Lacis and Oinas (1991), and Fu and Liou (1992) have undertaken proof of the validity of CKD and have tested its accuracy under a variety of atmospheric conditions. For flux calculations, errors due to the assumptions in CKD with respect to LBL results are generally on the order of 1%. Thus, the CKD method is a powerful technique for use in radiative transfer parameterizations in dynamic and climate models especially when multiple scattering of cloud and aerosol particles must be accounted for in flux and heating rate calculations. The accuracy of CKD will be further elaborated upon in Section 4.7. In the following, we discuss a simple numerical procedure for the sorting of absorption lines.

4.3.3   Numerical Procedures and Pertinent Results

Consider a spectral interval $\Delta\nu$ that contains numerous absorption lines. We may divide this interval into a subset of $\Delta\nu_i(i = 1, 2 \ldots, N)$ such that they are less than the line's half-width. In this case, the probability distribution function may be

written as

$$f(k) = \frac{1}{\Delta \nu} \frac{d\nu}{dk} = \frac{1}{\Delta \nu} \sum_j \left| \frac{\Delta \nu_j}{\Delta k} \right| \tag{4.3.11}$$

for a specific k, and $f(k)$ is 0 if the maximum of a line k_j (max) is smaller than k. The cumulative probability function is then

$$g(k) = \frac{1}{\Delta \nu} \sum_j \int_0^k \left| \frac{\Delta \nu_j}{\Delta k'} \right| dk'$$

$$= \frac{1}{\Delta \nu} \sum_j \int_0^k \Delta \nu_j(k) = \frac{n(0, k)}{N}, \tag{4.3.12}$$

where the total number of lines (computational points) $N = \Delta \nu / \delta$, with δ the mean line spacing defined by

$$\delta = \sum_{j=1}^N \nu_j / N, \tag{4.3.13}$$

and $n(0, k)$ denotes the number of lines (computational points) that contribute to k cumulatively.

For illustration purposes, let us consider three lines as displayed in Fig. 4.7. We may divide the absorption coefficient into 10 intervals. As shown, $n(0, \Delta k) = 4$, $n(0, 2\Delta k) = 4 + 6 = 10$, and so on. The total number $N = n(0, 9\Delta k) = 35$. From Eq. (4.3.12), we have $g(0) = 0$, $g(\Delta k) = 4/35$, $g(2\Delta k) = 10/35$, ..., and $g(9\Delta k) = 1$. The absorption coefficient $k(g)$ can now be expressed in the g-domain, as shown in Fig. 4.7, and is a monotonically increasing function. It follows that the computation of spectral transmittance may now be performed over the g-domain, with a relatively small number of computational points, instead of the ν-domain, which consists of highly fluctuating curves. Exercise 4.1 requires the analysis and computation of the g function and spectral transmittance for a group of water vapor lines in a 10 cm^{-1} spectral interval.

The foregoing procedure can be used to compute g functions for a given pressure and temperature, i.e., $g(k, p, T) = n(0, k; p, T)/N$. The effect of pressure and temperature on the absorption coefficient in the g-domain is shown in Fig. 4.8 using a spectral interval of 540–670 cm^{-1} in the 15 μm CO$_2$ band. For a given g, the absorption coefficient increases with increasing pressure and temperature. The pressure effect occurs through its dependence on the half-width, whereas the temperature effect is associated with both line strength and half-width. In numerical calculations of IR heating rates, it would be advantageous and practical to develop efficient parameterizations to account for the dependence of pressure and temperature on the absorption coefficient in the g-domain.
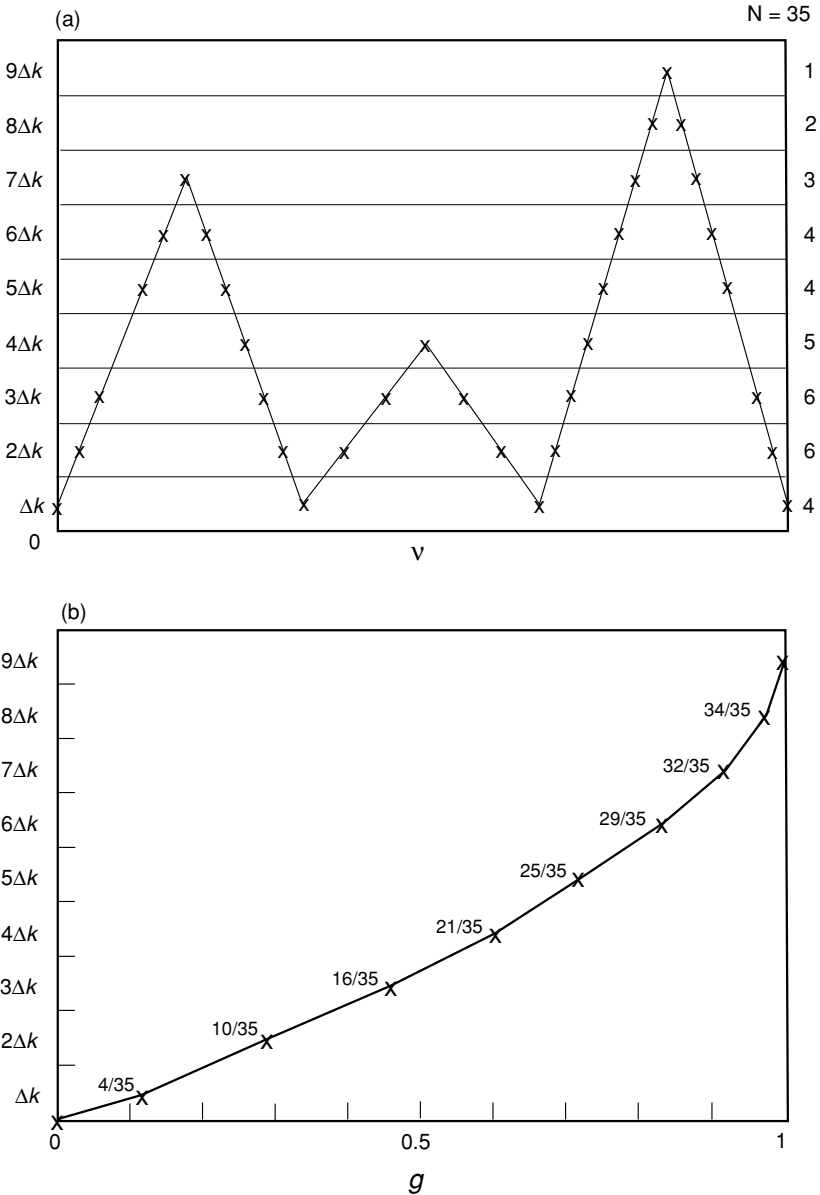
Figure 4.7 Illustration of the mapping of the absorption coefficient from the ν-domain to the g-domain. (a) Absorption coefficients of three lines in the ν-domain and division of the k-space into 10 equal intervals. The numbers are the data points in each interval with a total number of 35. (b) By definition, $g(j\Delta k) = n(0, j\Delta k)/N$, $j = 0, 1, \ldots, 9$. Thus, the data points in the ν-domain are transformed to the g-domain, where g is a monotonic increasing function.
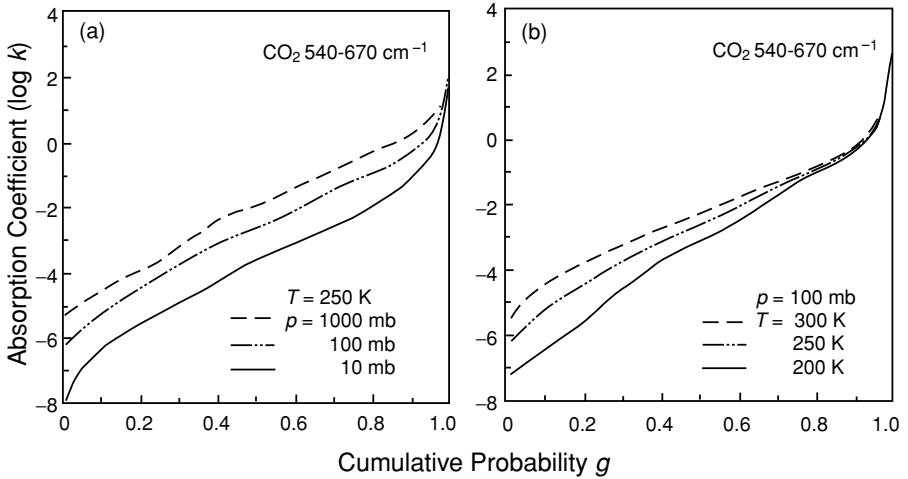
Figure 4.8   The absorption coefficient in $(cm\ atm)^{-1}$ as a function of the cumulative probability g for the CO_2 540–670 cm^{-1} spectral region (a) for a temperature of 250 K with three pressures and (b) for a pressure of 100 mb with three temperatures.

4.3.4   Line Overlap Consideration

Examination of the line displacement reveals that H_2O rotational lines and a portion of 15 μm CO_2 lines overlap significantly. Thus, an appropriate treatment must be undertaken to evaluate the spectral transmittance that contains the two gases. Let the mixing ratios for CO_2 and H_2O be q_c and q, respectively. Their respective optical depths are, therefore, given by

$$\tau_1 = q_c \int k_1(v, p, T)\rho\, dz, \qquad (4.3.14a)$$

$$\tau_2 = \int k_2(v, p, T)q\rho\, dz, \qquad (4.3.14b)$$

where ρ is the air density and q_c is a constant. The monochromatic transmittance for the two gases, by definition, can be written in the form

$$T_v(q, q_c) = e^{-(\tau_1 + \tau_2)} = T_v(q_c)T_v(q). \qquad (4.3.15)$$

Consider a small spectral interval such that $\Delta v \le 5$ cm^{-1}. The H_2O and CO_2 lines may be treated as statistically independent so that

$$\int_{\Delta v} [T_v(q) - T_{\bar{v}}(q)][T_v(q_c) - T_{\bar{v}}(q_c)]\frac{dv}{\Delta v} = 0. \qquad (4.3.16)$$

In this case, the spectral transmittance for the two gases is separable, leading to

$$T_{\bar{v}}(q, q_c) = T_{\bar{v}}(q_c)T_{\bar{v}}(q) = \int_{\Delta v} e^{-\tau_1}\frac{dv}{\Delta v} \cdot \int_{\Delta v} e^{-\tau_2}\frac{dv}{\Delta v}. \qquad (4.3.17a)$$

We may then apply the CKD method to each gas to obtain

$$T_{\bar{\nu}}(q, q_c) \cong \int_0^1 \int_0^1 \exp\left\{-\int [k(g_1)q_c + k(g_2)q]\rho\, dz\right\} dg_1\, dg_2. \qquad (4.3.17b)$$

The spectral transmittance now consists of two integrations. If M and N calculations are required for CO_2 and H_2O separately, then the total number of calculations would be $M \times N$, substantially increasing the computational requirement.

In the following, we wish to find a simplified approach to evaluate the spectral transmittance involving two overlap gases. We shall first consider the homogeneous condition such that the spectral transmittance defined in Eq. (4.3.17a) can be expressed by the definition of the k-distribution in the form

$$T_{\bar{\nu}}(q, q_c) = \int_0^\infty e^{-k_1 u_1} f_1(k_1) dk_1 \cdot \int_0^\infty e^{-k_2 u_2} f_2(k_2)\, dk_2, \qquad (4.3.17c)$$

where the path lengths $u_1 = q_c u$ and $u_2 = qu = u_1 q/q_c$, with u being the path length for air. With the definitions of u_1 and u_2, the two individual spectral transmittances can then be written as

$$T_{\bar{\nu}}(q_c) = T_1(u_1) = \int_0^\infty e^{-k_1 u_1} f_1(k_1)\, dk_1, \qquad (4.3.18a)$$

$$T_{\bar{\nu}}(q) = T_2(u_2) = \int_0^\infty e^{-k_2 u_2} f_2(k_2)\, dk_2. \qquad (4.3.18b)$$

Letting $k_2 = k'' q_c/q$ in Eq. (4.3.18b), we then have

$$T_2(u_1) = \int_0^\infty e^{-k'' u_1} f_2^*(k'')\, dk'', \qquad (4.3.18c)$$

where the probability distribution function

$$f_2^*(k'') = \frac{q_c}{q} f_2\left(\frac{q_c}{q}k''\right) = L^{-1}(T_2(u_1)). \qquad (4.3.19a)$$

From Eq. (4.3.18a), by letting $k_1 = k'$ for convenience of presentation, we also have

$$f_1(k') = L^{-1}(T_1(u_1)). \qquad (4.3.19b)$$

It follows that Eq. (4.3.17c) can be rewritten in the form

$$T_{\bar{\nu}}(q, q_c) = \int_0^\infty e^{-k' u_1} f_1(k')\, dk' \cdot \int_0^\infty e^{-k'' u_1} f_2^*(k'')\, dk''$$

$$= \int_0^\infty \int_0^\infty e^{-(k'+k'')u_1} f_1(k') f_2^*(k'')\, dk'\, dk''. \qquad (4.3.20a)$$

We may define $k = k' + k''$. By replacing k' with $k - k''$, interchanging the double integrals and their limits, and employing the convolution theorem of the Laplace

transform given by

$$f(k) = L^{-1}(T_{\bar{\nu}}(q, q_c)) = \int_0^k f_1(k - k'') f_2^*(k'') \, dk'', \tag{4.3.20b}$$

we can prove that

$$T_{\bar{\nu}}(q, q_c) = \int_0^\infty e^{-ku_1} f(k) \, dk = \int_0^1 e^{-k(g)u_1} \, dg. \tag{4.3.20c}$$

Because the mixing ratio for carbon dioxide q_c is constant, we may apply the correlated condition for an inhomogeneous path to Eq. (4.3.20c) to obtain

$$T_{\bar{\nu}}(q, q_c) \cong \int_0^1 \exp\left[-q_c \int k(g, T, p, q)\rho \, dz\right] dg, \tag{4.3.21a}$$

where the cumulative probability function is given by

$$g(k, p, T, q) = \int_0^k f(k, p, T, q) \, dk. \tag{4.3.21b}$$

The CKD for a single-mixture gas requires the same correlated assumptions as those for an individual gas, except that an additional variable, q, is needed. To facilitate the computation of $k(g, p, T, q)$, efficient parameterization can be developed for a number of pressure, temperature, and H_2O mixing ratios.

4.4   Band Models

Band models are traditional approaches that simplify the computation of the spectral transmittance. The atmosphere is assumed to be homogeneous so that analytical expressions may be developed. We first define spectral absorptance in the form

$$A_{\bar{\nu}}(u) = 1 - T_{\bar{\nu}} = \frac{1}{\Delta \nu} \int_{\Delta \nu} (1 - e^{-k_\nu u}) \, d\nu. \tag{4.4.1}$$

The quantity $A_{\bar{\nu}} \Delta \nu$ is referred to as the equivalent width $W(u)$. It is the width of an infinitely strong line of rectangular shape, which would be the same as the actual absorption of a single line (Fig. 4.9). The concept of equivalent width plays an important role in the development of band models.

4.4.1   A Single Line

Using a single Lorentz line, introducing two new variables, $x = Su/2\pi\alpha$ and $\tan y/2 = (\nu - \nu_0)/\alpha$, and extending the wavenumber integration from $-\infty$ to ∞, the spectral absorptance can be shown to be

$$A_{\bar{\nu}} = \frac{\alpha}{\Delta \nu} \int_{-\pi}^{\pi} \{1 - \exp[-x(1 + \cos y)]\} \, d(\tan y/2). \tag{4.4.2}$$
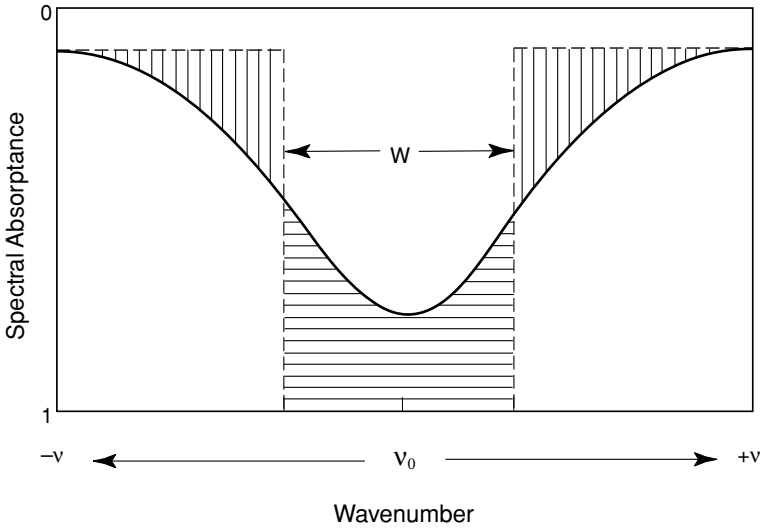
Figure 4.9   The definition of equivalent width, $W = A_{\bar{v}}\Delta\bar{v}$, where $A_{\bar{v}}$ represents the spectral absorptance for a spectral interval Δv. It is the width of an infinitely strong line of rectangular shape that is the same as the absorption of a single line.

Performing integration by parts and carrying out further trigonometric manipulations yield

$$A_{\bar{v}} = \frac{\alpha x e^{-x}}{\Delta v} \int_{-\pi}^{\pi} (e^{-x\cos y} - \cos y e^{-x\cos y})\,dy. \tag{4.4.3}$$

The integral representation of the Bessel function is given by

$$J_n(x) = \frac{i^{-n}}{\pi} \int_0^{\pi} e^{ix\cos\theta}\cos n\theta\,d\theta, \tag{4.4.4a}$$

and the modified Bessel function of the first kind of order n is

$$I_n(x) = i^{-n} J_n(ix), \tag{4.4.4b}$$

where $i = \sqrt{-1}$. Thus, in terms of the modified Bessel function, equivalent width is given by

$$W = A_{\bar{v}}\Delta v = 2\pi\alpha L(x) = 2\pi\alpha x e^{-x}[I_0(x) + I_1(x)], \tag{4.4.5}$$

where $L(x)$ is known as the Ladenburg and Reiche function.

In the case of weak-line absorption, either k_v or u is small, so that $k_v u \ll 1$. Spectral absorptance is approximately given by

$$A_{\bar{v}} \cong \frac{1}{\Delta v} \int_{-\infty}^{\infty} k_v u\,dv = \frac{Su}{\Delta v}, \tag{4.4.6}$$

based on the definition of line intensity, regardless of the line shape. Absorptance,

under the limits of weak-line approximations, is directly proportional to the path length and is called the region of *linear absorption*. On the other hand, if $Su/\pi\alpha \gg 1$, absorptance approaches 1 in the line center region. We may omit the half-width α in the denominator of the Lorentz profile so that

$$A_{\bar{\nu}}(u) = \frac{1}{\Delta\nu} \int_{-\infty}^{\infty} \left[1 - \exp\left(\frac{-S\alpha u}{\pi(\nu - \nu_0)^2} \right) \right] d\nu. \tag{4.4.7a}$$

Using the known integration formula (Exercise 3.6), we find

$$A_{\bar{\nu}}(u) = 2\sqrt{S\alpha u}/\Delta\nu. \tag{4.4.7b}$$

Absorptance is, therefore, proportional to the square root of the path length and is in the region of the so-called *square root absorption*. Approximations for the weak- and strong-line limits can also be derived directly from Eq. (4.4.5) (Exercise 4.2). Summarizing the preceding discussion, the equivalent width for a spectral line can be written as follows:

$$W = A_{\bar{\nu}}(u)\Delta\nu = \begin{cases} Su, & \text{weak-line,} \\ 2\sqrt{S\alpha u}, & \text{strong-line.} \end{cases} \tag{4.4.8}$$

These two limits are useful in the development of approximations for use in infrared radiative transfer calculations.

4.4.2 Regular Band Model

Inspection of the Q-branch of the 15 μm CO_2 band indicates that a single line may repeat itself periodically (or regularly) as shown in Fig. 4.6b. This configuration led Elsasser (1938) to the development of the regular band model. In this case, the absorption coefficient at a wavenumber displacement ν from the center of one particular line is then

$$k_\nu = \sum_{i=-\infty}^{\infty} \frac{S\alpha/\pi}{(\nu - i\delta)^2 + \alpha^2}, \tag{4.4.9}$$

where δ is the line spacing. From the Mittag–Leffler theorem (Whittaker and Watson, 1940), we can prove that this infinite sum can be expressed in terms of periodic and hyperbolic functions as follows:

$$k_\nu = \frac{S}{\delta} \frac{\sinh\beta}{\cosh\beta - \cos\gamma}, \tag{4.4.10}$$

where $\beta = 2\pi\alpha/\delta$ and $\gamma = 2\pi\nu/\delta$. By transforming the variables, the spectral transmittance may be expressed by

$$T_{\bar{\nu}}(u) = \frac{1}{\delta} \int_{-\delta/2}^{\delta/2} e^{-k_\nu u} d\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-k_\nu(\gamma)u} d\gamma. \tag{4.4.11a}$$

The differential value is

$$\frac{dT_{\bar{v}}(u)}{du} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-k_v u} k_v d\gamma. \tag{4.4.11b}$$

To derive an analytical expression, we define

$$\cos \rho = \frac{1 - \cosh \beta \cos \gamma}{\cosh \beta - \cos \gamma}. \tag{4.4.12a}$$

It follows from Eq. (4.4.10) that

$$d\rho = -\frac{\sinh \beta}{\cosh \beta - \cos \gamma} d\gamma = -k_v \frac{\delta}{S} d\gamma. \tag{4.4.12b}$$

Substituting Eqs. (4.4.10) and (4.4.12b) into Eq. (4.4.11b) leads to

$$\frac{dT_{\bar{v}}}{du} = -\frac{S}{2\pi\delta} \int_{-\pi}^{\pi} \exp\left(-\frac{Su}{\delta} \frac{\cosh \beta - \cos \rho}{\sinh \beta}\right) d\rho. \tag{4.4.13a}$$

We then define a new variable $y = Su/(\delta \sinh \beta)$ to obtain

$$\frac{dT_{\bar{v}}}{dy} = -\frac{\sinh \beta}{2\pi} \int_{-\pi}^{\pi} \exp(-y \cosh \beta - y \cos \rho) d\rho \tag{4.4.13b}$$

$$= -\sinh \beta e^{-y \cosh \beta} J_0(iy).$$

Since $T_{\bar{v}} = 0$ when u (or y) $\to \infty$, we have

$$T_{\bar{v}} = \int_0^{T_{\bar{v}}} dT_{\bar{v}} = \sinh \beta \int_y^{\infty} e^{-y \cosh \beta} J_0(iy) \, dy$$

$$= \int_z^{\infty} e^{-z \coth \beta} J_0(iz/\sinh \beta) \, dz, \tag{4.4.14}$$

where $z = y \sinh \beta$. This is the Elsasser transmittance that can be evaluated numerically. Further approximations and simplifications also can be made to this model. Since $\alpha \ll \delta$ and $\beta \to 0$, we find

$$\coth \beta \approx \frac{1}{\beta} + \frac{\beta}{3}, \quad \operatorname{csch} \beta \approx \frac{1}{\beta} - \frac{\beta}{6}, \tag{4.4.15a}$$

and

$$J_0(iz \operatorname{csch} \beta) \approx e^{z \operatorname{csch} \beta} / \sqrt{2\pi z \operatorname{csch} \beta} \tag{4.4.15b}$$

$$\approx \exp\left[z\left(\frac{1}{\beta} - \frac{\beta}{6}\right)\right] / \sqrt{2\pi z/\beta}.$$

With these approximations, Eq. (4.4.14) becomes

$$T_{\bar{v}} = \frac{1}{\sqrt{2\pi}} \int_z^{\infty} \sqrt{\beta/z} \, e^{-z\beta/2} dz. \tag{4.4.16}$$

Finally, we set $x^2 = z\beta/2$; the absorptance can now be expressed by

$$A_{\bar{\nu}} = 1 - \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-x^2} dx. \tag{4.4.17a}$$

By noting that $(2/\sqrt{\pi}) \int_0^\infty e^{-x^2} dx = 1$, we have

$$A_{\bar{\nu}} = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx = \mathrm{erf}(x) = \mathrm{erf}\left(\frac{\sqrt{\pi} S\alpha u}{\delta}\right). \tag{4.4.17b}$$

Values of $\mathrm{erf}(x)$ can be obtained from standard mathematical tables. This is referred to as the Elsasser model. For small values of x, $A_{\bar{\nu}} = 2x/\sqrt{\pi} = 2\sqrt{S\alpha u}/\delta$. This is the region of square root absorption denoted in Eq. (4.4.7b).

4.4.3 Statistical Band Model

On inspection of the water vapor rotational band, the only common feature over a 25 cm^{-1} range is the apparent random line positions (Goody, 1952, 1964). Hence, the absorption of a band with certain random properties should be considered. Let $\Delta\nu$ be a spectral interval consisting of n lines of mean distance δ, so that $\Delta\nu = n\delta$. Let $p(S_i)$ be the probability that the ith line has an intensity S_i, and let p be normalized such that

$$\int_0^\infty p(S_i)\, dS_i = 1, \quad i = 1, \dots, n. \tag{4.4.18}$$

We may assume that any line has an equal probability of being anywhere in the interval $\Delta\nu$. The mean transmittance is found by averaging the transmittance over all line positions and intensities. Hence,

$$\begin{aligned}
T_{\bar{\nu}} &= \frac{1}{(\Delta\nu)^n} \int_{\Delta\nu} d\nu_1 \cdots \int_{\Delta\nu} d\nu_n \\
&\quad \times \int_0^\infty p(S_1) e^{-k_1 u}\, dS_1 \cdots \int_0^\infty p(S_n) e^{-k_n u}\, dS_n,
\end{aligned} \tag{4.4.19a}$$

where k_n denotes the absorption coefficient for the nth line. Since all the integrals are alike, we have

$$\begin{aligned}
T_{\bar{\nu}} &= \left[\frac{1}{\Delta\nu} \int d\nu \int_0^\infty p(S) e^{-ku}\, dS \right]^n \\
&= \left[1 - \frac{1}{\Delta\nu} \int d\nu \int_0^\infty p(S)(1 - e^{-ku})\, dS \right]^n.
\end{aligned} \tag{4.4.19b}$$

The average equivalent width for n absorption lines may be defined by

$$\bar{W} = \int_0^\infty p(S) \int_{\Delta\nu} (1 - e^{-ku})\, d\nu\, dS. \tag{4.4.20}$$

Noting that $\Delta \nu = n\delta$, the spectral transmittance may then be written in terms of the averaged equivalent width in the form

$$T_{\bar{\nu}}(u) = \left[1 - \frac{1}{n} \left(\frac{\bar{W}}{\delta} \right) \right]^n . \tag{4.4.21a}$$

Since $\lim_{n \to \infty} (1 - x/n)^n \to e^{-x}$, we have

$$T_{\bar{\nu}}(u) = e^{-\bar{W}/\delta} . \tag{4.4.21b}$$

Let the lines be of different intensities, and consider a Poisson distribution for the probability of their intensities in the form

$$p(S) = \frac{1}{\bar{S}} e^{-S/\bar{S}}, \qquad \bar{S} = \int_0^\infty S p(S) \, dS, \tag{4.4.22}$$

where $\bar{S}$ is defined as the mean line intensity. We note that $p(S)$ is normalized to 1, as is required. Inserting this probability function into Eq. (4.4.20) for the average equivalent width, and writing $k_\nu = S f_\nu$ with f_ν the line-shape factor, we find

$$\bar{W} = \int_{\Delta\nu} \frac{\bar{S} f_\nu u}{1 + \bar{S} f_\nu u} \, d\nu . \tag{4.4.23}$$

Using the Lorentz line shape for f_ν and performing the wavenumber integration in the domain $(-\infty, \infty)$, without introducing significant errors for the integral confined in the $\Delta \nu$ interval, the spectral transmittance for randomly distributed Lorentz lines is given by

$$T_{\bar{\nu}}(u) = \exp\left[-\frac{\bar{S}}{\delta} u \left(1 + \frac{\bar{S}}{\alpha\pi} u \right)^{-1/2} \right] . \tag{4.4.24}$$

Thus the spectral transmittance for the random model can be expressed as a function of two parameters, $\bar{S}/\delta$ and $\bar{S}/\alpha\pi$, apart from the path length u. This is referred to as the *Goody random model*. For a given spectral interval $\Delta \nu$, these two parameters may be derived by fitting the random model with line-by-line data. For the computation of fluxes in the atmosphere, the device of the spectral interval must ensure that the Planck flux does not vary significantly.

On the basis of the preceding discussion, the average equivalent width for the random model is

$$\bar{W} = \bar{S} u \left(1 + \frac{\bar{S} u}{\alpha\pi} \right)^{-1/2} . \tag{4.4.25a}$$

The average equivalent width for n individual lines with equivalent widths of W_i is simply

$$\bar{W} = \sum_{j=1}^{n} \frac{W_j}{n} . \tag{4.4.25b}$$

The spectral transmittance given in Eq. (4.4.24) is general and should be valid under the limits of weak- and strong-line approximations. Thus, for weak-line approximations where $\bar{S}u/\alpha\pi \ll 1$, we have

$$\frac{\bar{W}}{\delta} \approx \frac{\bar{S}}{\delta}u. \tag{4.4.26}$$

However, by using the equivalent width for a spectral line denoted by the subscript j under the limits of weak-line approximations, we have

$$\bar{W} = \frac{1}{n}\sum_{j=1}^{n} W_j(\text{weak}) \approx \frac{1}{n}\sum_{j=1}^{n} S_j u. \tag{4.4.27}$$

From Eqs. (4.4.26) and (4.4.27), we obtain

$$\frac{\bar{S}}{\delta} = \sum_{j=1}^{n} \frac{S_j}{\Delta\nu} = a_{\bar{\nu}}, \tag{4.4.28}$$

where $a_{\bar{\nu}}$ is so defined. Following the same procedure under the limits of strong-line approximations, we find

$$\frac{\sqrt{\pi\alpha\bar{S}}}{\delta} = 2\sum_{j=1}^{n} \frac{\sqrt{S_j\alpha_j}}{\Delta\nu} = b_{\bar{\nu}}^*. \tag{4.4.29}$$

If we define $b_{\bar{\nu}} = \pi\alpha/\delta$, then $b_{\bar{\nu}}^* = \sqrt{a_{\bar{\nu}}b_{\bar{\nu}}}$. It follows that

$$b_{\bar{\nu}} = \left(2\sum_{j=1}^{n}\sqrt{S_j\alpha_j}\right)^2 \bigg/ \left(\Delta\nu\sum_{j=1}^{n} S_j\right). \tag{4.4.30}$$

Thus, the spectral transmittance may be written in the form

$$T_{\bar{\nu}}(u) = \exp[-a_{\bar{\nu}}u(1 + ua_{\bar{\nu}}/b_{\bar{\nu}})^{-1/2}]. \tag{4.4.31}$$

The band parameters $a_{\bar{\nu}}$ and $b_{\bar{\nu}}$ are functions of two variables: $\sum_j S_j$ and $\sum_j \sqrt{S_j\alpha_j}$. These variables can be computed from line-by-line data.

The spectral transmittance for the random model depends on two parameters, $a_{\bar{\nu}}$ and $a_{\bar{\nu}}/b_{\bar{\nu}}$. Examination of these parameters reveals that a further simplification may be obtained by defining a parameter referred to as the *generalized absorption coefficient*, $\ell_{\bar{\nu}}$, such that $a_{\bar{\nu}} = a\ell_{\bar{\nu}}$ and $a_{\bar{\nu}}/b_{\bar{\nu}} = b\ell_{\bar{\nu}}$, where the coefficients a and b are certain constants. Thus, a one-parameter representation of the spectral transmittance for the random model may be expressed by

$$T_{\bar{\nu}}(u) = \exp[-a\ell_{\bar{\nu}}u(1 + b\ell_{\bar{\nu}}u)^{-1/2}]. \tag{4.4.32}$$

The generalized absorption coefficient was originally developed by Elsasser (1942). Under the limits of strong- and weak-line approximations, we have

$$T_{\bar{\nu}}(u) \cong \begin{cases} \exp(-a\ell_{\bar{\nu}}u), & \text{weak-line}, \\ \exp(-c\ell_{\bar{\nu}}^*u^*), & \text{strong-line}, \end{cases} \tag{4.4.33}$$

where $c = a/\sqrt{b}$, $\ell_{\bar{v}}^* = \sqrt{\ell_{\bar{v}}}$, and $u^* = \sqrt{u}$. The transmittance is now expressed by a simple exponential function in terms of the generalized absorption coefficient, which can be obtained from the band parameters using statistical fitting procedures.

The random model just described utilizes the probability distribution function given in Eq. (4.4.22). In many cases, it has been found that this exponential intensity distribution substantially underestimates the number of low-intensity lines (Malkmus, 1967). The line intensity is governed by the Boltzmann factor in the form $S \sim \exp(-hcE/KT)$, where E is the lower energy level. Thus, dE/dS is proportional to S^{-1}. The probability distribution function $p(S)$ must be proportional to dn/dS, where n is the number density of energy levels. However, dn/dE is approximately constant since, in many cases, the energy levels are approximately equally spaced. It follows that $p(S) \sim (dE/dS)(dn/dE) \sim dE/dS \sim S^{-1}$. Thus, the S^{-1} dependence remains a dominating influence on the probability distribution function. For this reason, it is necessary to use a normalized probability distribution function for line intensity in the form

$$p(S) \sim \frac{1}{S} e^{-s/\bar{s}}. \tag{4.4.34a}$$

Using the normalized form of the probability distribution function, we can derive, after lengthy mathematical manipulations, the following spectral transmittance:

$$T_{\bar{v}}(u) = \exp\left\{-c_{\bar{v}}\left[(1 + d_{\bar{v}}u)^{1/2} - 1\right]\right\}, \tag{4.4.34b}$$

where the two parameters are defined by the line data as follows:

$$c_{\bar{v}} = 2\left(\sum_j \sqrt{\alpha_j S_j}\right)^2 \bigg/ \left(\Delta v \sum_j S_j\right),$$

$$d_{\bar{v}} = \left(\sum_j S_j\right)^2 \bigg/ \left(\sum_j \sqrt{\alpha_j S_j}\right)^2. \tag{4.4.34c}$$

This is referred to as the *Malkmus random model*. In terms of line parameters, we can prove that $c_{\bar{v}} = \pi\alpha/2\delta$, and $d_{\bar{v}} = 4\bar{S}/\pi\alpha$ (Exercise 4.8).

4.4.4   Application to Nonhomogeneous Atmospheres

In band models, it is assumed that the absorption coefficient is independent of temperature and pressure. To incorporate these effects in the calculation of spectral transmittance, we should first define the optical depth by

$$\tau = \int_u k_v(p, T)\, du. \tag{4.4.35}$$

The primary objective of approximate solutions for nonhomogeneous path lengths is to transform the transfer problem to that of a homogeneous path defined by a reference pressure, p_r, and a reference temperature, T_r, so that analytical exponential

functions can be employed for spectral transmittance calculations. In the limit of strong-line approximations denoted in Eq. (4.3.9), the wavenumber-dependent term can be factored out. Moreover, the half-width can be written as

$$\alpha(p, T) = \alpha(p_r, T_r) \left(\frac{p}{p_r} \right) \left(\frac{T_r}{T} \right)^n . \tag{4.4.36}$$

Thus, to a good approximation, we can prove that

$$k_\nu(p, T) \cong k_\nu(p_r, T_r) \left(\frac{p}{p_r} \right) \left(\frac{T_r}{T} \right)^n , \tag{4.4.37a}$$

where we have set

$$\sum_j \frac{S_j(T)\alpha_j(p_r, T_r)}{(\nu - \nu_{0j})^2} \Big/ \sum_j \frac{S_j(T_r)\alpha_j(p_r, T_r)}{(\nu - \nu_{0j})^2} \approx 1. \tag{4.4.37b}$$

The case in which the wavenumber, pressure, and temperature are decoupled is referred to as the *one-parameter scaling approximation*. It is particularly useful for application to the rotational band of water vapor (Chou and Arking, 1980).

From Eq. (4.4.37a), we may define a scaled path length such that

$$\tau = k_\nu(p_r, T_r)\tilde{u}, \tag{4.4.38a}$$

where

$$\tilde{u} = \int_u \left(\frac{p}{p_r} \right) \left(\frac{T_r}{T} \right)^n du. \tag{4.4.38b}$$

It follows that by replacing u with $\tilde{u}$, the spectral transmittance may be evaluated by using the absorption coefficient at a reference temperature and pressure. The scaling approximation for flux and cooling-rate calculations using $n = 1/2$ was originally proposed by Elsasser and Culbertson (1960).

Further, we may also search for an adjusted absorption coefficient such that

$$\tau = \int_u k_\nu(p, T) \, du = k_\nu(\tilde{p}, \tilde{T})u. \tag{4.4.39a}$$

Using the Lorentz line profile, we have

$$k_\nu(\tilde{p}, \tilde{T}) = \sum_j \tilde{S}_j \tilde{f}_{\nu j} = \sum_j \frac{\tilde{S}_j}{\pi} \frac{\tilde{\alpha}_j}{(\nu - \nu_{0j})^2 + \tilde{\alpha}_j^2}. \tag{4.4.39b}$$

Two adjusted parameters, $\tilde{S}$ and $\tilde{\alpha}$, are required to satisfy Eq. (4.4.39a). This is referred to as the *two-parameter approximation*.

In reference to Eq. (4.4.1), the average equivalent width of a nonhomogeneous path is

$$\bar{W} = A_{\bar{\nu}}(u)\Delta\nu = \int_{\Delta\nu} \left[1 - \exp\left(-\int_u \sum_j S_j f_{\nu j} \, du \right) \right] d\nu. \tag{4.4.40a}$$

In the weak-line limit, we have

$$\bar{W} \cong \int_{\Delta v} \int_u \sum_j S_j f_{vj} \, du \, dv = \int_u \sum_j S_j \, du, \qquad (4.4.40b)$$

regardless of the line shape. Moreover, using the Lorentz profile, the average equivalent width given in Eq. (4.4.40a) may be written in the form

$$\bar{W} = \int_{\Delta v} \left[1 - \exp\left(-\int_u \sum_j \frac{S_j}{\pi} \frac{\alpha_j}{(v - v_{0j})^2 + \alpha_j^2} \, du \right) \right] dv. \quad (4.4.40c)$$

In the strong-line limit, absorption near the line-center region is nearly saturated. The average equivalent width in this case is not very sensitive to the absorption coefficient. Hence, we may select any half-width, say α', to replace α in the denominator of Eq. (4.4.40c). To find the α' that also satisfies the weak-line limit, we must have

$$\bar{W} \cong \int_{\Delta v} \int_u \sum_j \frac{S_j}{\pi} \frac{\alpha_j}{(v - v_{0j})^2 + \tilde{\alpha}_j^2} \, du \, dv \cong \int_u \sum_j \frac{\alpha_j}{\tilde{\alpha}_j} S_j \, du, \quad (4.4.40d)$$

where Δv is set from $-\infty$ to ∞ for mathematical convenience, and $\tilde{\alpha}$ is the specific α' that fulfills the requirements of both the strong- and weak-line limits. Since $\alpha_j / \tilde{\alpha}_j \cong p/\tilde{p}$, combining Eqs. (4.4.40d) and (4.4.40b) gives

$$\tilde{p} = \int_u p \bar{S} \, du \Big/ \int_u \bar{S} \, du, \qquad (4.4.41)$$

where $\bar{S}$ is the mean line intensity. The effect of pressure on absorption is evident through the line half-width. The preceding scaling serves to adjust the pressure variation along the path.

Furthermore, from the equivalent width given in Eqs. (4.4.40a) and (4.4.40b), we may select an adjusted $\tilde{S}_j$ such that

$$\int_u \sum_j S_j \, du = u \sum_j \tilde{S}_j. \qquad (4.4.42)$$

This defines the scaled mean line intensity, regardless of the line shape, in the form

$$\tilde{\bar{S}} = \int \bar{S} \, du/u. \qquad (4.4.43a)$$

The effect of temperature on absorption is exerted through its dependence on the line strength. This part of scaling is to account for the nonisothermal path. Alternatively, we may select a reference $\bar{S}$, say $\bar{S}_r$, and adjust the path length u such that

$$\tilde{u} = \int_u \frac{\bar{S}}{\bar{S}_r} \, du. \qquad (4.4.43b)$$

Equations (4.4.41) and (4.4.43) constitute the so-called *Curtis–Godson* (CG) *approximation* for nonhomogeneous atmospheres, independently proposed by Curtis (1952)

and Godson (1953). Its validity has been checked by Walshaw and Rodgers (1963), who performed extensive cooling-rate calculations using a line-by-line integration program. Errors introduced by the CG approximation are less than a few percent for the H_2O rotational and 15 μm CO_2 bands. The CG approximation, however, is less satisfactory for the 9.6 μm O_3 band.

van de Hulst (1945) innovated a more general technique for application to a non-homogeneous atmosphere and demonstrated that the Voigt profile can be readily incorporated in this technique. Consider the cosine transformation of the optical depth in the form

$$\tilde{g}(t) = \int_{-\infty}^{\infty} \tau(\nu) \cos \nu t \, d\nu = \int_{u} g(t) \, du, \tag{4.4.44a}$$

where

$$\tau(\nu) = \frac{1}{\pi} \int_{0}^{\infty} \tilde{g}(t) \cos \nu t \, dt, \tag{4.4.44b}$$

$$g(t) = \int_{-\infty}^{\infty} k_\nu \cos \nu t \, d\nu. \tag{4.4.44c}$$

On substituting the expressions of the absorption coefficient for Lorentz, Doppler, and Voigt line shapes, we find

$$g(t) = \begin{cases} S \exp(-\alpha t), & \text{Lorentz} \\ S \exp(-\alpha_D^2 t^2/4), & \text{Doppler} \\ S \exp[-(\alpha t + \alpha_D^2 t^2/4)], & \text{Voigt.} \end{cases} \tag{4.4.45}$$

By using the Voigt profile and defining the scaled path length $\tilde{u}$ and the scaled line widths, $\tilde{\alpha}$ and $\tilde{\alpha}_D$, we obtain

$$\tilde{g}(t) = \int_{u} S \exp\left[-\left(\alpha t + \frac{\alpha_D^2 t^2}{4}\right)\right] du \equiv S_r \tilde{u} \exp\left[-\left(\frac{\tilde{\alpha} t + \tilde{\alpha}_D^2 t^2}{4}\right)\right]. \tag{4.4.46}$$

If we take the first two terms in the expansion and match t^n in the expansion for $n = 0$ and 1, we have

$$\tilde{u} = \int_{u} \left(\frac{S}{S_r}\right) du, \tag{4.4.47a}$$

$$\tilde{\alpha} = \frac{1}{\tilde{u}} \int_{u} \alpha \left(\frac{S}{S_r}\right) du. \tag{4.4.47b}$$

The Doppler width, α_D, is a function of temperature only. Thus, if the temperature does not vary significantly along the nonhomogeneous path, we may set $\tilde{\alpha}_D = \alpha_D$. Equations (4.4.47a) and (4.4.47b) are equivalent to the Curtis–Godson relations.

4.5   Broadband Approaches to Flux Computations

4.5.1   Broadband Emissivity

The essence of the broadband emissivity approach for the calculation of infrared fluxes and heating rates is to use temperature directly in terms of the Stefan–Boltzmann law instead of the Planck function. The early development of broadband emissivity methods was based on the concept of the radiation chart from which the flux integration may be carried out by using graphs and tables (Elsasser, 1942; Möller, 1943; Yamamoto, 1952).

Based on Eq. (4.2.10), the total upward and downward fluxes in the u coordinate may be written

$$F^\uparrow(u) = \int_0^\infty \pi B_\nu(T_s) T_\nu^f(u) d\nu + \int_0^\infty \int_0^u \pi B_\nu(u') \frac{dT_\nu^f(u - u')}{du'} du' \, d\nu, \quad (4.5.1a)$$

$$F^\downarrow(u) = \int_0^\infty \int_{u_1}^u \pi B_\nu(u') \frac{dT_\nu^f(u' - u)}{du'} du' \, d\nu. \quad (4.5.1b)$$

In order to express the upward and downward fluxes in terms of broadband emissivity, we may define isothermal broadband flux emissivity in the form

$$\epsilon^f(u, T) = \int_0^\infty \pi B_\nu(T) \left[1 - T_\nu^f(u)\right] \frac{d\nu}{\sigma T^4}. \quad (4.5.2)$$

Consider a plane-parallel atmosphere divided into a number of layers such that each of them may be thought of as an isothermal layer. We may then write

$$F^\uparrow(u) \cong \sigma T_s^4 [1 - \epsilon^f(u, T_s)] - \int_0^u \sigma T^4(u') \frac{d\epsilon^f(u - u', T(u'))}{du'} du', \quad (4.5.3a)$$

$$F^\downarrow(u) \cong \int_u^{u_1} \sigma T^4(u') \frac{d\epsilon^f(u' - u, T(u'))}{du'} du'. \quad (4.5.3b)$$

Computations of fluxes and cooling rates that use the broadband emissivity definition in Eqs. (4.5.3a,b) have been performed by numerous researchers.

In numerical calculations, Eq. (4.5.2) for broadband flux emissivity is expressed in terms of a finite sum over spectral bands, $\Delta\nu_i$, for three principal absorbers: CO_2, H_2O and O_3. Let $u_1(= u_w)$, $u_2(= u_c)$, and $u_3(= u_0)$ denote the path lengths for H_2O, CO_2 and O_3, respectively. Then we may write

$$\epsilon^f(u_j, T) = \sum_i \pi B_{\bar\nu, i}(T) \left[1 - T_{\bar\nu, i}^f(u_j)\right] \frac{\Delta\nu_i}{\sigma T^4}, \quad j = 1, 2, 3. \quad (4.5.4)$$

As a good approximation, we may replace the diffuse spectral transmittance by the spectral transmittance, $T_{\bar\nu}(u/\bar\mu)$, using the diffusivity factor $1/\bar\mu = 1.66$. The spectral transmittance can be generated from data via detailed line-by-line computations or band models for a homogeneous path. Figure 4.10 shows the broadband flux emissivities for water vapor and carbon dioxide as functions of the path length for a number
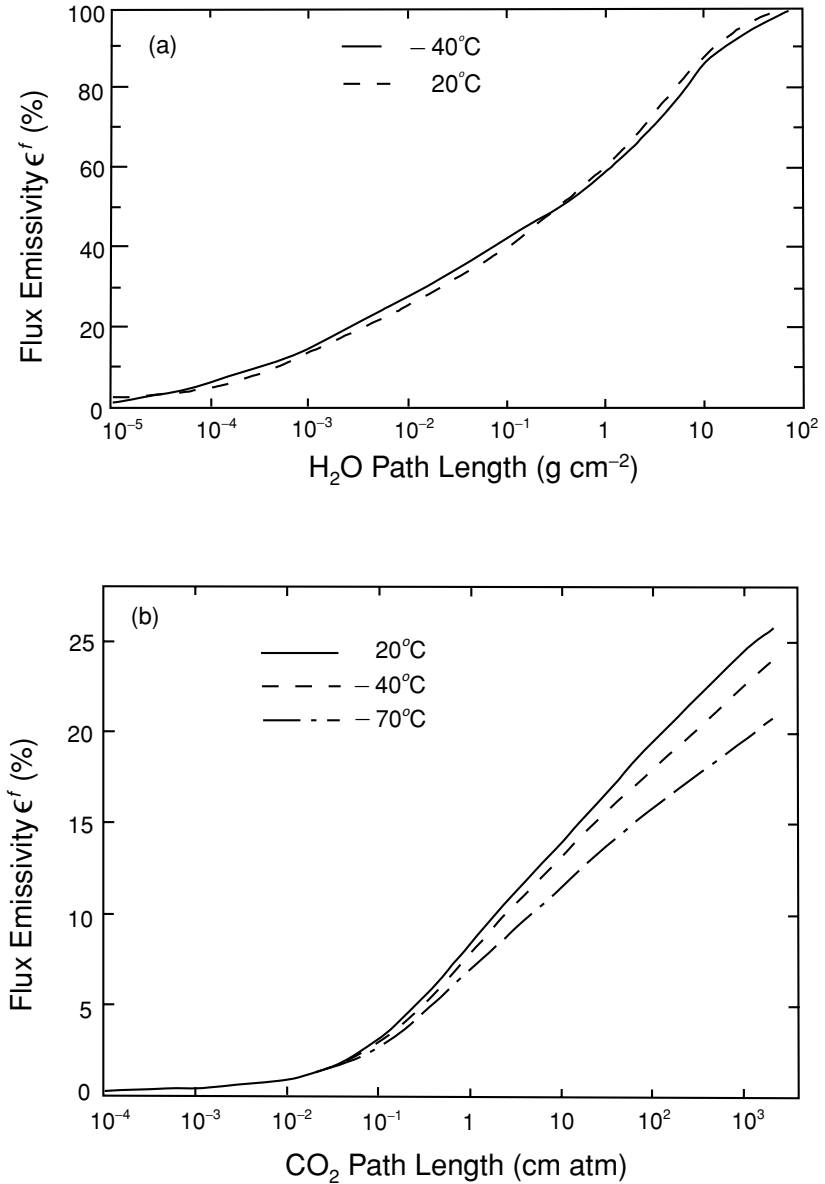
Figure 4.10   Broadband flux emissivity for (a) water vapor and (b) carbon dioxide covering the entire thermal infrared spectrum as a function of path length and temperature. Note that 1 g cm^{-2} = $2.24 \times 10^4/M$ cm atm, where M is the molecular weight of an individual gas.

of temperatures. For water vapor, the effect of temperature on emissivity is relatively small. For carbon dioxide, however, the temperature dependence of larger path lengths appears quite pronounced.

The total emission of an isothermal atmosphere is the sum of individual emissions due to various gases. However, the overlap of the H_2O rotational and 15 μm CO_2 absorption lines is significant. Thus, it is necessary to make a proper correction to circumvent the overestimation of H_2O and CO_2 emissions. The emissivity for the overlap region can be expressed exactly by

$$\epsilon^f(u_w, u_c, T) = \int_0^\infty \pi B_\nu(T)[1 - T_\nu(\bar{u}_w, \bar{u}_c)]\frac{d\nu}{\sigma T^4}, \qquad (4.5.5)$$

where $\bar{u}_w = u_w/\bar{\mu}$ and $\bar{u}_c = u_c/\bar{\mu}$. From the definition of monochromatic transmittance, its value for two absorbing gases is the product of the individual value for each absorbing gas. It follows that

$$T_\nu(\bar{u}_w, \bar{u}_c) = T_\nu(\bar{u}_w)T_\nu(\bar{u}_c). \qquad (4.5.6)$$

With this relation, which is valid only for monochromatic radiation, we may express the emissivity of the overlap region in terms of the individual emissivities in the form

$$\epsilon^f(u_w, u_c, T) = \epsilon^f(u_w, T) + \epsilon^f(u_c, T) - \Delta\epsilon^f(u_w, u_c, T), \qquad (4.5.7)$$

where the correction term is

$$\Delta\epsilon^f(u_w, u_c, T) = \int_0^\infty \pi B_\nu(T)[1 - T_\nu(\bar{u}_w)][1 - T_\nu(\bar{u}_c)]\frac{d\nu}{\sigma T^4}$$

$$\cong \sum_i \pi B_{\bar{\nu},i}(T) \int_{\Delta\nu_i} [1 - T_\nu(\bar{u}_w)][1 - T_\nu(\bar{u}_c)]\frac{d\nu}{\sigma T^4}. \qquad (4.5.8)$$

If the variation in either $T_\nu(\bar{u}_w)$ or $T_\nu(\bar{u}_c)$ is smaller than their products, we may carry out a wavenumber integration over either one to obtain

$$\Delta\epsilon^f(u_w, u_c, T) \cong \sum_i \pi B_{\bar{\nu},i}(T)[1 - T_{\bar{\nu},i}(\bar{u}_w)][1 - T_{\bar{\nu},i}(\bar{u}_c)]\frac{\Delta\nu_i}{\sigma T^4}. \quad (4.5.9)$$

This appears to be a good approximation and has been used to evaluate the H_2O–CO_2 overlap in the context of the broadband emissivity for flux computations. The computation of broadband flux emissivities for H_2O, CO_2, and O_3 has been correctly carried out by Staley and Jurica (1970, 1972) using the generalized absorption coefficients presented by Elsasser and Culbertson (1960). A review of various methods that can be used to construct the broadband emissivity for flux and heating rate calculations has been presented by Chou *et al.* (1991).

4.5.2   Newtonian Cooling Approximation

The general circulation of the middle atmosphere from ∼15 to 90 km is driven by differential vertical and horizontal radiative heating. Absorption of solar insolation in

this region is primarily generated by O_3 and O_2. This absorption is largely balanced by radiative cooling throughout much of the stratosphere and parts of the mesosphere produced principally by emission and absorption of thermal IR radiation due to CO_2 and O_3. For middle atmosphere applications, the shape and overlap of absorption lines, the variation in line intensity with temperature, and departures from LTE at high altitudes must be taken into account in the computations.

Many attempts have been made to develop accurate yet efficient methods to compute IR cooling rates in the middle atmosphere in connection with dynamic models. The most important aspect of radiative cooling from the viewpoint of dynamic circulation is its variation with changes in temperature. In the following, we introduce a method that is based on the cooling-to-space approximation widely used by modelers.

Consider the upward and downward fluxes given in Eqs. (4.2.10a) and (4.2.10b) for a spectral interval in which the variation in Planck fluxes can be neglected. In the height coordinate, we may write

$$F_{\bar{\nu}}^{\uparrow}(z) = \pi B_{\bar{\nu}}(0) T_{\bar{\nu}}^{f}(z) + \int_{0}^{z} \pi B_{\bar{\nu}}(z') \frac{d}{dz'} T_{\bar{\nu}}^{f}(z - z') \, dz', \qquad (4.5.10a)$$

$$F_{\bar{\nu}}^{\downarrow}(z) = \int_{z_{\infty}}^{z} \pi B_{\bar{\nu}}(z') \frac{d}{dz'} T_{\bar{\nu}}^{f}(z' - z) \, dz', \qquad (4.5.10b)$$

where z_{∞} denotes the height at TOA and $z = 0$ denotes the surface. The net flux at a given level is then

$$F_{\bar{\nu}}(z) = F_{\bar{\nu}}^{\uparrow}(z) - F_{\bar{\nu}}^{\downarrow}(z) = \pi B_{\bar{\nu}}(0) T_{\bar{\nu}}^{f}(z)$$
$$+ \int_{0}^{z_{\infty}} \pi B_{\bar{\nu}}(z') \frac{d}{dz'} T_{\bar{\nu}}^{f}(|z - z'|) \, dz'. \qquad (4.5.11)$$

The cooling rate for a spectral interval is then given by

$$\left(\frac{\partial T}{\partial t} \right)_{\bar{\nu}} = -\frac{1}{\rho C_p} \frac{d F_{\bar{\nu}}(z)}{dz} = -\frac{1}{\rho C_p}$$
$$\left(\pi B_{\bar{\nu}}(0) \frac{d}{dz} T_{\bar{\nu}}^{f}(z) + \int_{0}^{z_{\infty}} \pi B_{\bar{\nu}}(z') \frac{d}{dz'} \frac{d}{dz} T_{\bar{\nu}}^{f}(|z - z'|) \, dz' \right). \qquad (4.5.12)$$

Consider an atmosphere with an isothermal temperature profile such that local cooling rates are produced solely from the emission of a local layer. Under this condition we have

$$\left(\frac{\partial T}{\partial t} \right)_{\text{space}} = -\frac{1}{\rho C_p} \pi B_{\bar{\nu}}(z) \frac{d}{dz} T_{\bar{\nu}}^{f}(z_{\infty} - z). \qquad (4.5.13)$$

This is referred to as the *cooling-to-space approximation* in which the cooling rate is dependent on the local temperature but is independent of the temperatures of other levels. Although it is imperfect, this approximation gives reliable results under a number of conditions (Rodgers and Walshaw, 1966). The cooling rate may be expressed in terms of a cooling-to-space term that depends only on the temperature

at that level and a term representing the exchange of radiation between that level and all other levels. This latter term is a function of the entire temperature profile. Thus, we may write

$$Q = \left(\frac{\partial T}{\partial t}\right)_{ir} = \left(\frac{\partial T}{\partial t}\right)_{space} + \left(\frac{\partial T}{\partial t}\right)_{layer\ exchange}. \qquad (4.5.14)$$

Based on the preceding consideration, we may construct a simplified method for the calculation of cooling rates in terms of temperature perturbations. Let $T_0(z)$ denote the standard temperature and Q_0 its cooling rate profile. The deviation from this cooling profile is primarily related to temperature variation ΔT, particularly in a CO_2 atmosphere where the CO_2 mixing ratio is constant. Thus, we may write in finite-difference form:

$$\frac{\Delta Q}{\Delta T} = \frac{Q(T_0 + \Delta T) - Q(T_0 - \Delta T)}{2\Delta T} = a_0(z), \qquad (4.5.15)$$

where the term a_0 is called the *Newtonian cooling coefficient,* which, when multiplied by ΔT, gives the cooling rate deviation from a standard value Q_0. The preceding discussion provides the foundation for the Newtonian cooling approximation for the calculation of cooling rates in middle atmospheres, where the cooling-to-space approximation is most appropriate. The cooling rate may then be expressed by

$$Q(z) = Q_0(z) + \Delta Q(z) = Q_0(z) + a_0(z)[T(z) - T_0(z)].$$

The Newtonian cooling coefficient can be parameterized from the results computed by a line-by-line program, as illustrated in Dickinson (1973). Shown in Fig. 4.11 are the cooling-rate profiles produced by different types of CO_2, including the fundamental and all the first and second hot bands for $^{12}C^{16}O_2$ and other isotopic bands. It is evident that the hot bands are important contributors to cooling between 50 and 70 km.

4.6   Infrared Radiative Transfer in Cloudy Atmospheres

4.6.1   Fundamentals

In the thermal infrared, scattering, as well as emission and absorption, takes place within clouds. The basic IR radiative transfer equation for gaseous absorption and emission developed previously must, therefore, be modified to account for scattering processes.

Consider a plane-parallel cloud layer, and let the scattering coefficient for cloud particles be β_s and the absorption coefficient for cloud particles plus water vapor within the cloud be β_a. According to Kirchhoff's law, absorption is coupled with emission so that LTE is maintained. The source function in this case is the Planck function, B_ν. Let the source function associated with scattering be J_ν. The radiative
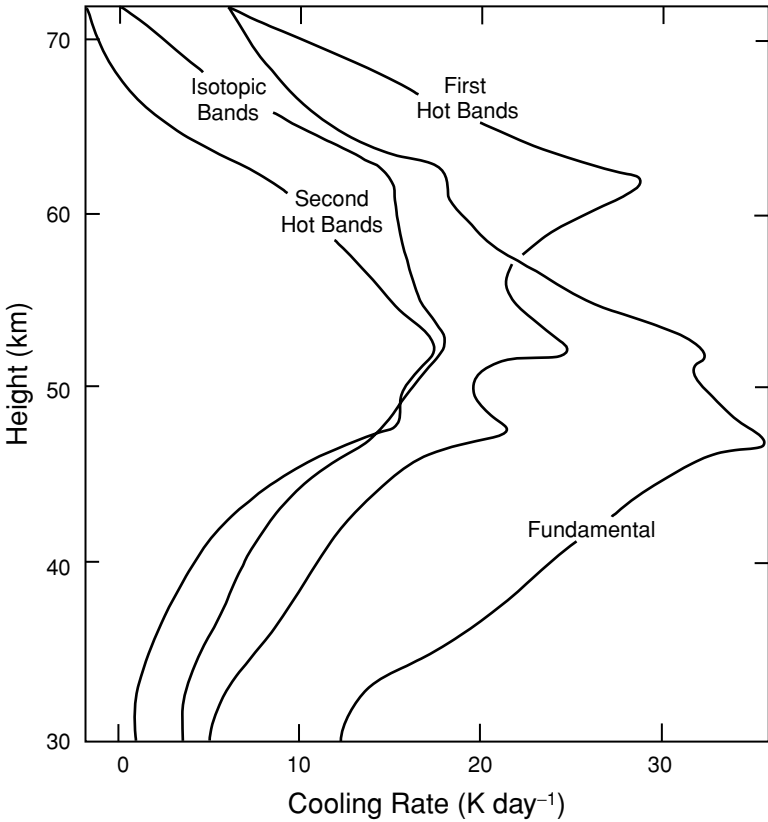
Figure 4.11   Infrared cooling rate given by different types of CO_2 bands for $^{12}C^{16}O_2$, including the fundamental, all the first and second hot bands, and all the fundamental bands of other isotopes (after Dickinson, 1973), based on the line data available at that time.

transfer equation may then be written as

$$
\begin{aligned}
\mu \frac{dI_\nu}{dz} &= -\beta_a(I_\nu - B_\nu) - \beta_s(I_\nu - J_\nu) \\
&= -\beta_e(I_\nu - S_\nu),
\end{aligned}
\tag{4.6.1}
$$

where the extinction coefficient $\beta_e = \beta_s + \beta_a$, and the source function involving absorption and scattering processes is

$$
S_\nu = (\beta_a B_\nu + \beta_s J_\nu)/\beta_e.
\tag{4.6.2a}
$$

It is an average of the two separate source functions, weighted by their respective absorption and scattering coefficients. The extinction coefficient is the inverse of the mean free path of a photon before scattering or absorption. Using the definition of the single-scattering albedo, $\tilde{\omega}_\nu = \beta_s/\beta_e$, in which a free path will end with a scattering

event, the source function may be expressed by

$$S_\nu = (1 - \tilde{\omega}_\nu)B_\nu + \tilde{\omega}_\nu J_\nu. \tag{4.6.2b}$$

The source function for scattering is associated with multiple scattering processes. In the thermal infrared, it suffices to take the azimuth-independent component:

$$J_\nu = \frac{1}{2} \int_{-1}^{1} P(\mu, \mu') I_\nu(\tau, \mu') \, d\mu', \tag{4.6.3a}$$

where the azimuth-independent phase function is defined by (see Section 3.4.1)

$$P(\mu, \mu') = \frac{1}{2\pi} \int_{0}^{2\pi} P(\cos \Theta) \, d\phi'. \tag{4.6.3b}$$

If the cloud as a whole is a blackbody, it would behave just like the earth's surface. In this case, radiation from below and above the cloud would not be able to penetrate the cloud. The emitted radiance at the cloud top or bottom is given by the Planck function. Most clouds that are composed of water droplets are black clouds, whereas clouds that are composed of ice crystals are generally non-black.

Shown in Fig. 4.12 is an infrared spectrum covering a spectral region from 9.1 to 17 μm for clear and various cloudy conditions measured from a high-spectral resolution infrared spectrometer aboard a high-flying ER-2 aircraft at about 20 km (see also Fig. 4.3). Strong absorption features are shown by ozone at 9.6 μm (1040 cm^{-1}) and by carbon dioxide at 15 μm (667 cm^{-1}), with weak lines of water vapor scattered in the 10–12 μm (1000–830 cm^{-1}) window of the spectrum. Except for thin cirrus containing small ice particles, clouds composed of water droplets such as low clouds or clouds containing large ice crystals behave like blackbodies or near blackbodies with little variations in the window. Thin cirrus, which are occasionally subvisual, contain small ice crystals with maximum dimensions ranging from about 5 to 20 μm. These clouds have features in the 10 μm window associated with the absorption coefficients of ice crystals that can be used for their identification. Infrared radiative transfer through thin cirrus in the window region must account for scattering processes in order to allow interpretation of the observed spectrum. In the following, we introduce approximations that can be employed to understand the transfer of infrared radiation in cloudy atmospheres.

4.6.2   Exchange of Infrared Radiation between Cloud and Surface

During the night, clouds play a critical role in modulating the surface temperature through thermal infrared radiation. To demonstrate this, we consider a configuration involving a surface with a temperature T_s. It is not a perfect blackbody but rather has an average emissivity ϵ_s. A cloud with a base temperature T_c and an average emissivity ϵ_c moves over the surface (e.g., snow). For simplicity, the effects of water vapor above and below the cloud are neglected in this discussion. Further, let the flux density emitted from the cloud and the surface be F_c and F_s, respectively, as shown in Fig. 4.13.
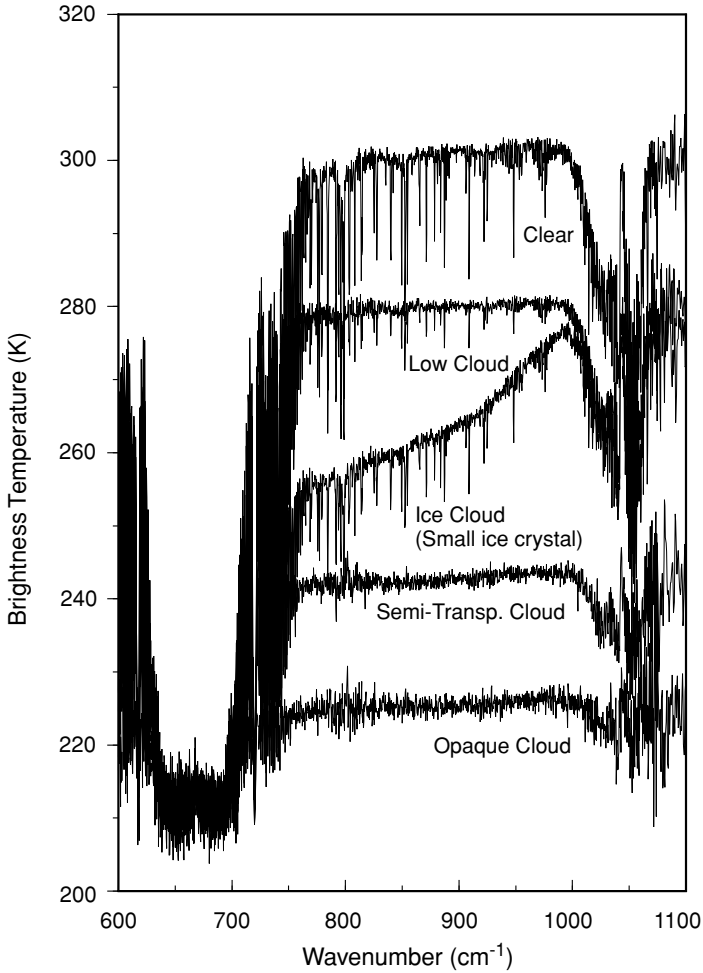
Figure 4.12 Spectra of brightness temperature observed from a high spectral resolution infrared spectrometer from the high-flying ER-2 aircraft over a domain, 37.1°–37.4° N, 95.0°–95.3° W, on April 21, 1996, indicating wavelength-dependent window brightness temperature changes according to various cloud types. The type of cloud indicated for each spectrum is identified from the Cloud Lidar System aboard the ER-2 (data taken from Smith *et al.,* 1998).

At the surface, contribution to the upward surface flux is produced by the emission from the surface plus the reflection of the flux emitted from the cloud base. Thus,

$$F_s^\uparrow = \epsilon_s \sigma T_s^4 + (1 - \epsilon_s) F_c^\downarrow. \tag{4.6.4a}$$

At the cloud base, contribution to the downward cloud flux is the sum of the emission
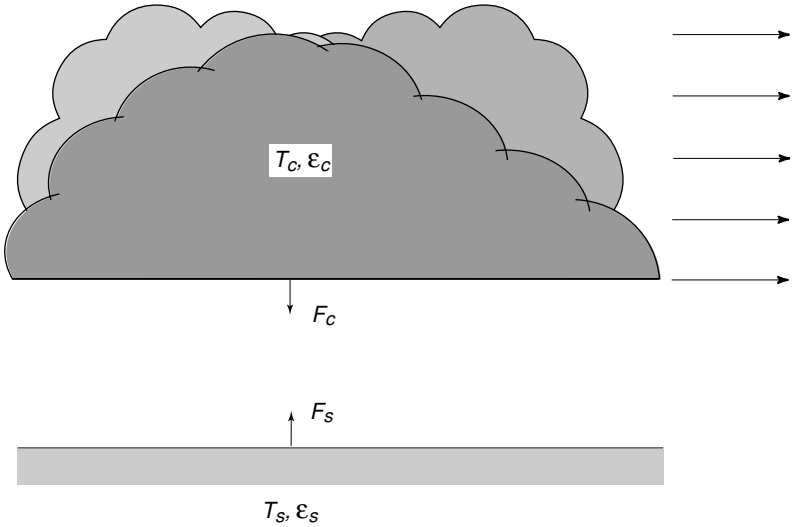
Figure 4.13   A simple configuration of a cloud moving over a surface. The surface temperature is modified by the presence of the cloud.

from the cloud and the reflection of the flux emitted from the surface. It follows that

$$F_c^\downarrow = \epsilon_c \sigma T_c^4 + (1 - \epsilon_c)F_s^\uparrow. \tag{4.6.4b}$$

In formulating the preceding equations, we have used the Stefan–Boltzmann and Kirchhoff laws. The solutions are

$$F_s^\uparrow = [(1 - \epsilon_s)\epsilon_c \sigma T_c^4 + \epsilon_s \sigma T_s^4]/[1 - (1 - \epsilon_s)(1 - \epsilon_c)], \tag{4.6.5a}$$

$$F_c^\downarrow = [(1 - \epsilon_c)\epsilon_s \sigma T_s^4 + \epsilon_c \sigma T_c^4]/[1 - (1 - \epsilon_c)(1 - \epsilon_s)]. \tag{4.6.5b}$$

Since the effects of water vapor between the cloud and the surface are neglected, the net flux in this case is then given by

$$\Delta F_n = F_s^\uparrow - F_c^\downarrow = \frac{\epsilon_c \epsilon_s}{1 - (1 - \epsilon_c)(1 - \epsilon_s)}\sigma\left(T_s^4 - T_c^4\right). \tag{4.6.6a}$$

If both the cloud and the surface are blackbodies, we can then define the cloud forcing as

$$\Delta F = \Delta F_n - \Delta F_{\text{clear}} = \sigma\left(T_s^4 - T_c^4\right) - \sigma T_s^4 = -\sigma T_c^4. \tag{4.6.6b}$$

Let the surface temperature increase due to the presence of the cloud be ΔT and assume that only infrared radiative processes take place in regulating the surface temperature. From the definition of the local heating rate, we should then have

$$\Delta T = \Delta t \left(-\frac{1}{\rho C_p}\frac{\Delta F}{\Delta z}\right). \tag{4.6.7}$$

The increase of the surface temperature ΔT is dependent on the time period Δt that the cloud remains over the snow surface and the net flux divergence $\Delta F / \Delta z$.

4.6.3   Two/Four-Stream Approximation

Combining Eqs. (4.6.1)–(4.6.3b) and using optical depth coordinates, the basic equation for the transfer of IR radiation in a scattering atmosphere may be written in the form

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - S(\tau, \mu), \qquad (4.6.8a)$$

where we have omitted the wavenumber subscript for simplicity of presentation. The source function defined in Eq. (4.6.2a) is given by

$$S(\tau, \mu) = \frac{\tilde{\omega}}{2} \int_{-1}^{1} P(\mu, \mu') I(\tau, \mu') d\mu' + (1 - \tilde{\omega}) B(\tau). \qquad (4.6.8b)$$

Equation (4.6.8a) can be solved exactly by means of the discrete-ordinates or adding methods for radiative transfer presented in Chapter 6. However, we introduce here a useful and accurate approximation utilizing the two-stream approximation discussed in Section 3.4.

The phase function can be expanded in Legendre polynomials P_ℓ based on the addition theorem for spherical harmonics, as shown in Appendix E. The azimuth-independent phase function defined in Eq. (4.6.3b) is given by

$$P(\mu, \mu') = \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell(\mu) P_\ell(\mu'). \qquad (4.6.9a)$$

In the context of the two-stream approximation, the upward and downward intensities are $I(\tau, +\mu_1) = I^\uparrow(\tau)$ and $I(\tau, -\mu_1) = I^\downarrow(\tau)$, where $\mu_1 = 1/\sqrt{3}$ based on the Gaussian quadrature. Moreover, the phase function in the limit of the two-stream approximation is given by

$$P(\mu_1, \pm \mu_1') = 1 \pm 3g\mu_1\mu_1' = 1 \pm g. \qquad (4.6.9b)$$

Thus, the integration in Eq. (4.6.8b) may be replaced by

$$\int_{-1}^{1} P(\mu, \mu') I(\tau, \mu') d\mu' = (1 - g) I^\downarrow(\tau) + (1 + g) I^\uparrow(\tau). \qquad (4.6.9c)$$

By expressing Eq. (4.6.8a) in terms of the upward and downward intensities, we have

$$\frac{dI^\uparrow(\tau)}{d\tau} = \gamma_1 I^\uparrow(\tau) - \gamma_2 I^\downarrow(\tau) - \gamma_3 B(\tau), \qquad (4.6.10a)$$

$$\frac{dI^\downarrow(\tau)}{d\tau} = \gamma_2 I^\uparrow(\tau) - \gamma_1 I^\downarrow(\tau) + \gamma_3 B(\tau), \qquad (4.6.10b)$$

where

$$\gamma_1 = [1 - \tilde{\omega}(1 + g)/2]/\mu_1, \ \gamma_2 = \tilde{\omega}(1 - g)/2\mu_1, \ \gamma_3 = (1 - \tilde{\omega})/\mu_1. \quad (4.6.10c)$$

Furthermore, we may parameterize the Planck function for a predivided layer in terms of the exponential function in the form

$$B(\tau) = B_0 e^{b\tau/\tau_1}, \quad (4.6.11)$$

where $b = \ln(B_1/B_0)$ with $B_0 = B(0)$ and $B_1 = B(\tau_1)$ being the Planck functions for temperatures at the top and bottom of the layer, respectively, and τ_1 is the optical depth of this layer. The parameterization is exact at the layer's boundaries.

The solutions for the preceding two first order differential equations subject to the inhomogeneous term defined in Eq. (4.6.11) can be derived in a straightforward but involved manner. They are given by

$$I^\uparrow(\tau) = K \, e^{-k(\tau_1 - \tau)} + H \, a \, e^{-k\tau} + Z^+ e^{b\tau/\tau_1}, \quad (4.6.12a)$$

$$I^\downarrow(\tau) = K a \, e^{-k(\tau_1 - \tau)} + H \, e^{-k\tau} + Z^- e^{b\tau/\tau_1}, \quad (4.6.12b)$$

where the eigenvalue of the solution and the similarity parameter are, respectively, defined by

$$k = \left(\gamma_1^2 - \gamma_2^2\right)^{1/2}, \quad a = \frac{\gamma_1 - k}{\gamma_2} = \frac{\gamma_2}{\gamma_1 + k}. \quad (4.6.12c)$$

The particular solution terms are

$$Z^\pm = \frac{B_0 \gamma_3 (\gamma_1 + \gamma_2 \pm b/\tau_1)}{k^2 - (b/\tau_1)^2}. \quad (4.6.12d)$$

The unknown coefficients K and H are to be determined from the radiation boundary conditions at the top and bottom of the layer. The preceding analysis constitutes the two-stream approximation for infrared radiative transfer in which we have retained the angular term $1/\mu_1$ in Eq. (4.6.10c).

Now, we return to Eqs. (4.6.8a,b) and solve the upward and downward intensities based on the integral technique illustrated in Eqs. (1.4.23) and (1.4.24). We have

$$I(0, \mu) = I(\tau_1, \mu)e^{-\tau_1/\mu} + \int_0^{\tau_1} S(\tau', \mu)e^{-\tau'/\mu} \frac{d\tau'}{\mu}, \quad (4.6.13a)$$

$$I(\tau_1, -\mu) = I(0, -\mu)e^{-\tau_1/\mu} + \int_0^{\tau_1} S(\tau', -\mu)e^{-(\tau_1 - \tau')/\mu} \frac{d\tau'}{\mu}. \quad (4.6.13b)$$

Analytic solutions are possible if specific forms for the source function terms are available. On substituting Eqs. (4.6.12a, b) into (4.6.8b) for the upward and downward intensities, respectively, and employing the two-term approximation for the integral term denoted in Eq. (4.6.9b), we obtain

$$S(\tau, \mu_1) = K \, v \, e^{-k(\tau_1 - \tau)} + H \, u \, e^{-k\tau} + Z_*^+ e^{b\tau/\tau_1}, \quad (4.6.14a)$$

$$S(\tau, -\mu_1) = K \, u \, e^{-k(\tau_1 - \tau)} + H \, v \, e^{-k\tau} + Z_*^- e^{b\tau/\tau_1}, \quad (4.6.14b)$$

where

$$v = (1 - \mu_1 k), u = a(1 + \mu_1 k),$$

$$Z_*^\pm = B_0(1 - \tilde{\omega})\left[1 + \frac{\tilde{\omega}(\gamma_1 + \gamma_2 \pm gb/\tau_1)}{\mu_1(k^2 - b^2/\tau_1^2)}\right]. \tag{4.6.14c}$$

Substitution of Eqs. (4.6.14a,b) into Eqs. (4.6.13a, b) yields

$$I^\uparrow(0, \mu) = I(\tau_1, \mu)e^{-\tau_1/\mu} + \frac{Kv}{(1 - \mu k)}(e^{-k\tau_1} - e^{-\tau_1/\mu})$$

$$+ \frac{Hu}{(1 + \mu k)}[1 - e^{-\tau_1(k+1/\mu)}] + \frac{Z_*^+}{(1 - b\mu/\tau_1)}(B_0 - B_1 e^{-\tau_1/\mu}), \tag{4.6.15a}$$

$$I^\downarrow(\tau_1, -\mu) = I(0, -\mu)e^{-\tau_1/\mu} + \frac{Ku}{(1 + \mu k)}[1 - e^{-\tau_1(k+1/\mu)}]$$

$$+ \frac{Hv}{(1 - \mu k)}(e^{-k\tau_1} - e^{-\tau_1/\mu}) + \frac{Z_*^-}{(1 + b\mu/\tau_1)}(B_1 - B_0 e^{-\tau_1/\mu}). \tag{4.6.15b}$$

By employing the two-stream approximation for the source function, we have demonstrated that the upward and downward intensities can be solved analytically. Moreover, for flux calculations, we may select two-stream in the upward direction and two-stream in the downward direction based on the double Gaussian quadrature such that the discrete directions $\mu = 0.2113248$ and 0.7886752, and the weight $a = 0.5$ for both directions. This constitutes the four-stream approximation in which the upward and downward fluxes are given by

$$F^\uparrow(\tau) = \int_0^1 I(\tau, \mu)\mu d\mu \cong \sum_{i=1}^2 I^\uparrow(\tau, \mu_i)\mu_i a_i, \tag{4.6.16a}$$

$$F^\downarrow(\tau) = \int_0^{-1} I(\tau, \mu)\mu d\mu \cong \sum_{i=1}^2 I^\downarrow(\tau, -\mu_i)\mu_i a_i. \tag{4.6.16b}$$

The methodology that combines the two-stream technique for the source function and the four-stream method for flux calculations is referred to as the *two/four-stream approximation* for radiative transfer. For infrared flux calculations, it suffices to use the diffusivity factor $1/\bar{\mu}$ of 1.66 to represent the inverse of the mean emergent angle of μ_1 that appears in the preceding equations. The use of 2 for $1/\bar{\mu}$ has also been suggested, which is more consistent with the four-stream approximation such that $\sum \mu_i/2 = \bar{\mu}$. However, in the context of the two/four-stream approximation, the use of 1.66 and 2 for $1/\bar{\mu}$ for flux calculations is similar. Further, to improve the representation of the phase function for the scattering of cloud particles based on a single parameter, namely the asymmetry factor, we may apply the delta-function adjustment based on the similarity principle for radiative transfer, introduced in Section 6.5.3, to adjust the optical depth, single-scattering albedo, and asymmetry factor. The delta two/four-stream approach has been shown to be an accurate and efficient means for the calculation of infrared

fluxes in cloudy conditions (Fu *et al.,* 1997; Toon *et al.,* 1989). Its accuracy, however, is less satisfactory for the calculation of solar flux transfer because of the strong anisotropic nature of the scattering phase function for cloud particles (see Section 6.5.4 for further discussion). In the following, we present some pertinent results for infrared cooling rates in typical clear and cloudy conditions and discuss the accuracy of the delta two/four-stream approximation.

4.7   Atmospheric Infrared Cooling Rates

Because the atmosphere loses radiative energy to space through thermal infrared emission, it is normally cooled by such processes. Thus, we speak of the infrared cooling rate (or negative heating rate) in the discussion of radiative transfer. Analogous to the solar heating rate defined in Eq. (3.5.3), we may define the infrared cooling rate in the form

$$\left(\frac{\partial T}{\partial t}\right)_{ir} = -\frac{1}{\rho C_p}\frac{dF(z)}{dz},\tag{4.7.1}$$

where the net infrared flux at a given height is defined by

$$F(z) = F^{\uparrow}(z) - F^{\downarrow}(z),\tag{4.7.2}$$

and the upward and downward infrared fluxes covering the entire thermal infrared spectrum have been defined in Eq. (4.2.11). In what follows, we present a number of representative infrared cooling rate profiles computed from the line-by-line (LBL) method and the correlated k-distribution (CKD) method.

Figures 4.14a, 4.14c, and 4.14e show the heating rate profiles computed using LBL integration for H_2O (0–2200 cm^{-1}), CO_2 (540–800 cm^{-1}), and O_3 (980–1100 cm^{-1}), respectively. Cooling rates produced by water vapor lines are seen in the troposphere and the middle atmosphere. A significant variability is shown for the four atmospheric humidity and temperature profiles used in the calculations. The maximum and minimum patterns located at about 50 km are related to the temperature inversion (see Fig. 3.1). In the lower atmosphere, cooling rates range from 1 to 2 K day^{-1}, and are primarily produced by the rotational band of water vapor. Carbon dioxide generates significant cooling rates in the middle atmosphere where the temperature profile is an important factor in determining flux exchanges. At about 50 km, a cooling rate of about 10 K day^{-1} is seen. The ozone cooling rate profile is governed by the ozone concentration, which has a maximum located at about 20–25 km, resulting in a small amount of positive heating.

Figures 4.14b, 4.14d, and 4.14f show the corresponding error profiles for the results computed using CKD. From the LBL results, it is clear that infrared cooling rates in
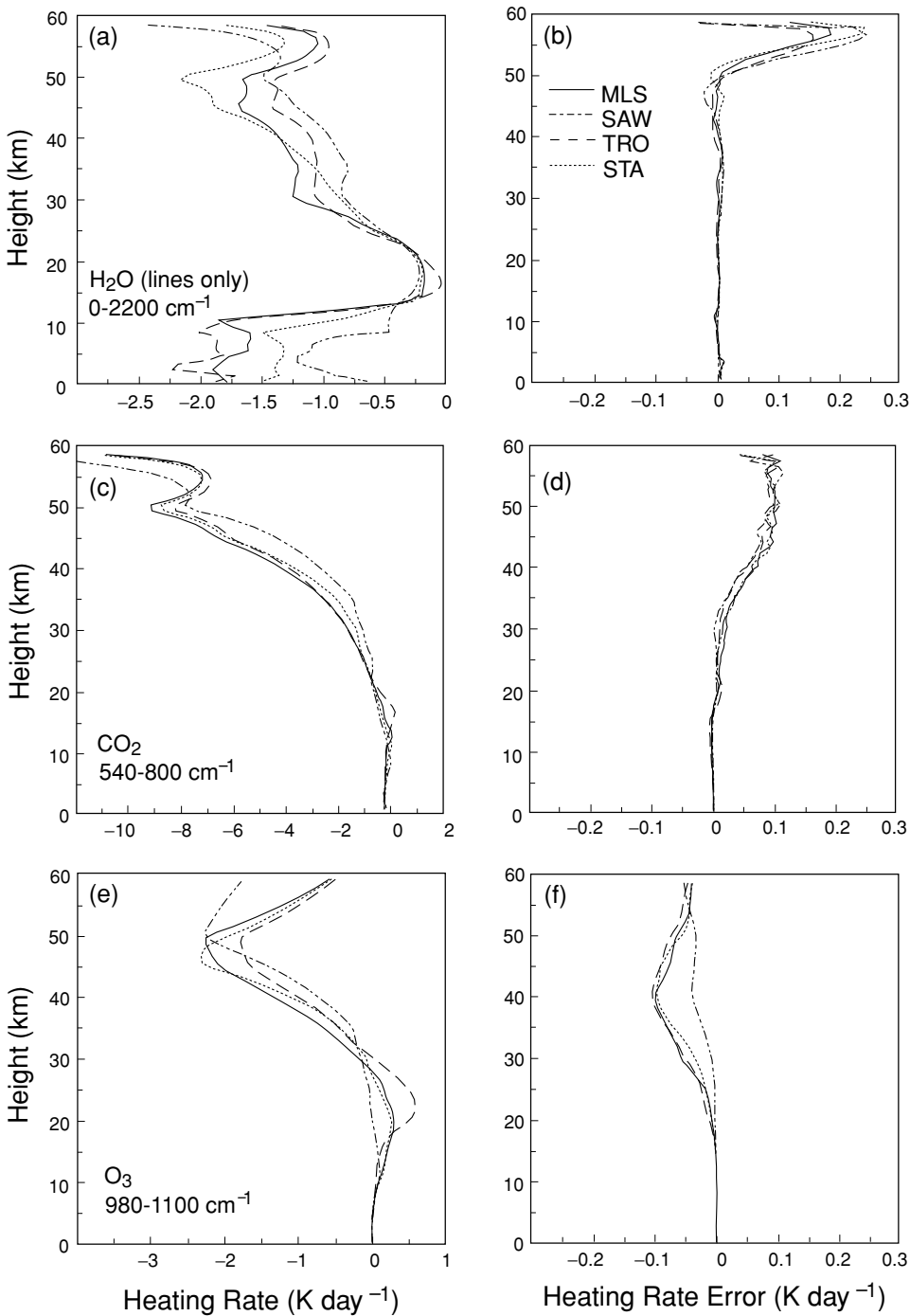
Figure 4.14 Heating rate profiles computed from LBL and the error profiles produced by CKD for the H_2O (a and b), CO_2 (c and d), and O_3 (e and f) bands in the infrared spectrum for the midlatitude summer (MLS), subarctic winter (SAW), tropical (TRO), and U.S. Standard (STA) atmospheres (data taken from Fu and Liou, 1992).

the troposphere are primarily produced by H_2O. In the stratosphere, H_2O, CO_2, and O_3 contribute $\sim$15%, 70%, and 15% of the total cooling in the vicinity of the 50–km region, respectively. For H_2O, errors in the heating rates calculated by CKD are less than $\sim$0.01 K day^{-1} in the troposphere and stratosphere, and less than $\sim$0.24 K day^{-1} above the stratopause, as shown in Fig. 4.14. The relatively large errors above the stratopause may be explained by the fact that at lower pressure regions the atmosphere is thinner and more transparent and as a result, distant layers would affect cooling more significantly. As shown in Fig. 4.14, errors in the CO_2 cooling rates produced by CKD in the troposphere and lower stratosphere are $\sim$0.01 K day^{-1}, while the errors above 30 km are less than $\sim$0.1 K day^{-1}. For O_3, the heating rates computed from CKD are also in excellent agreement with those from LBL. The differences are less than $\sim$0.1 K day^{-1}. CKD yields the most accurate results in the subarctic winter atmosphere due to the O_3 9.6-μm band. This can be explained by the larger amount of ozone in the subarctic winter atmosphere, which leads to stronger absorption and a more opaque atmosphere.

In Fig. 4.15, we display infrared cooling rate profiles for clear sky employing the midlatitude summer atmosphere, representing a relatively moist condition, and the subarctic winter atmosphere, representing a dry condition. The vertical resolution used in the calculation was 0.25 km. Calculations were made from a line-by-line equivalent multiple scattering program. Substantial differences are shown for these two profiles. For the former, cooling rates are on the order of 2 K day^{-1}, with the largest cooling of about 3 K day^{-1} near the surface produced by the rotational band of water vapor and continuous absorption principally in the 10 μm window. For the latter, cooling rates are only on the order of about 1 K day^{-1}. When clouds are present, significant cooling/heating takes place in the cloud region. In the low cloud condition, cloud top cooling reaches 35 K day^{-1}, generated from the flux exchange between the air above the cloud and the optically black cloud. Because the low cloud base (1 km) is close to the surface, cloud base heating is extremely small. In the middle cloud condition, however, a large cooling rate of about 60 K day^{-1} occurs at the cloud top coupled with a heating rate of about 16 K day^{-1}, resulting from the trapping of longwave flux emitted from the warmer surface and lower atmospheres. Because of a lower ice-water content in the case of high clouds, smaller cloud top cooling and cloud bottom warming, each with a value of about 4 K day^{-1}, are shown. These results clearly reveal the importance of clouds in the generation of cooling/heating rates that are directly related to their dynamic and physical processes in the atmosphere.

Finally, we present an example of infrared flux measurements from aircraft along with their theoretical interpretation. Measurement of the atmospheric heating/cooling rate from an aircraft platform is a difficult task. It requires calibrated broadband flux radiometers to perform upward and downward flux measurements during the course of aircraft operation. It is particularly difficult for the measurement of solar flux owing to the geometry of the sun with respect to the radiometer on the aircraft. Moreover, the heating/cooling rate is a result of net flux divergence, which is the difference of two close values. Because of the inherent uncertainty in radiometric instrumentation, heating/cooling rates derived from flux measurements usually contain substantial
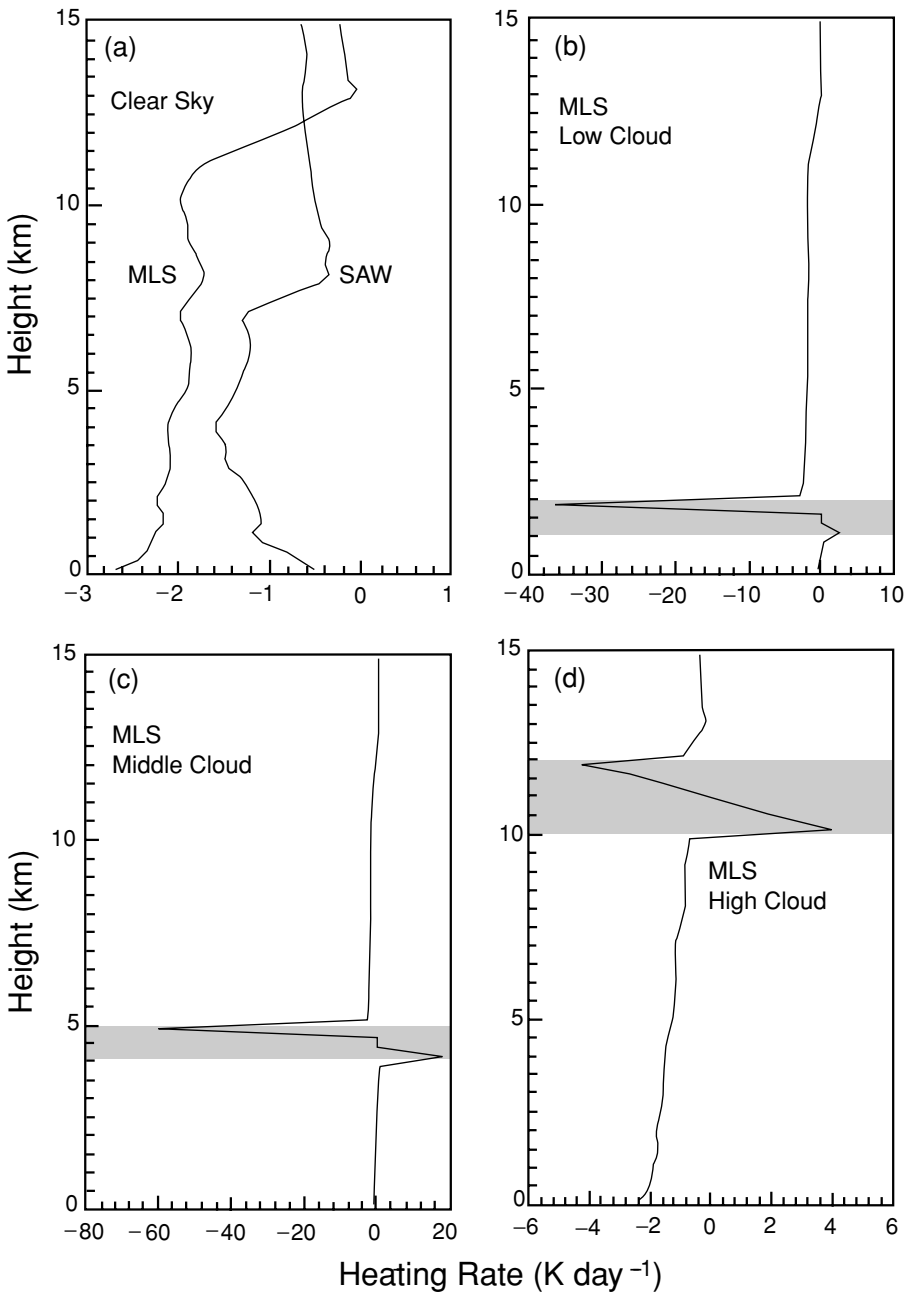
Figure 4.15 (a) Total cooling rates for the midlatitude summer (MLS) and subarctic winter (SAW) atmospheres in clear sky. (b), (c), and (d) display cooling rate profiles for the MLS atmosphere containing low, middle, and high clouds, respectively. The liquid (ice) water content and mean effective radius (size) for these clouds used in the calculations are (0.22 g m^{-3} and 5.89 μm), (0.28 g m^{-3} and 6.2 μm), and (0.0048 g m^{-3} and 41.5 μm), respectively. The locations of these clouds are shaded (data taken from Fu *et al.,* 1997).
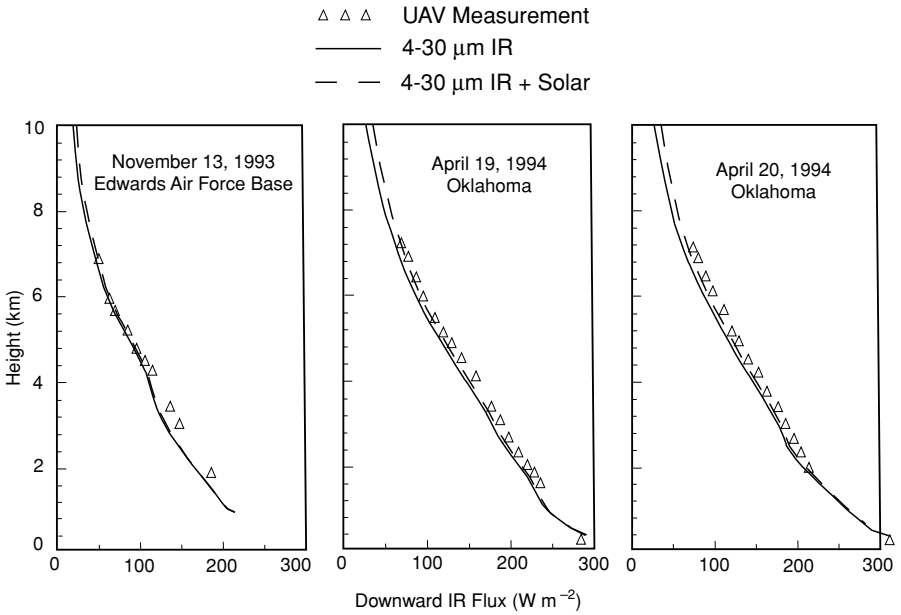
Figure 4.16 Downward infrared fluxes obtained from the measurements of unmanned aerospace vehicle (UAV) and from theoretical results for a spectral region covering 4–30 μm with and without including the solar radiation contribution (based on an interim report K. N. Liou submitted to P. Crowley and J. Vitko of the Department of Energy, June 1995).

uncertainty. In the following, we present downward infrared fluxes derived from radiometers on board an unmanned aerospace vehicle that was flown on a number of days and locations as indicated in Fig. 4.16. Downward infrared fluxes increase with decreasing altitude, signifying the emission contribution from the atmosphere. Theoretical results are computed by using the correlated k-distribution method along with observed temperature and moisture profiles collected from nearby soundings. There is a general agreement between the computed and measured downward infrared flux profiles, particularly when the contribution of solar flux is accounted for in the theoretical calculation. In Table 2.3, we show that there are about 11 W m^{-2} of solar energy contained in the wavelengths greater than 4 μm. This energy has been ignored in thermal infrared radiative transfer calculations when comparisons with radiometric measurements are made. Deviation in the lower atmosphere in the case of Edwards Air Force Base is due to uncertainty of the moisture profile data available from distant sounding.

In summary, to cross check theoretical heating/cooling calculations from the line-by-line or the correlated k-distribution models, we must have reliable and verifiable flux measurements from aircraft, an area that requires further advances in spaceborne radiometric technology.

Exercises

4.1 Table 4.1 lists the line-by-line data from HITRAN 96 (Rothman *et al.,* 1998) for a 10 cm^{-1} interval in the 1.38 μm H$_2$O band. (a) Using the Lorentz line shape, plot the absorption coefficient k_ν as a function of wavenumber in this interval. (b) Divide k_ν in the logarithmic scale into 50 equal intervals and compute the total cumulative number $n(0, k)$ of k_ν in progressive intervals. (c) The cumulative probability function is defined by $g(k) = n(0, k)/N$, where N is the total number such that $g(0) = 0$ and $g(50\Delta\log k) = 1$. Plot $k(g)$ in the g-domain.

Table 4.1

HITRAN 96 Line Data for a 10 cm^{-1} Interval in the 1.38 μm Band for $p = 1013$ mb and $T = 296$ K

ν_0 (cm^{-1})	S [cm^{-1}/(atm cm)]	α(cm^{-1})
7280.31512	4.194E-03	0.0704
7280.47400	8.872E-04	0.0846
7281.08200	3.764E-02	0.0994
7281.72912	4.033E-03	0.0602
7282.70531	5.673E-04	0.0752
7283.01859	1.132E-02	0.0680
7283.73107	1.710E-02	0.0710
7284.71668	2.401E-03	0.0702
7285.04497	4.275E-04	0.0866
7286.05083	4.732E-03	0.0683
7287.00300	6.990E-03	0.0886
7287.28900	2.285E-02	0.1020
7287.50218	2.877E-04	0.0685
7288.09091	6.882E-02	0.1002
7290.10832	3.226E-02	0.0872

Table 4.2

The Quadrature Points and Weights for Integrals in the g-Domain (0,1)

Quadrature points	Quadrature weights
3.20770E-02	8.13791E-02
1.60926E-01	1.71362E-01
3.61656E-01	2.22259E-01
5.88344E-01	2.22259E-01
7.89074E-01	1.71362E-01
9.17923E-01	8.13791E-02
9.53472E-01	8.69637E-03
9.66501E-01	1.63036E-02
9.83500E-01	1.63036E-02
9.96528E-01	8.69637E-03

(d) Compute the spectral transmittance from the line-by-line approach employing an interval of 0.01 cm^{-1} and from the k-distribution method using quadrature points and weights listed in Table 4.2. Use the path length u from 10^{-5} to 10 (g cm^{-2}). Compare the two results. Note that 1 (g cm^{-2}) is 2.24 $\times$ $10^4/M$ (atm cm), where M is the molecular weight of the gas.

4.2 (a) Show that the equivalent width of a single Lorentz line is given by $W = 2\pi\alpha L(x)$, where $L(x)$ is the Ladenberg and Reiche function, and $x = su/2\pi\alpha$.
(b) From the property of the Bessel function, show that the equivalent width is proportional to u and $\sqrt{u}$, respectively, in the limits of weak- and strong-line approximations.

4.3 Given the absorption coefficient expressed in the Elsasser regular band model [Eq. (4.4.10)], derive the probability function $f(k)$ for this case.

4.4 (a) From the Malkmus model defined in Eq. (4.4.34b), show that the probability distribution function can be derived from the inverse Laplace transform given by

$$f(k) = \frac{1}{2\sqrt{\pi}} c_{\bar{v}} d_{\bar{v}}^{1/2} k^{-3/2} \exp\left(c_{\bar{v}} - k/d_{\bar{v}} - c_{\bar{v}}^2 d_{\bar{v}}/4k\right).$$

(b) Show that the cumulative probability function is given by

$$g(k) = \frac{1}{2} e^{2c_{\bar{v}}} \mathrm{erfc}\left[\left(\frac{c_{\bar{v}}}{2}\right)^{1/2}\left(\frac{1}{y} + y\right)\right]$$
$$+ \frac{1}{2}\mathrm{erfc}\left[\left(\frac{c_{\bar{v}}}{2}\right)^{1/2}\left(\frac{1}{y} - y\right)\right],$$

where $y = [2k/(c_{\bar{v}} d_{\bar{v}})]^{1/2}$. The notation erfc is the complementary error function given by $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, and the error function is defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}\, dx.$$

(c) From (b), show that $dg(k)/dk = f(k)$.
(d) Derive the inverse Laplace transform of the spectral transmittance under the limits of weak- and strong-line approximations defined in Eq. (4.4.33).

4.5 Consider a simplified Voigt profile by combining a rectangular Doppler core (with a value of C) with Lorentzian wings as follows:

$$f_v(v) = \begin{cases} C, & |v| \leq v_0, \\ \alpha/\pi v^2, & |v| > v_0. \end{cases}$$

Derive the equivalent width for Goody's random model.

4.6 The half-width of a Lorentz line is proportional to pressure and can be expressed by $\alpha \cong \alpha_r(p/p_r)$, where α_r is the half-width at the reference pressure p_r. Show

that the optical depth may be expressed by

$$T_\nu = e^{-\tau} = \left[(\nu^2 + \alpha_1^2)/(\nu^2 + \alpha_2^2)\right]^\lambda,$$

where α_1 and α_2 are two integration limits and $\lambda = Sp_r q/(2\pi g\alpha_r)$, with q the mixing ratio and g the gravitational acceleration.

4.7 (a) With the strong-line approximation limit $\bar{S}u/\alpha \gg 1$, show that

$$T_{\bar{\nu}}(u) = \exp(-\sqrt{\pi \bar{S}\alpha u}/\delta).$$

This is referred to as the *square root approximation for the random model*. (b) Based on this approximation, show that the precipitable water in cloudy atmospheres may be determined from

$$PW = (c/m)\left[\ln(F_{\bar{\nu}}/F_{0,\bar{\nu}})\right]^2,$$

where c is a constant related to the band and known atmospheric parameters, m denotes the air mass, which is related to pressure, and $F_{\bar{\nu}}$ and $F_{0,\bar{\nu}}$ represent the observed solar flux in the 0.94 μm band at the ground and at the top of the atmosphere, respectively. This is the principle of the sunphotometer for the measurement of precipitable water.

4.8 Using the probability function

$$P(S) = \frac{C}{S} e^{-s/\bar{s}},$$

where C is a normalization factor, show that the average equivalent width is given by

$$\bar{W} = c_{\bar{\nu}}\left[(1 + d_{\bar{\nu}}u)^{1/2} - 1\right],$$

where $c_{\bar{\nu}} = \pi\alpha/2\delta$ and $d_{\bar{\nu}} = 4\bar{S}/\pi\alpha$. In the derivation, use the following integration formulas:

$$\int_0^\infty e^{-ax}\frac{1 - e^{-x}}{x}dx = \ln(1 + 1/a),$$

$$\int \ln(x^2 + b)\,dx = x\ln(x^2 + b) - 2x + 2\sqrt{b}\tan^{-1}(x/\sqrt{b}).$$

Based on the procedure outlined in Section 4.4.3, determine the coefficients $c_{\bar{\nu}}$ and $d_{\bar{\nu}}$ from the line-by-line data.

4.9 From the definition of the Voigt profile, derive Eq. (4.4.45). In the derivation, set $\cos \nu t = \cos[(\nu - \nu') + \nu']t$.

4.10 The spectral transmittance defined in Eq. (4.3.20a) can be written in the form

$$T_{\bar{\nu}}(q, q_c) = \int_0^\infty f_2^*(k'')dk'' \int_{k''}^\infty e^{-ku_1} f_1(k - k'')\,dk.$$

By changing the order of integrals such that the resulting areas of the integrations remain the same, derive Eq. (4.3.20c) via the convolution theorem of the Laplace transform given in Eq. (4.3.20b).

4.11 Derive Eqs. (4.6.12a,b) from Eqs. (4.6.10a,b). This is the two-stream approximation for thermal infrared radiation transfer.

Suggested Reading

Elsasser, W. M., and Culbertson, M. F (1960). Atmospheric radiation tables. *Meteorol. Monog*. **4** (23), 1–43. This monograph provides essential information on the transfer of infrared radiation and broadband emissivity values.

Goody, R. M., and Yung, Y. L. (1989). *Atmospheric Radiation. Theoretical Basis,* 2nd ed. Oxford University Press, New York. Chapters 3–5 contain authoritative and fundamental discussions of the theory of gaseous absorption and band models.

Kondratyev, K. Ya. (1969). *Radiation in the Atmosphere* (International Geophysics Series, Vol. 12). Academic Press, New York. Gaseous absorption and infrared radiative transfer are discussed in Chapters 3 and 9.

Liou, K. N. (1992). *Radiation and Cloud Processes in the Atmosphere. Theory, Observation, and Modeling*. Oxford University Press, New York. Chapter 2 provides a more detailed presentation of the band models and emissivity parameterizations.

Chapter 5 | Light Scattering by Atmospheric Particulates

5.1 Morphology of Atmospheric Particulates

The earth's atmosphere contains various types of particulates ranging from aerosols, water droplets, and ice crystals to raindrops, snowflakes, and hailstones. They are produced by a number of physical and dynamic processes, which govern their formation and growth in the atmosphere. In association with the discussion of the fundamentals of light scattering and absorption by particulates, we present in the following an overview of the morphology of pertinent atmospheric particulates, with a specific emphasis on aerosols and cloud particles.

In Section 3.1.2, we briefly discussed the sources of aerosol particles in the atmosphere. Aerosols of natural origin include dust from arid and semiarid regions, particles from sea spray over the oceans, volcanic debris, smoke from forest fires, extraterrestrial or interplanetary dust, and small particles produced by the chemical reactions of natural gases. Man-made aerosols are produced by particles directly emitted during combustion processes and particles formed from emitted gases.

Aerosols are usually classified in terms of their location and type. The continental type is subdivided into clear (rural, forest), average, desert (background, windy), and urban/industrial conditions. The aerosol components involved in this type are water-soluble particles, dustlike particles, soot, and minerals. The maritime type is grouped into clear, mineral, and polluted conditions, involving sea salt, mineral, sulfate, and soot particles. The polar type is subdivided into polluted, clear Arctic, and clear Antarctic conditions, which contain soot, mineral, sea salt, and sulfate particles. Because of the diversity of aerosol types in various regions, it appears that they must be governed in part by the transport process associated with atmospheric circulation.

The dustlike substances are mineral dust formed in nondesert locations and are representative of soil conditions. Water-soluble substances refer to the parts of aerosols that are soluble in water and consist of a mixture of sulfate, nitrate, and organic compounds, the mixing ratio of which varies with the source. The term "soot" represents all carbonaceous materials that are a product either of direct particle emission into the atmosphere by combustion processes or of the transformation of combustion-related gases to particles. Sulfate particles are products of the oxidation of sulfur-bearing gases

that are generated both naturally and anthropogenically. Natural sulfate is produced mostly by biogenic sources, whereas its anthropogenic counterpart stems primarily from coal and oil burning, smelting, petroleum refining, and transportation. The diverse tropospheric aerosol types have been conveniently grouped into five basic categories: dustlike soil, soot, sulfate, sea salt, and organic aerosols.

The size distributions of atmospheric aerosols are complex, but they have frequently been divided into two size classes, representing two primary formation mechanisms. Particles with diameters larger than about 1 μm are produced by the breakup and suspension of bulk materials by the wind (e.g., sea salt and soil dust). Fine particles smaller than 1 μm are usually formed by combustion or chemical conversions of the gaseous precursors into liquid or solid products. The size spectrum of aerosols in the atmosphere has been the subject of extensive research in the past four decades, and a detailed discussion is beyond the scope of this text. However, for the purpose of illustration, Fig. 5.1 depicts well-known measured size distributions of natural aerosols in reference to the logarithmic scale (Junge, 1963). Aerosols with radii smaller than 0.1 μm are collectively referred to as *Aitken nuclei*. For radii between 0.1 and 1.0 μm, the aerosols are referred to as *large nuclei*. Aerosols with radii larger than 1.0 μm are called *grant nuclei*. It is clear that the majority of atmospheric aerosols have sizes on the order of 0.1 μm, but aerosols as large as 10 to 20 μm have also been observed. To facilitate the remote sensing of aerosols, a number of analytic expressions have been developed to represent their size distributions. These include the Junge power law (Fig. 5.1 and Section 7.2.1), the gamma distribution [Eq. (5.3.44)], the log-normal distribution [Eq. (5.3.45)], and their modifications.

The radiative properties of atmospheric aerosols are critically dependent on their refractive indices as functions of wavelength. Both the real and imaginary parts of aerosols have been measured in the laboratory and tabulated in terms of dustlike, water-soluble, soot, oceanic, sulfate, mineral, and water substances covering a spectrum of wavelengths (e.g., d'Almeida *et al.,* 1991). In the solar visible region, absorption due to oceanic and sulfate particles is relatively small. A large absorption occurs for mineral, dustlike, and water-soluble particles, and particularly for soot. In a wet environment, aerosols interact with the ambient water vapor, the process of which affects their size, shape, and chemical composition, and consequently their optical properties. The humidity effects on aerosol size, shape, and composition are intricate and are directly related to the formation of water droplets.

Electron microscopic photographs of aerosols show a great variety of shapes ranging from quasi-spherical to highly irregular geometries. Solid aerosols are generally irregular. Figure 5.2 displays an example of microsized dust particles from the Sahara desert collected in Israel. The shape of aerosols also depends on the relative humidity. Many spherical aerosols have internal inclusions and/or attachments (see Fig. 5.28). A large number of submicrometer particles from combustion and biomass burnings are in the form of clusters and aggregates. Some aerosols are effective condensation and ice nuclei upon which cloud particles are formed in the atmosphere.
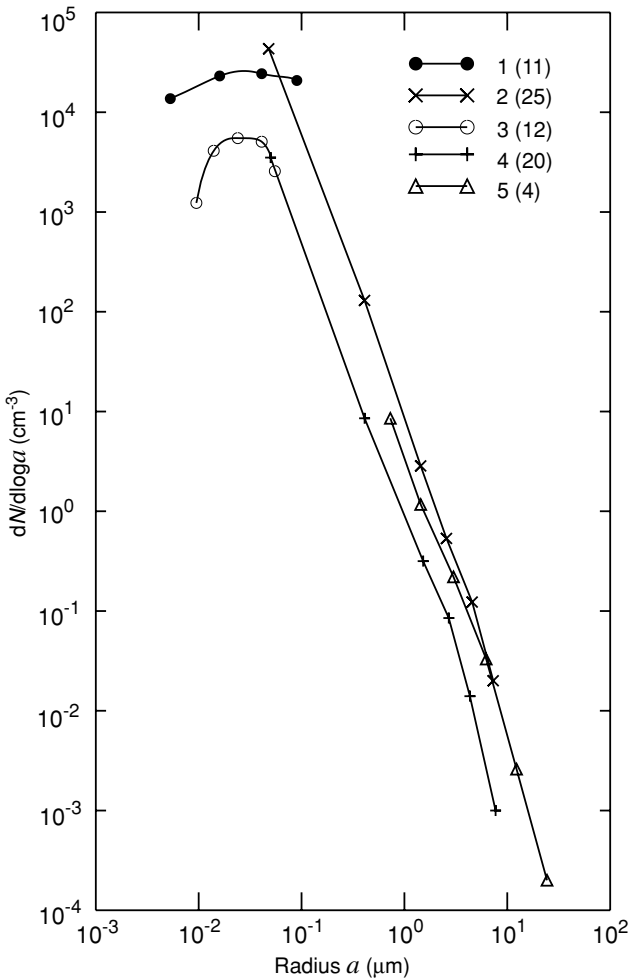
Figure 5.1 Size distribution of natural aerosols. Curves 1, 2, and 5 were measurements at Frankfurt/Main. Curve 1 represents ion counts converted to nuclei numbers. Curve 2 denotes data from impactors, and curve 5 is the average sedimentation data. Curves 3 and 4 were simultaneous measurements at Zugspitze, 3 km above sea level. The figures in parentheses are the number of individual measurements (data taken from Junge, 1963).

Clouds, which are composed of water droplets and/or ice crystals, are conventionally classified in terms of their position and appearance in the atmosphere. In midlatitudes, clouds with base heights of about ∼6 km are defined as high clouds and are commonly referred to as cirrus clouds. The group of low clouds with base heights below ∼2 km include stratus and cumulus. In between high and low clouds are a group called middle clouds consisting of altocumulus and altostratus. Clouds with significant vertical developments such as those occurring in the tropics are named cumulonimbus.
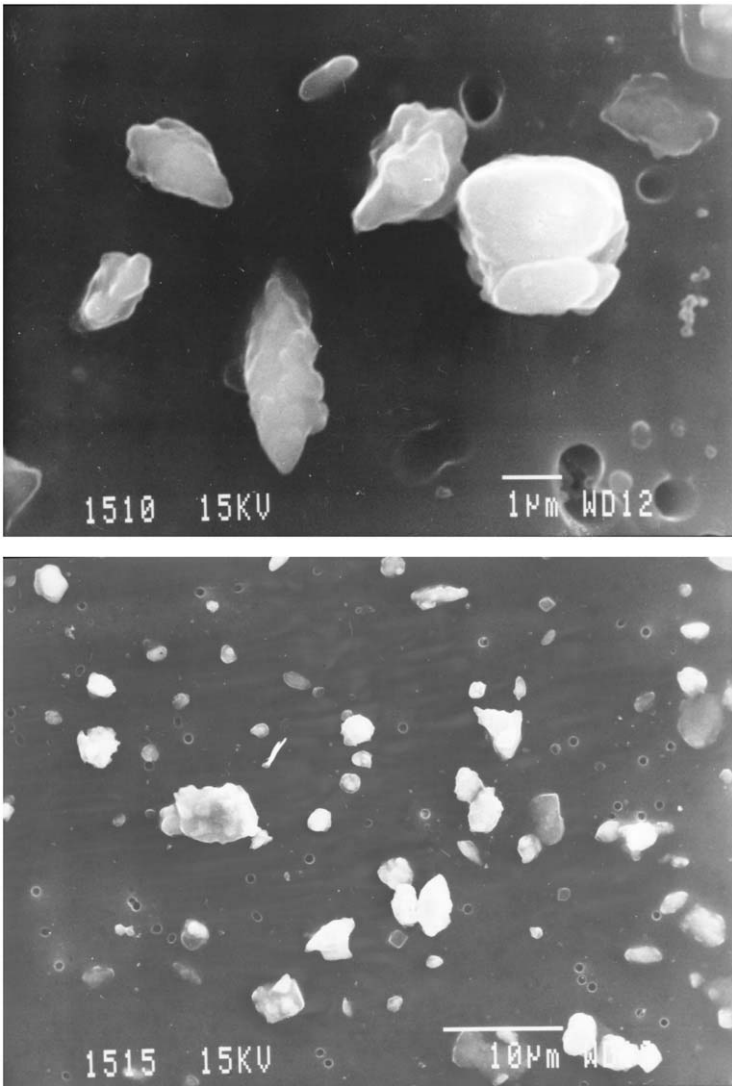
Figure 5.2   Microsized dust particles from the Sahara desert collected in Israel in 1998. Note that the scales for upper and lower diagrams are 1 and 10 μm, respectively (courtesy of Yoram Kaufman).

The microphysical state of a cloud is determined by the dispersion of particle sizes and their phase (ice or liquid). Based on extensive aircraft observations, low clouds and some middle clouds, because of temperature stratification, are generally composed of spherical water droplets with sizes ranging from ~1 μm to 20 μm. A typical water droplet is on the order of 5 μm. Some middle clouds with temperatures warmer than about $-20°$C contain supercooled water droplets that coexist with ice particles.

Small water droplets are spherical in nature because of the requirement of surface pressure to hold the water molecules together. However, collision and coalescence of a fortunate larger water droplet (on the order of 20 μm) with other water droplets can produce millimeter- and centimeter-sized raindrops, as is evident from our daily experience. The falling raindrops in a dynamic flow field deviate from the spherical shape and have been investigated by Pruppacher and Pitter (1971) and Wang (1982). The latter author has developed mathematical expressions to define the shape of falling graupel, hailstones, and some deformed raindrops. The nonsphericity of these hydrometeors is important in the development of radar backscattering techniques for the quantitative detection of precipitation and thunderstorms based on light scattering theory.

Ice crystals in the atmosphere are primarily present in cirrus clouds, as well as in the top portion of some middle clouds. The formation, maintenance, and dissipation of cirrus clouds are principally associated with large-scale synoptic features and disturbances. In the tropics, they are related to deep-cumulus outflows. The ice crystal shapes depend on temperature and relative humidity as well as whether they undergo collision and coalescence processes in the clouds. Weickmann (1948) observed that at humidities approaching water saturation, ice crystals in cirrus castellatus, cirrocumulus, and cirrus generating cells have prismatic skeleton shapes that occur in hollow and cluster crystals, referred to as *bullet rosettes.* Between ice and water saturation, ice crystals grow in the form of prisms in cirrus filosus and cirrus densus. In cirrostratus, where relative humidities are close to ice saturation, ice crystals are primarily individual with full crystals such as columns, prisms, and plates.

In midlatitude cirrus, where substantial aircraft observations have been conducted, ice crystal shapes are generally classified as a function of temperature (Heymsfield and Platt, 1984; Heymsfield and Iaquinta, 2000). Hollow columns and hexagonal plates are the most abundant types near the cloud top for most cirrus clouds. Spatial ice crystals such as bullet rosettes are the predominant forms above about $-40°$C, while hollow or solid columns prevail below about $-50°$C. In between these temperatures, convective cirrus contain predominantly spatial crystal forms, whereas stable cirrus are primarily composed of hollow columns. Figure 5.3 illustrates a spectrum of ice crystal sizes and shapes as a function of height in a typical midlatitude cirrus. It is evident that at the cloud top, pristine and small columns and plates are predominant. At the cloud bottom, however, irregular bullet rosettes and aggregates are produced, presumably due to the collision and coalescence associated with vertical mixing and gravitational pulling. Because ice crystal size and shape vary greatly with time and space, a presentation of representative values for the purpose of remote sensing and climate applications is a difficult task. Nevertheless, typical ice-crystal size distributions for midlatitude cirrus clouds have been developed on the basis of a number of intensive field observations of cirrus clouds from aircraft platforms, including the First ISCCP Regional Experiments in 1986 and 1991, the European cirrus experiment in 1989, and the Subsonic Aircraft: Contrail and Cloud Effect Special Study in 1996. These distributions are shown in Fig. 5.4. Displayed are five representative
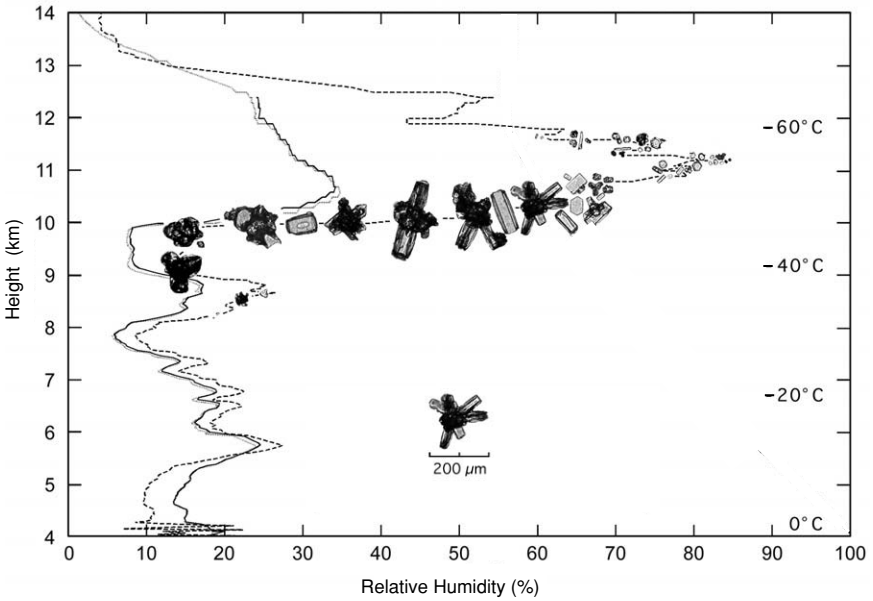
Figure 5.3   Ice crystal size and shape as a function of height and relative humidity captured by a replicator balloon sounding system in Marshall, Colorado, on November 10, 1994. The relative humidity was measured by cryogenic hygrometer (dashed line) and Vaisala RS80 instruments (solid line and dots). Also shown is temperature as a function of height (courtesy of Andrew Heymsfield of the National Center for Atmospheric Research).

size distributions spanning a mean effective ice crystal size range from 10 to 124 μm defined by

$$D_e = \int_{L_{\min}}^{L_{\max}} V \, n(L) \, dL \bigg/ \int_{L_{\min}}^{L_{\max}} A \, n(L) \, dL, \qquad (5.1.1)$$

where V is the volume of an ice crystal, L is the maximum dimension, and A is the geometric projected area of an ice crystal on a surface perpendicular to the incident beam. For the purpose of calculating D_e, we have assumed $V \approx LD^2$ and $A \approx LD$, where D is the ice-crystal width. From available replicator and optical probe observations, the majority of ice-crystal shapes are aggregate, bullet rosettes, hollow columns, and plates, as displayed in Fig. 5.3.

Observations of ice-crystal size and shape distributions have been extremely limited for tropical cirrus. During the Central Equatorial Pacific Experiment conducted in 1993, aircraft microphysical measurements were carried out by a two-dimensional optical probe (30–300 μm) and a video ice particle sampler (<30 μm). Based on available measurements, ice-crystal sizes in the tropics range from about 10 μm to 2000 μm with four predominant shapes: bullet rosettes, aggregates, hollow columns, and plates, similar to those occurring in midlatitudes. Smaller ice-crystal sizes generally occur in cloud-top (colder temperature) conditions, whereas larger ice-crystal
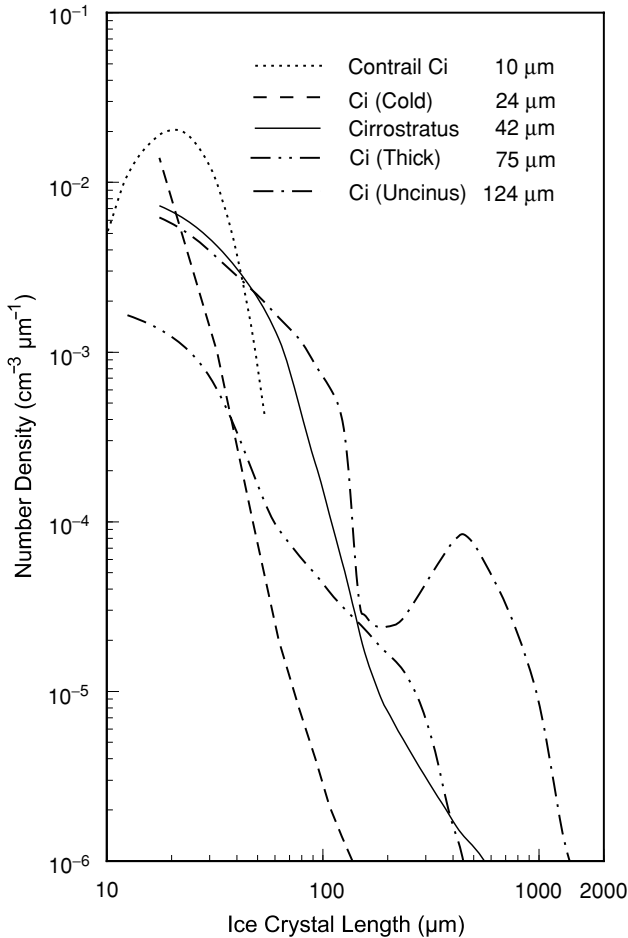
Figure 5.4 Representative ice-crystal size distributions for midlatitude cirrus clouds covering a range of mean effective ice-crystal sizes from 10 μm (Contrail), 24 μm (Cold), 42 μm (Cs), 75 μm (Thick), to 124 μm (Uncinus) (data from Heymsfield and Platt, 1984; First ISCCP Regional Experiment, 1991; Liou *et al.,* 1998).

sizes are associated with warmer temperatures and/or the developing stage of clouds associated with convection (Rolland *et al.,* 2000).

Ice-crystal shape and size data in arctic cirrus and stratiform clouds associated with frontal systems was collected during the First ISCCP Regional Experiment Arctic Cloud Experiment. Ice crystals with sizes larger than 40 μm were collected, and inspection of their shapes shows a combination of pristine and irregular types, including solid and hollow columns, prisms, plates, aggregates, and branched particles (Korolev *et al.,* 1999). The extensive collection of ice particles at a surface station in the Antarctic illustrated the prevalence of long needle ice-crystal types (Grenfell and

Sizes and Size Parameters for Atmospheric Particulates in Visible, Infrared, and Microwave Wavelengths

Type	Size (a)	Size parameters ($2\pi a/\lambda$)		
		$\lambda_v(0.5\,\mu m)$	$\lambda_{ir}(10\,\mu m)$	$\lambda_m(1\,cm)$
Aerosol (S*/NS†)	$\lesssim 1\,\mu m$	1.26×10^1	6.3×10^{-1}	6.3×10^{-4}
Water droplet (S)	$\sim 10\,\mu m$	1.26×10^2	6.3×10^0	6.3×10^{-3}
Ice crystal (NS)	$\sim 10^2\,\mu m$	1.26×10^3	6.3×10^1	6.3×10^{-2}
Raindrop (NS)	$\sim 1\,mm$	1.26×10^4	6.3×10^2	6.3×10^{-1}
Snowflake (hailstone) (NS)	$\sim 1\,cm$	1.26×10^5	6.3×10^3	6.3×10^0

*Spherical; †nonspherical.

Warren, 1999). Clearly, then, ice crystals vary substantially in size and shape from the tropics to midlatitudes to the polar regions.

In Table 5.1, we summarize typical sizes of atmospheric particulates in terms of an equivalent radius (for nonspherical particles) and size parameters (Section 1.1.4) with respect to representative wavelengths in the solar, infrared, and microwave regions. Except in the microwave region, all the particle sizes are comparable to or larger than solar and infrared wavelengths. In these cases, the dipole mode of the electric field, which leads to the development of the Rayleigh scattering theory (Section 3.3.1), is not applicable.

In Section 3.3.2, we briefly introduced the Lorenz–Mie theory, the geometric optics approach, and the anomalous diffraction theory of light scattering by particles. This chapter is a continuation of the discussion of atmospheric scattering. We first present a complete description of the Lorenz–Mie theory of light scattering by spheres, which is applicable to spherical aerosols and cloud droplets. Next, we introduce the fundamentals of the geometric optics approach in terms of diffraction and geometric reflection and refraction. Finally, we present contemporary developments in the study of light scattering by nonspherical ice crystals and aerosols.

5.2   Lorenz–Mie Theory of Light Scattering by Spherical Particles

5.2.1   Electromagnetic Wave Equation and Solution

We shall first introduce the fundamental Maxwell equations for the electromagnetic field. The state of excitation that is established in space by the presence of electric charges is said to constitute an electromagnetic field. It is represented by two vectors **E** and **B,** called the electric vector and magnetic induction, respectively. It is necessary to introduce a second set of vectors involving the electric current density **j,** the electric displacement **D,** and the magnetic vector **H** to describe the effects of the electromagnetic field on material objects. At every point where the physical properties of the medium are continuous in its neighborhood, the space and time derivatives of these

five vectors can be related by Maxwell's equations as follows:

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c} \mathbf{j}, \tag{5.2.1}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \tag{5.2.2}$$

$$\nabla \cdot \mathbf{D} = 4\pi\rho, \tag{5.2.3}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{5.2.4}$$

where t denotes time, c the velocity of light, and ρ the density of charge. Equation (5.2.3) may be regarded as a defining equation for the electric charge density ρ, and Eq. (5.2.4) implies that no free magnetic poles exist.

From Eq. (5.2.1), since $\nabla \cdot \nabla \times \mathbf{H} = 0$, a dot-product operation leads to

$$\nabla \cdot \mathbf{j} = -\frac{1}{4\pi} \nabla \cdot \frac{\partial \mathbf{D}}{\partial t}. \tag{5.2.5}$$

Hence, differentiating Eq. (5.2.3) with respect to t leads to

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0. \tag{5.2.6}$$

This is the equation of continuity in an electromagnetic field.

To allow a unique determination of the field vectors from a given distribution of current and charges, the preceding equations must be supplemented by relationships describing the behavior of substances under the influence of the field. These relationships are given by

$$\mathbf{j} = \sigma \mathbf{E}, \tag{5.2.7}$$

$$\mathbf{D} = \varepsilon \mathbf{E}, \tag{5.2.8}$$

$$\mathbf{B} = \mu \mathbf{H}, \tag{5.2.9}$$

where σ is the specific conductivity, ε the permittivity, and μ the magnetic permeability.

We shall now confine our attention to the field where there are no charges ($\rho = 0$) and current ($|\mathbf{j}| = 0$), and to the medium which is homogeneous so that ε and μ are constants. Thus, the Maxwell equations reduce to

$$\nabla \times \mathbf{H} = \frac{\varepsilon}{c} \frac{\partial \mathbf{E}}{\partial t}, \tag{5.2.10}$$

$$\nabla \times \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t}, \tag{5.2.11}$$

$$\nabla \cdot \mathbf{E} = 0, \tag{5.2.12}$$

$$\nabla \cdot \mathbf{H} = 0. \tag{5.2.13}$$

Equations (5.2.10)–(5.2.13) will be used in the following to derive the electromagnetic wave equation. Note that Eqs. (5.2.12) and (5.2.13) can be obtained directly from Eqs. (5.2.10) and (5.2.11) by carrying out the dot operation.

We shall consider a plane electromagnetic wave in a periodic field with a circular frequency ω so that we may write

$$\mathbf{E} \rightarrow \mathbf{E}e^{i\omega t}, \tag{5.2.14}$$

$$\mathbf{H} \rightarrow \mathbf{H}e^{i\omega t}. \tag{5.2.15}$$

On the basis of these transformations, Eqs. (5.2.10) and (5.2.11) become

$$\nabla \times \mathbf{H} = ikm^2\mathbf{E}, \tag{5.2.16}$$

$$\nabla \times \mathbf{E} = -ik\mathbf{H}, \tag{5.2.17}$$

where $k = 2\pi/\lambda (= \omega/c)$ is the wavenumber denoting the propagation constant in vacuum, λ is the wavelength in vacuum, $m = \sqrt{\varepsilon}$ is the complex refractive index of the medium at the frequency ω, and $\mu \approx 1$ is the permeability of air. Note that the wavenumber k used in this chapter differs from the wavenumber ν (cm^{-1}) defined in Section 1.1 by a factor of 2π.

We then perform the curl operation of Eq. (5.2.17) to obtain

$$\nabla \times \nabla \times \mathbf{E} = -ik\nabla \times \mathbf{H}. \tag{5.2.18}$$

Moreover, by noting that $\nabla \cdot \nabla \times \mathbf{E} = 0$ and that $\nabla \cdot \mathbf{E} = 0$, we get

$$\nabla^2\mathbf{E} = -k^2m^2\mathbf{E}. \tag{5.2.19}$$

In a similar way, from Eqs. (5.2.16) and (5.2.13), we derive the following relationship:

$$\nabla^2\mathbf{H} = -k^2m^2\mathbf{H}. \tag{5.2.20}$$

Equations (5.2.19) and (5.2.20) reveal that the electric vector and magnetic induction in a homogeneous medium satisfy the vector wave equation in the form

$$\nabla^2\mathbf{A} + k^2m^2\mathbf{A} = 0, \tag{5.2.21}$$

where $\mathbf{A}$ can be either $\mathbf{E}$ or $\mathbf{H}.$

Now, if ψ satisfies the scalar wave equation

$$\nabla^2\psi + k^2m^2\psi = 0, \tag{5.2.22}$$

vectors $\mathbf{M}_\psi$ and $\mathbf{N}_\psi$ in spherical coordinates (r, θ, ϕ) defined by

$$\mathbf{M}_\psi = \nabla \times [\mathbf{a}_r(r\psi)] = \left(\mathbf{a}_r\frac{\partial}{\partial r} + \mathbf{a}_\theta\frac{1}{r}\frac{\partial}{\partial \theta} + \mathbf{a}_\phi\frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\right) \times [\mathbf{a}_r(r\psi)]$$

$$= \mathbf{a}_\theta\frac{1}{r\sin\theta}\frac{\partial(r\psi)}{\partial \phi} - \mathbf{a}_\phi\frac{1}{r}\frac{\partial(r\psi)}{\partial \theta}, \tag{5.2.23}$$

$$mk\mathbf{N}_\psi = \nabla \times \mathbf{M}_\psi$$

$$= \mathbf{a}_r r\left[\frac{\partial^2(r\psi)}{\partial r^2} + m^2k^2(r\psi)\right] + \mathbf{a}_\theta\frac{1}{r}\frac{\partial^2(r\psi)}{\partial r\partial \theta} + \mathbf{a}_\phi\frac{1}{r\sin\theta}\frac{\partial^2(r\psi)}{\partial r\partial \phi}$$

$$\tag{5.2.24}$$

satisfy the vector wave equation defined in Eq. (5.2.21) subject to Eq. (5.2.22). The terms $\mathbf{a}_r$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are unit vectors in spherical coordinates. To obtain Eq. (5.2.24), we used Eq. (5.2.29), which is defined in the following.

Assuming that u and v are two independent solutions of the scalar wave equation defined in Eq. (5.2.22), then the electric and magnetic field vectors expressed by

$$\mathbf{E} = \mathbf{M}_v + i\mathbf{N}_u, \tag{5.2.25}$$

$$\mathbf{H} = m(-\mathbf{M}_u + i\mathbf{N}_v) \tag{5.2.26}$$

satisfy Eqs. (5.2.16) and (5.2.17). Employing Eqs. (5.2.23) and (5.2.24), $\mathbf{E}$ and $\mathbf{H}$ can be written explicitly as follows:

$$\mathbf{E} = \mathbf{a}_r \frac{i}{mk} \left[\frac{\partial^2(ru)}{\partial r^2} + m^2 k^2 (ru) \right] + \mathbf{a}_\theta \left[\frac{1}{r \sin\theta} \frac{\partial(rv)}{\partial \phi} + \frac{i}{mkr} \frac{\partial^2(ru)}{\partial r \partial \theta} \right]$$
$$+ \mathbf{a}_\phi \left[-\frac{1}{r} \frac{\partial(rv)}{\partial \theta} + \frac{1}{mkr \sin\theta} \frac{\partial^2(ru)}{\partial r \partial \phi} \right], \tag{5.2.27}$$

$$\mathbf{H} = \mathbf{a}_r \frac{i}{k} \left[\frac{\partial^2(rv)}{\partial r^2} + m^2 k^2 (rv) \right] + \mathbf{a}_\theta \left[-\frac{m}{r \sin\theta} \frac{\partial(ru)}{\partial \phi} + \frac{i}{kr} \frac{\partial^2(rv)}{\partial r \partial \theta} \right]$$
$$+ \mathbf{a}_\phi \left[\frac{m}{r} \frac{\partial(ru)}{\partial \theta} + \frac{i}{kr \sin\theta} \frac{\partial^2(rv)}{\partial r \partial \phi} \right]. \tag{5.2.28}$$

The scalar wave equation defined in Eq. (5.2.22) may be written in spherical coordinates in the form

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial^2 \psi}{\partial \phi^2} + k^2 m^2 \psi = 0. \tag{5.2.29}$$

This equation is separable by defining

$$\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi). \tag{5.2.30}$$

Upon substituting Eq. (5.2.30) into Eq. (5.2.29) and dividing the entire equation by $\psi(r, \theta, \phi)$, we obtain

$$\frac{1}{r^2} \frac{1}{R} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{1}{\Theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial \Theta}{\partial \theta} \right)$$
$$+ \frac{1}{r^2 \sin^2\theta} \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \phi^2} + k^2 m^2 = 0. \tag{5.2.31}$$

We then multiply Eq. (5.2.31) by $r^2 \sin^2 \theta$ to obtain

$$\left[\sin^2\theta \frac{1}{R} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + \sin\theta \frac{1}{\Theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial \Theta}{\partial \theta} \right) \right.$$
$$\left. + k^2 m^2 r^2 \sin^2 \theta \right] + \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \phi^2} = 0. \tag{5.2.32}$$

Since the first three terms in this equation consist of the variables r and θ, but not ϕ, the only possibility that Eq. (5.2.32) can be valid is when

$$\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2} = \text{const} = -\ell^2, \tag{5.2.33}$$

where we set the constant equal to $-\ell^2$ (ℓ denotes an integer) for mathematical convenience. In view of Eqs. (5.2.32) and (5.2.33), it is also clear that

$$\sin^2\theta\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + \sin\theta\frac{1}{\Theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right) + k^2m^2r^2\sin^2\theta - \ell^2 = 0. \tag{5.2.34}$$

On dividing Eq. (5.2.34) by $\sin^2\theta$, we obtain

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right) + k^2m^2r^2 + \frac{1}{\sin\theta}\frac{1}{\Theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Theta}{\partial\theta}\right) - \frac{\ell^2}{\sin^2\theta} = 0. \tag{5.2.35}$$

Thus, we must have

$$\frac{1}{R}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) + k^2m^2r^2 = \text{const} = n(n+1), \tag{5.2.36}$$

$$\frac{1}{\sin\theta}\frac{1}{\Theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) - \frac{\ell^2}{\sin^2\theta} = \text{const} = -n(n+1) \tag{5.2.37}$$

in order to satisfy Eq. (5.2.35), where n is an integer. The selection of the constant here is also for mathematical convenience. Rearranging Eqs. (5.2.33), (5.2.36), and (5.2.37) leads to

$$\frac{d^2(rR)}{dr^2} + \left[k^2m^2 - \frac{n(n+1)}{r^2}\right](rR) = 0, \tag{5.2.38}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \left[n(n+1) - \frac{\ell^2}{\sin^2\theta}\right]\Theta = 0, \tag{5.2.39}$$

$$\frac{d^2\Phi}{d\phi^2} + \ell^2\Phi = 0. \tag{5.2.40}$$

The single-value solution for Eq. (5.2.40) is given by

$$\Phi = a_\ell\cos\ell\phi + b_\ell\sin\ell\phi, \tag{5.2.41}$$

where a_ℓ and b_ℓ are arbitrary constants. Equation (5.2.39) is the well-known equation for spherical harmonics. For convenience we introduce a new variable $\mu = \cos\theta$ so that

$$\frac{d}{d\mu}\left[(1-\mu^2)\frac{d\Theta}{d\mu}\right] + \left[n(n+1) - \frac{\ell^2}{1-\mu^2}\right]\Theta = 0. \tag{5.2.42}$$

The solutions of Eq. (5.2.42) can be expressed by the associated Legendre polynomials (spherical harmonics of the first kind) in the form

$$\Theta = P_n^\ell(\mu) = P_n^\ell(\cos\theta). \tag{5.2.43}$$

Finally, in order to solve the remaining equation (5.2.38), we set

$$kmr = \rho, \qquad R = (1/\sqrt{\rho})Z(\rho), \tag{5.2.44}$$

to obtain

$$\frac{d^2Z}{d\rho^2} + \frac{1}{\rho}\frac{dZ}{d\rho} + \left[1 - \frac{(n+\frac{1}{2})^2}{\rho^2}\right]Z = 0. \tag{5.2.45}$$

The solution of this equation can be expressed by the general cylindrical function of order $n + \frac{1}{2}$ and is given by

$$Z = Z_{n+\frac{1}{2}}(\rho). \tag{5.2.46}$$

Thus, the solution of Eq. (5.2.38) is then

$$R = \frac{1}{\sqrt{kmr}}Z_{n+\frac{1}{2}}(kmr). \tag{5.2.47}$$

Upon combining Eqs. (5.2.41), (5.2.43), and (5.2.47), the elementary wave functions at all points on the surface of a sphere can be expressed by

$$\psi(r, \theta, \phi) = \frac{1}{\sqrt{kmr}}Z_{n+\frac{1}{2}}(kmr)P_n^\ell(\cos\theta)(a_\ell \cos\ell\phi + b_\ell \sin\ell\phi). \tag{5.2.48}$$

Each cylindrical function denoted in Eq. (5.2.47) may be expressed as a linear combination of two cylindrical functions of standard type, i.e., the Bessel function $J_{n+\frac{1}{2}}$ and the Neumann function $N_{n+\frac{1}{2}}$. We define

$$\psi_n(\rho) = \sqrt{\pi\rho/2}J_{n+\frac{1}{2}}(\rho), \qquad \chi_n(\rho) = -\sqrt{\pi\rho/2}N_{n+\frac{1}{2}}(\rho). \tag{5.2.49}$$

The functions ψ_n are regular in every finite domain of the ρ plane including the origin, whereas the functions χ_n have singularities at the origin $\rho = 0$ at which they become infinite. Thus, we may use ψ_n, but not χ_n, to represent the wave inside the sphere. On utilizing the definitions in Eq. (5.2.49), Eq. (5.2.47) can be rewritten in the form

$$rR = c_n\psi_n(kmr) + d_n\chi_n(kmr), \tag{5.2.50}$$

where c_n and d_n are arbitrary constants. Equation (5.2.50) now represents the general solution of Eq. (5.2.38). It follows that the general solution of the scalar wave equation (5.2.29) can then be expressed by

$$r\psi(r, \theta, \phi) = \sum_{n=0}^{\infty}\sum_{\ell=-n}^{n} P_n^\ell(\cos\theta)[c_n\,\psi_n(kmr)$$
$$+ d_n\,\chi_n(kmr)](a_\ell \cos\ell\phi + b_\ell \sin\ell\phi). \tag{5.2.51}$$

Note that the electric and magnetic field vectors of electromagnetic waves can be subsequently derived from Eqs. (5.2.27) and (5.2.28).

Moreover, when $c_n = 1$ and $d_n = i$, we have

$$\psi_n(\rho) + i\chi_n(\rho) = \sqrt{\pi\rho/2}\,H_{n+\frac{1}{2}}^{(2)}(\rho) = \xi_n(\rho), \qquad (5.2.52)$$

where $H_{n+\frac{1}{2}}^{(2)}$ is the half-integral-order Hankel function of the second kind, which has the property of vanishing at infinity in the complex plane. This function is therefore suitable for the representation of the scattered wave.

5.2.2   Formal Scattering Solution

Having solved the vector wave equation, we can now discuss the scattering of a plane wave by a homogeneous sphere. For simplicity, we shall assume that outside the medium is vacuum ($m = 1$), that the material of the sphere has an index of refraction m, and that the incident radiation is linearly polarized. We select the origin of a rectangular system of coordinates at the center of the sphere, with the positive z axis along the direction of propagation of the incident wave. If the amplitude of the incident wave is normalized to unity, the incident electric and magnetic field vectors are given by

$$\mathbf{E}^i = \mathbf{a}_x e^{-ikz}, \qquad \mathbf{H}^i = \mathbf{a}_y e^{-ikz}, \qquad (5.2.53)$$

where $\mathbf{a}_x$ and $\mathbf{a}_y$ are unit vectors along the x and y axes, respectively.

The components of any vector, say $\mathbf{a}$, in the Cartesian system (x,y,z) may be transformed to the spherical polar coordinates (r,θ,ϕ) in the forms

$$x = r\sin\theta\cos\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\theta. \qquad (5.2.54)$$

Based on the geometric relationship shown in Fig. 5.5, we find

$$\mathbf{a}_r = \mathbf{a}_x \sin\theta\cos\phi + \mathbf{a}_y \sin\theta\sin\phi + \mathbf{a}_z \cos\theta,$$

$$\mathbf{a}_\theta = \mathbf{a}_x \cos\theta\cos\phi + \mathbf{a}_y \cos\theta\sin\phi - \mathbf{a}_z \sin\theta,$$

$$\mathbf{a}_\phi = -\mathbf{a}_x \sin\phi + \mathbf{a}_y \cos\phi, \qquad (5.2.55)$$

where $\mathbf{a}_x$, $\mathbf{a}_y$, and $\mathbf{a}_z$ are unit vectors in Cartesian coordinates, while $\mathbf{a}_r$, $\mathbf{a}_\theta$, and $\mathbf{a}_\phi$ are unit vectors in spherical coordinates.

In view of the preceding analysis, the electric and magnetic field vectors of the incident wave may be written in the forms

$$E_r^i = e^{-ikr\cos\theta}\sin\theta\cos\phi,$$

$$E_\theta^i = e^{-ikr\cos\theta}\cos\theta\cos\phi,$$

$$E_\phi^i = -e^{-ikr\cos\theta}\sin\phi. \qquad (5.2.56)$$

$$H_r^i = e^{-ikr\cos\theta}\sin\theta\sin\phi,$$

$$H_\theta^i = e^{-ikr\cos\theta}\cos\theta\sin\phi,$$

$$H_\phi^i = e^{-ikr\cos\theta}\cos\phi. \qquad (5.2.57)$$
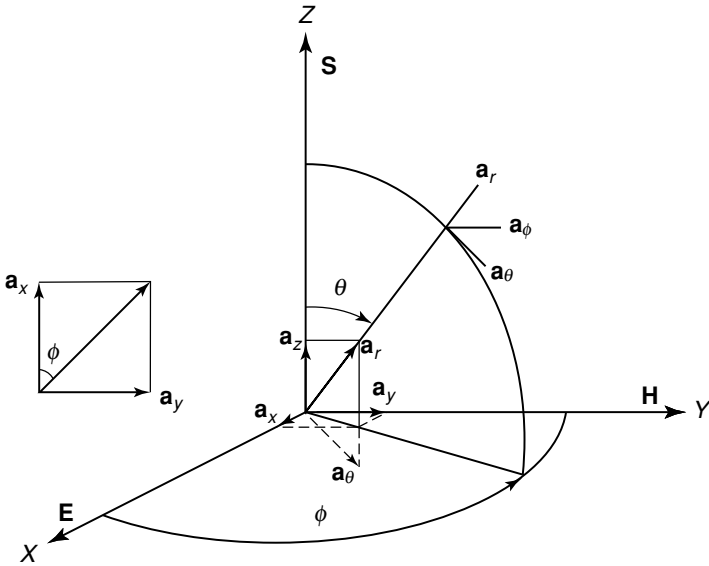
Figure 5.5   Transformation of rectangular to spherical coordinates. **S** is the Poynting vector; **a** is an arbitrary unit vector; θ and ϕ are zenith and azimuthal angles; and **E** and **H** are electric and magnetic vectors placed on the X and Y axes, respectively.

On the basis of Bauer's formula (Watson, 1944), the first factor on the right-hand side of this equation may be expressed in the following differentiable series of Legendre polynomials:

$$e^{-ikr\cos\theta} = \sum_{n=0}^{\infty}(-i)^n(2n+1)\frac{\psi_n(kr)}{kr}P_n(\cos\theta), \qquad (5.2.58)$$

where ψ_n is defined in Eq. (5.2.49). Moreover, we have the following mathematical identities:

$$e^{-ikr\cos\theta}\sin\theta = \frac{1}{ikr}\frac{\partial}{\partial\theta}(e^{-ikr\cos\theta}), \qquad (5.2.59)$$

$$\frac{\partial}{\partial\theta}P_n(\cos\theta) = -P_n^1(\cos\theta), \qquad P_0^1(\cos\theta) = 0. \qquad (5.2.60)$$

Equation (5.2.60) relates the Legendre polynomial P_n to the associated Legendre polynomial P_n^1.

To determine the potentials u and v, we require only one of the components in Eq. (5.2.27). The first of them is ($m = 1$)

$$E_r^i = e^{-ikr\cos\theta}\sin\theta\cos\phi = \frac{i}{k}\left[\frac{\partial^2(ru^i)}{\partial r^2} + k^2(ru^i)\right]. \qquad (5.2.61)$$

In view of Eqs. (5.2.58)–(5.2.60), we have

$$e^{-ikr\cos\theta}\sin\theta\cos\phi = \frac{1}{(kr)^2}\sum_{n=1}^{\infty}(-i)^{n-1}(2n+1)\psi_n(kr)P_n^1(\cos\theta)\cos\phi.$$

$$(5.2.62)$$

Accordingly, we take a trial solution in Eq. (5.2.61) by using an expanding series of a similar form:

$$ru^i = \frac{1}{k}\sum_{n=1}^{\infty}\alpha_n\psi_n(kr)P_n^1(\cos\theta)\cos\phi. \qquad (5.2.63)$$

Upon substituting Eqs. (5.2.62) and (5.2.63) into Eq. (5.2.61) and comparing coefficients, we obtain

$$\alpha_n\left[k^2\psi_n(kr)+\frac{\partial^2\psi_n(kr)}{\partial r^2}\right] = (-i)^n(2n+1)\frac{\psi_n(kr)}{r^2}. \qquad (5.2.64)$$

In Eq. (5.2.50), since $\chi_n(kr)$ become infinite at the origin through which the incident wave must pass, we may let $c_n = 1$ and $d_n = 0$. It follows that

$$\psi_n(kr) = rR \qquad (5.2.65)$$

is a solution of Eq. (5.2.38) (with $m = 1$),

$$\frac{d^2\psi_n}{dr^2}+\left(k^2-\frac{\alpha}{r^2}\right)\psi_n = 0, \qquad (5.2.66)$$

provided that $\alpha = n(n+1)$. Comparing Eq. (5.2.66) to (5.2.64), we find

$$\alpha_n = (-i)^n\frac{2n+1}{n(n+1)}. \qquad (5.2.67)$$

Utilizing similar procedures, v^i can be derived from Eq. (5.2.28). Thus, for incident waves outside the sphere, we have

$$ru^i = \frac{1}{k}\sum_{n=1}^{\infty}(-i)^n\frac{2n+1}{n(n+1)}\psi_n(kr)P_n^1(\cos\theta)\cos\phi,$$

$$rv^i = \frac{1}{k}\sum_{n=1}^{\infty}(-i)^n\frac{2n+1}{n(n+1)}\psi_n(kr)P_n^1(\cos\theta)\sin\phi. \qquad (5.2.68)$$

In order to match u^i and v^i with those of the internal and scattered waves whose potentials have been derived in Eq. (5.2.51), the latter must be expressed in a series of a similar form but with arbitrary coefficients. For internal waves, because the function $\chi_n(kmr)$ becomes infinite at the origin, only the function $\psi_n(kmr)$ may be used. Thus, for internal waves we have

$$ru^t = \frac{1}{mk}\sum_{n=1}^{\infty}(-i)^n\frac{2n+1}{n(n+1)}c_n\psi_n(kmr)P_n^1(\cos\theta)\cos\phi,$$

$$rv^t = \frac{1}{mk}\sum_{n=1}^{\infty}(-i)^n\frac{2n+1}{n(n+1)}d_n\psi_n(kmr)P_n^1(\cos\theta)\sin\phi. \qquad (5.2.69)$$

For scattered waves, these two functions must vanish at infinity. The Hankel functions expressed in Eq. (5.2.52) possess precisely this property. Thus, for scattered waves we have

$$ru^s = \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} a_n \xi_n(kr) P_n^1(\cos\theta) \cos\phi,$$

$$rv^s = \frac{1}{k} \sum_{n=1}^{\infty} (-i)^n \frac{2n+1}{n(n+1)} b_n \xi_n(kr) P_n^1(\cos\theta) \sin\phi. \tag{5.2.70}$$

The coefficients a_n, b_n, c_n, and d_n must be determined from the boundary conditions at the surface of the sphere; that is, the tangential components of **E** and **H** are continuous across the spherical surface $r = a$ such that

$$E_\theta^i + E_\theta^s = E_\theta^t, \qquad H_\theta^i + H_\theta^s = H_\theta^t,$$

$$E_\phi^i + E_\phi^s = E_\phi^t, \qquad H_\phi^i + H_\phi^s = H_\phi^t. \tag{5.2.71}$$

In view of Eqs. (5.2.27), (5.2.28), and (5.2.68)–(5.2.70), apart from common factors and differentiations with respect to θ and ϕ, which are the same for the wave inside and outside the sphere, both of the field components E_θ and E_ϕ contain the expressions v and $\partial(ru)/m\partial r$. It is also clear that the components H_θ and H_ϕ contain mu and $\partial(rv)/\partial r$. Equation (5.2.71) implies that these four expressions must be continuous at $r = a$. Consequently,

$$\frac{\partial}{\partial r}\left[r(u^i + u^s)\right] = \frac{1}{m}\frac{\partial}{\partial r}(ru^t), \qquad u^i + u^s = mu^t,$$

$$\frac{\partial}{\partial r}\left[r(v^i + v^s)\right] = \frac{\partial}{\partial r}(rv^t), \qquad v^i + v^s = v^t. \tag{5.2.72}$$

From Eq. (5.2.72) we have

$$m[\psi_n'(ka) - a_n\xi_n'(ka)] = c_n\psi_n'(kma),$$

$$[\psi_n'(ka) - b_n\xi_n'(ka)] = d_n\psi_n'(kma),$$

$$[\psi_n(ka) - a_n\xi_n(ka)] = c_n\psi_n(kma),$$

$$m[\psi_n(ka) - b_n\xi_n(ka)] = d_n\psi_n(kma), \tag{5.2.73}$$

where the prime denotes differentiation with respect to the argument. On eliminating c_n and d_n, we obtain the coefficients for the scattered waves in the forms

$$a_n = \frac{\psi_n'(y)\psi_n(x) - m\psi_n(y)\psi_n'(x)}{\psi_n'(y)\xi_n(x) - m\psi_n(y)\xi_n'(x)},$$

$$b_n = \frac{m\psi_n'(y)\psi_n(x) - \psi_n(y)\psi_n'(x)}{m\psi_n'(y)\xi_n(x) - \psi_n(y)\xi_n'(x)}, \tag{5.2.74}$$

where $x = ka$, and $y = mx$. The solutions for c_n and d_n contain the same respective denominators as those of a_n and b_n but with $m[\psi_n'(x)\xi_n(x) - \psi_n(x)\xi_n'(x)]$ as a common numerator. At this point, solution of the scattering of electromagnetic waves by a

sphere with a radius of a and an index of refraction of m is complete. The electric and magnetic field vectors expressed in Eqs. (5.2.27) and (5.2.28) at any point inside or outside the sphere are now expressed in terms of the known mathematical functions given by Eqs. (5.2.68)–(5.2.70). For the sake of simplicity, we have assumed up to this point that the suspending medium is a vacuum. We may now let the outside medium and the sphere have refractive indices m_2 (real part) and m_1 (maybe complex), respectively. By replacing m by m_1/m_2 and the wavenumber k by $m_2 k$ (vacuum), the results given in Eq. (5.2.74) can be generalized to cases where a sphere is suspended in a medium.

5.2.3 The Far-Field Solution and Extinction Parameters

In the following, consider the scattered field at very large distances from the sphere. It is noted that for practical applications, light scattering observations are normally made in the far-field zone. In the far field, the Hankel functions denoted in Eq. (5.2.52) reduce to the form

$$\xi_n(kr) \approx i^{n+1} e^{-ikr}, \qquad kr \gg 1. \tag{5.2.75}$$

With this simplification, Eq. (5.2.70) becomes

$$ru^s \approx -\frac{i e^{-ikr} \cos \phi}{k} \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} a_n P_n^1(\cos \theta),$$

$$rv^s \approx -\frac{i e^{-ikr} \sin \phi}{k} \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} b_n P_n^1(\cos \theta). \tag{5.2.76}$$

The three components of the electric and magnetic field vectors in Eqs. (5.2.27) and (5.2.28) are given by

$$E_r^s = H_r^s \approx 0,$$

$$E_\theta^s = H_\phi^s \approx \frac{-i}{kr} e^{-ikr} \cos \phi \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left[a_n \frac{dP_n^1(\cos \theta)}{d\theta} + b_n \frac{P_n^1(\cos \theta)}{\sin \theta} \right],$$

$$E_\phi^s = H_\theta^s \approx \frac{i}{kr} e^{-ikr} \sin \phi \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} \left[a_n \frac{P_n^1(\cos \theta)}{\sin \theta} + b_n \frac{dP_n^1(\cos \theta)}{d\theta} \right].$$

$$\tag{5.2.77}$$

We find that the radial components E_r^s and H_r^s may be neglected in the far-field zone. To simplify Eq. (5.2.77), we define two *scattering functions* in the forms

$$S_1(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [a_n \pi_n(\cos \theta) + b_n \tau_n(\cos \theta)],$$

$$S_2(\theta) = \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} [b_n \pi_n(\cos \theta) + a_n \tau_n(\cos \theta)], \tag{5.2.78}$$

where

$$\pi_n(\cos\theta) = \frac{1}{\sin\theta} P_n^1(\cos\theta),$$

$$\tau_n(\cos\theta) = \frac{d}{d\theta} P_n^1(\cos\theta). \tag{5.2.79}$$

Thus, we may write

$$E_\theta^s = \frac{i}{kr} e^{-ikr} \cos\phi \, S_2(\theta),$$

$$-E_\phi^s = \frac{i}{kr} e^{-ikr} \sin\phi \, S_1(\theta). \tag{5.2.80}$$

These fields represent an outgoing spherical wave where amplitude and the state of polarization are functions of the scattering angle θ. It is convenient to define the *perpendicular* and *parallel* components of the electric field as E_r and E_l, respectively. In reference to Fig. 5.6, the scattered perpendicular and parallel electric fields are given by

$$E_r^s = -E_\phi^s, \qquad E_l^s = E_\theta^s. \tag{5.2.81}$$

Also, the normalized incident electric vector [see Eq. (5.2.53)] may be decomposed into perpendicular and parallel components as follows:

$$E_r^i = e^{-ikz} \sin\phi, \qquad E_l^i = e^{-ikz} \cos\phi. \tag{5.2.82}$$
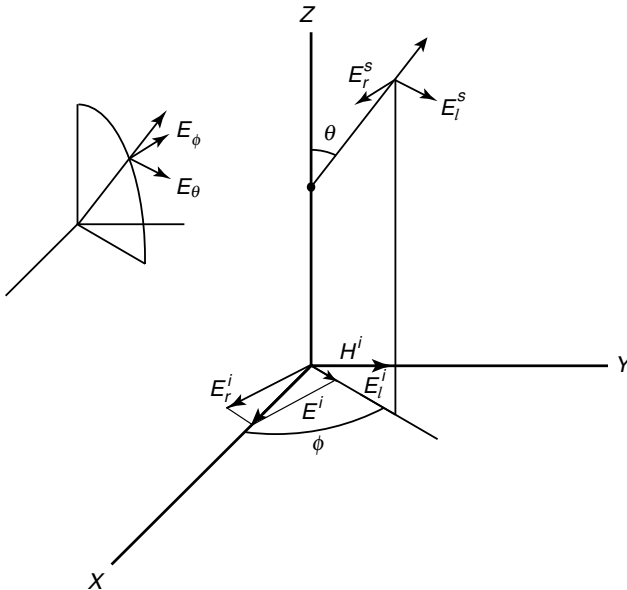
Figure 5.6   Decomposition of the incident (i) and scattered (s) electric vectors into perpendicular (r) and parallel (l) components in Cartesian (x, y, z) and spherical (r, θ, ϕ) coordinates.

Equation (5.2.80) can then be expressed by

$$\begin{bmatrix} E_l^s \\ E_r^s \end{bmatrix} = \frac{e^{-ikr + ikz}}{ikr} \begin{bmatrix} S_2(\theta) & 0 \\ 0 & S_1(\theta) \end{bmatrix} \begin{bmatrix} E_l^i \\ E_r^i \end{bmatrix}, \tag{5.2.83}$$

representing the fundamental equation for the study of a light beam scattered by spheres including polarization.

The scattered intensity components in the far-field zone can now be written in terms of the incident intensity components in the form

$$I_l^s = I_l^i \frac{i_2}{k^2 r^2}, \qquad I_r^s = I_r^i \frac{i_1}{k^2 r^2}, \tag{5.2.84}$$

where

$$i_1(\theta) = |S_1(\theta)|^2, \qquad i_2(\theta) = |S_2(\theta)|^2, \tag{5.2.85}$$

referred to as the *intensity functions* for the perpendicular and parallel components, respectively. Each of these components of scattered light can be thought of as emanating from that component of the incident light beam polarized in the same direction. The computational problem involved in Lorenz–Mie scattering is to evaluate i_1 and i_2 as functions of the scattering angle, index of refraction, and particle size parameter.

In the far-field zone, we wish to evaluate the reduction of the incident energy due to the absorption and scattering of light by a sphere. For this purpose we may consider incident light polarized linearly in the perpendicular direction. Based on Eq. (5.2.83), the scattered electric field is given by

$$E_r^s = \frac{e^{-ikr + ikz}}{ikr} S_1(\theta) E_r^i. \tag{5.2.86}$$

Further, we consider a point (x, y, z) in the forward direction, i.e., $\theta \approx 0$. In the far field, since $x(y) \ll z$, we have

$$r = \left(x^2 + y^2 + z^2\right)^{1/2} \approx z + \frac{x^2 + y^2}{2z}. \tag{5.2.87}$$

Superimposing incident and scattered electric fields in the forward direction yields

$$E_r^i + E_r^s \approx E_r^i \left[1 + \frac{S_1(0)}{ikz} e^{-ik(x^2 + y^2)/2z}\right]. \tag{5.2.88}$$

The far-field combined flux density in the forward direction is then proportional to

$$\left|E_r^i + E_r^s\right|^2 \approx \left|E_r^i\right|^2 \left\{1 + \frac{2}{kz} \text{Re}\left[\frac{S_1(0)}{i} e^{-ik(x^2 + y^2)/2z}\right]\right\}, \tag{5.2.89}$$

where Re[ ] represents the real part of the argument.

We may now integrate the combined flux density over the cross-section area of a sphere whose radius is a to obtain the total power of the combined image as follows:

$$\frac{1}{\left|E_r^i\right|^2} \int\int \left|E_r^i + E_r^s\right|^2 dx\, dy = \pi a^2 + \sigma_e, \qquad (5.2.90)$$

where the first term on the right-hand side of Eq. (5.2.90) represents the cross-section area of the sphere. The physical interpretation of the second term, σ_e, is that the total light received in the forward direction is reduced by the presence of the sphere, and that the amount of this reduction is as if an area equivalent to σ_e of the object had been covered up. The double integral over $dx\, dy$ from which σ_e is defined contains two Fresnel integrals. If the limits are assumed to extend to infinity, we obtain

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-ik(x^2+y^2)/2z} dx\, dy = \frac{2\pi z}{ik}. \qquad (5.2.91)$$

It follows that the extinction cross section is given by

$$\sigma_e = (4\pi/k^2)\mathrm{Re}[S(0)]. \qquad (5.2.92)$$

In the forward direction, we have

$$S_1(0) = S_2(0) = S(0) = \frac{1}{2}\sum_{n=1}^{\infty}(2n+1)(a_n + b_n). \qquad (5.2.93)$$

There is only one $S(0)$ because of the symmetry of the forward scattering in which the extinction is independent of the state of polarization of the incident light. It should be noted that Eq. (5.2.92) is valid only when the sphere is isotropic and homogeneous. Furthermore, the extinction efficiency of a sphere with a radius of $r = a$ is defined by

$$Q_e = \frac{\sigma_e}{\pi a^2} = \frac{2}{x^2}\sum_{n=1}^{\infty}(2n+1)\mathrm{Re}(a_n + b_n), \qquad (5.2.94)$$

where $x = ka$ is the size parameter defined in Section 1.1.4.

To derive the scattering cross section, we use Eq. (5.2.80). The flux density of the scattered light in an arbitrary direction is given by

$$F(\theta, \phi) = \frac{F_0}{k^2 r^2}\left[i_2(\theta)\cos^2\phi + i_1(\theta)\sin^2\phi\right] \qquad (5.2.95)$$

with F_0 representing the incident flux density. The total flux (or power) of the scattered light is, therefore,

$$f = \int_0^{2\pi}\int_0^{\pi} F(\theta, \phi)r^2 \sin\theta\, d\theta\, d\phi, \qquad (5.2.96)$$

where $\sin\theta\, d\theta\, d\phi$ is the differential solid angle $d\Omega$, and $r^2\, d\Omega$ denotes the differential area. The scattering cross section is then given by

$$\sigma_s = \frac{f}{F_0} = \frac{\pi}{k^2}\int_0^{\pi}[i_1(\theta) + i_2(\theta)]\sin\theta\, d\theta. \qquad (5.2.97)$$

As in the extinction case, we may define the scattering efficiency of a sphere as

$$Q_s = \frac{\sigma_s}{\pi a^2} = \frac{1}{x^2} \int_0^\pi [i_1(\theta) + i_2(\theta)] \sin \theta \, d\theta. \tag{5.2.98}$$

We note the following orthogonal and recurrence properties of the associated Legendre polynomials (see Appendix E):

$$\int_0^\pi \left(\frac{dP_n^1}{d\theta} \frac{dP_m^1}{d\theta} + \frac{1}{\sin^2 \theta} P_n^1 P_m^1 \right) \sin \theta \, d\theta = \begin{cases} 0, & \text{if } n \neq m \\ \dfrac{2n(n+1)}{2n+1} \dfrac{(n+1)!}{(n-1)!}, & \text{if } n = m, \end{cases} \tag{5.2.99}$$

and

$$\int_0^\pi \left(\frac{P_n^1}{\sin \theta} \frac{dP_m^1}{d\theta} + \frac{P_m^1}{\sin \theta} \frac{dP_n^1}{d\theta} \right) \sin \theta \, d\theta = \left[P_n^1(\theta) P_m^1(\theta) \right]_0^\pi = 0. \tag{5.2.100}$$

The scattering efficiency can be evaluated with the help of these two equations to yield

$$Q_s = \frac{2}{x^2} \sum_{n=1}^\infty (2n+1)(|a_n|^2 + |b_n|^2). \tag{5.2.101}$$

Finally, based on the energy conservation principle, the absorption cross section and efficiency of a sphere can be evaluated from the following relationships:

$$\sigma_a = \sigma_e - \sigma_s,$$
$$Q_a = Q_e - Q_s. \tag{5.2.102}$$

For an absorbing sphere, it is convenient to define the index of refraction as $m = m_r - im_i$, with m_r and m_i representing the real and imaginary parts of the refractive index, respectively.

Figure 5.7 shows the scattering efficiency factor Q_s as a function of the size parameter x for a real index of refraction of 1.5 and several values of the imaginary part. For $m_i = 0$, i.e., a perfect reflector, there is no absorption. In this case, $Q_s (= Q_e)$ shows a series of major maxima and minima and ripples. The major maxima and minima are due to interference of light diffracted and transmitted by the sphere, whereas the ripple arises from edge rays that are grazing and traveling around the sphere, spewing off energy in all directions. Q_s (or Q_e) increases rapidly when the size parameter reaches about 5 and approaches an asymptotic value of 2. This implies that a large particle removes from the incident beam exactly twice the amount of light that it can intercept. The light removed from the incident light beam includes the diffracted component, which passes by the particle, plus the light scattered by reflection and refraction within the particle, which will be discussed in the following section. Both the ripples and the major maxima and minima damp out as absorption within the particle increases.
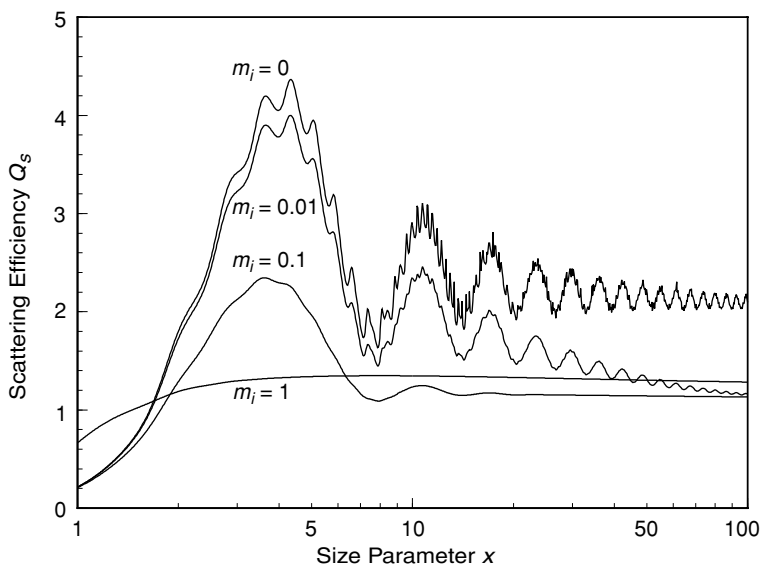
Figure 5.7 Efficiency factor for scattering, Q_s, as a function of the size parameter $x = 2\pi a/\lambda$. The real part of the refractive index used is $m_r = 1.5$, with results shown for four values of the imaginary part m_i.

5.2.4 Scattering Phase Matrix for Spherical Particles

To describe the complete polarization properties of electromagnetic waves, a set of four quantities referred to as the *Stokes parameters*, first introduced by Stokes (1852), has been defined. Because the intensity is proportional to the absolute square of the electric field, we may define, upon neglecting a constant of proportionality, the following four parameters:

$$I = E_l E_l^* + E_r E_r^*, \tag{5.2.103a}$$

$$Q = E_l E_l^* - E_r E_r^*, \tag{5.2.103b}$$

$$U = E_l E_r^* + E_r E_l^*, \tag{5.2.103c}$$

$$V = -i(E_l E_r^* - E_r E_l^*), \tag{5.2.103d}$$

where an asterisk denotes the complex conjugate value and $i = \sqrt{-1}$. I, Q, U, and V are real quantities that satisfy $I^2 = Q^2 + U^2 + V^2$. Further discussion of the Stokes parameters is presented in Section 6.6.1, which deals with multiple scattering of polarized light.

We may express the incident and scattered electric vectors given by Eq. (5.2.83) in terms of their intensity components. Letting the subscript 0 denote the incident

component, we can show that

$$
\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{\mathbf{F}}{k^2 r^2} \begin{bmatrix} I_0 \\ Q_0 \\ U_0 \\ V_0 \end{bmatrix},
\tag{5.2.104}
$$

where the matrix

$$
\mathbf{F} = \begin{bmatrix} \frac{1}{2}(M_2 + M_1) & \frac{1}{2}(M_2 - M_1) & 0 & 0 \\ \frac{1}{2}(M_2 - M_1) & \frac{1}{2}(M_2 + M_1) & 0 & 0 \\ 0 & 0 & S_{21} & -D_{21} \\ 0 & 0 & D_{21} & S_{21} \end{bmatrix},
\tag{5.2.105}
$$

and its components are defined by

$$
M_{1,2} = S_{1,2}(\theta) S_{1,2}^*(\theta),
$$
$$
S_{21} = [S_1(\theta) S_2^*(\theta) + S_2(\theta) S_1^*(\theta)]/2,
$$
$$
-D_{21} = [S_1(\theta) S_2^*(\theta) - S_2(\theta) S_1^*(\theta)] i/2.
\tag{5.2.106}
$$

The matrix $\mathbf{F}$ is called the *transformation matrix* of the light scattering of a single sphere.

In conjunction with the transformation matrix, we may define a parameter referred to as the *scattering phase matrix* such that

$$
\frac{\mathbf{F}(\theta)}{k^2 r^2} = C\mathbf{P}(\theta).
\tag{5.2.107}
$$

The coefficient C can be determined from normalization of the first matrix element in the form

$$
\int_0^{2\pi} \int_0^\pi \frac{P_{11}(\theta)}{4\pi} \sin\theta \, d\theta \, d\phi = 1.
\tag{5.2.108}
$$

It follows that from Eqs. (5.2.107) and (5.2.108), we obtain

$$
C = \frac{1}{2k^2 r^2} \int_0^\pi \frac{1}{2}[M_1(\theta) + M_2(\theta)] \sin\theta \, d\theta = \frac{1}{4k^2 r^2} \int_0^\pi [i_1(\theta) + i_2(\theta)] \sin\theta \, d\theta.
\tag{5.2.109}
$$

Moreover, from the scattering cross section defined in Eq. (5.2.97), we obtain the coefficient in the form

$$
C = \sigma_s / (4\pi r^2).
\tag{5.2.110}
$$

Thus,

$$
\frac{P_{11}}{4\pi} = \frac{1}{2k^2 \sigma_s}(i_1 + i_2) = \frac{1}{2}\left(\frac{P_1}{4\pi} + \frac{P_2}{4\pi}\right),
\tag{5.2.111a}
$$

$$
\frac{P_{12}}{4\pi} = \frac{1}{2k^2 \sigma_s}(i_2 - i_1) = \frac{1}{2}\left(\frac{P_2}{4\pi} - \frac{P_1}{4\pi}\right),
\tag{5.2.111b}
$$

$$\frac{P_{33}}{4\pi} = \frac{1}{2k^2\sigma_s}(i_3 + i_4), \tag{5.2.111c}$$

$$-\frac{P_{34}}{4\pi} = \frac{i}{2k^2\sigma_s}(i_4 - i_3), \tag{5.2.111d}$$

where

$$i_j = S_j S_j^* = |S_j|^2, \qquad j = 1, 2, \tag{5.2.112a}$$

$$i_3 = S_2 S_1^*, \qquad i_4 = S_1 S_2^*. \tag{5.2.112b}$$

The scattering phase matrix for a single homogeneous sphere is then

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & 0 & 0 \\ P_{12} & P_{11} & 0 & 0 \\ 0 & 0 & P_{33} & -P_{34} \\ 0 & 0 & P_{34} & P_{33} \end{bmatrix}. \tag{5.2.113}$$

In general, if no assumption is made about the shape and position of the scatterer, the scattering phase matrix consists of 16 nonzero elements (see Section 5.4.3). For a single sphere, the independent elements reduce to only four. Graphs of P_{11} (the phase function), P_{12}, P_{33}, and P_{34} as functions of the scattering angle for a real part of the refractive index of 1.5 and a size parameter of 60, are shown in Fig. 5.8. The phase functions of a Lorenz–Mie particle are characterized by strong forward scattering, while large backscattering is noticeable as well. These scattering patterns consist of rapid fluctuation due to interference effects, depending on the size parameter. It is clear that the scattering behavior of a Lorenz–Mie particle differs significantly from that of a Rayleigh molecule (Section 3.3.1). Since a spherical particle is symmetrical with respect to the incident light, its scattering pattern is symmetrical in the intervals $(0°, 180°)$ and $(180°, 360°)$. Thus, we may present the Lorenz–Mie scattering phase function in a polar diagram similar to those displayed in Fig. 1.4. The element P_{33} has the same behavior as the phase function P_{11}, but P_{12} and P_{34} show negative values resulting from the differences in the two intensity functions.

The preceding discussion has been concerned with the scattering of electromagnetic waves by a single homogeneous sphere. We shall now extend the discussion to a sample of cloud or aerosol particles so that practical equations for the calculation of extinction parameters and phase functions can be derived. In what follows, particles are assumed to be sufficiently far from each other and that the distance between them is much greater than the incident wavelength. In this case, the scattering by one particle can be treated independently without reference to that of other particles. Consequently, intensities scattered by a sample of particles may be added without regard to the phase of the scattered waves. This particular scattering phenomenon is referred to as *independent scattering* (see also Section 1.1.4). It is in the context of the independent scattering concept that we present the following discussion.
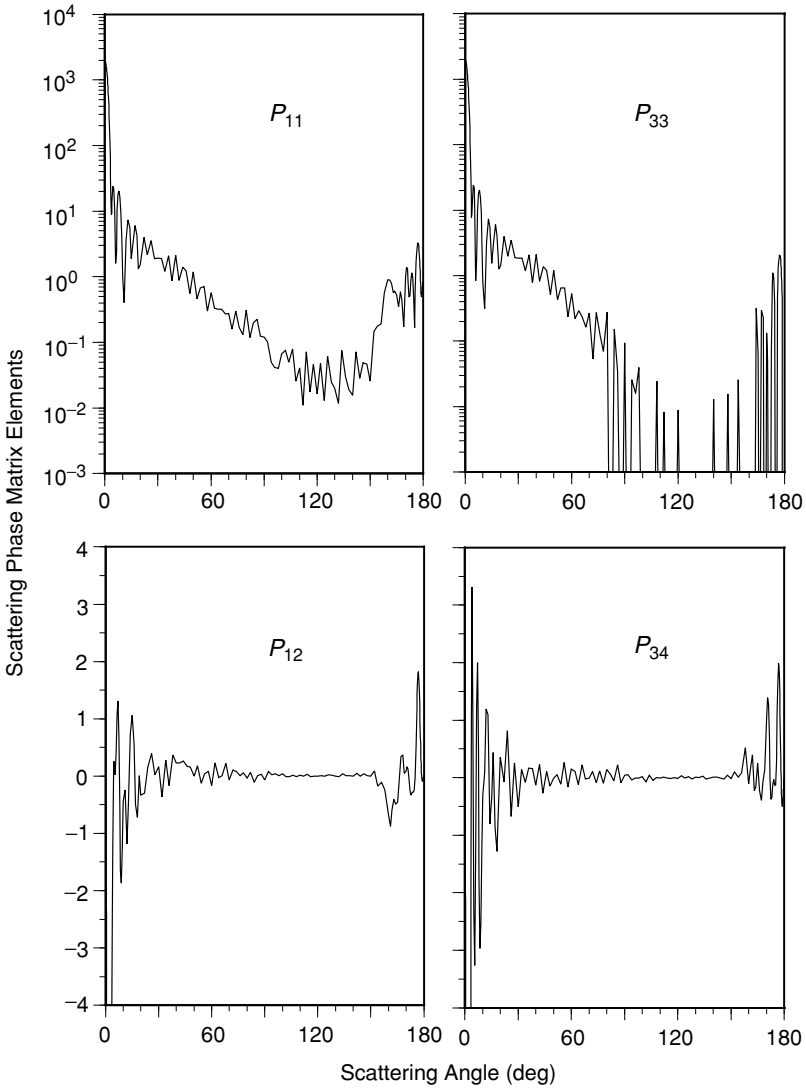
Figure 5.8   Scattering phase matrix elements P_{11}, P_{12}, P_{33}, and P_{34} as functions of the scattering angle for a refractive index $m_r = 1.5$ and a size parameter $x = 60$, computed from the Lorenz–Mie theory.

Consider a sample of cloud particles whose size spectrum can be described by $n(a)$ (in units of cm^{-3} μm^{-1}). Assuming that the size range of particles is from a_1 to a_2, the total number of particles is then given by

$$N = \int_{a_1}^{a_2} n(a)\,da. \tag{5.2.114}$$

In connection with the particle size distribution, the extinction and scattering coefficients (in units of per length) are defined, respectively, as follows:

$$\beta_e = \int_{a_1}^{a_2} \sigma_e(a)n(a)\,da, \tag{5.2.115}$$

$$\beta_s = \int_{a_1}^{a_2} \sigma_s(a)n(a)\,da. \tag{5.2.116}$$

Finally, the single-scattering albedo for a sample of particles is defined by

$$\tilde{\omega} = \beta_s/\beta_e, \tag{5.2.117}$$

which represents the percentage of a light beam that undergoes the scattering event. The remaining part of this section defines the scattering phase matrix for a sample of spherical particles.

Since the scattering phase matrix is a nondimensional physical parameter denoting the scattered intensity and polarization state for a sample of particles in the radius range (a_1, a_2), it is independent of the particle size distribution $n(a)$. Hence, we may rearrange Eq. (5.2.111a) and perform particle size integration to obtain

$$\frac{P_{11}}{4\pi} \int_{a_1}^{a_2} \sigma_s n(a)da = \frac{1}{2k^2} \int_{a_1}^{a_2} [i_1(a) + i_2(a)]n(a)\,da. \tag{5.2.118}$$

From Eq. (5.2.115), we find

$$\frac{P_{11}}{4\pi} = \frac{1}{2k^2\beta_s} \int_{a_1}^{a_2} [i_1(a) + i_2(a)]n(a)\,da. \tag{5.2.119}$$

Similarly, we have

$$\frac{P_{12}}{4\pi} = \frac{1}{2k^2\beta_s} \int_{a_1}^{a_2} [i_2(a) - i_1(a)]n(a)\,da, \tag{5.2.120}$$

$$\frac{P_{33}}{4\pi} = \frac{1}{2k^2\beta_s} \int_{a_1}^{a_2} [i_3(a) + i_4(a)]n(a)\,da, \tag{5.2.121}$$

$$-\frac{P_{34}}{4\pi} = \frac{i}{2k^2\beta_s} \int_{a_1}^{a_2} [i_4(a) - i_3(a)]n(a)\,da. \tag{5.2.122}$$

Note here that the intensity functions $i_j(j = 1, 2, 3, 4)$ are functions of the particle radius a, the index of refraction m, the incident wavelength λ, and the scattering angle θ.

5.3   Geometric Optics

The laws of geometric optics may be used to compute the angular distribution of scattered light when a plane electromagnetic wave is incident on a particle much larger than the wavelength of the incident light. This approach is based on the assumption

that the light may be thought of as consisting of separate localized rays that travel along straight-line paths; it is an asymptotic approach that becomes increasingly accurate in the limit as the size-to-wavelength ratio approaches infinity. Processes involving geometric optics include rays externally reflected by the particle and rays refracted into the particle; the latter rays may be absorbed in the particle or they may emerge from it after possibly suffering several internal reflections. Hence, the total energy scattered and absorbed by the particle is equal to that impinging on the cross section of the particle in reference to the incident beam.

Particles much larger than the incident wavelength also scatter light by means of diffraction, a process by which energy is removed from the light wave passing by the particle. The diffraction is concentrated in a narrow lobe in the forward direction, and like geometrical reflection and refraction, it contains an amount of energy equal to that incident on the cross section of the particle. In the far field, the diffracted component of the scattered light may be approximated by the Fraunhofer diffraction theory. The diffraction pattern depends only upon the shape of the cross section of the particle. The geometric configuration of different contributions to the light scattered by a large sphere was illustrated in Fig. 3.14. In the following, we provide the theoretical foundations of the treatment of geometric optics and diffraction.

5.3.1 Diffraction

The theoretical development of diffraction begins with Babinet's principle, which states that the diffraction pattern in the far field, i.e., Fraunhofer diffraction, from a circular aperture is the same as that from an opaque disk or sphere of the same radius. Let the z axis be in the direction of propagation of the incident light, and let the wave disturbance be sought at a distance P from the geometric aperture A. In reference to Fig. 5.9, the distances from point P to point $O'(x, y)$ on the aperture area and to the origin O are denoted as r and r_0, respectively. Thus, the phase difference of the disturbance at P for waves passing through points O and O' is given by (see Fig. 5.9)

$$\delta = k(r - r_0) = k(x \cos \phi + y \sin \phi) \sin \theta, \tag{5.3.1}$$

where $k = 2\pi/\lambda$, and λ is the wavelength.

In the far field, the light-wave disturbance at P can be derived from the Fraunhofer diffraction theory (Born and Wolf, 1975) and is given by

$$u_p = -\frac{i u_0}{r\lambda} \int \int_A e^{-ikr} dx\, dy, \tag{5.3.2}$$

where u_0 represents the disturbance in the original wave at point 0 on the plane wave front whose wavelength is λ. Using Eq. (5.3.1), we obtain

$$u_p = -\frac{i u_0}{r\lambda} e^{-ikr_0} \int \int_A \exp[-ik(x \cos \phi + y \sin \phi) \sin \theta]\, dx\, dy. \tag{5.3.3}$$
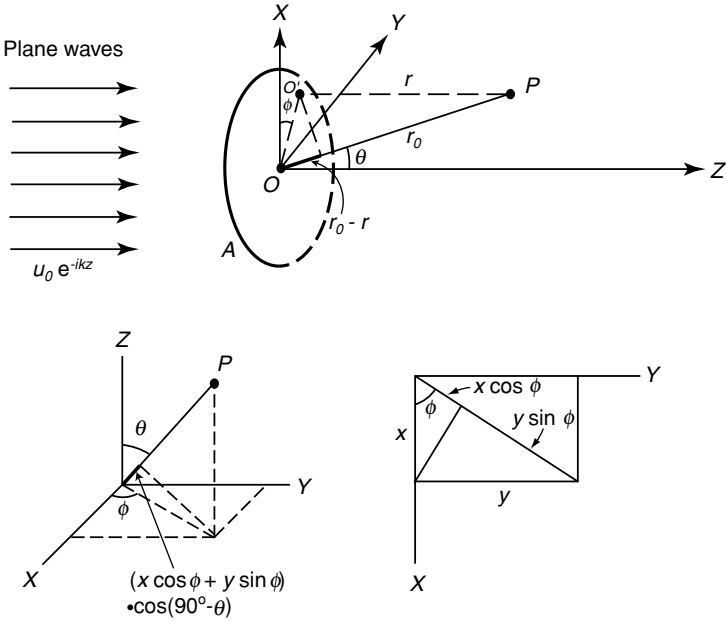
Figure 5.9 Diffraction of plane waves by a circular aperture with a geometrical area A evaluated at position P. The geometrical relationships between the phase difference and the coordinate systems are also shown. All the notations are defined in the text.

For a circular aperture, we may transfer rectangular coordinates to polar coordinates (ρ, ψ) such that $x = \rho \cos \psi$ and $y = \rho \sin \psi$. Thus,

$$u_p = -\frac{i u_0}{r\lambda} e^{-ikr_0} \int_0^a \int_0^{2\pi} \exp[-ik\rho \cos(\psi - \phi) \sin \theta] \rho \, d\rho \, d\psi. \qquad (5.3.4)$$

First, we note that the zero-order Bessel function is defined by

$$J_0(y) = \frac{1}{2\pi} \int_0^{2\pi} e^{iy \cos \alpha} d\alpha. \qquad (5.3.5)$$

It follows that

$$u_p = -\frac{i u_0}{r\lambda} e^{-ikr_0} 2\pi \int_0^a J_0(k\rho \sin \theta) \rho \, d\rho. \qquad (5.3.6)$$

Second, we also note a well-known recurrence relationship involving Bessel functions given by

$$\frac{d}{dy}[y J_1(y)] = y J_0(y), \qquad (5.3.7)$$

from which we have

$$\int_0^y y' J_0(y')dy' = y J_1(y).$$ (5.3.8)

Thus, from Eqs. (5.3.6) and (5.3.8), we obtain

$$u_p = -\frac{iu_0}{r\lambda}e^{-ikr_0}A\frac{2J_1(x\sin\theta)}{x\sin\theta},$$ (5.3.9)

where the geometric shadow area $A = \pi a^2$ and the size parameter $x = ka$. Hence, the scattered intensity in terms of the incident intensity $I_0 = |u_0|^2$ is given by

$$I_p = |u_p|^2 = I_0\frac{i_p}{k^2r^2},$$ (5.3.10)

where the angular intensity function for diffraction analogous to the Lorenz–Mie theory for a single sphere is defined by

$$i_p = \frac{x^4}{4}\left[\frac{2J_1(x\sin\theta)}{x\sin\theta}\right]^2.$$ (5.3.11)

It is clear that diffraction depends only on the particle size parameter and is independent of the index of refraction.

Figure 5.10 shows a plot of $D^2 = [2J_1(y)/y]^2$ versus y. It has a principal maximum of 1 at $y = 0$ (i.e., $\theta = 0$) and oscillates with gradually diminishing amplitude as y
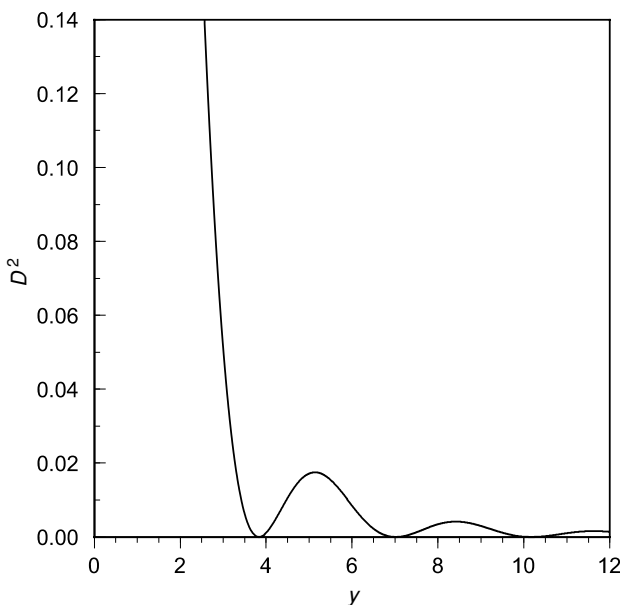
Figure 5.10   The relative diffraction pattern (see text for the definition) as a function of the parameter $y = x\sin\theta$.

increases. When $J_1(y) = 0$, $D^2 = 0$, giving the minima of the diffraction pattern. The positions of the maxima are defined by values of y that satisfy

$$\frac{d}{dy}[J_1(y)/y] = 0. \tag{5.3.12}$$

Table 5.2 lists these maxima and minima. The minima, or dark rings, can be approximated by

$$y = ka \, \sin\theta = (n + 0.22)\pi, \quad n = 1, 2, \ldots, \tag{5.3.13}$$

or

$$\sin\theta = (n + 0.22)\lambda/(2a). \tag{5.3.14}$$

The first maximum at $y = 0$ usually is obscured by the finite size of the source. Thus, the first observable maximum diffraction ring is for $y = 5.136$.

The diffraction theory for a single sphere developed above can be used to explain the optical phenomenon known as a *corona*. A corona is associated with the illumination frequently observed near the sun, the moon, or other luminous objects when they are seen through a mist or thin cloud. They are usually observed in the form of circles, or near circles, concentric with the luminous body. A corona is usually very bright and of a white or bluish-white color with a reddish or brownish tinge. The colors are diluted with a great deal of white light. As many as four corona rings have been recorded, though only the first ring has been frequently observed around the sun and the moon. A corona may be produced by thin clouds when such clouds are composed of particles of almost equal size, a condition called *monodisperse*. Applying the preceding diffraction theory, one may evaluate the angular positions of the corona if the wavelength of the visible sunlight and the mean particle size are known. Based on Eq. (5.3.14), it is evident that red color, having the longer wavelength in the visible, must be seen in the outer ring of the corona with blue and green colors inside the ring. Also note that the angular width of the corona depends on the diameter of the particles involved.

Table 5.2

The First Few Maxima and Minima of the Diffraction Pattern

y	D^2	Max or min
0	1	Max
3.832	0	Min
5.136	0.0175	Max
7.016	0	Min
8.417	0.0042	Max
10.174	0	Min
11.620	0.0016	Max

5.3.2   Geometric Reflection and Refraction

When a plane wave impinges on a boundary separating two homogeneous media of different optical properties, it is split into two waves: a transmitted wave entering into the second medium and a reflected wave propagating back into the first medium. From the part of the wave that hits the surface plane, we may isolate a narrow beam that is much smaller than the surface. Such a beam is commonly called a *ray* as in geometric optics. Let v_1 and v_2 be the velocities of propagation in the two media ($v_1 > v_2$), and let θ_i and θ_t be the angles corresponding to the incident and refracted waves. Referring to Fig. 5.11, we find (see also the subsection on geometric optics in Section 3.3.2.)

$$\sin \theta_i / \sin \theta_t = v_1/v_2 = m, \tag{5.3.15}$$

where m is the index of refraction for the second medium with respect to the first medium. This is Snell's law, which relates the incident and refracted angles through the index of refraction.

Let $\mathbf{E}^i$ be the electric vector of the incident field. As shown in Fig. 5.11, the components of the incident electric field vector perpendicular (r) and parallel (l) to the
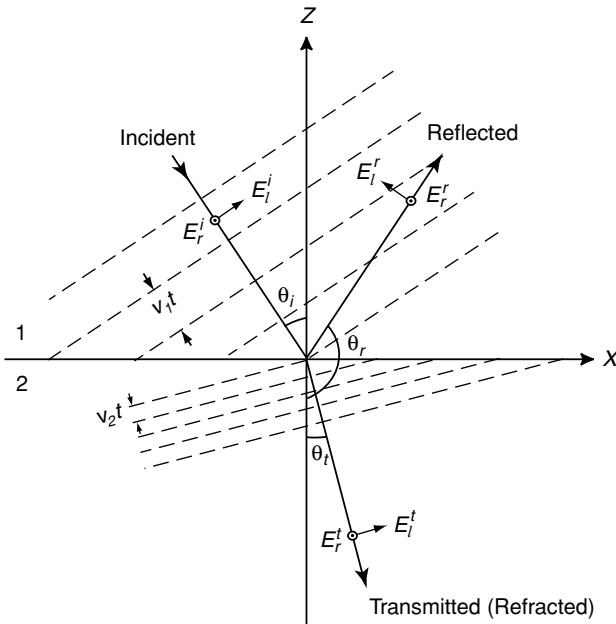
Figure 5.11 Illustration of the reflection and refraction of a plane wave. The choice of the positive directions for the parallel components (l) of the electric vectors is indicated in the diagram. The perpendicular components are at right angles into the plane of reference. v_1t and v_2t denote the distance in media 1 and 2, respectively, and other notations are defined in the text. See also Fig. 3.14a.

plane containing the incident and refracted fields mapped in rectangular coordinates are given by

$$E_x^i = -E_l^i \cos \theta_i,$$
$$E_y^i = E_r^i,$$
$$E_z^i = E_l^i \sin \theta_i. \tag{5.3.16}$$

From the Maxwell equations, the relationship between the electric and magnetic vectors can be shown to be $\mathbf{H} = \sqrt{\varepsilon}\, \mathbf{a} \times \mathbf{E}$, or $\mathbf{E} = -\sqrt{1/\varepsilon}\, \mathbf{a} \times \mathbf{H}$, where $\mathbf{a}$ is a unit vector in the direction of propagation. Thus, the components of the magnetic vector are as follows ($\mu \approx 1$, $m = \sqrt{\varepsilon}$):

$$H_x^i = -E_r^i \cos \theta_i m_1,$$
$$H_y^i = -E_l^i m_1,$$
$$H_z^i = E_r^i \sin \theta_i m_1, \tag{5.3.17}$$

where m_1 is the refractive index of the first medium with respect to vacuum.

Likewise, if $\mathbf{E}^t$ and $\mathbf{E}^r$ denote the transmitted (refracted) and reflected electric vectors, respectively, we find the following relationships:

$$E_x^t = -E_l^t \cos \theta_t, \qquad H_x^t = -E_r^t \cos \theta_t m_2,$$
$$E_y^t = E_r^t, \qquad H_y^t = -E_l^t m_2,$$
$$E_z^t = E_l^t \sin \theta_t, \qquad H_z^t = E_r^t \sin \theta_t m_2, \tag{5.3.18}$$

$$E_x^r = -E_l^r \cos \theta_r, \qquad H_x^r = -E_r^r \cos \theta_r m_1,$$
$$E_y^r = E_r^r, \qquad H_y^r = -E_l^r m_1,$$
$$E_z^r = E_l^r \sin \theta_r, \qquad H_z^r = E_r^r \sin \theta_r m_1, \tag{5.3.19}$$

where m_2 is the refractive index of the second medium with respect to vacuum, and $\theta_r = 180° - \theta_i$.

The boundary conditions require that the tangential components of $\mathbf{E}$ and $\mathbf{H}$ be continuous. Hence, we must have

$$E_x^i + E_x^r = E_x^t, \qquad H_x^i + H_x^r = H_x^t,$$
$$E_y^i + E_y^r = E_y^t, \qquad H_y^i + H_y^r = H_y^t. \tag{5.3.20}$$

On substituting all the electric and magnetic components into Eq. (5.3.20), we obtain the following four relationships:

$$\cos \theta_i \left(E_l^i - E_l^r \right) = \cos \theta_t E_l^t, \tag{5.3.21a}$$

$$E_r^i + E_r^r = E_r^t, \tag{5.3.21b}$$

$$m_1 \cos\theta_i \left(E_r^i - E_r^r\right) = m_2 \cos\theta_t E_r^t, \qquad (5.3.21c)$$

$$m_1 \left(E_l^i + E_l^r\right) = m_2 E_l^t. \qquad (5.3.21d)$$

On the basis of these equations, solutions of the electric components of the reflected and transmitted waves in terms of the incident wave are given by

$$E_l^r = R_l E_l^i, \quad E_l^t = T_l E_l^i, \quad E_r^r = R_r E_r^i, \quad E_r^t = T_r E_r^i, \qquad (5.3.22)$$

where the amplitude coefficients are

$$R_r = \frac{\cos\theta_i - m\cos\theta_t}{\cos\theta_i + m\cos\theta_t}, \quad R_l = \frac{m\cos\theta_i - \cos\theta_t}{m\cos\theta_i + \cos\theta_t}, \qquad (5.3.23a)$$

$$T_r = \frac{2\cos\theta_i}{\cos\theta_i + m\cos\theta_t}, \quad T_l = \frac{2\cos\theta_i}{m\cos\theta_i + \cos\theta_t}, \qquad (5.3.23b)$$

with $m = m_2/m_1$, the refractive index of the second medium with respect to the first medium. Equations (5.3.23a) and (5.3.23b) are referred to as *Fresnel formulas* (Fresnel, 1823). When absorption is involved, the amplitude coefficients become much more complicated, but they can be derived by means of a straightforward mathematical analysis (see Exercise 5.8). Interested readers may wish to refer to Section 5.4.1 for further discussion.

From the Poynting vector, $\mathbf{S} = (c/4\pi)\,\mathbf{E}\times\mathbf{H}$, and the relation between $\mathbf{E}$ and $\mathbf{H}$ denoted previously, we find that the flux density $|\mathbf{S}| = (c/4\pi)\sqrt{\varepsilon}|\mathbf{E}|^2(\mu = 1)$. Thus, the amount of energy incident, reflected, and transmitted on a unit area of the boundary per unit time is

$$F^i = |\mathbf{S}^i|\cos\theta_i = (c/4\pi)m_1|\mathbf{E}^i|^2\cos\theta_i, \qquad (5.3.24a)$$

$$F^r = |\mathbf{S}^r|\cos\theta_r = (c/4\pi)m_1|\mathbf{E}^r|^2\cos\theta_r, \qquad (5.3.24b)$$

$$F^t = |\mathbf{S}^t|\cos\theta_t = (c/4\pi)m_2|\mathbf{E}^t|^2\cos\theta_t. \qquad (5.3.24c)$$

Therefore, the reflected and transmitted portions of the energy in two polarization components, with respect to the incident energy, are proportional to $R_{r,l}^2$ and $T_{r,l}^2 m\cos\theta_t/\cos\theta_i$, respectively. It can be proven that $R_{r,l}^2 + T_{r,l}^2 m\cos\theta_t/\cos\theta_i = 1$, in fulfillment of the energy conservation principle. It follows that the transmitted (or refracted) parts of the energy can be written simply as $(1 - R_{r,l}^2)$.

Consider now a large sphere and let $p = 0$ for the external reflection, $p = 1$ for two refractions, and $p \geq 2$ for internal reflections. We may define the amplitude coefficients as follows:

$$\varepsilon_1 = R_r \qquad\qquad\qquad \text{for} \quad p = 0,$$

$$\varepsilon_1 = (1 - R_r^2)^{1/2}(-R_r)^{p-1}(1 - R_r^2)^{1/2} \quad \text{for} \quad p \geq 1, \qquad (5.3.25)$$

where $-R_r$ denotes the amplitude coefficient for an internal reflection. These definitions also apply to the index $l(2)$ of the other polarization component.

Next, we discuss the effect of curvature on the reflected and refracted intensity. We consider a finite pencil of light characterized by $d\theta_i$ and $d\phi$, with ϕ being the azimuthal

angle. Let I_0 denote the incident intensity of the light pencil plane-polarized in one of the two main directions. The flux of energy contained in this pencil is, therefore, $I_0 a^2 \cos \theta_i \sin \theta_i d\theta_i d\phi$, where a is the radius of the sphere. This flux of energy is divided by successive reflection and refraction. The emergent pencil spreads into a solid angle, $\sin \theta d\theta d\phi$, at a large distance, r, from the sphere. As a result, the scattered intensity is given by

$$I_r = \frac{\varepsilon_1^2 I_0 a^2 \cos \theta_i \sin \theta_i d\theta_i d\phi}{r^2 \sin \theta d\theta d\phi}. \tag{5.3.26}$$

The pencil of light emergent from the sphere is characterized by a small range $d\theta$ around the scattering angle θ. In reference to Fig. 5.12, the total deviation from the original direction is

$$\theta' = 2(\theta_i - \theta_t) + 2(p - 1)(\pi/2 - \theta_t). \tag{5.3.27a}$$

The scattering angle defined in the interval $(0, \pi)$ may be expressed by

$$\theta' = 2\pi n - q\theta, \tag{5.3.27b}$$

where n is an integer and $q = +1$ or -1. Hence,

$$\frac{d\theta}{d\theta_i} = \left| \frac{d\theta'}{d\theta_i} \right| = 2 - 2p \frac{\cos \theta_i}{m \cos \theta_t}. \tag{5.3.28}$$

Moreover, we define the divergence factor due to the curvature effect in the form

$$D = \frac{\cos \theta_i \sin \theta_i}{\sin \theta d\theta/d\theta_i}. \tag{5.3.29}$$

Thus, from Eq. (5.3.26), which is also valid for index l, we obtain

$$I_{r,l} = \frac{I_0}{k^2 r^2} x^2 \varepsilon_{1,2}^2 D. \tag{5.3.30}$$

In comparison with the Lorenz-Mie scattering theory, we find

$$i_{1,2} = x^2 \varepsilon_{1,2}^2 D. \tag{5.3.31}$$

The rainbow is probably the best-known phenomenon of atmospheric optics. It has inspired art and mythology in all peoples and has been a challenge for mathematical physicists. We see a rainbow in the sky usually on summer afternoons after a rain shower is over. Rainbows are produced by the geometric reflections of the sun's rays within raindrops. The sun's rays undergo minimum deviation within the drops and generate the maximum intensity at a specific angle that is much stronger than that at neighboring angles.

To evaluate the angles at which rainbows are formed, we return to the equation denoting the total deviation from the original direction. The minimum deviation of a bundle of rays may be found by differentiating Eq. (5.3.27a) with respect to the incident angle and setting the result equal to zero. Thus,

$$\frac{d\theta'}{d\theta_i} = 0 = 2\left(1 - p \frac{d\theta_t}{d\theta_i}\right). \tag{5.3.32}$$

(a)

(b)
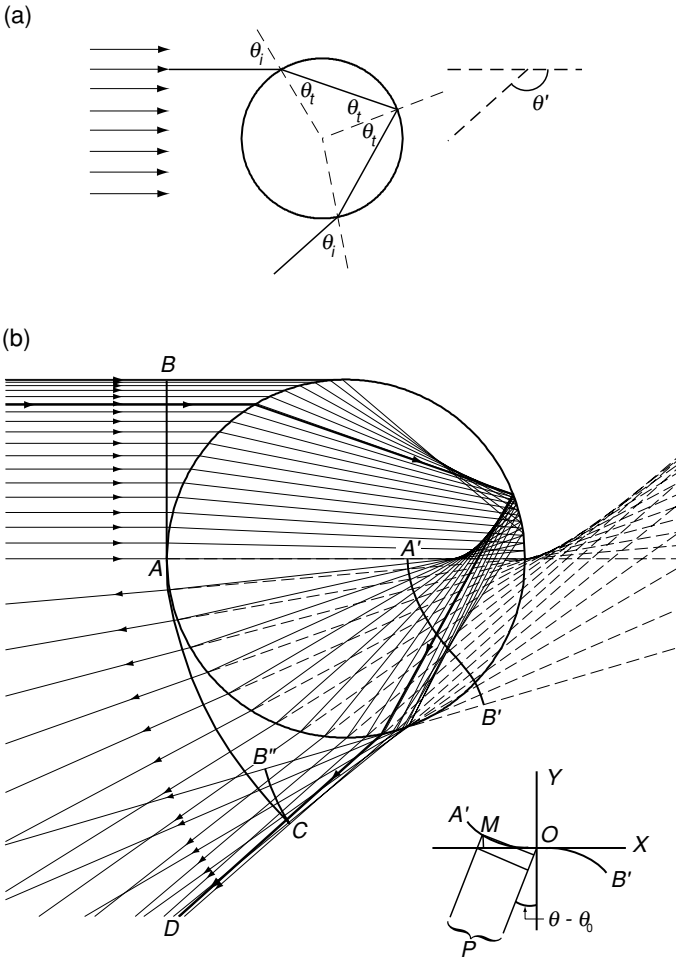
Figure 5.12   (a) Geometrical reflection and refraction by a sphere and the definition of the deviation angle θ'. The incident and refracted angles are denoted by θ_i and θ_t, respectively. (b) Refraction and reflection of light rays with the plane wave front denoted by AB. The heavy line represents the ray that undergoes minimum deviation, referred to as the Descartes ray. The line $A'B'$ denotes the outgoing wave front near the Descartes ray. The line ACB'' denotes the outgoing wave front outside the sphere. The attached diagram shows the phase difference at the wave front $A'B'$, as described in the text (data provided by Yoshihide Takano of the University of California, Los Angeles.)

Furthermore, differentiating Snell's law leads to

$$\frac{d\theta_t}{d\theta_i} = \frac{\cos\theta_i}{m\cos\theta_t}.$$

(5.3.33a)

It follows that

$$p\cos\theta_i = m\cos\theta_t.$$

(5.3.33b)

On eliminating the refracted angle θ_t from Eq. (5.3.33b) and Snell's equation, we obtain the incident angle at which the minimum deviation takes place as follows:

$$\cos^2 \theta_i = (m^2 - 1)/(p^2 - 1), \qquad p \geq 2. \qquad (5.3.34)$$

Once the refractive index has been given, we may evaluate the incident angle corresponding to the minimum deviation for a given p. The refracted angle and the scattering angle may also be subsequently calculated from Snell's law and Eqs. (5.3.27a) and (5.3.27b), respectively.

Table 5.3 lists the incident and scattering angles for rainbows with various indices of refraction. Other than rainbows, minimum deviation can also occur at $\theta = 180°$, known as *glory*. The condition for producing a glory associated with the $p = 2$ rays is that the refractive index must be between $\sqrt{2}$ and 2. Since the refractive index for water drops is about 1.33, glory cannot be produced by means of geometric ray-tracing. The fact that glory is often observed in clouds reveals one of the greatest discrepancies between geometric optics and the Lorenz–Mie theory. It has been suggested that the glory phenomenon is produced by the backscattering from edge rays apparently connected with surface waves generated on a sphere. Primary ($p = 2$) and secondary ($p = 3$) rainbows frequently are observed in the atmosphere. Because of variation of the index of refraction of water with respect to the incident visible wavelengths, various color sequences are produced. Exercise 5.7 requires the calculation of the color position for primary and secondary rainbows. When rainbows are very pronounced, supernumerary rainbows often become visible. They are produced by interference phenomena accompanying refraction of the light in the raindrop and cannot be explained by the geometric ray-tracing approach.

The preceding analysis only defines the angles at which rainbows can be produced. However, geometric optics cannot define rainbow intensity. It is evident from Eq. (5.3.30) that when minimum deviations occur, the intensity approaches infinity. The geometric optics approximation assumes that the wave fronts near any point are sufficiently characterized by their normals and local radii of curvature. Such an approximation breaks down near the rainbow. The next highest approximation can be

Table 5.3
Incident (θ_i) and Scattering (θ) Angles for the Glory and Rainbows

	Glory for $p = 2$		Rainbow for $p = 2$		Rainbow for $p = 3$		Rainbow for $p = 4$		Rainbow for $p = 5$	
m	θ_i	θ	θ_i	θ	θ_i	θ	θ_i	θ	θ_i	θ
1.10	—	—	75°	84°	81°	139°	83°	170°	85°	119°
1.33	—	—	60	137	72	130	77	43	80	42
1.45	87°	180°	53	152	68	102	74	4	78	92
1.50	83	180	50	157	67	93	73	9	77	109
1.54	79	180	47	161	66	86	72	19	76	121
1.75	60	180	34	173	59	58	68	60	73	175
2.00	33	180	0	180	52	35	63	94	69	140

obtained by using a cubic wave front, which leads to the Airy theory. It should be noted, however, that the intensity results derived from geometric optics are accurate in the vicinity of the rainbow angles, and with appropriate extrapolations, intensity patterns at the rainbow angles may be approximated. We shall demonstrate this later when comparing the ray optics and Lorenz–Mie approaches.

In the following, we introduce the Airy theory for rainbows. Since a raindrop is assumed to be spherical, it suffices to use only a single plane containing the center of the drop and the luminous object, and to trace rays incident on one quadrant of the intersection circle. In reference to Fig. 5.12b, let AB be the wave front of parallel incident rays above the ray that passes through the center of the drop, referred to as the axial ray, and consider rays that undergo only one internal reflection. The heavy line denotes the course of the ray of minimum deviation, referred to as the Descartes ray, for a water drop having an index of refraction of 4/3. Because deviations of the incident rays that are between the Descartes and axial rays are greater than that of the Descartes ray, their exits must lie between those two specific rays. Likewise, the rays that are located between the Descartes and edge rays must also have more deviations than the Descartes ray. As a consequence, they emerge from the drop beyond the Descartes ray and eventually come between this ray and the axial ray. The internally reflected light rays are therefore diffuse and weakened, except near the direction of minimum deviation, and are confined to the region between this direction and the axial ray. The wave front is now described by ACB''.

The outgoing wave front near the Descartes ray is in the form of $A'B'$. Through tedious geometric analysis and numerous approximations (e.g., see Humphreys, 1954), it can be shown that the wave front in Cartesian coordinates is related by the cubic equation in the form

$$y = hx^3/(3a^2), \tag{5.3.35a}$$

where a is the radius of the drop and

$$h = \frac{(p^2 + 2p)^2}{(p+1)^2(m^2-1)}\left[\frac{(p+1)^2 - m^2}{m^2 - 1}\right]^{1/2}. \tag{5.3.35b}$$

Equation (5.3.35b) represents a curve closely coincident with the portion of the wave front by which rainbows are produced.

To evaluate the intensity and its variation with angular distance from the ray of minimum deviation, we consider the diagram depicted in Fig. 5.12b, and let O be the point of inflection of the outgoing cubic wave front near a drop. Let θ_0 and θ be the deviation angles for the Descartes ray and the neighboring rays, respectively, and let P be a distant point in the direction $\theta - \theta_0$ from the Descartes ray. We find that the phase difference of the disturbance at P for waves passing through points O and M is given by

$$\delta = k[x\sin(\theta - \theta_0) - y\cos(\theta - \theta_0)]$$
$$= k\left[x\sin(\theta - \theta_0) - \frac{h}{3a^2}x^3\cos(\theta - \theta_0)\right], \tag{5.3.36}$$

where $k = 2\pi/\lambda$. The amplitude of the wave disturbance u_p is then proportional to the integration of all the possible vibrations due to phase differences along the x axis as

$$u_p \sim \int_{-\infty}^{\infty} \exp\left\{-ik\left[x\sin(\theta - \theta_0) - \frac{h}{3a^2}x^3\cos(\theta - \theta_0)\right]\right\} dx. \qquad (5.3.37a)$$

It suffices to use the cosine representation, and if we let

$$(2h/3a^2\lambda)x^3\cos(\theta - \theta_0) = t^3/2, \qquad (2x/\lambda)\sin(\theta - \theta_0) = zt/2, \qquad (5.3.38a)$$

the amplitude is now given by

$$u_p \sim 2\left[\frac{3a^2\lambda}{4h\cos(\theta - \theta_0)}\right]^{1/3} f(z). \qquad (5.3.37b)$$

The intensity $I = u_p^2$, where the *rainbow integral* due to Airy is defined by

$$f(z) = \int_0^{\infty} \cos\frac{\pi}{2}(zt - t^3)dt. \qquad (5.3.38a)$$

From Eq. (5.3.38a), we obtain

$$z^3 = \frac{48a^2}{h\lambda^2}\frac{\sin^3(\theta - \theta_0)}{\cos(\theta - \theta_0)}, \qquad (5.3.38b)$$

which, for small values of $\theta - \theta_0$, is proportional to $(\theta - \theta_0)^3$. Thus, for a small departure from the Descartes ray, we have

$$z \approx \left(\frac{48a^2}{h\lambda^2}\right)^{1/3}(\theta - \theta_0). \qquad (5.3.38c)$$

Table 5.4 lists maximum and minimum values of z and $f^2(z)$ for a given wavelength and drops of a definite size. Note that the first maximum (main rainbow) does not coincide with $z = 0$, the geometric position of the primary rainbow ($\theta = 138.0°$). Also note that the absolute intensity derived from the Airy theory may be obtained by comparing the result from the geometric optics method with the value of $f^2(z)$ for large z.

The validity of the Airy theory for rainbows was investigated by van de Hulst (1957), who indicated that it is an accurate approach when $x \geq 5000$ and the deviation from the rainbow angle $|\varepsilon| \leq 0.5°$. Nussenzveig (1979, and the reference papers therein) presented a more advanced mathematical analysis based on the asymptotic behavior of the exact Lorenz–Mie solution for the physical explanation of rainbows and glory. The two scattering functions denoted in Eq. (5.2.78) can be expressed in the form

$$S_j(x, \theta) = S_{j,0}(x, \theta) + \sum_{p=1}^{N} S_{j,p}(x, \theta) + \text{remainder}, \quad j = 1, 2, \qquad (5.3.39a)$$

Table 5.4

Maxima and Minima of the Rainbow Integral

Maxima			Minima	
Number	z	$f^2(z)$	Number	z
1	1.0845	1.005	1	2.4955
2	3.4669	0.615	2	4.3631
3	5.1446	0.510	3	5.8922
4	6.5782	0.450	4	7.2436
5	7.8685	0.412	5	8.4788
6	9.0599	0.384	6	9.6300
7	10.1774	0.362	7	10.7161
8	11.2364	0.345	8	11.7496
9	12.2475	0.330	9	12.7395
10	13.2185	0.318	10	13.6924

where the first term on the right-hand side of Eq. (5.3.39a) is associated with the direct reflection from the surface of a sphere, and the second term is associated with transmission after $(p - 1)$ internal reflections at the surface. Equation (5.3.39a) is referred to as the *Debye expansion*. The terms in this equation can be transformed in a complex domain based on the Poisson sum formula, referred to as the *Watson transformation,* in the form

$$\sum_{\ell=0}^{\infty} \varphi(\Lambda, \mathbf{r}) = \sum_{m=-\infty}^{\infty} (-1)^m \int_0^{\infty} \varphi(\Lambda, \mathbf{r}) \exp(2im\pi \Lambda) d\Lambda, \qquad (5.3.39b)$$

where φ is a certain function, $\Lambda = \ell + 1/2$, ℓ is the order of expansion in the scattering functions, and $\mathbf{r}$ is a position vector. In this manner, the real rays are mapped into a complex domain, allowing the expression of the scattering functions $S_j (j = 1, 2)$ in terms of the Fresnel coefficients (geometric optics) and the Airy integral, which is called the *complex angular momentum theory* of the rainbow. This theory can be used to generate the intensity peaks produced in rainbow angles, including polarization, associated with internal reflections.

In the preceding discussion, we pointed out that because of its limitations, the principle of geometric optics cannot be utilized to explain the formation of the glory feature, which is often observed on mountaintops and from aircraft, because of the required backscattering geometry. Following the complex angular momentum theory, the tangential rays may be thought of as composing complex angular momentum that can tunnel into the regions not accessible to geometric ray-tracing as forms of surface waves on the edge of a sphere. These rays can then undergo critical refraction to the inside at a point, total reflection at a suitable point inside, and refraction to the outside at an opposite symmetric point. Subsequently, these rays travel as surface waves in the backscattering direction to the observer, giving rise to the glory phenomenon. It turns out that the glory is produced by the ray path undergoing two shortcuts associated with the $p = 2$ Debye term in Eq. (5.3.39b).

5.3.3   Geometric Optics, Lorenz–Mie Theory, and Representative Results

In order to compare the scattering results derived from the geometric optics approximation with those from the Lorenz–Mie theory, we define the gain G relative to an isotropic scatterer. This gain is defined as the ratio of the scattered intensity to the intensity that would be found in any direction if the particle scattered the incoming energy isotropically. Thus, the average gain over the entire solid angle is 1 such that

$$\frac{1}{4\pi} \int_{4\pi} G_{1,2}(\theta)d\Omega = 1. \tag{5.3.40}$$

Isotropic scattering implies that the incident energy $I_0\pi a^2$ to a sphere with radius a is uniformly distributed over its surface $4\pi r^2$. Consequently,

$$I_{1,2}(\theta) = \frac{I_0\pi a^2}{4\pi r^2} G_{1,2}(\theta). \tag{5.3.41}$$

From Eq. (5.2.84), we have

$$G_{1,2} = 4i_{1,2}/x. \tag{5.3.42a}$$

In a similar manner, the gain due to diffraction can be written as follows:

$$G^f = 4i_p/x, \tag{5.3.42b}$$

where i_p is given in Eq. (5.3.11). The gain due to diffraction is the same for the perpendicular and parallel components.

   The total gain caused by diffraction and geometric reflection and refraction can now be expressed by

$$G_{1,2}^t = G^f + \sum_{p=0}^{N} G_{1,2}^p, \tag{5.3.43}$$

where the geometric reflection and refraction are represented by the index p, including external reflection ($p = 0$), refraction ($p = 1$), and internal reflection ($p \geq 2$).

   The foregoing discussion neglects the different contributions to the scattered intensity caused by the phase interferences produced by various ray components. In the case of large particles, phase interferences give rise to rapidly oscillating intensities as a function of the scattering angle as shown in Fig 5.8. However, if particles are randomly located and separated by distances much larger than the incident wavelength, a condition referred to as *polydisperse,* the intensities from separate particles may be added without regard to the phase because the numerous maxima and minima are then lost in the integration over particle size. Hence, it is reasonable to neglect the phase altogether in adding the intensities for diffraction, reflection, and refraction for a sample of large particles of various sizes.

   Figure 5.13 compares phase function P_{11} and the degree of linear polarization computed from the Lorenz–Mie theory to the corresponding results from the geometric optics approach for typical refractive indices of 1.33 and 1.50, representing water drops and aerosols, respectively, in the visible spectrum. The degree of linear
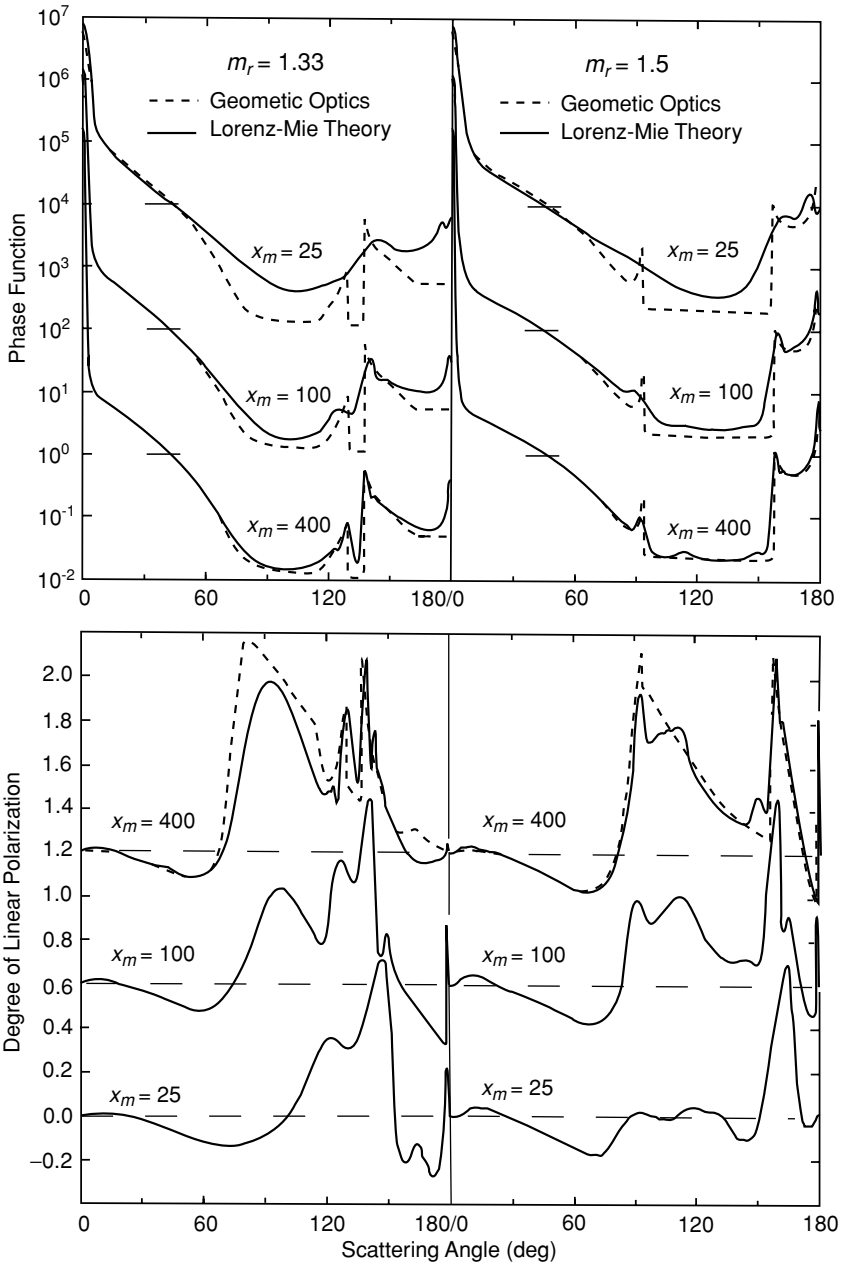
Figure 5.13 Comparison of geometric optics and the Lorenz–Mie theory for phase function P_{11} and degree of linear polarization. Two refractive indices are shown along with three size distributions. The vertical scale applies to the lowermost curves. The other curves are successively displaced upward by factors of 10^2 and shifts of 0.6 for the phase function and the degree of linear polarization, respectively (after Liou and Hansen, 1971).

polarization is defined by $-Q/I$ from Eq. (5.2.104); for incident unpolarized sunlight (see Section 6.6.1 for further discussion), it is given by $-P_{12}/P_{11}$. Lorenz–Mie calculations were made for size parameters of 25, 100, and 400. The droplet size distribution employed in the Lorenz–Mie and diffraction calculations to smooth out rapid oscillating patterns is a gamma function given by

$$n(x) = Cx^6 \exp(-6x/x_m), \tag{5.3.44}$$

where C is an arbitrary constant and x_m is the modal parameter. There is close agreement between geometric optics and the Lorenz–Mie theory when the size parameter is as large as 400. An exception is the glory feature for $m_r = 1.33$, which, as discussed earlier, does not occur in geometric optics results. Most of the discrepancies and their variations with the size parameter can be qualitatively understood in terms of the increasing inapplicability of the localization principle for decreasing size parameters. This causes the light in the individual features to be blurred over a wider range of angles than predicted by geometric optics. The secondary rainbow is quite smooth at $x_m = 100$ and is lost at $x_m = 25$, while the primary rainbow is still visible. The number of rainbows visible in the intensity pattern thus gives some indication of the particle size. The rainbow, in addition to being smoothed out, tends to move away from its geometric optics location as the size parameter decreases. For Lorenz–Mie scattering with $x_m = 400$, the small secondary peaks on the less steep side of the rainbows are supernumerary bows, which are caused by interference phenomena and hence are not rendered by geometric optics, which neglects phase interference effects. There is also a small but noticeable discrepancy in the diffraction peak. The higher value for Lorenz–Mie scattering could be produced by surface waves that scatter in the forward direction. The lower figure compares the degree of linear polarization, which contains much stronger imprints of most of the features occurring in the scattered light, such as rainbows, supernumerary bows, the glory, and external reflections, all of which produce positive polarization.

In Fig. 5.14, we illustrate comparisons of theoretical and experimental scattering phase function and polarization patterns. These scattering patterns are derived from measurements of a dense water cloud in a cold chamber, utilizing a He–Ne (0.6328 μm) laser light. The curves depicted in this figure comprise five successive nephelometer scans (10°–175°) that were each normalized at 10° scattering angles and then averaged, with the standard deviations of the averages shown as vertical bars. The measured cloud droplet size, using a continuous-impactor–replicator device, displayed a modal diameter of 2 μm and a maximum diameter of 10 μm. This size spectrum was fitted with a zeroth order log-normal distribution in the form

$$n(a) = \exp\left[-(\log a - \log a_m)^2/(2\sigma_0^2)\right]\Big/\sqrt{2\pi}\,\sigma_0 a, \tag{5.3.45}$$

where a_m (= 2 μm) denotes the modal diameter, and σ_0 (= 0.275) is the geometric mean standard deviation. Lorenz–Mie scattering calculations were performed employing this size distribution and the results were compared to the measured data.
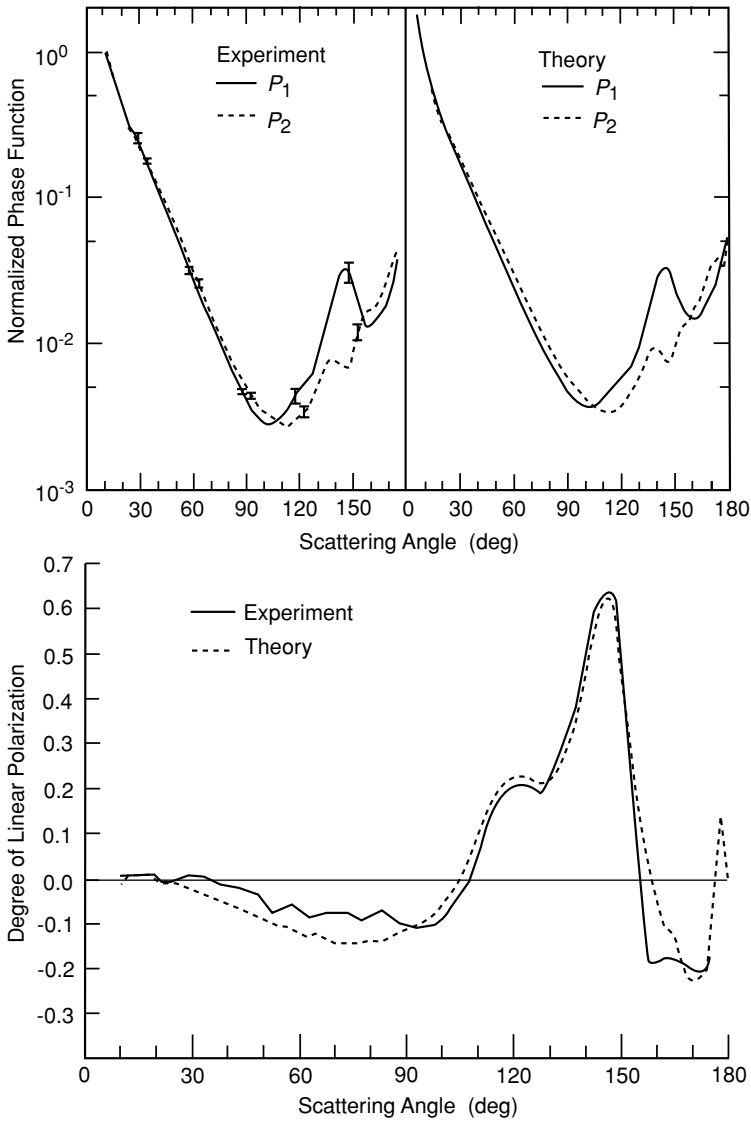
Figure 5.14   Comparison of the normalized phase function components P_1 and P_2 [see Eq. (5.2.111a) for the definitions] and the degree of linear polarization from water cloud experiments and Lorenz–Mie scattering theory, using the same cloud droplet size distribution (after Sassen and Liou, 1979).

The close agreement between the theory and measurements is illustrated in Fig. 5.14. Because of the relatively small size of cloud droplets (the modal size parameter $x_m \approx$ 20), both theory and measurements reveal that the secondary cloudbow is absent, and the primary cloudbow reaches a maximum at about 146°. Close agreement is

also found for the linear polarization pattern, especially in the vicinity of the primary cloudbow where large positive polarization values are shown.

Lorenz–Mie scattering calculations can now routinely be carried out for any size parameters based on more efficient formulations and vector structures of computers (see, e.g., Wiscombe, 1980). In the following, we present some representative results for phase functions, linear polarizations, average extinction cross sections, and single-scattering albedos for water clouds computed from the Lorenz–Mie theory. Using the droplet size distribution denoted in Eq. (5.3.44), the phase functions and the degree of linear polarization patterns for the 0.63, 1.6, 3.7, and 10 μm wavelengths, typical of the image channels, e.g., the Advanced Very High Resolution Radiometer (AVHRR) onboard NOAA satellites, are presented in Fig. 5.15. At 0.63 μm the feature shown at a scattering angle of about 140° is the well-known primary rainbow that arises from rays undergoing one internal reflection in the cloud droplets. This cloudbow decreases in strength at 1.6 μm because of higher absorption and the smaller size parameters involved. At 0.63 μm, the minor maximum located at about 120° is the secondary rainbow, which is produced by rays undergoing two internal reflections. The maximum that occurs at about 180° for 0.63 μm is the glory pattern. The glory feature moves toward smaller scattering angles at 1.6 μm. At 3.7 and 10 μm, the rainbow and glory features vanish because of the significant absorption by cloud droplets. Also
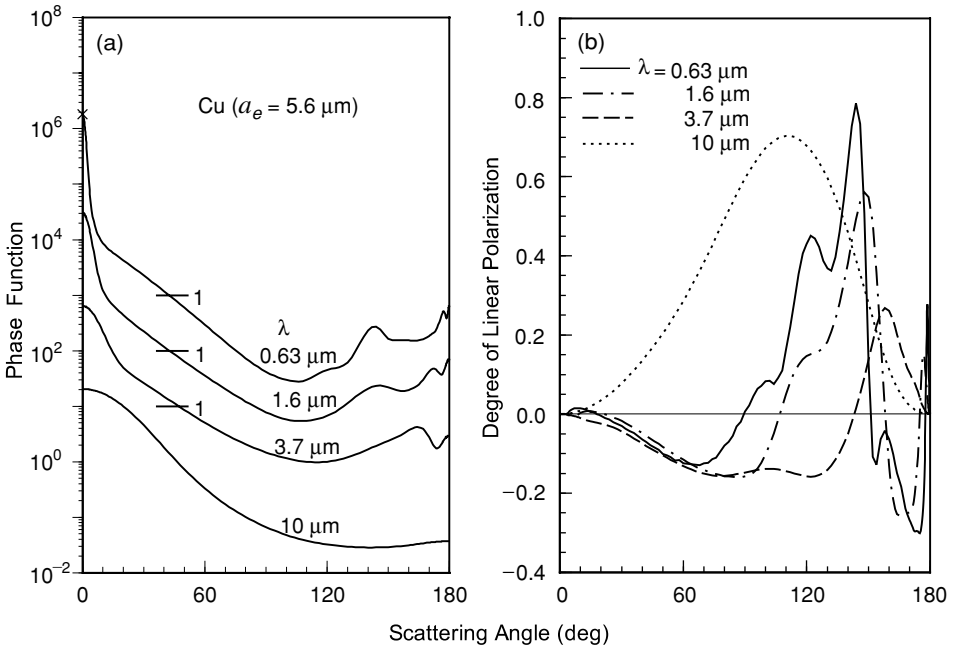
Figure 5.15   Phase function (a) and degree of linear polarization (b) for incident wavelengths of 0.63, 1.6, 3.7, and 10 μm involving the droplet size distribution representative of cumulus and stratus clouds. For the phase function, the vertical scale applies to the lowest curve, while the upper curves are displayed upward by a factor of 10. The symbol $\times$ denotes the diffraction peak for the 0.63 μm wavelength.

noted is the reduction of the Fraunhofer diffraction peak when the size parameter is decreased. The polarization patterns for 0.63 μm contain many strong imprints of scattered light in the directions of the rainbow and glory angles. The positions of the primary rainbows for 1.6 and 3.7 μm shift to larger scattering angles because of the effect of the refractive index and size parameter. For the 10 μm wavelength, scattering is primarily associated with external reflection, which produces a maximum positive polarization at about 90° scattering angle.

Figure 5.16 displays the average extinction cross section, single-scattering albedo, and asymmetry factor for typical cumulus clouds with a droplet size distribution
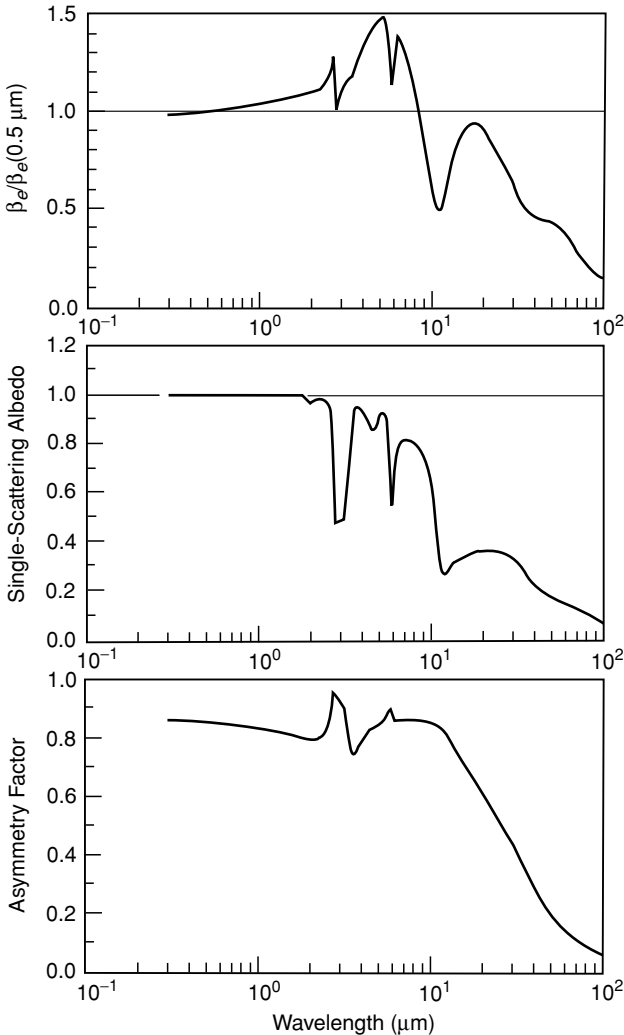
Figure 5.16   Normalized extinction coefficient [$\beta_e(0.5\ \mu m) = 41.8\ \text{km}^{-1}$], single-scattering albedo, and asymmetry factor for the droplet size distribution of typical low cumulus and stratus clouds, covering the wavelengths from 0.3 to 100 μm.

similar to the gamma function denoted in Eq. (5.3.44). The average extinction cross section is normalized with respect to its value at 0.5 μm. In the visible wavelengths, the average size parameter is sufficiently large that the extinction efficiency is approximately equal to 2, based on the optical theorem described previously. Extinction efficiency varies with the size parameter and refractive index. The single-scattering albedo generally resembles the maximum and minimum patterns in the imaginary refractive index for water. Three large minima occur at about 3, 6, and 10 μm. For wavelengths longer than 10 μm, the size parameter effect becomes more important. The asymmetry factor remains about the same, with values of about 0.82–0.86, for wavelengths up to about 10 μm, with exceptions at about 3 and 6 μm; at these wavelengths, diffraction predominates and produces strong forward scattering. Because of the reduction of the size parameter, the asymmetry factor decreases drastically for wavelengths longer than 10 μm.

5.4   Light Scattering by Ice Crystals: A Unified Theory

The scattering of light by spheres can be solved by the exact Lorenz-Mie theory presented in Section 5.2, and computations can be performed for the size parameters that are practical for atmospheric applications. However, an exact solution for the scattering of light by nonspherical ice crystals covering all sizes and shapes that occur in the earth's atmosphere, as displayed in Section 5.1, does not exist in practical terms. It is unlikely that one specific method can be employed to resolve all of the scattering problems associated with nonspherical ice crystals. In the following, we present a unified theory for light scattering by ice crystals by means of a combination of the geometric optics and finite-difference time domain methods.

5.4.1   Geometric Optics for Ice Crystals

As discussed in Section 5.3, the principles of geometric optics are the asymptotic approximations of the fundamental electromagnetic theory, valid for light scattering computations involving a target whose dimension is much larger than the incident wavelength. The geometric optics method has been employed to identify the optical phenomena occurring in the atmosphere, such as halos, arcs, and rainbows. In addition, it is currently the only practical approach for the solution of light scattering by large nonspherical particles. In this section, we present the conventional and improved approaches, the methodology dealing with absorption in the context of geometric ray tracing, and the numerical implementation of the Monte Carlo method.

5.4.1.1   Conventional Approach

When the size of an ice crystal is much larger than the incident wavelength, we may consider a light beam as consisting of a bundle of separate parallel rays that undergo reflection and refraction outside and inside the ice crystal with propagation directions determined by the Snell law at the surface. The total field is assumed to consist of the diffracted rays and the reflected and refracted rays. The diffracted rays pass around

the ice crystal, while the rays impinging on the ice crystal undergo local reflection and refraction, referred to as *Fresnelian interaction*. The energy that is carried by the diffraction and the Fresnelian rays is assumed to be the same as the energy that is intercepted by the ice crystal cross section projected along the incident direction. The intensity of the scattered light within the small scattering-angle interval $\Delta\theta$ in the scattering direction θ in the far field can be computed from the summation of the intensity contributed by each individual ray emerging from the direction between $\theta + \Delta\theta/2$ and $\theta - \Delta\theta/2$. It is usually assumed that the interference is smoothed out when the particles are randomly oriented. In this case, the extinction efficiency (the ratio of the extinction cross section to the average projected area of the particle) of the scatterer is 2. On the basis of Babinet's principle, diffraction by a scatterer may be regarded as that by an opening on an opaque screen perpendicular to the incident light, which has the same geometric shape as the projected cross section of the scatterer. The Fraunhofer diffraction approach can then be employed to compute the diffraction component for hexagonal ice particles (Cai and Liou, 1982).

In the geometric ray-tracing method, the directions of the rays are determined first. In reference to Fig. 5.17, these rays can be defined by the following unit vectors:

$$\mathbf{e}_p^r = \mathbf{x}_p - 2(\mathbf{x}_p \cdot \mathbf{n}_p)\mathbf{n}_p, \quad p = 1, 2, 3, \ldots, \tag{5.4.1a}$$
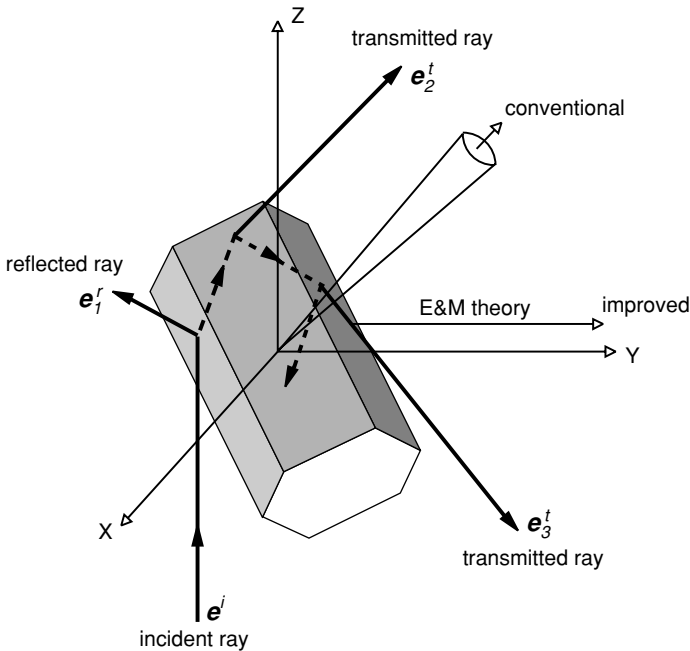
Figure 5.17 Geometry of ray tracing involving an ice crystal in three-dimensional space. Conventional and improved methods are also indicated in the diagram. The vector **e** is a directional cosine associated with incident, reflected, and transmitted rays.

$$\mathbf{e}_p^t = \frac{1}{m_p}\left\{\mathbf{x}_p - (\mathbf{x}_p \cdot \mathbf{n}_p)\mathbf{n}_p - \left[m_p^2 - 1 + (\mathbf{x}_p \cdot \mathbf{n}_p)^2\right]^{1/2}\mathbf{n}_p\right\}, \quad p = 1, 2, 3, \dots ,$$

$$(5.4.1b)$$

$$\mathbf{x}_p = \begin{cases} \mathbf{e}^i & p = 1, \\ \mathbf{e}_1^t, & p = 2, \\ \mathbf{e}_{p-1}^r, & p \geq 3, \end{cases} \tag{5.4.1c}$$

where the superindices *i, r,* and *t* denote incident, reflected, and transmitted rays; $m_p = m$ for $p = 1$ and $m_p = 1/m$ for $p > 1$, with m being the refractive index; and $\mathbf{n}_p$ denotes a unit vector normal to the surface. When $m_p^2 < 1 - (\mathbf{x}_p \cdot \mathbf{n}_p)^2$, total reflection occurs so that the refracted ray does not exist. The electric fields for two polarization components associated with the rays can be computed from the Fresnel formulas [see Eq. (5.3.23)]. By summing the energy of the rays that emerge in a given direction within a preset small scattering-angle interval, the phase function can be obtained for this part of the scattering process. Let the normalized phase functions for the parts of reflection and refraction, and diffraction be P_{11}^r and P_{11}^d, respectively. The normalized phase function is then $P_{11} = (1 - f_d)P_{11}^r + f_d P_{11}^d$, where $f_d = 1/[2\tilde{\omega}(1 - f_\delta)]$ is the delta transmission associated with the $0°$ refraction produced by two parallel prismatic faces, and $\tilde{\omega}$ is the single-scattering albedo, which can be determined from the absorption of individual rays and the constant extinction efficiency.

5.4.1.2   IMPROVED GEOMETRIC OPTICS APPROACH

The laws of geometric optics are applicable to light scattering by a particle whose size is much larger than the incident wavelength so that rays can be localized. In addition to the requirement of the localization principle, the conventional geometric ray-tracing technique assumes that the energy attenuated by the scatterer may be decomposed into equal extinction from diffraction and Fresnel rays. Moreover, the Fraunhofer diffraction formulation used in geometric ray-tracing does not account for the vector property of the electromagnetic field. Finally, direct calculations of the far field by ray-tracing will produce a discontinuous distribution of the scattered energy, such as the delta transmission noted by Takano and Liou (1989a).

To circumvent a number of shortcomings in the conventional geometric optics approach, an improved method has been developed (Yang and Liou, 1995, 1996a). The concept is simple in that the energies determined from geometric ray-tracing at the particle surface are collected and mapped to the far field based on the exact internal geometric ray-tracing. This differs from the conventional approach, which collects energies produced by geometric reflections and refractions directly at the far field through a prescribed solid angle.

The tangential components of the electric and magnetic fields on a surface S that encloses the scatterer can be used to determine the equivalent electric and magnetic currents for the computation of the scattered far field on the basis of the *electromagnetic equivalence theorem* (Schelkunoff, 1943). In this theorem, the electromagnetic field detected by an observer outside the surface would be the same as if the scatterer

were removed and replaced by the equivalent electric and magnetic currents given by

$$\mathbf{J} = \mathbf{n}_s \times \mathbf{H}, \tag{5.4.2a}$$

$$\mathbf{M} = \mathbf{E} \times \mathbf{n}_s, \tag{5.4.2b}$$

where $\mathbf{n}_s$ is the outward unit vector normal to the surface. For the far-field region, we have

$$\mathbf{E}^s(\mathbf{r}) = \frac{e^{ikr}}{ikr}\frac{k^2}{4\pi}\left(\frac{\mathbf{r}}{r}\right)\int\int_S\left[\mathbf{M}(\mathbf{r}') + \left(\frac{\mathbf{r}}{r}\right)\times\mathbf{J}(\mathbf{r}')\right]\exp\left(-ik\mathbf{r}\cdot\frac{\mathbf{r}'}{r}\right)d^2\mathbf{r}', \tag{5.4.3}$$

where $\mathbf{r}/r$ denotes the scattering direction, $\mathbf{r}$ is the reference position vector, $\mathbf{r}'$ is the position vector of the source point, k is the wavenumber, and $i = \sqrt{-1}$. The far-field solution can also be determined by a volume integral involving the internal field.

By means of geometric ray-tracing, the electric field on the surface of a particle can be evaluated after the successive application of Fresnel reflection and refraction coefficients parallel and perpendicular to a defined reference plane at the point of interaction taking into account the path length in the three-dimensional geometry. If an ice crystal shape is of great complexity, as with an aggregate, the surface can be defined as a cubic box so that the computation of the electric field can be conducted on a regularly shaped surface. The electric field of the illuminated and shadowed sides can be defined as follows:

$$\mathbf{E}(\mathbf{r}) = \begin{cases}\mathbf{E}_a(\mathbf{r}) + \mathbf{E}_b(\mathbf{r}), & \mathbf{r} \in \text{ illuminated side,} \\ \mathbf{E}_b(\mathbf{r}), & \mathbf{r} \in \text{ shadowed side,}\end{cases} \tag{5.4.4a}$$

where

$$\mathbf{E}_a(\mathbf{r}) = \mathbf{E}_i(\mathbf{r}) + \mathbf{E}_1^r(\mathbf{r}), \tag{5.4.4b}$$

$$\mathbf{E}_b(\mathbf{r}) = \sum_{p=2}^{\infty}\mathbf{E}_p^t(\mathbf{r}). \tag{5.4.4c}$$

In these equations, $\mathbf{E}_i$ is the incident electric field, $\mathbf{E}_1^r$ is the electric field for external reflection, and $\mathbf{E}_p^t$ are the electric fields produced by two refractions and internal reflections ($p \geq 2$). Because the transverse electromagnetic wave condition is implied in ray-tracing, the magnetic field for each reflection and refraction can be obtained from

$$\mathbf{H}_p^{r,t}(\mathbf{r}) = \mathbf{e}_p^{r,t} \times \mathbf{E}_p^{r,t}(\mathbf{r}), \qquad \text{for } \mathbf{r} \in \text{outside the particle.} \tag{5.4.5}$$

In practice, the mapping of the near field solution to the far field can be done in its entirety for $\mathbf{E}_a$ in Eq. (5.4.4b). However, for $\mathbf{E}_b$ in Eq. (5.4.4c), the mapping is done ray by ray and the results will include the diffraction pattern. Full account of phase interferences is taken in this mapping process in the determination of the phase function.

In accord with the conservation principle for electromagnetic energy concerning the Poynting vector (Jackson, 1975), the extinction and absorption cross sections of the particle can be derived as follows:

$$\sigma_e = \text{Im}\left\{\frac{k}{|\mathbf{E}_i|^2}\int\int\int_V (\varepsilon - 1)\mathbf{E}(\mathbf{r}')\cdot\mathbf{E}_i^*(\mathbf{r}')d^3\mathbf{r}'\right\}, \qquad (5.4.6a)$$

$$\sigma_a = \frac{k}{|\mathbf{E}_i|^2}\int\int\int_V \varepsilon_i\mathbf{E}(\mathbf{r}')\cdot\mathbf{E}^*(\mathbf{r}')d^3\mathbf{r}', \qquad (5.4.6b)$$

where the asterisk denotes the complex conjugate, ε_i is the imaginary part of the permittivity, and V is the particle volume.

Finally, when the ray-tracing technique is applied to obtain the surface field, one must properly account for the area elements from which the externally reflected and transmitted localized waves make a contribution to the surface field. If the cross section of the incident localized wave is $\Delta\sigma_i$, the area of the particle surface for external reflection is

$$\Delta\sigma_1^r = -\Delta\sigma_i(\mathbf{n}_i \cdot \mathbf{e}^i)^{-1}. \qquad (5.4.7a)$$

For the transmitted rays, the area is given by

$$\Delta\sigma_p^t = -\Delta\sigma_i(\mathbf{n}_1 \cdot \mathbf{e}_1^t)\big[(\mathbf{n}_1 \cdot \mathbf{e}^i)(\mathbf{n}_p \cdot \mathbf{e}_p^t)\big]^{-1}, \qquad p = 2, 3, 4, \ldots, \qquad (5.4.7b)$$

where all unit vectors have been defined in Eqs. (5.4.la)–(5.4.lc). The radius of the cross section of a ray should be on the order of k^{-1} so that the phase change over the ray cross section is not significant and permits taking proper account of the phase interference of the localized waves by using the phase information at the centers of the rays. Because the phase variation over the ray cross section can be neglected, the numerical results are not sensitive to the shape of the ray cross sections. We may use a circular shape in the calculations.

5.4.1.3   ABSORPTION EFFECTS IN GEOMETRIC OPTICS

The geometric optics approach that has been used in the past generally assumes that the effect of absorption within the particle on the propagating direction of a ray can be neglected so that the refracted angle and the ray path length can be computed from Snell's law and the geometry of the particle. This approach is correct if absorption is weak, as in the case of ice and water at most solar wavelengths. For strong absorption cases, rays refracted inside the particle are almost totally absorbed, so that the geometric optics method can also be used to compute diffraction and external reflection as long as the particle size is much larger than the incident wavelength. Although the preceding argument is physically correct in the limits of weak and strong absorption, we shall consider the general absorption effect in the context of geometric optics based on the fundamental electromagnetic wave theory. Note that the effect of the complex refractive index on geometric optics has been formulated only for the Fresnel coefficients (Stratton, 1941; Born and Wolf, 1975).
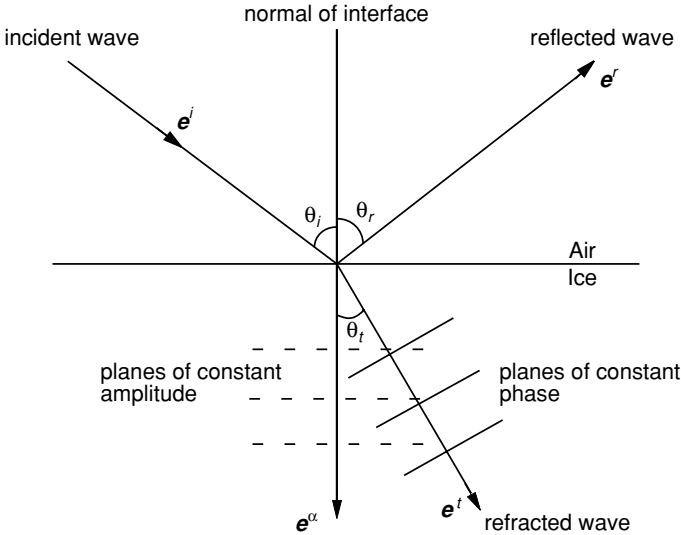
Figure 5.18   Geometric ray-tracing in a medium with absorption. The planes of constant amplitude of the refracted wave are parallel to the interface, whereas the direction of the phase propagation for the inhomogeneous wave inside the medium is determined via Snell's law.

Consider the propagation of the incident wave from air into ice (Fig. 5.18). The wave vectors associated with the incident and reflected waves are real because these waves, which are outside the ice medium, must have the same properties. However, the wave vector of the refracted wave is complex; this is referred to as the inhomogeneity effect. These wave vectors can be presented by

$$\mathbf{k}_i = k \, \mathbf{e}^i, \quad \mathbf{k}_r = k \, \mathbf{e}^r, \quad \mathbf{k}_t = k_t \, \mathbf{e}^t + i k_\alpha \, \mathbf{e}^\alpha, \tag{5.4.8}$$

where $\mathbf{e}^i$, $\mathbf{e}^r$, $\mathbf{e}^t$, and $\mathbf{e}^\alpha$ are unit vectors; the subscripts i, r, and t denote the incident, reflected, and refracted waves respectively; $k = 2\pi/\lambda$, in which λ is the wavelength in air; and k_t and k_α are two real parameters that determine the complex wave vector of the refracted wave. For nonabsorptive cases, k_α is zero. The corresponding electric vectors can be expressed by

$$\mathbf{E}_i(\mathbf{r}, t) = \mathbf{A}_i \exp[i(k\mathbf{r} \cdot \mathbf{e}^i - \omega t)], \tag{5.4.9a}$$

$$\mathbf{E}_r(\mathbf{r}, t) = \mathbf{A}_r \exp[i(k\mathbf{r} \cdot \mathbf{e}^r - \omega t)], \tag{5.4.9b}$$

$$\mathbf{E}_t(\mathbf{r}, t) = \mathbf{A}_t \exp[i(k_t \mathbf{r} \cdot \mathbf{e}^t + i k_\alpha \mathbf{r} \cdot \mathbf{e}^\alpha - \omega t)], \tag{5.4.9c}$$

where $\mathbf{A}_i$, $\mathbf{A}_r$, and $\mathbf{A}_t$ are the amplitudes and ω is the circular frequency. Further, we define the following parameters:

$$N_r = \frac{k_t}{k}, \quad \tilde{N}_i = \frac{k_\alpha}{k}. \tag{5.4.10}$$

At the interface of the two media, at which the position vector is denoted as $\mathbf{r}_s$, the phases of the wave vibration must be the same for the incident, reflected, and refracted waves. Thus, from Eqs. (5.4.8) and (5.4.10), we obtain

$$\mathbf{e}^i \cdot \mathbf{r}_s = \mathbf{e}^r \cdot \mathbf{r}_s = N_r(\mathbf{e}^t \cdot \mathbf{r}_s) + i\tilde{N}_i(\mathbf{e}^\alpha \cdot \mathbf{r}_s). \tag{5.4.11}$$

Because the wave vectors for the incident and reflected waves are real, we must have

$$\mathbf{e}^i \cdot \mathbf{r}_s = \mathbf{e}^r \cdot \mathbf{r}_s = N_r(\mathbf{e}^t \cdot \mathbf{r}_s), \qquad \mathbf{e}^\alpha \cdot \mathbf{r}_s = 0. \tag{5.4.12}$$

Based on the geometry defined by Eq. (5.4.12), a generalized form of the Snell law can be derived and is given by

$$\sin \theta_i = \sin \theta_r, \quad \sin \theta_t = \frac{\sin \theta_i}{N_r}, \tag{5.4.13}$$

where θ_i, θ_r, and θ_t denote the incident, reflected, and refracted angles, respectively (Fig. 5.18). The vector $\mathbf{e}_\alpha$ in Eq. (5.4.12) is normal to the interface of the two media. It follows that the planes of constant amplitude of the refracted wave are parallel to the interface. To determine N_r and N_i, we use the electric field of the refracted wave, which must satisfy the wave equation in the form

$$\nabla^2 \mathbf{E}_t(\mathbf{r}, t) - \frac{(m_r + im_i)^2}{c^2} \frac{\partial^2 \mathbf{E}_t(\mathbf{r}, t)}{\partial t^2} = 0, \tag{5.4.14}$$

where c is the speed of light in vacuum and m_r and m_i are the real and imaginary parts of the refractive index, respectively. Substituting Eq. (5.4.9c) into Eq. (5.4.14) and using Eq. (5.4.10) lead to

$$N_r^2 - \tilde{N}_i^2 = m_r^2 - m_i^2, \quad N_r\tilde{N}_i \cos \theta_t = m_r m_i. \tag{5.4.15}$$

Let $N_i = \tilde{N}_i \cos \theta_t$. Then, from Eqs. (5.4.13) and (5.4.15), we obtain

$$N_r = \frac{\sqrt{2}}{2} \left\{ m_r^2 - m_i^2 + \sin^2 \theta_i + \left[(m_r^2 - m_i^2 - \sin^2 \theta_i)^2 + 4m_r^2 m_i^2 \right]^{1/2} \right\}^{1/2}, \tag{5.4.16a}$$

$$N_i = \frac{m_r m_i}{N_r}. \tag{5.4.16b}$$

These two parameters are referred to as the adjusted real and imaginary refractive indices.

After determining N_r and N_i, the refracted wave given in Eq. (5.4.9c) can be rewritten in the form

$$\mathbf{E}_t(\mathbf{r}, t) = \mathbf{A}_t \exp(-kN_i l_a) \exp\left[i(kN_r \mathbf{e}^t \cdot \mathbf{r} - \omega t)\right], \tag{5.4.17}$$

where $l_a = (\mathbf{e}^\alpha \cdot \mathbf{r}) / \cos \theta_t$ is the distance of the propagation of the refracted wave along the direction $\mathbf{e}^t$. It is clear that the direction of the phase propagation for the inhomogeneous wave inside the medium is determined by N_r via Snell's law, whereas the attenuation of the wave amplitude during the wave propagation is determined by N_i. Consequently, the refracted wave can be traced precisely. The Fresnel reflection and refraction coefficients defined in Eq. (5.3.23) in terms of the adjusted real and imaginary refractive indices can then be written as follows:

$$R_l = \frac{N_r \cos \theta_i - \cos \theta_t}{N_r \cos \theta_i + \cos \theta_t}, \quad T_l = \frac{2 \cos \theta_i}{N_r \cos \theta_i + \cos \theta_t}, \tag{5.4.18a}$$

$$R_r = \frac{\cos \theta_i - N_r \cos \theta_t}{\cos \theta_i + N_r \cos \theta_t}, \quad T_r = \frac{2 \cos \theta_i}{\cos \theta_i + N_r \cos \theta_t}, \tag{5.4.18b}$$

where the subscripts l and r denote the parallel and perpendicular polarized components, respectively.

5.4.1.4   MONTE CARLO METHOD FOR RAY TRACING

Use of the Monte Carlo method in connection with geometric ray-tracing was first developed by Wendling *et al.* (1979) for hexagonal ice columns and plates. Takano and Liou (1995) further innovated a hit-and-miss Monte Carlo method to trace photons in complex ice crystals, including the contributions of absorption and polarization.

Let a bundle of parallel rays, representing a flow of photons, be incident on a crystal from a direction denoted by a set of two angles with respect to the crystal's principal axis. Consider a plane normal to this bundle of incident rays and the geometric shadow of a crystal projected onto this plane. Further, let a rectangle (defined by X and Y) enclose this geometric shadow such that the center of this rectangle coincides with the center of the crystal. One of the sides, X, is parallel to the geometric shadow of the crystal's principal axis. A point (x_i, y_i), is selected inside this rectangle using random numbers, RN, whose range is from 0 to 1 such that

$$x_i = X \left(\text{RN} - \frac{1}{2} \right), \tag{5.4.19a}$$

$$y_i = Y \left(\text{RN} - \frac{1}{2} \right). \tag{5.4.19b}$$

In this manner, x_i is from $-X/2$ to $X/2$, whereas y_i is from $-Y/2$ to $Y/2$. If the point is inside the geometric shadow, it is regarded as an incident point on the crystal. Otherwise, it is disregarded. If there are more than two crystal planes for a photon, the point closer to the light source is regarded as the incident point. The coordinates of an incident point (x_i, y_i) can be transformed into the coordinates (x, y, z) with respect to the body-framed coordinate system using the method described by Takano and Asano

(1983) for efficient geometric ray-tracing procedures. Once the incident coordinates are determined, the photons are traced with a hit-and-miss Monte Carlo method. The Fresnel reflection coefficients, R_l and R_r, are first calculated and compared with a random number, RN. If $(|R_l|^2 + |R_r|^2)/2$ is greater than RN, the photon is reflected. Otherwise, it is transmitted. When a photon traverses a particle, it can be absorbed. One can account for absorption by means of a stochastic procedure. When a photon enters a crystal, an absorption path length, l_a, is generated with a random number such that

$$\text{RN} = \exp(-2kN_i l_a), \quad \text{i.e., } l_a = -\ln\left(\frac{\text{RN}}{2kN_i}\right). \tag{5.4.20}$$

The random number represents the probability of the transmission of a photon. The absorption path length, l_a, denotes the distance traversed by a photon in the crystal before the photon is absorbed. An actual path length, l, between an incident point and the next internal incident point can then be calculated on the basis of Snell's law and the specific ice-crystal geometry. The transmission is then given by $T = \exp(-2kN_i l)$. If $T \leq \text{RN} \leq 1$, then the photons associated with these RNs are absorbed. Similarly, if l is greater than l_a, then the photon is absorbed. Otherwise, it is transmitted without absorption. This procedure is repeated whenever photons travel inside the crystal.

After a photon is transmitted out of the crystal or reflected externally, it can reenter the crystal depending on the crystal's shape. In this case, a new incident direction can be calculated using the direction cosine of the scattered beam. The new incident coordinates can also be determined from the new incident direction and the coordinates of an emergent point of the photon on the crystal surface. The foregoing procedure is repeated until the photon escapes from the crystal. When a photon reenters the crystal, the scattering angle and the scattering matrix are computed with respect to the original incident direction. In the conventional method, the number of scattered photons per unit solid angle, $2\pi \sin\theta\,\Delta\theta$, is counted as the phase function. The single-scattering albedo is obtained from the ratio of the number of scattered photons to the number of incident photons. The Monte Carlo method allows us to treat complicated ice crystals effectively and can be employed in connection with the improved geometric ray-tracing approach.

The surfaces of ice crystals may not be exactly smooth, particularly if they undergo collision processes. Also, a careful examination of some polycrystalline ice crystals reveals rough structures on the surfaces. Halo and arc patterns that are absent from some cirrus clouds could be caused by deviations of the ice-crystal surfaces from defined hexagonal structures. Incorporation of some aspects of ice crystal surface roughness in geometric ray-tracing can be made by following the idea developed by Cox and Munk (1954) for wavy sea surfaces. A rough surface may be thought of as consisting of a number of small facets that are locally planar and randomly tilted from the flat surface. We may use a two-dimensional Gaussian probability function

to define the surface tilt as follows:

$$p(z_x, z_y) = \frac{1}{\pi \sigma^2} \exp\left(-\frac{z_x^2 + z_y^2}{\sigma^2}\right), \tag{5.4.21a}$$

with

$$z_x = \frac{\partial z}{\partial x} = \tan \theta \cos \phi, \tag{5.4.21b}$$

$$z_y = \frac{\partial z}{\partial y} = \tan \theta \sin \phi, \tag{5.4.21c}$$

where z_x and z_y are the slopes defined for a facet of rough surface along two orthogonal directions, θ and ϕ are the local polar angles defining the position of the tilt of the surface facet, and σ is a parameter controlling the degree of roughness. In general, surface roughness of ice particles has the effect of smoothing out the scattering maxima that occur in the phase function.

5.4.2   Introduction to the Finite-Difference Time Domain Method

To circumvent the inherent shortcomings in the geometric optics approach due to the requirement of localization of light rays, we introduce a specific numerical method referred to as the *finite-difference time domain* (FDTD) method for light scattering by small ice crystals. Details of this method will be elaborated on in the discussion of light scattering by nonspherical aerosols in Section 5.5.1. For the continuity of this presentation, however, we shall address the physical fundamentals of the methodology.

The FDTD technique is a direct implementation of the Maxwell curl equations to solve the temporal variation of electromagnetic waves within a finite space containing the scatterer. The three-dimensional scatterer must be discretized into a number of suitably selected rectangular cells, referred to as grid meshes, in which the optical properties are defined. Discretizations are subsequently carried out for the Maxwell curl equations by using the finite difference approximation in both time and space. The propagation and scattering of the excited wave in the time domain can be simulated from the discretized equations by a method of time-marching iterations.

In numerical computations, scattering of the electromagnetic wave by a particle must be confined to finite space. In the application of the FDTD technique, therefore, it is necessary to impose artificial boundaries so that the simulated field within the truncated region would be the same as that in the unbounded case. Implementation of an efficient absorbing boundary condition to suppress spurious reflections is an important aspect of the FDTD method associated with numerical stability and computer time and memory requirements.

The solution of the finite difference analog of the Maxwell curl equations is in the time domain. To obtain the frequency response of the scattering particle, we require an appropriate transformation. The discrete Fourier transform technique can be employed to obtain the frequency spectrum of the time-dependent signals if a Gaussian pulse

is used as an initial excitation. Correct selection of the pulse is required to avoid numerical aliasing and dispersion.

Finally, mapping of the near-field results to the far field must be performed to derive the scattering and polarization properties of the particle. A surface integration or a volume integration technique can be employed to obtain the far-field solution. Numerical calculations of the FDTD method pose some fundamental problems, including the staircasing effect in approximating the particle shape and the absorbing boundary condition used to truncate the computational domain. In Section 5.5.1, we will demonstrate that the FDTD approach can be applied to size parameters smaller than about 20 with adequate accuracy.

5.4.3   Scattering Phase Matrix for Nonspherical Ice Particles

Consider a particle of arbitrary shape and size. The scattered electric field at a distance r from the particle must be related to the two components of the incident electric field (E_l^i, E_r^i). In the far field, the two-by-two amplitude matrix that transforms the incident electric vector into the scattered electric vector may be written in the form

$$\begin{bmatrix} E_l^s \\ E_r^s \end{bmatrix} = \frac{\exp(-ikr + ikz)}{ikr} \begin{bmatrix} S_2 & S_3 \\ S_4 & S_1 \end{bmatrix} \begin{bmatrix} E_l^i \\ E_r^i \end{bmatrix}, \tag{5.4.22}$$

where z is the vertical direction in the Cartesian coordinates and S_j ($j = 1, 2, 3, 4$) are the amplitude functions. For spherical particles, $S_3 = S_4 = 0$, as shown in Eq. (5.2.83).

In terms of the Stokes parameters defined in Eq. (5.2.103), we find

$$\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{\mathbf{F}}{k^2 r^2} \begin{bmatrix} I_0 \\ Q_0 \\ U_0 \\ V_0 \end{bmatrix}, \tag{5.4.23}$$

where the subscript 0 denotes the incident beam and the transformation matrix is given by

$$\mathbf{F} = \begin{bmatrix} \frac{1}{2}(M_2 + M_3 + M_4 + M_1) & \frac{1}{2}(M_2 - M_3 + M_4 - M_1) & S_{23} + S_{41} & -D_{23} - D_{41} \\ \frac{1}{2}(M_2 + M_3 - M_4 - M_1) & \frac{1}{2}(M_2 - M_3 - M_4 + M_1) & S_{23} - S_{41} & -D_{23} + D_{41} \\ S_{24} + S_{31} & S_{24} - S_{31} & S_{21} + S_{34} & -D_{21} + D_{34} \\ D_{24} + D_{31} & D_{24} - D_{31} & D_{21} + D_{34} & S_{21} - S_{34} \end{bmatrix}. \tag{5.4.24}$$

Expressions for the matrix elements in terms of the electric fields can be derived from the definition of the Stokes parameters and are given by

$$M_k = |S_k|^2, \tag{5.4.25a}$$

$$S_{kj} = S_{jk} = (S_j S_k^* + S_k S_j^*)/2, \tag{5.4.25b}$$

$$-D_{kj} = D_{jk} = (S_j S_k^* - S_k S_j^*)i/2, \quad j, k = 1, 2, 3, 4. \tag{5.4.25c}$$

The preceding matrix elements are real numbers. In radiative transfer, it is conventional to define the scattering phase matrix, $\mathbf{P}$, such that its first element is normalized to unity as follows:

$$\int_0^{2\pi} \int_0^\pi \frac{P_{11}(\theta)}{4\pi} \sin\theta \, d\theta \, d\phi = 1, \qquad (5.4.26)$$

where θ and ϕ denote the scattering and azimuthal angles, respectively.

Next, we define the scattering cross section, which represents the amount of incident flux that is removed from the original direction as a result of a single-scattering event such that this flux is distributed isotropically throughout the area of a sphere whose radius is R and whose center is the scatterer. The scattering cross section is related to the first element of the scattering phase matrix in the form

$$\sigma_s = \frac{1}{k^2} \int_0^{2\pi} \int_0^\pi \left(\frac{1}{2} \sum_{k=1}^4 M_k\right) \sin\theta \, d\theta \, d\phi. \qquad (5.4.27)$$

Using Eq. (5.4.27), the scattering phase matrix may be defined in terms of the transformation matrix as follows:

$$\frac{\mathbf{P}}{4\pi} = \frac{1}{\sigma_s k^2}\mathbf{F}. \qquad (5.4.28)$$

The Stokes parameters can then be expressed in the form

$$\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \Omega_{\text{eff}} \frac{\mathbf{P}}{4\pi} \begin{bmatrix} I_0 \\ Q_0 \\ U_0 \\ V_0 \end{bmatrix}, \qquad (5.4.29)$$

where $\Omega_{\text{eff}} = \sigma_s/r^2$, denoting the effective solid angle associated with scattering. If no assumption is made about the shape and position of the scatterer, the scattering phase matrix consists of 16 nonzero elements:

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{bmatrix}. \qquad (5.4.30)$$

The preceding discussion is concerned with the scattering of light by a nonspherical particle. For a group of nonspherical particles, its scattering property is determined by the orientation and size of individual particles with respect to the incident light beam. For continuity of the present discussion, we shall consider a sample of nonspherical particles of the same size randomly oriented in space. The scattering phase matrix can then be expressed by

$$\mathbf{P}(\theta) = \frac{1}{2\pi\sigma_s} \int_0^{2\pi} \int_0^{\pi/2} \mathbf{P}'(\alpha', \gamma')\sigma_s'(\alpha', \gamma') \sin\alpha' d\alpha' d\gamma', \qquad (5.4.31a)$$

where α' and γ' are the orientation angles of a nonspherical particle with respect to the incident light beam (see Fig. 6.17 and Section 6.7.1 for further discussion), and

$\mathbf{P}'$ denotes the scattering phase matrix for a single particle. In this case, the scattering phase matrix is a function of the scattering angle only. Further, the scattering cross section defined in Eq. (5.4.27) for randomly oriented nonspherical particles can be written in the form

$$\sigma_s = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} \sigma_s'(\alpha', \gamma') \sin \alpha' \, d\alpha' \, d\gamma'. \tag{5.4.31b}$$

Again, σ_s' denotes the scattering cross section for a single particle. For a symmetric hexagonal crystal, the azimuthal orientation angle γ' ranges from 0 to $\pi/6$ and an analytic expression can be obtained in this case (Exercise 5.14).

If the scatterers are randomly oriented in space such that every scatterer has a plane of symmetry, the law of reciprocity can be applied (Perrin, 1942; van de Hulst, 1957). Thus, we can reverse the directions of the incident and scattered polarized beams, and achieve the same results. It follows that the amplitude functions (S_3, S_4) in Eq. (5.4.23a) must be equivalent to $(-S_3, -S_4)$. Consequently, the following six relationships between the phase matrix elements are valid: $P_{12} = P_{21}$, $P_{13} = -P_{31}$, $P_{14} = P_{41}$, $P_{23} = -P_{32}$, $P_{24} = P_{42}$, and $P_{34} = -P_{43}$ (Exercise 5.15). Moreover, consider an incident light beam described by negative ellipticity and orientation angles (see Section 6.6.1) and whose Stokes parameters are $(I_0, Q_0, -U_0, -V_0)$. The scattered beam from an isotropic medium composed of randomly oriented scatterers must have the same form for the Stokes parameters: $(I, Q, -U, -V)$. The (I, Q) components by definition are invariant to the change in the incident Stokes parameters from (U, V) to $(-U, -V)$. Using Eq. (5.4.29), four relationships can be derived such that $P_{13} = P_{14} = P_{23} = P_{24} = 0$, and $P_{31} = P_{32} = P_{41} = P_{42} = 0$. It follows that the scattering phase matrix reduces to six independent elements in the form

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & 0 & 0 \\ P_{12} & P_{22} & 0 & 0 \\ 0 & 0 & P_{33} & P_{34} \\ 0 & 0 & -P_{34} & P_{44} \end{bmatrix}. \tag{5.4.32}$$

For spherical scatterers, $S_3 = S_4 = 0$. Thus, $P_{22} = P_{11}$ and $P_{44} = P_{33}$. Consequently, there are only four independent scattering phase matrix elements (Section 5.2.4).

Consider now a spectrum of nonspherical particles randomly oriented in space with a size distribution given by $n(L)$, where L is the major axis of a nonspherical particle. In the case of ice particles or aerosols, we use a parameter referred to as the aspect ratio $L/2a$, where $2a$ denotes the width, to define the particle size. We may employ a mean effective size in a manner defined in Eq. (5.1.1) in association with light scattering calculations. The scattering phase matrix associated with a suitable volume, which contains a spectrum of nonspherical particles defined by a size distribution $n(L)$, can be obtained from

$$\mathbf{P}(\theta) = \int_{L_1}^{L_2} \mathbf{P}(\theta, L) \sigma_s(L) n(L) dL \bigg/ \int_{L_1}^{L_2} \sigma_s(L) n(L) dL, \tag{5.4.33a}$$

where L_1 and L_2 are the lower and upper limits of the particle length, and the scattering cross section for a spectrum of nonspherical particles of different sizes and shapes is

given by

$$\sigma_s = \int_{L_1}^{L_2} \sigma_s(L)n(L)\,dL/N, \tag{5.4.33b}$$

where

$$N = \int_{L_1}^{L_2} n(L)\,dL \tag{5.4.33c}$$

is the total number of nonspherical particles. A similar definition can be derived for the extinction cross section. It follows that a single-scattering albedo for a spectrum of nonspherical particles is given by $\tilde{\omega} = \sigma_s/\sigma_e$.

As stated previously, the condition under which the scattering phase matrix is composed of six independent elements is that the nonspherical particles are randomly oriented in space in such a manner that every one of them has a plane of symmetry based on which the law of reciprocity can be applied: that the incident and scattered beams are reversible. In Section 5.1, we illustrated examples of typical ice-crystal sizes and shapes. Assuming that they are randomly oriented in space, it is very difficult, if not impossible, to prove that every one of these ice crystals has a plane of symmetry that would obey the law of reciprocity. Nevertheless, numerical calculations for light scattering by ice crystals carried out by Takano and Liou (1989a, 1995) based on the geometric ray-tracing approach show that the scattering phase matrix elements P_{13}, P_{14}, P_{23}, P_{24}, P_{31}, P_{32}, P_{41}, and P_{42} are practically zeros and that $P_{12} = P_{21}$ and $P_{34} = -P_{43}$. Light scattering and radiative transfer dealing with the possibility of ice crystals oriented horizontally in the atmosphere are discussed further in Section 6.7.1. In the following, we present the essence of the unified theory for light scattering by ice crystals, comparison to available measurements, and some representative single-scattering results.

5.4.4   Presentation of a Unified Theory for Light Scattering by Ice Crystals

5.4.4.1   The Essence of the Unified Theory

In recognition of the inherent limitations of each methodology for light scattering calculations, it is unlikely that one specific method can be satisfactorily employed to resolve the intricate scattering problems involving nonspherical ice crystals of all size parameters and shapes. However, by unifying the geometric optics and FDTD methods described earlier, calculations of light scattering and absorption by ice crystals covering all sizes and shapes that commonly occur in the atmosphere can be performed with adequate precision. This approach is referred to as the unified theory for light scattering by ice crystals (Liou *et al.,* 2000). In Fig. 5.19, we demonstrate this theory in terms of the extinction efficiency as a function of size parameter for randomly oriented columns of a uniform size. The length-to-width ratio for the column is 6 and the incident wavelength used is 0.63 μm. The FDTD results are presented for size parameters from 1 to 30 for comparison purposes. The results computed from the improved geometric optics method cover size parameters from 1 to 1000. Significant deviations between these two methods occur when the size parameter is
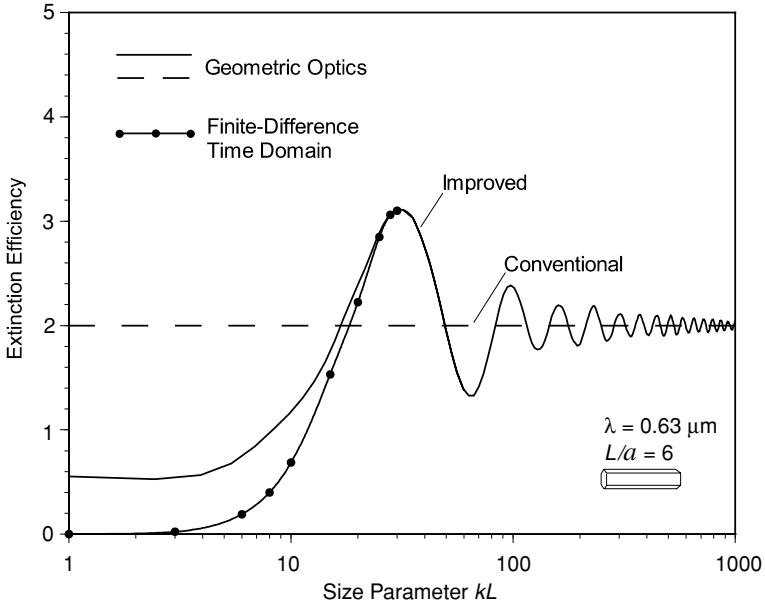
Figure 5.19   Presentation of a unified theory for light scattering by ice crystals using the extinction efficiency as a function of size parameter as an example. The solid line represents results computed from the improved geometric optics method, while the dots are results computed from the finite-difference time domain method. The dashed line represents the conventional extinction efficiency of 2 based on the principle of geometric optics. The presentations use a wavelength of 0.63 μm and randomly oriented ice columns of a uniform size identified in the figure.

smaller than about 15. This comparison highlights the increasing inapplicability of the localization principle inherent in geometric ray-tracing for the calculation of the electric field at the surface of a scatterer with a small size parameter. Based on the optical theorem noted previously, the extinction efficiency is equal to 2 regardless of the size parameter in the conventional geometric optics method. A series of systematic comparisons has been carried out for the phase function, single-scattering albedo, and extinction cross section computed by the improved geometric optics and FDTD methods for solid and hollow columns, plates, bullet rosettes, and aggregates. The improved geometric optics method was shown to produce acceptable accuracy for the single-scattering parameters for these ice crystal shapes with size parameters on the order of 15–20. Also, in these calculations, the peaks in the 22° and 46° halos generated by the geometric ray-tracing were found to be substantially smaller when the size parameter is smaller than about 60. Some of the cirrus clouds that do not produce the well-known halos and arcs could involve small ice crystals and/or the deviation of large ice crystals from the hexagonal geometries caused by the growth process.

The unified theory for light scattering by ice crystals is a powerful tool for the calculation of the phase function and other optical parameters. Figure 5.20 displays various commonly observed ice-crystal habits generated by computer and their associated phase functions. Eleven ice-crystal types are shown, including solid and
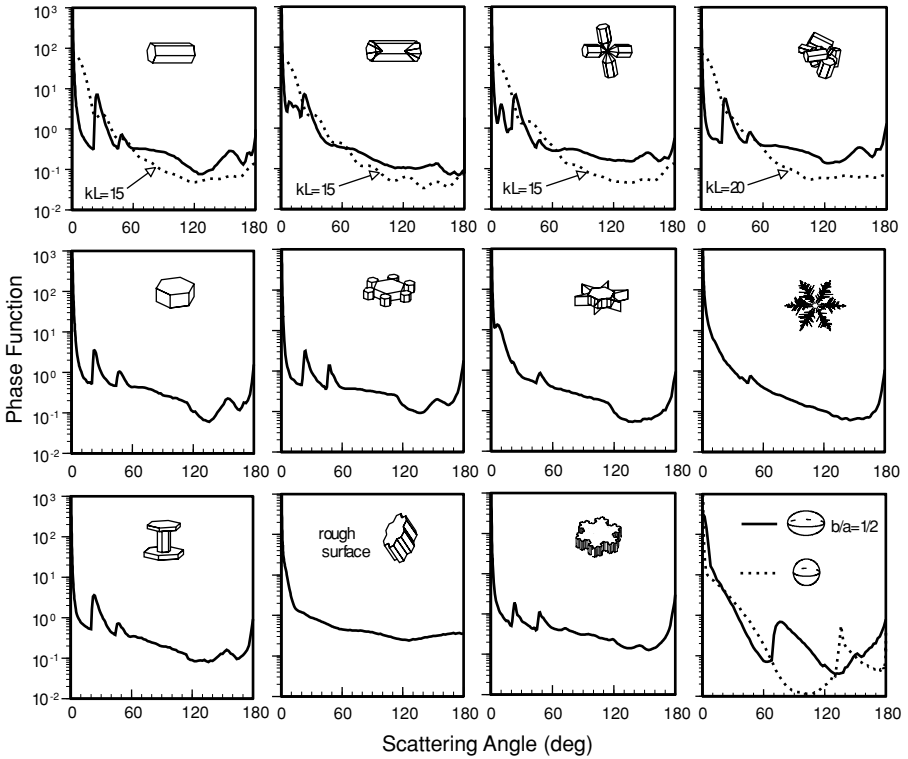
Figure 5.20   Eleven ice-crystal habits commonly occurring in cirrus clouds generated by computer program, along with their associated phase function patterns for the 0.63 μm wavelength. The solid lines are results computed from the geometric ray-tracing method; the dashed lines in the upper panel are results generated from the finite-difference time domain technique for which the size parameters are indicated in the diagrams. Also shown for comparison are phase functions for a spheroid and a sphere (after Liou *et al.,* 2000).

hollow columns, single and double plates and a plate with attachments, a bullet rosette with four branches, an aggregate composed of eight columns, a snowflake, a dendrite, and an ice particle with rough surfaces defined by a two-dimensional Gaussian probability function. Also displayed are an ice sphere and an ice spheroid whose phase functions are computed by the exact theory. The phase function for the sphere shows a maximum at the rainbow angle of about 138° and a broad minimum at the side-scattering directions 60–120°, whereas the spheroid displays a maximum pattern in this scattering-angle range. Both phase functions deviate significantly from those of ice particles. The patterns for the snowflake, the dendrite, and the plate with attachments were produced by a fractal shape-generation program. The phase function patterns for the hexagonally based ice crystals, such as solid column, single and double plates, bullet rosette, and column aggregate, all display a strong 22° halo peak (two refractions through a 60° prism angle), a halo peak at 46° (two refractions

through a 90° prism angle), and a maximum at about 160° (one and two internal reflections). The backscattering peak for these ice crystals is produced by external and internal reflections. The bullet also exhibits a peak at about 7°. For plates with small attachments and snowflakes, the 22° and 46° halo peaks are also produced from geometric ray tracing. In the case of the hollow column, the 46°, 160°, and backscattering maxima are absent because of the lack of scattering material. The 22° and 160° maxima vanish for simple and complicated dendrites, but the backscattering is still strong because of internal reflection contributions. The results for small ice crystals computed from the FDTD method are illustrated in the top four panels overlying those computed from geometric ray tracing. Phase function differences between large and small ice crystals are clearly illustrated in the graphs.

5.4.4.2   Theory versus Measurement and Representative Results

Measurements of the scattering and polarization patterns of ice crystals have been performed in cold chambers (e.g., Sassen and Liou, 1979; Volkovitsky *et al.,* 1980). Desirable ice crystal sizes and shapes, however, are difficult to generate and sustain for a period of time long enough to perform light-scattering experiments. Barkey *et al.* (1999) conducted an experimental light-scattering program using hexagonal ice-like crystals measured in the analog manner so that optical experiments could be performed over a relatively long period of time for complex-shaped particles. The experiment consisted of a polarized laser beam at $\lambda = 0.63$ μm and an array of 36 highly sensitive photodiode detectors arranged between the scattering angles of 2.8° and 177.2° mounted in a linear array on a half-dome, which could be rotated to vary the azimuthal angle. After careful calibration and signal acquisition, this system was first used to measure the phase functions of a glass sphere and a glass fiber configured to scatter light like an infinite cylinder. The experimental results closely match those computed from the Lorenz–Mie theory. Subsequently, sodium fluoride (NaF crystal), which has an index of refraction (1.33) close to that of ice in the visible, was adopted for the light scattering experiment. The NaF crystal could be machined to a desirable shape with a size on the order of a millimeter. It was mounted on top of a small pedestal, and its orientation position was controlled by a rotator. Angular integrations in the experiment could follow the computational procedures that are carried out in theory.

Figure 5.21a shows a comparison between measurements and theory for an aggregate that was assembled from NaF columns with small glass-fiber attachments glued onto small holes. To simulate random orientation, a 1° increment was used for all possible orientation angles. Results derived from the measurements and theory are shown to be in general agreement. Several discrepancies should be noted. Particularly, the experimental results are lower than the theory in backscattering directions, which are dominated by internal reflections. This difference could be caused by absorption by the small glass fibers and glues that connect the columns. Comparison results for a rough-surface plate are shown in Fig. 5.21b. All eight sides were sanded with small scratches evenly distributed across the crystal surface. Between the 25° and 180° scattering angles, the measurements closely follow the theoretical results. For scattering angles less than 20°, however, the experimental results are higher. The
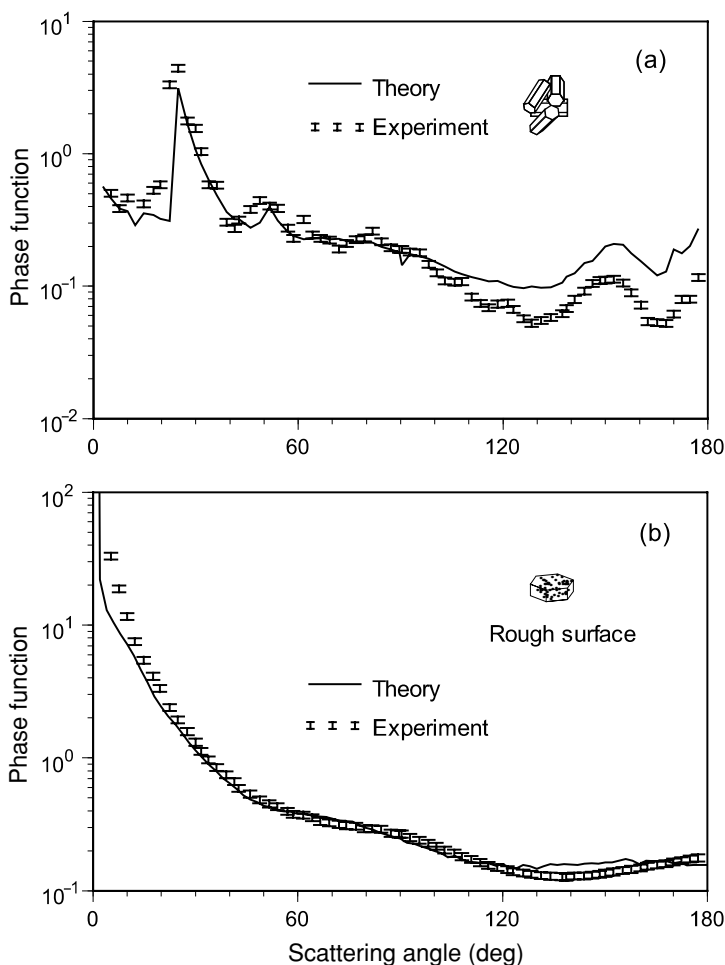
Figure 5.21   Phase functions for (a) randomly oriented aggregate and (b) rough surface plate crystals made from NaF whose index of refraction is 1.33 in the visible. The experiment used a polarized laser beam at 0.63 μm as the light source and controlled the positions of the detector and the crystal by automatic mechanical devices (Barkey *et al.,* 1999). The theoretical results were derived from the geometric-ray-tracing/Monte-Carlo method.

scanning electron photomicrographs reveal features on the roughened crystal surface on the order of 0.5–1 μm. More light could have been scattered through them as compared to the defined cross-sectional area used in diffraction calculations.

The electrodynamic levitation technique has also been used to suspend and grow an individual ice crystal for light-scattering experiments (Bacon *et al.,* 1998). The apparatus consists of an electrodynamic balance with an internally mounted thermal diffusion chamber, a laser beam, a 1024-element linear photodiode array, and two cameras to provide top and side views of the ice crystal. Shown in Fig. 5.22 are
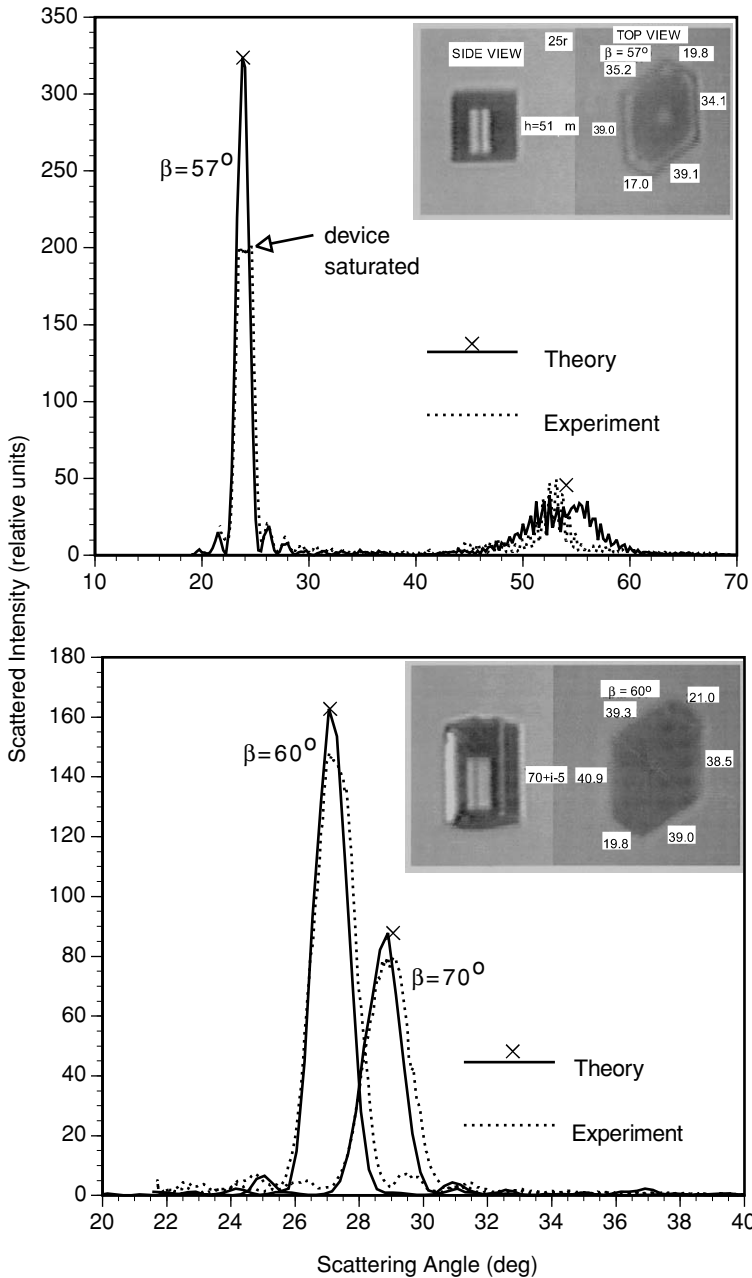
Figure 5.22 Phase function measurements using a laser beam at 0.63 μm and a large number of photodiode detector array for a single ice crystal suspended by the electrodynamic levitation technique (Bacon *et al.*, 1998). The sizes and shapes are determined from the top and side views of two cameras. The angle β denotes the ice-crystal orientation with respect to the incident laser beam. Theoretical results are computed from the conventional ($\times$) and improved geometric ray-tracing methods. Two ice-crystal cases are shown (after Liou *et al.*, 2000).
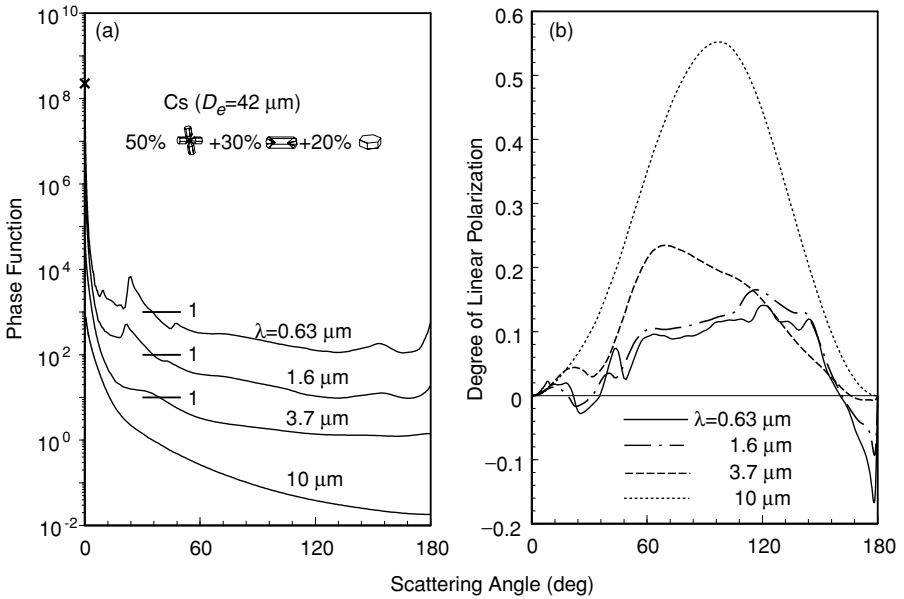
Figure 5.23   (a) Phase function and (b) degree of linear polarization for a typical cirrostratus with a mean effective ice crystal size of 42 μm composed of 50% bullet rosettes/aggregates, 30% hollow columns, and 20% plates, a shape model based on replicator and optical probe measurements. Four remote sensing wavelengths are displayed. For size parameters less than 15, the finite-difference time domain method is employed in the calculations. In the phase function, the vertical scale is applied to the lowest curve, while the upper curves are displayed upward by a factor of 10. The symbol $\times$ denotes the diffraction peak for the 0.63 μm wavelength.

the experimental results for two ice-crystal sizes and shapes defined by the depicted photos. Theoretical results computed from conventional geometric ray-tracing, which does not account for phase interferences, show discrete maxima associated with the halo pattern. The improved geometric optics method generates closely matched patterns, with the exception of some deviations in the 55° scattering angle region in the top diagram. Differences between the theoretical and experimental results can be attributed to the uncertainty of the measurement of the ice-crystal size ($\sim$ 4 μm) and of the computation of the near field based on the geometric ray-tracing approximation.

In Fig. 5.23 we illustrate the phase function and degree of linear polarization for a representative cirrostratus ice crystal size distribution (Section 5.1) having a mean effective ice-crystal size of 42 μm. Based on replicator and optical probe measurements, a cirrus cloud model composed of 50% bullet rosettes/aggregates, 30% hollow columns, and 20% plates is used in this presentation. Four remote sensing wavelengths of 0.63, 1.6, 3.7 and 10 μm are employed in the calculations. As pointed out in Section 5.3.3, these wavelengths are typical of the image channels on meteorological satellites. The scattering and polarization results for other remote-sensing wavelengths, 0.86 μm and 10.9 and 11 μm, are similar to those for 0.63 and 10 μm,

respectively. In the phase function, the 0.63 and 1.6 μm wavelengths display a peak at the 22° halo position because of the basic hexagonal structure of the rosettes, columns, and plates used in the calculations. Because of ice absorption, all the features associated with hexagonal ice crystals vanish at 3.7 and 10 μm, except for the forward diffraction peak. The linear polarization patterns for single scattering show a number of maxima features at the 0.63 and 1.6 μm wavelengths: a slight negative polarization at the 22° peak, a positive polarization at the 46° peak, and a strong negative polarization close to the backscattering angle. From about 60° to 140°, about 10% polarization is observed. The case of 10 μm exhibits a maximum at about 90° produced principally by external reflection and limited internal reflections of the light beam.

The spectral single-scattering parameters for ice crystals in terms of the asymmetry factor, single-scattering albedo, and extinction coefficient covering the spectral intervals of 0.2–5 μm and 8–13 μm are presented in Fig. 5.24 using the same cirrus cloud model shown in Fig. 5.23. Three mean effective ice-crystal sizes ranging from 10 μm (contrail cirrus), to 42 μm (typical cirrostratus), to 124 μm (cirrus uncinus), were selected for this illustration. The extinction coefficient is normalized in reference to the value at 0.5 μm. In the visible and near IR wavelengths, the average size parameter is sufficiently large that the extinction efficiency is approximately equal to 2 on the basis of the optical theorem, except in the vicinity of about 3 μm where a significant dispersion of the real part of the refractive index m_r for ice occurs (see also Fig. D.1). The single-scattering albedo pattern mimics the variability of the imaginary refractive index for ice (Liou, 1992, Fig. 5.1) with a large minimum located at about 3 μm. The single-scattering albedo decreases as the mean effective size increases, with an exception in the vicinity of about 2.85 μm, referred to as the *Christiansen effect*. This effect occurs when the real part of the refractive index approaches 1, while the corresponding imaginary part is substantially large, resulting in the domination of absorption. At the wavelength of about 10.9 μm, m_r for ice is close to 1.08, which is the smallest in the window region. However, m_i increases in the window wavelengths, the combination of which leads to a smaller value for extinction for the smallest size parameter of 10 μm. We also notice that the single-scattering albedo displays a significant drop at 10.9 μm. Because of this unique property of ice crystals, inference of their size appears feasible by using the window wavelengths that exist in present and future satellite systems (see Section 7.4.5, the subsection on information content in the IR line spectrum, for additional discussion).

5.5   Light Scattering by Nonspherical Aerosols

In Section 5.1, we showed that atmospheric aerosols exhibit various shapes with a typical size less than about 1 μm. The Lorenz–Mie theory of light scattering by spheres presented in Section 5.2 cannot be applied to nonspherical aerosol particles, nor can the geometric optics approach introduced in Section 5.3 be used for scattering and absorption calculations for aerosols because of their small size parameters. Many
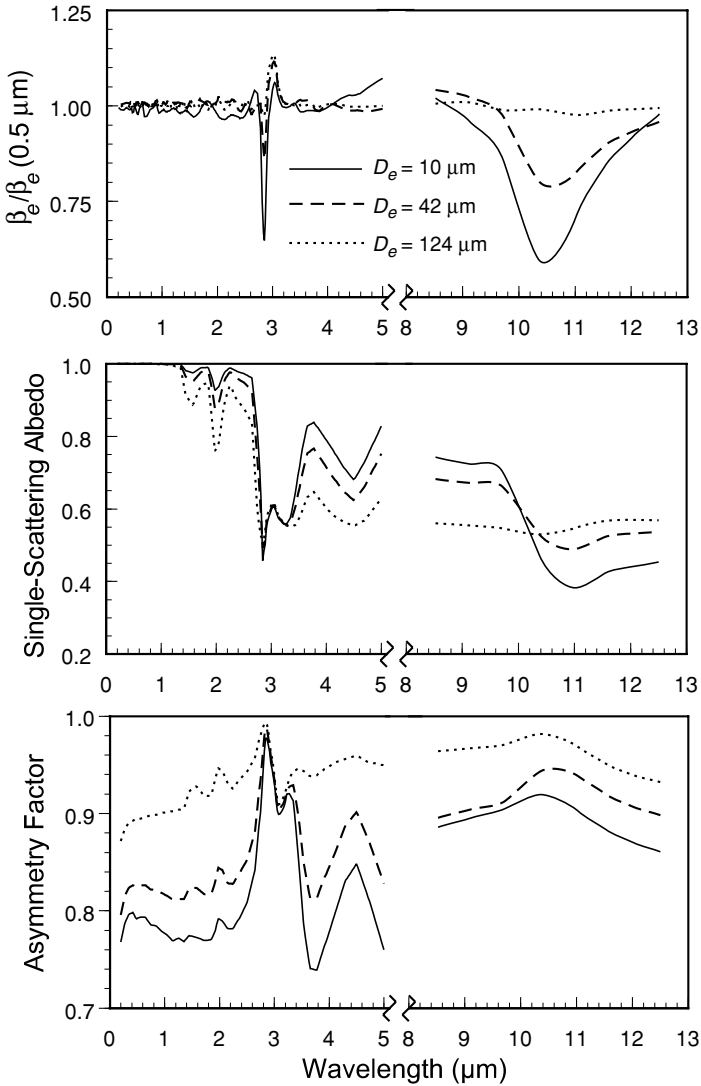
Figure 5.24   Normalized extinction coefficient (in reference to the value at 0.5 μm), single-scattering albedo, and asymmetry factor as functions of wavelength for three ice-crystal size distributions with mean effective sizes of 10, 42, and 124 μm and the shape model defined in Fig. 5.23. The extinction coefficient values at $\lambda = 0.5\,\mu$m are 0.3036, 0.2082, and 1.979 km^{-1} for $D_e = 10$, 42, and 124 μm, respectively. The spectral intervals covered are from 0.2 to 5 μm in the solar spectrum, and from 8 to 13 μm in the thermal infrared window.

approaches have been developed for the calculation of light scattering by nonspherical particles because of the need for precise scattering information in optics, geophysics, remote sensing, astrophysics, engineering, medicine, and biology. In the following sections, we introduce two contemporary methods that have been developed for the calculation of light scattering and absorption by nonspherical aerosols.

5.5.1 Finite-Difference Time Domain Method

The finite-difference time domain (FDTD) technique has been shown to be an efficient computational method of solving for the interaction of electromagnetic waves with scatterers, particularly those with complicated geometry and inhomogeneous composition. In this method, the space containing a scattering particle is discretized by using a grid mesh, and the existence of the particle is represented by assigning suitable electromagnetic constants in terms of permittivity, permeability, and conductivity to the grid points (Fig. 5.25). Because it is not necessary to impose the electromagnetic boundary conditions at the particle surface, the FDTD approach with appropriate and minor modifications can be applied to the solution of light scattering by various small nonspherical and inhomogeneous particles, such as irregular ice crystals and aerosols with inclusions.

The FDTD method was developed and pioneered by Yee (1966), but it did not receive significant recognition until high-quality absorbing boundary conditions were developed in the 1980s. Through the persistent efforts of a number of electrical engineers and computational physicists (e.g., Taflove, 1995), several advantages of the FDTD method have now become widely recognized. In recent years, the FDTD technique has been used to solve for the interaction between targets and electromagnetic
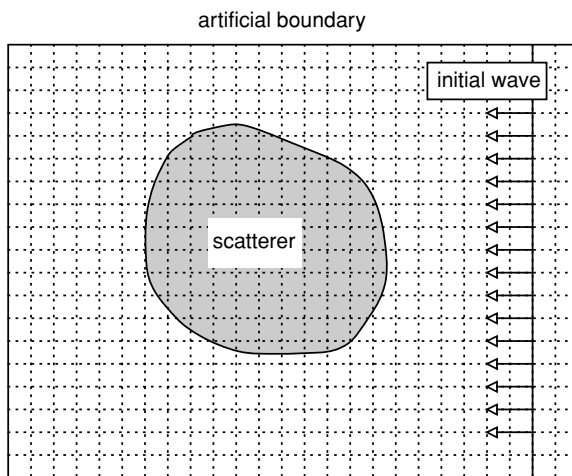
Figure 5.25 A conceptual diagram for the computation of the near field for the scattering of a nonspherical particle illuminated by plane waves by means of the finite-difference time domain (FDTD) method.

waves involving such problems as antenna scattering, numerical modeling of microstrip structures, and electromagnetic absorption by human tissues. Application of this method to the solution of the scattering and polarization properties of atmospheric nonspherical particles has also been carried out by Yang and Liou (1996b, 2000).

As described in Section 5.4.2, the FDTD technique uses Maxwell's time-dependent curl equations, denoted in Eqs. (5.2.10) and (5.2.11), to solve for the temporal variation of electromagnetic waves within a finite space that contains the scattering object. In this case, we write

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\varepsilon(\mathbf{r})}{c} \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t}, \tag{5.5.1a}$$

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{1}{c} \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t}, \tag{5.5.1b}$$

where ε is the permittivity of the dielectric medium, usually a complex variable, and c is the speed of light in vacuum. In Eq. (5.5.1b), the permeability, μ, has been assumed to be unity because cloud and aerosol particles in the atmosphere and many other scattering targets are mostly nonferromagnetic materials. We should select a harmonic time-dependent factor of $\exp(-i\omega t)$ for the electromagnetic wave in the frequency domain so as to have a positive imaginary part of the refractive index. Thus, we have

$$\varepsilon = \varepsilon_r + i\varepsilon_i, \tag{5.5.2a}$$

$$\varepsilon_r = m_r^2 - m_i^2, \quad \varepsilon_i = 2m_r m_i, \tag{5.5.2b}$$

where $i = \sqrt{-1}$ and m_r and m_i are the real and imaginary parts of the refractive index, respectively. When the medium is absorptive, i.e., m_i is nonzero, a complex calculation is required for Eq. (5.5.1a). To circumvent the complex operation, it is desirable to have an equivalent expression for Eq. (5.5.1a) in which the complex permittivity can be avoided. We may introduce an effective real permittivity, ε', and conductivity, σ, and rewrite the first Maxwell curl equation in a source-dependent form as follows:

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\varepsilon'}{c} \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} + \frac{4\pi}{c} \mathbf{J}(\mathbf{r}, t), \tag{5.5.3a}$$

where the second term on the right-hand side of the equation is due to the contribution of the effective electric current given by

$$\mathbf{J}(\mathbf{r}, t) = \sigma \mathbf{E}(\mathbf{r}, t). \tag{5.5.3b}$$

Transforming Eqs. (5.5.1a) and (5.5.3a) to the equivalent equations in the frequency domain yields

$$\nabla \times \mathbf{H}(\mathbf{r}) = -ik\varepsilon \mathbf{E}(\mathbf{r}), \tag{5.5.4}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -ik(\varepsilon' + i4\pi\sigma/kc)\mathbf{E}(\mathbf{r}), \tag{5.5.5}$$

where $k = \omega/c$ is the wavenumber of the electromagnetic wave in free space. From a comparison of Eqs. (5.5.4) and (5.5.5), it is clear that the two equations are equivalent if the following conditions hold: $\varepsilon_r = \varepsilon'$ and $\varepsilon_i = 4\pi\sigma/kc$. It follows that Eq. (5.5.1a) can be rewritten as follows:

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\varepsilon_r}{c}\left[\frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} + \tau\mathbf{E}(\mathbf{r}, t)\right], \tag{5.5.6}$$

where $\tau = kc\varepsilon_i/\varepsilon_r$. At this point we can use Eqs. (5.5.1b) and (5.5.6) to construct the finite-difference analog of the Maxwell curl equations, which involves only real calculations. First, we require discretizing the equation in the time domain. To do this, we rewrite Eq. (5.5.6) in the form

$$\frac{\partial[\exp(\tau t)\mathbf{E}(\mathbf{r}, t)]}{\partial t} = \exp(\tau t)\frac{c}{\varepsilon_r}\nabla \times \mathbf{H}(\mathbf{r}, t). \tag{5.5.7}$$

Integrating Eq. (5.5.7) over the time interval of $[n\Delta t, (n + 1)\Delta t]$, we obtain

$$\exp[\tau(n + 1)\Delta t]\mathbf{E}^{n+1}(\mathbf{r}) - \exp(\tau n\,\Delta t)\mathbf{E}^n(\mathbf{r})$$

$$= \int_{n\Delta t}^{(n+1)\Delta t} \exp(\tau t)\frac{c}{\varepsilon_r}\nabla \times \mathbf{H}(\mathbf{r}, t)dt$$

$$\approx \Delta t \exp[\tau(n + \tfrac{1}{2})\Delta t]\frac{c}{\varepsilon_r}\nabla \times \mathbf{H}^{n+\frac{1}{2}}(\mathbf{r}). \tag{5.5.8}$$

If follows that

$$\mathbf{E}^{n+1}(\mathbf{r}) = \exp(-\tau\Delta t)\mathbf{E}^n(\mathbf{r}) + \exp(-\tau\Delta t/2)\frac{c\Delta t}{\varepsilon_r}\nabla \times \mathbf{H}^{n+\frac{1}{2}}(\mathbf{r}). \tag{5.5.9}$$

Likewise, for the magnetic field we have

$$\mathbf{H}^{n+\frac{1}{2}}(\mathbf{r}) = \mathbf{H}^{n-\frac{1}{2}}(\mathbf{r}) - c\Delta t\nabla \times \mathbf{E}^n(\mathbf{r}). \tag{5.5.10}$$

In the foregoing equations the superscript n denotes that the associated field is evaluated at the time step $t = n\Delta t$.

The space containing the scattering particle must be discretized by a number of grid cells. We may use cubic cells in the 3D case and evaluate the components of the electric and magnetic fields of a cell at the staggered locations as shown in Fig. 5.26. The advantage of such staggered positions is that the electromagnetic boundary conditions are guaranteed at the interfaces of the cells, so that the tangential components of the E field and the normal components of the H field are continuous at the interfaces. A spatial location in the discretized space is denoted by the indices $(I, J, K) = (I\Delta x, J\Delta y, K\Delta z)$, and any variable as a function of space and time is defined as

$$F^n(I, J, K) = F(I\Delta x, J\Delta y, K\Delta z, n\Delta t), \tag{5.5.11}$$

in which Δx, Δy, and Δz are the cell dimensions along the x, y, and z axes, respectively. The permittivity must be homogeneous within each cell. For a given cell with
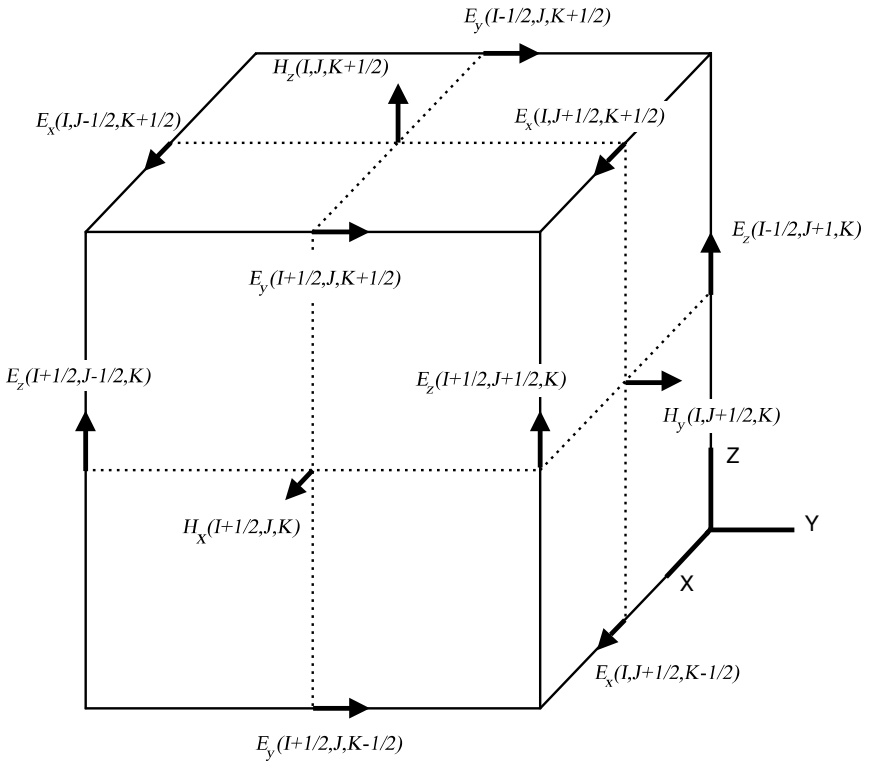
Figure 5.26 Locations of various field components on a cubic cell. The **E** field is tangential to the surface, whereas the **H** field is normal to the surface. The notations I, J, and K are indices assigned to the cell.

its center located at a lattice index (I, J, K), the mean permittivity can be evaluated on the basis of the Maxwell–Garnett rule via

$$\frac{\bar{\varepsilon}(I, J, K) - 1}{\bar{\varepsilon}(I, J, K) + 2} = \frac{1}{\Delta x \Delta y \Delta z} \iiint\limits_{\text{cell}(I, J, K)} \frac{\varepsilon(x, y, z) - 1}{\varepsilon(x, y, z) + 2} \, dx \, dy \, dz. \qquad (5.5.12)$$

Using the mean permittivity produces smaller staircasing errors than those produced by a sharp step-by-step approximation of nonspherical geometry.

The general forms of the finite-difference analogs of Eqs. (5.5.1) and (5.5.6) can then be expressed for each Cartesian component. In practice, the truncation errors of the finite-difference analog of Maxwell's curl equations are of second order in both time and space. The time-marching iteration can be applied directly by employing initial values for the E and H fields. The values of Δx, Δy, Δz, and Δt cannot be specified arbitrarily but must satisfy the Courant–Friedrichs–Levy condition such that

$$c\Delta t \leq \frac{1}{\sqrt{1/\Delta x^2 + 1/\Delta y^2 + 1/\Delta z^2}}. \qquad (5.5.13)$$

Moreover, the spatial increments must also be smaller than about 1/20 of the incident wavelength so that the phase variation of the waves is negligible over the distance of the cell dimensions.

The numerical implementation of the FDTD technique requires the imposition of an appropriate absorbing boundary condition, which is critical for the stability of the numerical computations and the reliability of the results. In addition, the white space between the boundary and the scatterer required by a specific boundary condition is an important factor in determining the required computational effort. Many techniques have been developed for the imposition of absorbing boundary conditions in the application of the FDTD technique to electromagnetic scattering problems. A recent advancement in this area is a novel numerical technique, referred to as the *perfectly matched layer* (PML) boundary condition, developed by Berenger (1994). Calculation of the absorption of the outgoing wave by the PML method is based on the absorption by a medium located at the outermost layers in the computational domain backed by a perfectly conducting surface. The conventional technique based on an absorbing medium is to specifically define the wave impedance of the medium so that it matches that of the free space. Such a simple matching approach produces substantial nonzero reflections when a scattered wave impinges on the absorbing medium obliquely. To overcome the disadvantage of the conventional method, the absorbing medium can be selected such that the wave decay due to absorption is imposed on the field components parallel to their boundary layers. To achieve this goal, each Cartesian component of the electromagnetic field is split into two parts as follows:

$$(E_x, E_y, E_z) = \left[(E_{x2} + E_{x3}), (E_{y1} + E_{y3}), (E_{z1} + E_{z2}) \right],$$
$$(H_x, H_y, H_z) = \left[(H_{x2} + H_{x3}), (H_{y1} + H_{y3}), (H_{z1} + H_{z2}) \right], \qquad (5.5.14)$$

where the subscript 1, 2, or 3 denotes the component of the electric (or magnetic) field that is associated with the spatial differential of the magnetic (or electric) field component along the *x, y,* and *z* directions, respectively. With the split field components, discretized scalar equations that govern the propagation of electromagnetic waves can be replaced by a set of 12 equations for numerical computations. Numerical experiments have shown that the spurious reflection produced by the PML boundary conditions is much smaller than that generated by other analytical absorption boundary conditions derived from the wave equation.

The values of the near field computed by the preceding FDTD algorithm are in the time domain. A transformation of the time-dependent field values to their corresponding counterparts in the frequency domain is required to obtain the single-scattering properties. The transformation algorithm depends on the kind of initial wave that is used. We may select an incident pulse (e.g., Gaussian) as the initial excitation in the computation. The width of the pulse can be properly selected to avoid numerical dispersion caused by the finite-difference approximation. In this manner, an incident pulse with respect to a fixed shape and size of the particle can provide the results for a number of size parameters simultaneously. Further, the frequency of the simulated field can be obtained by the discrete Fourier transform. Let *f* be a component of the

field and its value at the time step n be f_n. Then, the time variation of f can be written as

$$f(t) = \sum_{n=0}^{N} f_n \delta(t - n\Delta t), \qquad (5.5.15)$$

where δ is the Dirac delta function and the maximum time step N is chosen such that the field in the time domain is reduced to a small value. The corresponding spectrum in the frequency domain is given by

$$F(k) = \int_{-\infty}^{\infty} \left[\sum_{n=0}^{N} f_n \delta(t - n\Delta t) \right] \exp(ikct) \, dt = \sum_{n=0}^{N} f_n \exp(ikcn\Delta t), \quad (5.5.16)$$

where k is the wavenumber in vacuum. To avoid aliasing and numerical dispersion and to obtain a correct frequency spectrum, one must band the maximum wavenumber or the minimum wavelength for the region within which the frequency response of the scattering is evaluated. In any finite-difference equation, it is required that the wavelength of a simulated wave be larger than the grid size. Finally, the field values in the frequency domain obtained by this procedure must be normalized by the Fourier transform of the incident wave at the center of the grid mesh so that the frequency response of the scattering particle will return to a unit incident harmonic wave.

In order to compute the scattering and absorption quantities, we must transform the near field determined from the FDTD algorithm and the discrete Fourier transform technique to the far field. We may follow an approach that uses a volume integration method for nonconducting scatterers. In the far field (or radiation zone), $k\mathbf{r} \to \infty$, the scattered far field for the electric vector may be written in terms of a volume integral as follows:

$$E_s(\mathbf{r}) = \frac{k^2 \exp(ikr)}{4\pi r} \int\int\int_V [\varepsilon(\mathbf{r}') - 1] \left\{ \mathbf{E}(\mathbf{r}') - \mathbf{n}[\mathbf{n} \cdot \mathbf{E}(\mathbf{r}')] \right\} \exp(-ik\mathbf{n} \cdot \mathbf{r}') \, d^3\mathbf{r}'.$$
$$(5.5.17)$$

To compute the scattering phase matrix, the scattered field given by Eq. (5.5.17) must be expressed in terms of the amplitude matrix. Because the scattered field is a transverse wave with respect to the scattering direction, it can be decomposed into the components parallel and perpendicular to the scattering plane in the form

$$\mathbf{E}_s(\mathbf{r}) = \alpha E_{s,\alpha}(\mathbf{r}) + \beta E_{s,\beta}(\mathbf{r}), \qquad (5.5.18)$$

where α and β are the unit vectors parallel and perpendicular to the scattering plane, respectively, and satisfy $\mathbf{n} = \beta \times \alpha$. Expressing Eq. (5.5.18) in matrix form yields

$$\begin{pmatrix} E_{s,\alpha}(\mathbf{r}) \\ E_{s,\beta}(\mathbf{r}) \end{pmatrix} = \frac{k^2 \exp(ikr)}{4\pi r} \int\int\int_V [\varepsilon(\mathbf{r}') - 1] \begin{pmatrix} \alpha \cdot \mathbf{E}(\mathbf{r}') \\ \beta \cdot \mathbf{E}(\mathbf{r}') \end{pmatrix} \exp(-ik\mathbf{n} \cdot \mathbf{r}') \, d^3\mathbf{r}'$$

$$= \frac{\exp(ikr)}{r} \mathbf{S} \begin{pmatrix} E_{o,\alpha} \\ E_{o,\beta} \end{pmatrix}, \qquad (5.5.19)$$

where $\mathbf{S}$ is a 2×2 amplitude scattering matrix and $E_{o,\alpha}$ and $E_{o,\beta}$ are the incident E-field components defined with respect to the scattering plane. In the FDTD method, the incident wave is defined with respect to the incident coordinate system given by $E_{o,x}$ and $E_{o,y}$. Based on the geometry implied by Eq. (5.5.18), we have

$$\begin{pmatrix} E_{o,\alpha} \\ E_{o,\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \cdot \mathbf{x} & -\boldsymbol{\beta} \cdot \mathbf{y} \\ \boldsymbol{\beta} \cdot \mathbf{y} & \boldsymbol{\beta} \cdot \mathbf{x} \end{pmatrix} \begin{pmatrix} E_{o,y} \\ E_{o,x} \end{pmatrix}, \tag{5.5.20}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the unit vectors along the x and y axes, respectively. To obtain the scattering properties of the particle with complete polarization information, we can select two incident cases: (a) $E_{o,x} = 1$ and $E_{o,y} = 0$; and (b) $E_{o,x} = 0$ and $E_{o,y} = 1$. We can then define the following quantities:

$$\begin{pmatrix} F_{\alpha,x} \\ F_{\beta,x} \end{pmatrix} = \frac{k^2}{4\pi} \int \int \int_V [\varepsilon(\mathbf{r}') - 1] \begin{pmatrix} \boldsymbol{\alpha} \cdot \mathbf{E}(\mathbf{r}') \\ \boldsymbol{\beta} \cdot \mathbf{E}(\mathbf{r}') \end{pmatrix} \exp(-ik\mathbf{n} \cdot \mathbf{r}') d^3\mathbf{r}' \bigg|_{E_{o,x}=1, E_{o,y}=0}, \tag{5.5.21a}$$

$$\begin{pmatrix} F_{\alpha,y} \\ F_{\beta,y} \end{pmatrix} = \frac{k^2}{4\pi} \int \int \int_V [\varepsilon(\mathbf{r}') - 1] \begin{pmatrix} \boldsymbol{\alpha} \cdot \mathbf{E}(\mathbf{r}') \\ \boldsymbol{\beta} \cdot \mathbf{E}(\mathbf{r}') \end{pmatrix} \exp(-ik\mathbf{n} \cdot \mathbf{r}') d^3\mathbf{r}' \bigg|_{E_{o,x}=0, E_{o,y}=1}. \tag{5.5.21b}$$

Using Eqs. (5.5.19)–(5.5.21b) along with some algebraic manipulations, it can be proven that the amplitude matrix is given by

$$\begin{pmatrix} S_2 & S_3 \\ S_4 & S_1 \end{pmatrix} = \begin{pmatrix} F_{\alpha,y} & F_{\alpha,x} \\ F_{\beta,y} & F_{\beta,x} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \cdot \mathbf{x} & \boldsymbol{\beta} \cdot \mathbf{y} \\ -\boldsymbol{\beta} \cdot \mathbf{y} & \boldsymbol{\beta} \cdot \mathbf{x} \end{pmatrix}. \tag{5.5.22}$$

The scattering phase matrix can subsequently be defined following the procedure outlined in Section 5.4.3. For nonspherical aerosols oriented randomly in space, the scattering phase matrix has a block-diagonal structure with eight nonzero elements of which only six are independent (Section 5.4.1). Finally, the expressions for extinction and absorption cross sections have been given in Eqs. (5.4.6a) and (5.4.6b), respectively.

Accuracy of the FDTD method can be checked with the exact solutions for infinite circulars and spheres. In general, when the size of the grid cells is on the order of $1/20$ of the incident wavelength, its solutions are in excellent agreement with their corresponding analytical counterparts. Improvement of the accuracy can be achieved by decreasing the ratio of the grid size to the incident wavelength, but only at the expense of increasing the computational effort required. Based on comprehensive numerical experiments, the FDTD method can achieve reliable results for nonspherical aerosol size parameters smaller than about 15–20 (Yang and Liou, 2000). In what follows, we present a number of representative scattering results for nonspherical aerosols.

As illustrated in Section 5.1, aerosols in the atmosphere exhibit a variety of shapes ranging from quasispheres to highly irregular geometries. In addition, aerosols usually appear as a mixed product of different compositions involving dustlike,
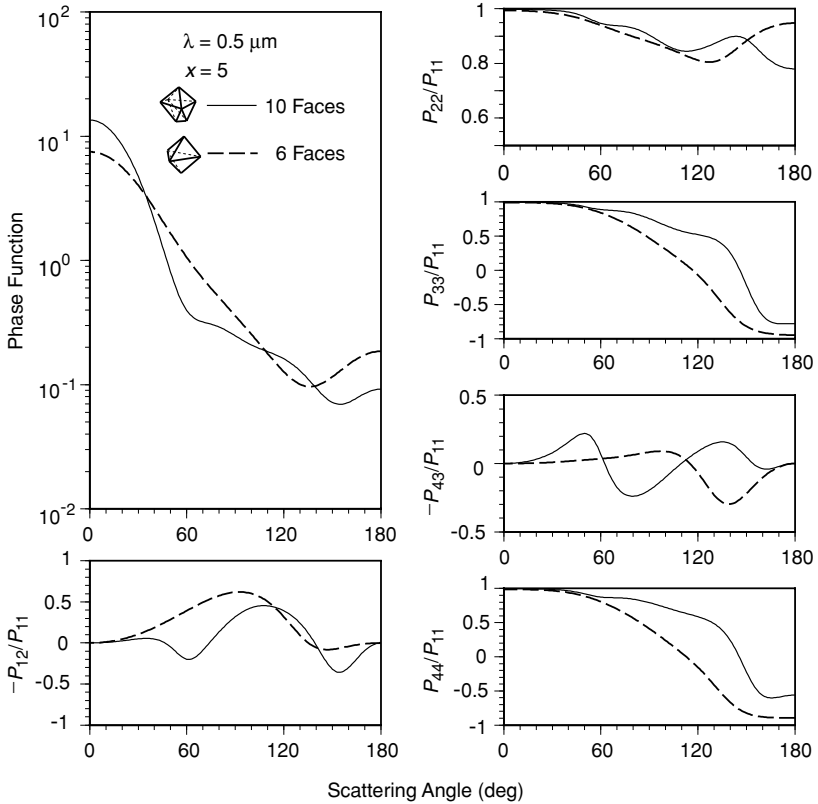
Figure 5.27   Scattering phase matrix elements for two types of randomly oriented dust aerosols. The size parameter used is 5 and the real and imaginary parts of the refractive index are 1.53 and 0.008, respectively, for a wavelength of 0.5 μm (data taken from Yang and Liou, 2000).

water-soluble, soot, oceanic, sulfate, mineral, water, and organic materials. To understand the scattering characteristics of aerosols, various representative aerosol geometries and inhomogeneous compositions are defined for light-scattering computations based on the FDTD method.

The left panels of Fig. 5.27 illustrate the phase function and the degree of linear polarization at $\lambda = 0.5\,\mu$m for two randomly oriented dustlike aerosol shapes, one with 10 faces and the other with six, constructed by a computer graphics program. The size parameters of these irregular aerosols are specified in terms of the dimensions of their peripheral spheres. Although the two polyhedrons have the same size parameters, the particle with 10 faces scatters more energy in the forward direction than its 6-face counterpart, because the volume of the former is larger. However, the reverse is true for backscattering. Note that the ratio of the extinction cross sections for these two aerosol shapes is 3.92. Dustlike aerosols are absorptive in the visible wavelength as indicated by the single-scattering albedos of 0.9656 and 0.9626 for the two polyhedral

geometries with 10 and 6 faces, respectively. The other panels of Fig. 5.27 show the remaining scattering phase matrix elements associated with the polarization state of the scattered wave. It is evident that the structure of aerosol geometry has a substantial impact on the polarization configuration.

Black carbon or soot aerosols generated from the incomplete combustion of fossil fuel and biomass burning can serve as condensation nuclei or become outside attachments to water droplets, a possibility relevant to cloud absorption. To illustrate the applicability of the FDTD method to light scattering by these types of aerosols, we present in Fig. 5.28 the phase functions and the degrees of linear polarization for four aerosol models defined in the diagram. The prime and double prime denote
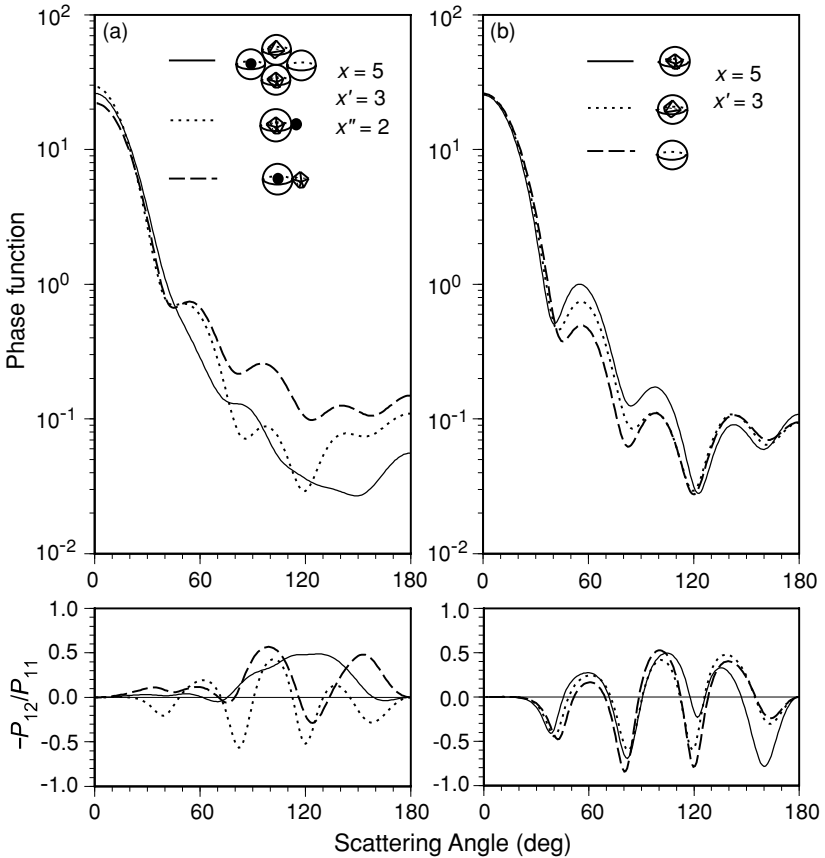
Figure 5.28   Phase functions and degrees of linear polarization for five aerosol geometric configurations using a wavelength of 0.5 μm. In (a), the size parameter for the largest water aerosol is 5, while the dust and spherical inclusions (or attachments) are 3 and 2, respectively. In (b), the inclusions are 10- and 6-face dust particles. The refractive indices used in the FDTD calculations are $(1.335, 1.0 \times 10^{-9})$ for water, $(1.53, 7.8 \times 10^{-3})$ for mineral dusts, and $(1.75, 0.45)$ for soot. Results for spheres computed from the Lorenz-Mie theory are shown in (b) for comparison purposes (data taken from Yang and Liou, 2000).

that the associated parameters are for mineral/dustlike and soot components, respectively, whereas the corresponding unprimed parameters are for water parts of the compounded particles. Polyhedral particles and sphere clusters produce smoother angular scattering patterns in comparison to the cases involving spheres with inclusions and/or attachments. In the case of the latter, the spherical parts of the compounded aerosols dominate the scattering properties. It is clear that the phase functions and polarization patterns of polyhedral and cluster aerosols are substantially different from those of homogeneous spheres, also shown in the diagram.

5.5.2   *T*-Matrix Method

The *T*-matrix method for light-scattering calculations is based on the expansion of the incident and scattered fields in vector spherical wave functions and relating these expansions by means of a *T*-matrix for computing electromagnetic scattering by single, homogeneous nonspherical particles. The *T*-matrix approach was initially introduced by Waterman (1971) and has been shown to be an efficient method for scattering calculations involving rotationally-symmetric nonspherical particles, such as spheroids, cylinders, two-sphere clusters, and Chebyshev particles (Mishchenko *et al.,* 2000). This method has been used for the analysis of light scattering by nonspherical aerosols, particularly with respect to the investigation of deviations of single-scattering results from those computed from the spherical assumption.

The essence of the *T*-matrix method begins with the expansion of the incident and scattered fields involving the scattering of a plane electromagnetic wave by a single particle in the forms

$$\mathbf{E}^i(\mathbf{R}) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \left[a_{mn}\mathbf{M}_{mn}^*(k\mathbf{R}) + b_{mn}\mathbf{N}_{mn}^*(k\mathbf{R}) \right], \qquad (5.5.23a)$$

$$\mathbf{E}^s(\mathbf{R}) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \left[p_{mn}\mathbf{M}_{mn}(k\mathbf{R}) + q_{mn}\mathbf{N}_{mn}(k\mathbf{R}) \right], \qquad (5.5.23b)$$

where $\mathbf{M}_{mn}$ and $\mathbf{N}_{mn}$ are certain vector spherical wave functions involving spherical Hankel functions; $\mathbf{M}_{mn}^*$ and $\mathbf{N}_{mn}^*$ are defined by similar functions, except spherical Hankel functions are replaced by spherical Bessel functions; a_{mn} and b_{mn} are expansion coefficients of the plane incident wave; and p_{mn} and q_{mn} are expansion coefficients of the scattered wave. Because of the linearity of Maxwell's equations and boundary conditions, the relationship between the scattered and incident coefficients must be linear and can be expressed by a *T*-matrix as follows:

$$p_{mn} = \sum_{\ell=1}^{\infty} \sum_{k=-\ell}^{\ell} \left[T_{mnk\ell}^{11} a_{k\ell} + T_{mnk\ell}^{12} b_{k\ell} \right], \qquad (5.5.24a)$$

$$q_{mn} = \sum_{\ell=1}^{\infty} \sum_{k=-\ell}^{\ell} \left[T_{mnk\ell}^{21} a_{k\ell} + T_{mnk\ell}^{22} b_{k\ell} \right]. \qquad (5.5.24b)$$

In matrix form, we write

$$\begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} = \mathbf{T} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{T}^{11} & \mathbf{T}^{12} \\ \mathbf{T}^{21} & \mathbf{T}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}. \tag{5.5.25}$$

Equation (5.5.25) forms the foundation of the T-matrix approach. If the T-matrix for a given scatterer is known, the scattered field defined in Eq. (5.5.23b) can then be evaluated. Consequently, the amplitude matrix defined in Eq. (5.4.23a) can be obtained. The T-matrix depends only on the physical and geometric characteristics of the scattering particle, including the refractive index, size, shape, and orientation with respect to the incident light beam.

The general approach to determining the T-matrix is to find the field scattered by an object bounded by a closed surface in terms of an integral equation that involves the unknown surface field on the exterior of this surface. The determination of the internal field must then be defined by

$$\mathbf{E}^{\mathrm{int}}(\mathbf{R}) = \sum_{n=1}^{\infty} \sum_{m=-n}^{n} \left[c_{mn} \mathbf{M}_{mn}^{*}(mk\mathbf{R}) + d_{mn} \mathbf{N}_{mn}^{*}(mk\mathbf{R}) \right], \tag{5.5.26}$$

where m is the refractive index of the particle relative to that of the surrounding medium, and c_{mn} and d_{mn} are unknown coefficients. By using boundary conditions at the surface of the scatterer that require the continuity of the tangential components of the electric and magnetic fields, we obtain the following matrix equation:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^{11} & \mathbf{Q}^{12} \\ \mathbf{Q}^{21} & \mathbf{Q}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}. \tag{5.5.27}$$

The elements of the $\mathbf{Q}$ matrix are surface integrals of products of the vector spherical wave functions that depend only on the particle's size, shape, and refractive index. Thus, by inverting this matrix equation, the unknown expansion coefficients of the internal field, $\mathbf{c}$ and $\mathbf{d}$, are expressed in terms of the known expansion coefficients of the incident field, $\mathbf{a}$ and $\mathbf{b}$. Moreover, from the boundary condition and the relationships among the incident, scattered, and surface fields, one can show that

$$\begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} = - \begin{bmatrix} \mathbf{Q}^{*11} & \mathbf{Q}^{*12} \\ \mathbf{Q}^{*21} & \mathbf{Q}^{*22} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}, \tag{5.5.28}$$

where the $\mathbf{Q^*}$ elements are also given by integrals over the particle's surface and depend only on the particle's characteristics. A comparison between Eqs. (5.5.25) and (5.5.28) leads to the solution of the T-matrix as follows:

$$\mathbf{T} = -\mathbf{Q}^{*}\mathbf{Q}^{-1}. \tag{5.5.29}$$

The preceding procedure for computing the T-matrix for single homogeneous scatterers in reference to the particle's coordinates is referred to as the *extended boundary condition method* (EBCM). The EBCM technique has been shown to have numerical stability problems in calculations for particles with very large real and/or imaginary parts of the refractive index, large size parameters (>20), and/or extreme geometries such as spheroids with large axial ratios. More efficient approaches that overcome the
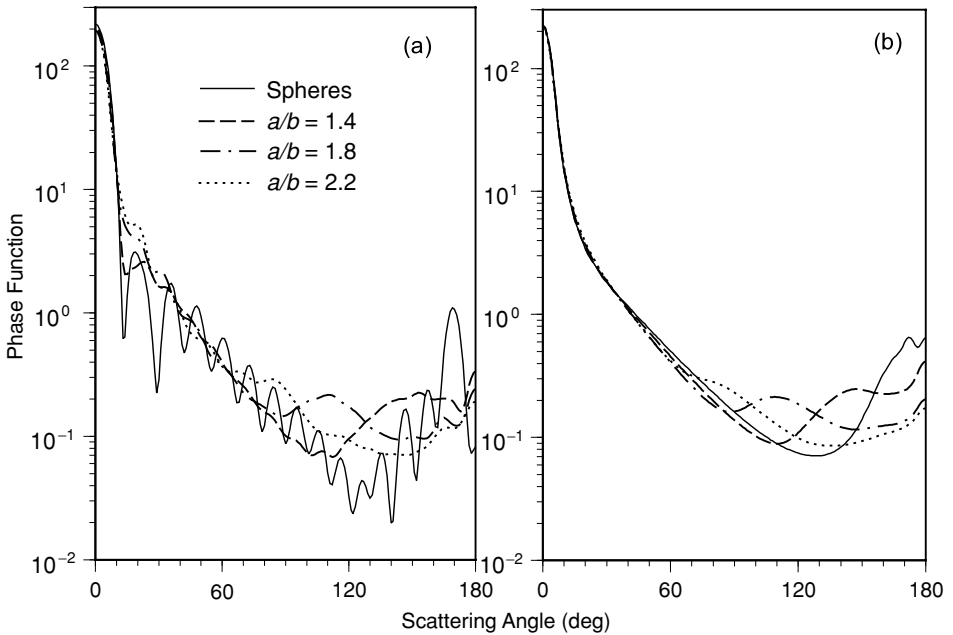
Figure 5.29   *T*-matrix computations of the phase function versus the scattering angle for monodisperse and polydisperse spheres and randomly oriented spheroids with a refractive index of $1.53 + 0.008i$. The wavelength is 0.443 μm. (a) Results for monodisperse spheres with a radius of 1.163 μm and surface-equivalent prolate spheroids with aspect ratios *a/b*, increasing from 1.4 to 2.2. (b) Similar computations but for a log-normal size distribution with a mean effective radius of 1.163 μm and an effective variance of 0.168 (data provided by Michael Mishchenko of the NASA Goddard Institute for Space Studies).

numerical instability problem in computing the *T*-matrix for highly elongated particles have also been developed. With computer coding improvements, the *T*-matrix method can be applied to symmetric homogeneous nonspherical particles with size parameters as large as 180. Interested readers may wish to refer to Mishchenko *et al.* (2000) for details.

Examples of *T*-matrix computations of the phase function as a function of the scattering angle for randomly oriented monodisperse and polydisperse spheroids and spheres are presented in Fig. 5.29. Panel (a) displays results for spheres with a radius of 1.163 μm and surface-equivalent prolate spheroids with aspect ratios ranging from 1.4 to 2.2. Panel (b) shows similar computations, but for a log-normal size distribution with an effective radius of 1.163 μm, illustrating the deviation in the phase function of nonsphericity from the spherical assumption.

Finally, comparisons of the phase functions of spheres computed from the Lorenz–Mie theory, of spheroids computed from the *T*-matrix approach, and of convex and concave particles computed from the FDTD method are presented in Fig. 5.30. These shapes are used to resemble aerosols with a refractive index of $1.38 + i3.9 \times 10^{-9}$ and a maximum size parameter of 10 with three aspect ratios. For $a/b = 1$
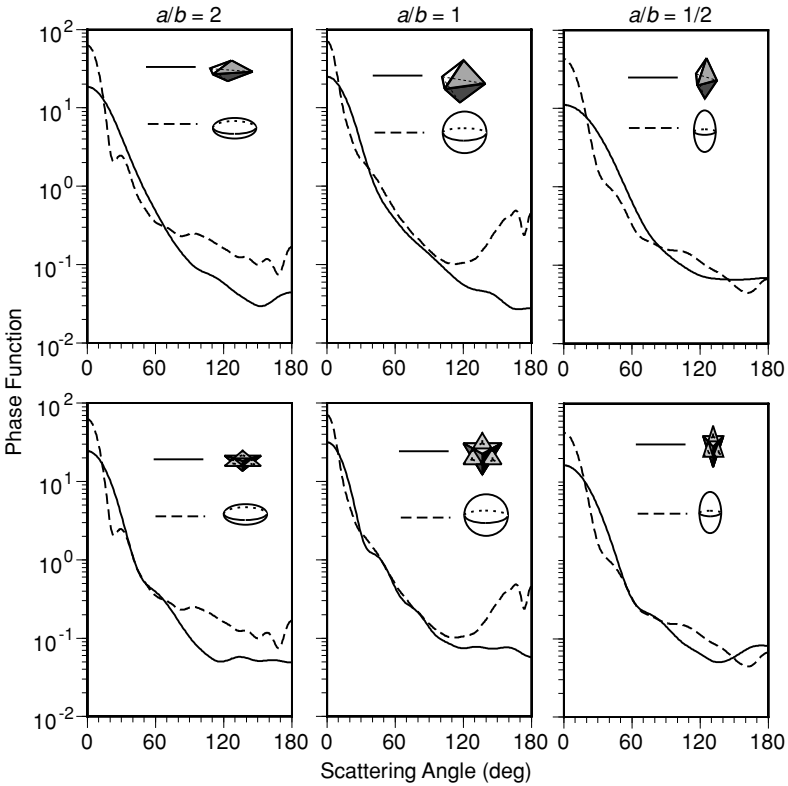
Figure 5.30   Comparison of the phase functions of oceanic aerosol particles with various shapes. For the spherical case, i.e., $a/b = 1$, the power law size distribution is employed to smooth out the resonant fluctuations. The maximum size parameter used is 10.

(a is the semimajor axis and b is the semiminor axis), the phase functions of nonspherical particles are substantially smaller than those of spheres in the scattering angular region $120°$–$180°$; in particular, the differences in the results for spherical and nonspherical particles are pronounced in the backscattering direction. For $a/b = 2$ and $a/b = 1/2$, we also see substantial differences between smooth spheroids and irregular convex and concave particles. Since the phase function is the critical parameter that is required in the satellite retrieval of aerosol optical depth (Section 7.3.3), selection of a proper and reliable value for each aerosol type must be undertaken to ensure the accuracy and precision of the retrieval.

5.5.3   Note on Light-Scattering Measurements for Nonspherical Aerosols

Development of integrating nephelometers for measurement of the scattering cross section of aerosols has been a subject of considerable research in the past (Heintzenberg and Charlson, 1996). The successful development of this instrument is significant

because if an accurate and reliable extinction cross section can be determined, the absorption cross section and hence the single-scattering albedo can be obtained to assess the climatic radiation forcing produced by aerosols.

Measurements of light scattering by aerosols as a function of wavelength and/or angular distribution for sizing purposes based on Lorenz–Mie results have been an interdisciplinary research subject, particularly related to instrument technology. However, angular scattering measurements of nonspherical aerosols, including polarization, have been extremely limited. The optical technique generally uses a laser beam and a sample of known aerosol types generated in the laboratory. Early laser scattering experiments with measurements of the scattering phase matrix have been reported by Holland and Gagne (1970) for randomly oriented microsized quartz crystals with flat platelike shapes. More recently, laboratory measurements of the scattering and polarization properties of aerosol particles have been presented by Volten *et al.* (1999). In angular scattering experiments, a He–Ne laser beam with a wavelength of 0.633 μm is usually used as a light source, which passes through a linear polarizer and travels through the electro-optic modulator, which can be rotated about its long axis. The light scattered in a particular direction passes through a quarter-wave plate and an analyzer and is then recorded by a photomultiplier detector. In this manner, the four-by-four scattering phase matrix can be measured. The sample aerosols are disturbed by air motion such that random orientation is anticipated. Thus, there are only six independent matrix elements. The scattering angular range covered is generally limited by the detector size with typical values ranging from 5° to 175° (see also Sections 5.4.4 and 5.3.3). The reliability of the experimental setup can be tested by employing spherical particles for which the measured data can be checked against results computed from the Lorenz–Mie theory.

An example of the scattering phase matrix elements for three types of mineral aerosols, feldspar, red clay, and quartz, with mean sizes 3.0, 5.1, and 9.7 μm, respectively, is shown in Fig. 5.31. Except for quartz, the refractive indices are basically unknown. The phase function (P_{11}) is normalized to 1 at the scattering angle of 30° in this presentation because of the lack of measured data from the scattering angle ranges 0–5° and 175–180°. The phase functions for these three aerosol types are extremely similar. The degree of linear polarization ($-P_{12}/P_{11}$) shows that the smallest particles have the largest maximum value. Close to the backscattering direction, negative polarization is seen. The remaining elements display different patterns for red clay than for the two other aerosol types, presumably because of its absorption property. A precise determination of the composition, size, and shape of aerosols based on scattering and polarization measurements, particularly from space, is critically important in order to reduce the uncertainties of the role aerosols play in global climate change.

As a result of the difficulty in manufacturing micrometer-sized aerosol particles with known optical properties and obtaining precise orientation information to compare to theoretical results, the microwave analog technique has been employed for light-scattering measurements involving particles with small size parameters. This technique usually involves incident microwave radiation from a transmitting antenna that passes a polarizer and is scattered by the object. The scattered radiation then
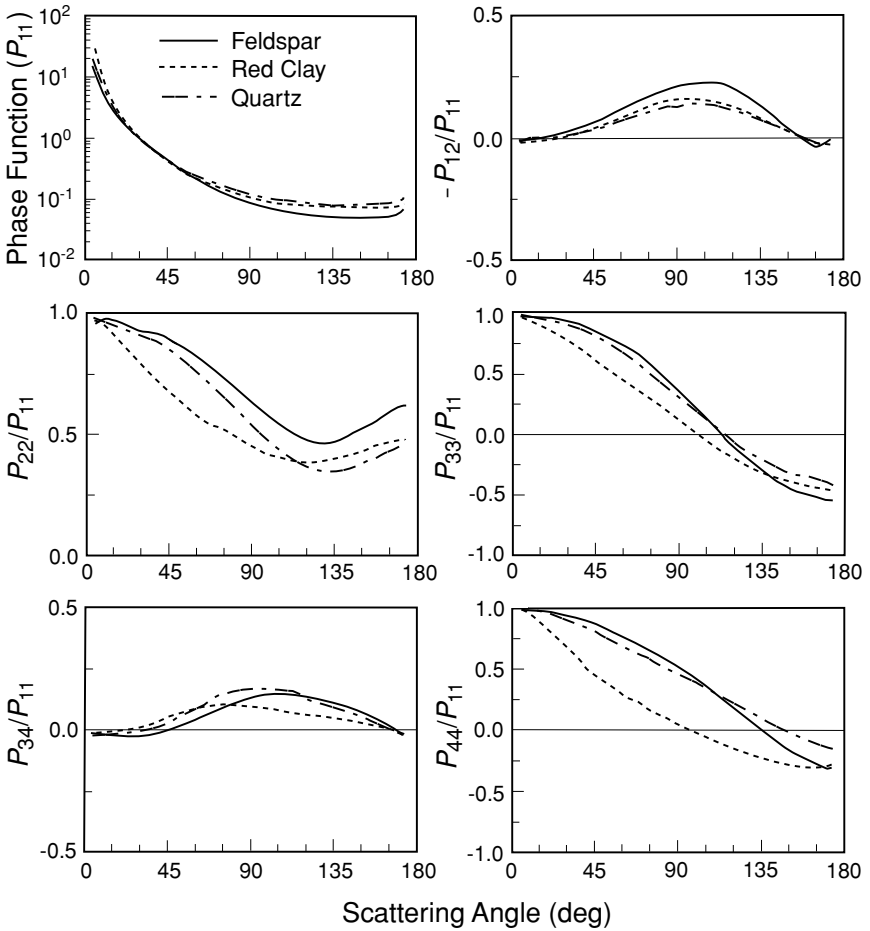
Figure 5.31   Measured scattering matrix elements versus scattering angle in a laboratory setting using a laser beam with $\lambda = 0.633$ μm. The solid, dotted, and dash-dotted lines correspond to the results for feldspar, red clay, and quartz, respectively (data taken from Hovenier, 2000).

goes through the other polarizer and is measured by a receiving antenna. The method allows wide coverage of scattering angles, including the exact forward direction, as well as a greater degree of control of the object's size, shape, and orientation than conventional optical measurements. Figure 5.32 illustrates an example of phase function comparison between the theoretical results computed from the FDTD method and the experimental data derived from a microwave analog experiment. The measurements involved randomly oriented convex and concave particles with a refractive index of $m = 1.5 + i0.005$ and size parameters ranging from 5.9 to 17.8. In the FDTD calculation, a mixture of 50% convex and 50% concave particle shape was used. The phase function for this combination of particle shape appears to match the experimental results. Also shown for comparison is the phase function for the equivalent volume
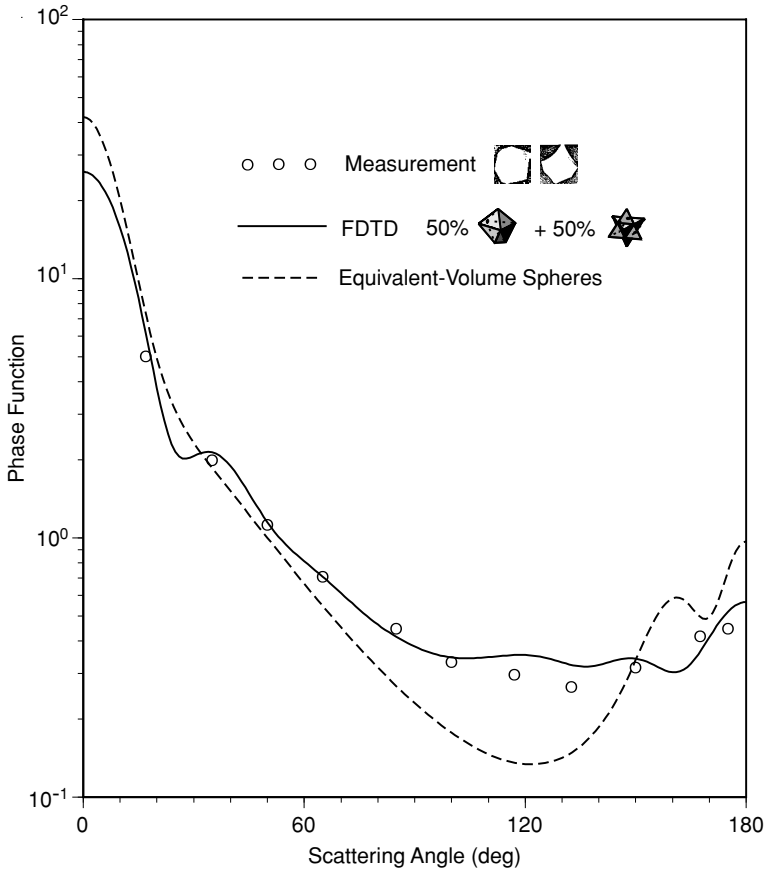
Figure 5.32   Phase function measured by the microwave analog technique (Zerull, 1976) and computed from the FDTD method for randomly oriented convex and concave particles with a refractive index $m = 1.5 + i0.005$ and a size parameter range from 5.9 to 17.8.

sphere computed from the Lorenz–Mie theory. Large deviations are evident between spherical and nonspherical particles, particularly in side scattering directions and the directions associated with the sun–satellite geometry.

Exercises

5.1   Based on the definitions of $\mathbf{M}_\psi$ and $\mathbf{N}_\psi$ in Eqs. (5.2.23) and (5.2.24), show that

$$\boldsymbol{\nabla} \times \mathbf{N}_\psi = mk\mathbf{M}_\psi,$$

and prove that

$$\boldsymbol{\nabla}^2\mathbf{N}_\psi + k^2m^2\mathbf{N}_\psi = 0, \quad \boldsymbol{\nabla}^2\mathbf{M}_\psi + k^2m^2\mathbf{M}_\psi = 0.$$

5.2 From the radial component of the magnetic vector

$$H_r^i = e^{ik\cos\theta}\sin\theta\sin\phi = \frac{i}{k}\left[\frac{\partial^2(rv^i)}{\partial r^2} + k^2(rv^i)\right],$$

show that

$$rv^i = \frac{1}{k}\sum_{n=1}^{\infty}(-i)^n\frac{2n+1}{n(n+1)}\psi_n(kr)P_n^1(\cos\theta)\sin\phi.$$

5.3 The electric and magnetic field vectors in a homogeneous medium satisfy the following vector wave equation:

$$\nabla^2\mathbf{A} + k^2m^2\mathbf{A} = 0.$$

If ψ satisfies the scalar wave equation

$$\nabla^2\psi + k^2m^2\psi = 0,$$

(a) show that vectors $\mathbf{M}_\psi$ and $\mathbf{N}_\psi$ in cylindrical coordinates (r, ϕ, z) defined by

$$\mathbf{M}_\psi = \nabla\times(\mathbf{a}_z\psi),\quad mk\mathbf{N}_\psi = \nabla\times\mathbf{M}_\psi$$

satisfy the vector wave equation, where $\mathbf{a}_z$ is a unit vector in the z direction.
(b) Also prove that

$$\mathbf{E} = \mathbf{M}_v + i\mathbf{N}_u,\quad \mathbf{H} = m(-\mathbf{M}_u + i\mathbf{N}_v)$$

satisfy the Maxwell equations, where u and v are solutions of the scalar wave equation. Write out the expressions for $\mathbf{E}$ and $\mathbf{H}$ in terms of u and v.

5.4 The scalar wave equation in cylindrical coordinates is given by

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\psi}{\partial\phi^2} + \frac{\partial^2\psi}{\partial z^2} + m^2k^2\psi = 0.$$

Utilizing the method of separation of variables, show that the solution can be written as

$$\psi_n(r, \phi, z) = e^{i\omega t}Z_n(jr)e^{in\phi}e^{-ihz},$$

where h is an arbitrary constant, n is an integer, $j = (m^2k^2 - h^2)^{1/2}$, and Z_n is any Bessel function of order n.

5.5 Derive the explicit form of the transformation matrix $\mathbf{F}$ associated with the Stokes parameters in terms of $S_j(j = 1, 2, 3, 4)$, defined in Eq. (5.4.22).

5.6 Using the Fresnel formulas, show explicitly that the transmitted and reflected portions of the energy for the two polarization components are conserved. Also compute the incident angle θ_i at which $R_l = 0$. This angle is called the polarizing or *Brewster* angle, under which the electric vector of the reflected light has no component in the plane of incidence.

5.7 The refractive indices of water at the wavelengths of red (0.656 μm) and violet (0.405 μm) light are 1.332 and 1.344, respectively. (a) Describe the color

sequence of the corona. Find the radius of cloud particles that produce a secondary white corona with a $10°$ angular radius about the sun. (b) Describe the mechanism of cloudbow formation. Find the scattering angles for the primary and secondary cloudbows at these two wavelengths. (c) Find the angular radii of the rings of the halos formed by prism angles of 60 and $90°$. What will be the widths of the rings?

5.8 When absorption is involved, the refractive index $m = m_r - im_i$. Prove that the Fresnel reflection coefficients are given by

$$|R_r|^2 = \frac{(\cos\theta_i - u)^2 + v^2}{(\cos\theta_i + u)^2 + v^2},$$

$$|R_l|^2 = \frac{\left[(m_r^2 - m_i^2)\cos\theta_i - u\right]^2 + (2m_r m_i \cos\theta_i - v)^2}{\left[(m_r^2 - m_i^2)\cos\theta_i + u\right]^2 + (2m_r m_i \cos\theta_i + v)^2},$$

$$u^2 = \frac{1}{2}\left\{m_r^2 - m_i^2 - \sin^2\theta_i + \left[(m_r^2 - m_i^2 - \sin^2\theta_i)^2 + 4m_r^2 m_i^2\right]^{1/2}\right\},$$

$$v^2 = \frac{1}{2}\left\{-(m_r^2 - m_i^2 - \sin^2\theta_i) + \left[(m_r^2 - m_i^2 - \sin^2\theta_i)^2 + 4m_r^2 m_i^2\right]^{1/2}\right\}.$$

To derive these equations, let $m\cos\theta_t = u + iv$ in Eq. (5.3.23) and use the law of refraction $\sin\theta_i = m\sin\theta_t$.

5.9 From the definition of the Stokes parameters in Eq. (5.2.103), show that the scattering phase matrix for Rayleigh scattering [Eq. (3.3.6)] is given by

$$\mathbf{P} = \frac{3}{2}\begin{bmatrix} \frac{1}{2}(1+\cos^2\Theta) & -\frac{1}{2}\sin^2\Theta & 0 & 0 \\ -\frac{1}{2}\sin^2\Theta & \frac{1}{2}(1+\cos^2\Theta) & 0 & 0 \\ 0 & 0 & \cos\Theta & 0 \\ 0 & 0 & 0 & \cos\Theta \end{bmatrix}.$$

5.10 Consider a group of ice plates randomly oriented horizontally. (a) Show that the incident angle of a geometric ray θ_i is related to the elevation angle ε_i and azimuthal angle ϕ_i by $\cos\theta_i = \cos\varepsilon_i \cos\phi_i$. (b) Prove that the general Snell law can be expressed by $m_r \sin\varepsilon_t = \sin\varepsilon_i$, and $m_r(\cos\varepsilon_t/\cos\varepsilon_i)\sin\phi_t = \sin\phi_i$, where (ε_t, ϕ_t) are corresponding refracted angles and m_r is the real index of refraction. (c) Find the angular distance between the sun and two bright spots on its plane, referred to as *sundogs,* using a sun's elevation angle of $30°$. Sketch a diagram of sundogs showing the angular and azimuthal distances and the width of red and violet colors.

5.11 The projection of a hexagonal column onto a horizontal plane is given by a rectangle. Based on the Fraunhofer diffraction theory presented in Eq. (5.3.3), show that the diffraction pattern for this aperture is proportional to

$$\frac{\sin^2(ak\sin\theta\cos\phi)}{(ak\sin\theta\cos\phi)^2} \cdot \frac{\sin^2[(L/2)k\sin\theta\sin\phi]}{[(L/2)k\sin\theta\sin\phi]^2},$$

where L is the length of the column, a is the half-width, and other notations are as defined in Eq. (5.3.3).

5.12 From the vector equation for geometric ray-tracing, show the individual components for external reflection, two refractions, and one internal reflection based on the directional cosine in polar coordinates.

5.13 Derive the adjusted real and imaginary refractive indices given in Eqs. (5.4.16 a,b) from Eqs. (5.4.9c), (5.4.13), and (5.4.14).

5.14 Show that the extinction cross section for a sample of randomly oriented hexagonal ice crystals of the same size is equal to half of the surface area of a hexagonal cylinder. In carrying out this exercise, first derive the following expression for the geometric cross section area of an arbitrarily oriented hexagon:

$$G(\alpha, \beta) = 3\sqrt{3}a^2 \sin \alpha/2 + 2aL \cos \alpha \cos (\pi/6 - \beta),$$

where α and β are orientation angles of a hexagon, L is the length, and $2a$ denotes the width.

5.15 Consider a sample of randomly oriented ice particles each of which has a plane of symmetry and to which the law of reciprocity is applicable. (a) Derive the six relationships for the scattering phase matrix elements by changing the signs of (S_3, S_4), in Eq. (5.4.22a). (b) Derive the additional four relationships by changing the signs of (U, V) in Eq. (5.4.29). (c) Prove that Eq. (5.4.32) is true.

Suggested Reading

Born, M., and Wolf, E. (1975). *Principles of Optics*. Pergamon Press, New York. Chapter 13, Section 13.5, gives a tractable approach to the solution of the vector wave equation for a homogeneous sphere utilizing the Debye potentials up to the point of the derivation of the scattered electric and magnetic fields. The book also presents an elegant treatment of basic and advanced material regarding electromagnetic scattering, geometric optics, and diffraction.

Greenler, R. (1980). *Rainbows, Halos, and Glories*. Cambridge University Press, Cambridge, U.K. A beautiful book that provides fascinating color pictures of various types of halos and arcs.

Kerker, M. (1969). *The Scattering of Light and Other Electromagnetic Radiation*. Academic Press, New York. Chapter 3 gives a comprehensive history of the solution for scattering by a homogeneous sphere. This book also provides a practical discussion of the scattering of light by spheres and cylinders in conjunction with the particle-size determination from scattering information.

Liou, K. N. (1992). *Radiation and Cloud Processes in the Atmosphere: Theory, Observation, and Modeling*. Oxford University Press, New York. Chapter 5 contains pertinent material and the theoretical foundations of light scattering by water droplets and ice crystals.

Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., eds. (2000). *Light Scattering by Nonspherical Particles,* Academic Press, San Diego. This monograph provides a comprehensive and up-to-date overview of the subject of light scattering by nonspherical particles and its applications to science and technology.

Stratton, J. A. (1941). *Electromagnetic Theory.* McGraw-Hill, New York. Chapters 7 and 9 present advanced and organized materials for the solution of vector wave equations in spherical coordinates. (Stratton's text was the first to present this material.) Basic Bessel, Neumann, and Hankel functions and Legendre polynomials are also discussed in some detail in the book.

van de Hulst, H. C. (1957). *Light Scattering by Small Particles.* Wiley, New York. Chapters 8, 9, and 12 provide authoritative analyses of diffraction, Lorenz–Mie scattering theory, and geometric optics, respectively. In Chapters 2–4, fundamental scattering equations are derived by means of physical insight and postulations.

Chapter 6 | Principles of Radiative Transfer in Planetary Atmospheres

6.1 Introduction

6.1.1 A Brief History of Radiative Transfer

The subject of radiative transfer covers a variety of fields, including astrophysics, applied physics and optics, planetary sciences, atmospheric sciences, and meteorology, as well as various engineering disciplines. The notation that we use generally follows that developed by astrophysicists who pioneered the field at the beginning of the 20th century. The work of Schuster (1905) on the investigation of the transfer of radiation through a foggy atmosphere appears to be the first paper discussing the importance of multiple scattering. Schuster included the contribution of both the upward and downward light beams in the formulation, as cited in Section 3.4.2. This is the origin of the two-stream approximation for radiative transfer (Section 6.5.2). The two-beam concept was also employed by Schwarzschild (1906) to explain the limb darkening of the sun by substituting the two-beam solution into an integral equation. In this manner, a continuous distribution of radiation fields could be determined (see Section 4.6.3). In his effort to understand the physical structure of the interior of a star, Eddington (1916) developed an expansion of intensity using two-term Legendre polynominals, leading to the so-called Eddington's approximation for radiative transfer (Section 6.5.2). Furthermore, Schwarzschild (1914) introduced the concept that a medium could experience emission as well as absorption in the context of thermodynamic equilibrium, a subject concerning the transfer of thermal infrared radiation in molecular atmospheres assuming scattering can be neglected (Section 1.4.4). These pioneering papers in radiative transfer appear in a book volume in which Schwarzschild's two papers were translated into English (Menzel, 1966).

Prior to 1950, the subject of radiative transfer was studied principally by astrophysicists, although it was also an important research area in applied physics and nuclear engineering associated with neutron transport. In his landmark book, Chandrasekhar (1950) presented the subject of radiative transfer in plane-parallel atmospheres as a branch of mathematical physics and developed numerous solution methods and techniques, including the consideration of polarization. The principle of radiative transfer

has also been extensively employed by planetary scientists, particularly in association with the remote sensing of planetary atmospheres by means of spectroscopy and polarimetry (Chamberlain and Hunten, 1987). In the first edition of this text, we introduced the subject of radiative transfer with reference to approximation methods and their application to the remote sensing of atmospheric composition and structure.

The underlying principles of radiative transfer have been firmly established in the papers and books cited earlier. However, new avenues, such as the analytic solution for the four-stream approximation, the delta-function adjustment for the phase function, and the efficient incorporation of line absorption in multiple scattering atmospheres, have been explored in recent years. In addition, radiative transfer in clouds consisting of nonspherical and spatially oriented ice crystals that occur in the upper troposphere, and radiative transfer in clouds that are finite and inhomogeneous in nature are both contemporary research subjects.

In Sections 1.1.4, 1.4.4, and 1.4.5, we introduced the concept of multiple scattering and presented the basic equations for radiative transfer. We also derived the source function in terms of the phase function and single-scattering albedo in Section 3.4.1. Following these discussions, we should first discuss here the fundamentals and exact solutions for radiative transfer, including the discrete-ordinates method, the principles of invariance, and the adding method. We will prove that the last two are, in principle, equivalent, and point out that the discrete-ordinates and adding methods are similar in terms of numerical calculations. Subsequently, we present various approximations for radiative transfer and the subject of radiative transfer including polarization. Finally, we discuss a number of advanced topics not covered by the plane-parallel assumption. In what follows, we present a number of fundamental equations for the plane-parallel condition and define the associated physical terms.

6.1.2   Basic Equations for the Plane-Parallel Condition

On the basis of Eqs. (3.4.5) and (3.4.6), the basic scalar equation for the transfer of radiation in plane-parallel atmospheres can be written in the form

$$\mu \frac{dI(\tau, \mathbf{\Omega})}{d\tau} = I(\tau, \mathbf{\Omega}) - \frac{\tilde{\omega}}{4\pi} \int_{4\pi} I(\tau, \mathbf{\Omega}')P(\mathbf{\Omega}, \mathbf{\Omega}') \, d\Omega'$$

$$- \frac{\tilde{\omega}}{4\pi} F_{\odot} P(\mathbf{\Omega}, -\mathbf{\Omega}_0)e^{-\tau/\mu_0} + (1 - \tilde{\omega})B[T(\tau)], \qquad (6.1.1)$$

where $\mathbf{\Omega} = (\mu, \phi)$ and $d\Omega = d\mu \, d\phi$ (other notations are defined in Section 3.4.1).

As shown in Section 3.4.1, the phase function may be numerically expanded in Legendre polynomials with a finite number of terms, N. And, in view of the definition of the scattering angle in Eq. (3.4.7), the phase function may be expressed by

$$P(\mu, \phi; \mu', \phi') = \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell[\mu\mu' + (1 - \mu^2)^{1/2}(1 - \mu'^2)^{1/2} \cos(\phi' - \phi)].$$

$$(6.1.2)$$
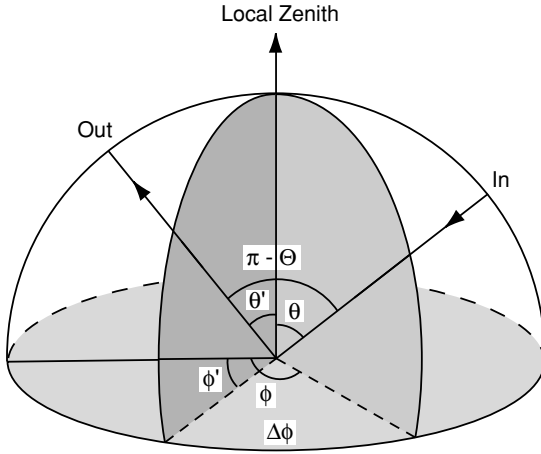
Figure 6.1 Relation of scattering (Θ), zenith (θ, θ'), and azimuthal angles (ϕ, ϕ') in a spherical atmosphere. In the discussion of multiple scattering of a light beam, the notation Θ is usually used to denote the scattering angle, while θ is used for the emergent angles. Note that in Chapter 5 dealing with single-scattering processes, θ is employed as the scattering angle.

The geometric relationship of the scattering, zenith, and azimuthal angles is shown in Fig. 6.1. The Legendre polynomials for the argument shown in Eq. (6.1.2) can be expanded by the addition theorem for spherical harmonics (see Appendix E) to give

$$P(\mu, \phi; \mu', \phi') = \sum_{m=0}^{N} \sum_{\ell=m}^{N} \tilde{\omega}_\ell^m P_\ell^m(\mu) P_\ell^m(\mu') \cos m(\phi' - \phi), \qquad (6.1.3a)$$

where

$$\tilde{\omega}_\ell^m = (2 - \delta_{0,m}) \tilde{\omega}_\ell \frac{(\ell - m)!}{(\ell + m)!} \qquad (\ell = m, \ldots, N, \quad 0 \le m \le N), \quad (6.1.3b)$$

$$\delta_{0,m} = \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{otherwise}, \end{cases} \qquad (6.1.3c)$$

and P_ℓ^m denote the associated Legendre polynomials.

In view of the phase function expansion, we may also expand the intensity in the form

$$I(\tau; \mu, \phi) = \sum_{m=0}^{N} I^m(\tau, \mu) \cos m(\phi_0 - \phi). \qquad (6.1.4)$$

On inserting Eqs. (6.1.3a) and (6.1.4) into Eq. (6.1.1) and noting the orthogonality of the associated Legendre polynomials, Eq. (6.1.1) splits up into $(N + 1)$ independent

equations:

$$\mu \frac{dI^m(\tau, \mu)}{d\tau} = I^m(\tau, \mu) - (1 + \delta_{0,m})\frac{\tilde{\omega}}{4} \sum_{\ell=m}^{N} \tilde{\omega}_\ell^m P_\ell^m(\mu) \int_{-1}^{1} P_\ell^m(\mu') I^m(\tau, \mu') \, d\mu'$$

$$- \frac{\tilde{\omega}}{4\pi} \sum_{\ell=m}^{N} \tilde{\omega}_\ell^m P_\ell^m(\mu) P_\ell^m(-\mu_0) F_\odot e^{-\tau/\mu_0}$$

$$+ \delta_{0,m}(1 - \tilde{\omega}) B[T(\tau)] \qquad (m = 0, 1, \ldots, N). \tag{6.1.5}$$

Each equation may be solved independently for I^m. Consequently, from Eq. (6.1.4), I may be determined.

For $m = 0$, the intensity expressed in Eq. (6.1.4) corresponds to the azimuthal-independent case. We shall omit the superscript 0 for simplicity and rewrite Eq. (6.1.5) to yield

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{\tilde{\omega}}{2} \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell(\mu) \int_{-1}^{1} P_\ell(\mu') I(\tau, \mu') \, d\mu'$$

$$- \frac{\tilde{\omega}}{4\pi} \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell(\mu) P_\ell(-\mu_0) F_\odot e^{-\tau/\mu_0} + (1 - \tilde{\omega}) B[T(\tau)]. \tag{6.1.6}$$

For scattering atmospheres, the diffuse upward and downward solar flux densities are given, respectively, by

$$F_{\text{dif}}^\uparrow(\tau) = \int_0^{2\pi} \int_0^1 I(\tau; \mu, \phi) \mu \, d\mu \, d\phi, \qquad \mu \geq 0, \tag{6.1.7a}$$

$$F_{\text{dif}}^\downarrow(\tau) = \int_0^{2\pi} \int_0^{-1} I(\tau; \mu, \phi) \mu \, d\mu \, d\phi, \qquad \mu \leq 0. \tag{6.1.7b}$$

Thus, by noting that

$$\int_0^{2\pi} \cos m(\phi_0 - \phi) \, d\phi = 0, \qquad m \neq 0, \tag{6.1.8}$$

in Eq. (6.1.4), we obtain the upward and downward solar flux densities as follows:

$$F_{\text{dif}}^{\uparrow\downarrow}(\tau) = 2\pi \int_0^{\pm 1} I(\tau, \mu) \mu \, d\mu. \tag{6.1.9}$$

Consequently, for calculations of solar fluxes in the atmosphere, the azimuthal dependence of the intensity expansion can be neglected, and Eq. (6.1.6) is sufficient for radiation studies.

Moreover, for azimuthal independent cases, we may define the phase function

$$P(\mu, \mu') = \frac{1}{2\pi} \int_0^{2\pi} P(\mu, \phi; \mu', \phi') \, d\phi'. \tag{6.1.10}$$

In view of the phase function expansion represented by Eq. (6.1.3a), we have

$$P(\mu, \mu') = \sum_{\ell=0}^{N} \tilde{\omega}_\ell \, P_\ell(\mu) P_\ell(\mu').$$

(6.1.11)

By virtue of this equation, the azimuthally independent transfer equation for diffuse radiation expressed in Eq. (6.1.6) can be rewritten as follows:

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{\tilde{\omega}}{2} \int_{-1}^{1} I(\tau, \mu') P(\mu, \mu') \, d\mu'$$

$$- \frac{\tilde{\omega}}{4\pi} F_\odot P(\mu, -\mu_0) e^{-\tau/\mu_0} + (1 - \tilde{\omega}) B[T(\tau)].$$

(6.1.12)

We use the positive and negative μ's to denote the upward and downward light beams, as is evident in Eqs. (6.1.7a) and (6.1.7b). Hence the μ_0's that denote the direct solar radiation component are negative values. However, a positive μ_0 has been used for convenience, and $-\mu_0$ represents the fact that the direct solar radiation is downward. For the transfer of terrestrial infrared radiation in scattering atmospheres that are in local thermodynamic equilibrium, the direct solar term involving $F_\odot$ does not appear (Section 4.6).

For solar radiation, the equation of radiative transfer, however, only describes the diffuse component, i.e., light beams scattered more than once, and so we must include the direct component to account for the downward radiation. This is given by the simple Beer–Bouguer–Lambert law for extinction in the form

$$F_{\text{dir}}^{\downarrow}(\tau) = \mu_0 F_\odot e^{-\tau/\mu_0}.$$

(6.1.13)

The total upward and downward flux densities at a given τ are, respectively,

$$F^{\uparrow}(\tau) = F_{\text{dif}}^{\uparrow}(\tau) = 2\pi \int_{0}^{1} I(\tau, \mu) \mu \, d\mu,$$

(6.1.14a)

$$F^{\downarrow}(\tau) = F_{\text{dif}}^{\downarrow}(\tau) + F_{\text{dir}}^{\downarrow}(\tau) = 2\pi \int_{0}^{-1} I(\tau, \mu) \mu \, d\mu + \mu_0 F_\odot e^{-\tau/\mu_0}.$$

(6.1.14b)

The net flux density for a given level is, therefore,

$$F(\tau) = F^{\downarrow}(\tau) - F^{\uparrow}(\tau).$$

(6.1.14c)

(See also Section 3.5 for a discussion of solar net flux and heating rate.) In the presentation of multiple scattering in planetary atmospheres, the thermal infrared emission term will be omitted.

6.2   Discrete-Ordinates Method for Radiative Transfer

The discrete-ordinates method for radiative transfer was elegantly developed by Chandrasekhar (1950) for application to the transfer of radiation in planetary atmospheres. Liou (1973a) demonstrated that the discrete-ordinates method is a useful and

powerful method for the computation of radiation fields in aerosol and cloudy atmospheres. The method involves the discretization of the basic radiative transfer equation and the solution of a set of first-order differential equations. With the advance in numerical techniques for solving differential equations, the discrete-ordinates method has been found to be efficient and accurate for calculations of scattered intensities and fluxes. In presenting the fundamentals of this method, we first discuss the case involving isotropic scattering, including the law of diffuse reflection for semi-infinite atmospheres. That is followed by the general solution for anisotropic scattering. Finally, we present the application of the discrete-ordinates method for radiative transfer to nonhomogeneous atmospheres.

6.2.1   General Solution for Isotropic Scattering

For simplicity in introducing the discrete-ordinates method for radiative transfer, we shall first assume isotropic scattering, i.e., the scattering phase function $P(\mu, \phi; \mu', \phi') = 1$. In this case, the azimuthally independent intensity may be defined by

$$I(\tau, \mu) = \frac{1}{2\pi} \int_0^{2\pi} I(\tau; \mu, \phi) \, d\phi. \tag{6.2.1}$$

By neglecting the emission term, the equation of transfer given in Eq. (6.1.12) becomes

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{\tilde{\omega}}{2} \int_{-1}^{1} I(\tau, \mu') \, d\mu' - \frac{\tilde{\omega} F_\odot}{4\pi} e^{-\tau/\mu_0}. \tag{6.2.2}$$

Now replacing the integral by a summation, according to Gauss's formula (see Section 6.5.2 for definition) and setting $I_i = I(\tau, \mu_i)$, we obtain

$$\mu_i \frac{dI_i}{d\tau} = I_i - \frac{\tilde{\omega}}{2} \sum_j I_j a_j - \frac{\tilde{\omega} F_\odot}{4\pi} e^{-\tau/\mu_0}, \qquad i = -n, \dots, n, \tag{6.2.3}$$

where Σ_j denotes summation from $-n$ to n, i.e., $2n$ terms.

The solution of Eq. (6.2.3) may be derived by seeking first the general solution for the homogeneous part of the differential equation and then adding a particular solution. For the homogeneous part of the differential equation, we set

$$I_i = g_i e^{-k\tau}, \tag{6.2.4}$$

where g_i and k are constants. Substituting Eq. (6.2.4) into the homogeneous part of Eq. (6.2.3), we find

$$g_i(1 + \mu_i k) = \frac{\tilde{\omega}}{2} \sum_j a_j g_j. \tag{6.2.5}$$

This implies that g_i must be in the following form with a constant L:

$$g_i = L/(1 + \mu_i k). \tag{6.2.6}$$

With the expression of g_i given by Eq. (6.2.6), we obtain the characteristic equation for the determination of the eigenvalue k as follows:

$$1 = \frac{\tilde{\omega}}{2} \sum_j \frac{a_j}{1 + \mu_j k} = \tilde{\omega} \sum_{j=1}^n \frac{a_j}{1 - \mu_j^2 k^2}. \tag{6.2.7}$$

For $\tilde{\omega} < 1$, Eq. (6.2.7) admits $2n$ distinct nonzero eigenvalues, which occur in pairs as $\pm k_j (j = 1, \ldots, n)$. Thus, the general solution for the homogeneous part is

$$I_i = \sum_j \frac{L_j}{1 + \mu_i k_j} e^{-k_j \tau}. \tag{6.2.8}$$

For a particular solution, we assume the following form:

$$I_i = \frac{\tilde{\omega} F_\odot}{4\pi} h_i e^{-\tau/\mu_0}, \tag{6.2.9}$$

where h_i are constants. Inserting Eq. (6.2.9) into Eq. (6.2.3) leads to

$$h_i(1 + \mu_i/\mu_0) = \frac{\tilde{\omega}}{2} \sum_j a_j h_j + 1. \tag{6.2.10}$$

Hence, the constants h_i must be in the form

$$h_i = \gamma/(1 + \mu_i/\mu_0), \tag{6.2.11}$$

where γ can be determined from Eq. (6.2.10) with the form

$$\gamma = 1 \left/ \left[1 - \tilde{\omega} \sum_{j=1}^n a_j \left/ \left(1 - \mu_j^2/\mu_0^2\right) \right. \right] \right. . \tag{6.2.12}$$

Adding the general and particular solutions, we obtain

$$I_i = \sum_j \frac{L_j}{1 + \mu_i k_j} e^{-k_j \tau} + \frac{\tilde{\omega} F_\odot \gamma}{4\pi(1 + \mu_i/\mu_0)} e^{-\tau/\mu_0}, \qquad i = -n, \ldots, n. \tag{6.2.13}$$

The unknown coefficients of proportionality L_j are determined from the boundary conditions imposed.

The next step is to introduce Chandrasekhar's H function to replace the constant γ. Consider the function

$$T(z) = 1 - \frac{\tilde{\omega} z}{2} \sum_j \frac{a_j}{z + \mu_j} = 1 - \tilde{\omega} z^2 \sum_{j=1}^n \frac{a_j}{z^2 - \mu_j^2}, \tag{6.2.14}$$

which is a polynomial of degree $2n$ in z. We then compare this equation with the characteristic equation (6.2.7) and find that $z = \pm 1/k_j$ for $T(z) = 0$. Thus, we must have

$$\prod_{j=1}^n \left(z^2 - \mu_j^2\right) T(z) = \text{const} \prod_{j=1}^n \left(1 - k_j^2 z^2\right), \tag{6.2.15}$$

since the two polynomials of degree $2n$ have the same zeros. For $z = 0$, we find that

$$\text{const} = \prod_{j=1}^{n} \left(-\mu_j^2 \right).$$

Thus,

$$T(z) = (-1)^n \mu_1^2 \ldots \mu_n^2 \prod_{j=1}^{n} \left(1 - k_j^2 z^2 \right) \bigg/ \prod_{j=1}^{n} \left(z^2 - \mu_j^2 \right). \qquad (6.2.16)$$

The H function is defined by

$$H(\mu) = \frac{1}{\mu_1 \ldots \mu_n} \frac{\prod_{j=1}^{n}(\mu + \mu_j)}{\prod_{j=1}^{n}(1 + k_j \mu)}. \qquad (6.2.17)$$

In terms of the H function, we have

$$\gamma = 1/T(\mu_0) = H(\mu_0) H(-\mu_0). \qquad (6.2.18)$$

It follows that the complete solution to the isotropic, nonconservative radiative transfer equation in the nth approximation can now be expressed by

$$I_i = \sum_j \frac{L_j}{1 + \mu_i k_j} e^{-k_j \tau} + \frac{\tilde{\omega} F_{\odot} H(\mu_0) H(-\mu_0)}{4\pi (1 + \mu_i/\mu_0)} e^{-\tau/\mu_0}. \qquad (6.2.19)$$

Figure 6.2 illustrates a distribution of eigenvalues for isotropic scattering with a single-scattering albedo $\tilde{\omega} = 0.95$ using four discrete streams. The characteristic equation denoted in Eq. (6.2.7) can be written in the form

$$f(k) = 1 - \tilde{\omega} \sum_{j=1}^{n} \frac{a_j}{1 - \mu_j^2 k^2}. \qquad (6.2.20)$$

For $\tilde{\omega} \neq 0$, $f(k_j) \to \pm\infty$, as $k_j \to \mu_j^{-1}$. In this figure, the same intervals between each μ_j^{-1} were divided so that lines across the zeros can be clearly identified. The eigenvalues occur in pairs and there exists one, and only one, eigenvalue in each interval that can be mathematically proven. The eigenvalue in the discrete-ordinates method for radiative transfer may be physically interpreted as an effective extinction coefficient that, when multiplied by the normal optical depth, represents an effective optical path length in each discrete stream.

For conservative scattering, $\tilde{\omega} = 1$, we note that the characteristic equation (6.2.7) admits two zero eigenvalues, namely, $k^2 = 0$. Based on the relation

$$\sum_{j=-n}^{n} a_j \mu_j^\ell = \int_{-1}^{1} \mu^\ell \, d\mu = 2\delta_\ell/(2\ell + 1), \qquad \delta_\ell = \begin{cases} 1 & \text{even,} \\ 0 & \text{odd,} \end{cases}$$

we can show that

$$I_i = b(\tau + \mu_i + Q) \qquad (6.2.21)$$
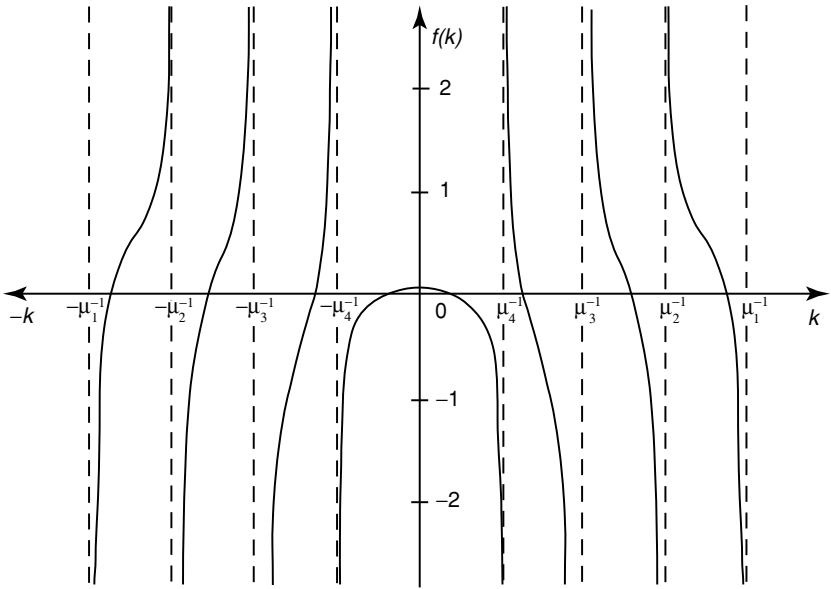
Figure 6.2   A typical distribution of eigenvalues (intercepts on the k-axis) for isotropic scattering with a single-scattering albedo of 0.95 for equal intervals between each μ_j^{-1} ($j = \pm 1, \pm 2, \pm 3, \pm 4$).

satisfies the homogeneous part of the differential equation, where b and Q are two arbitrary constants of integration. Thus, the complete solution to the isotropic radiative transfer equation in the nth approximation may be written as follows:

$$I_i = \sum_{j=-(n-1)}^{n-1} \frac{L_j}{1 + \mu_i k_j} e^{-k_j \tau} + (\tau + \mu_i) L_{-n} + L_n + \frac{F_\odot H(\mu_0) H(-\mu_0)}{4\pi(1 + \mu_i/\mu_0)} e^{-\tau/\mu_0}.$$

$$(6.2.22)$$

6.2.2   The Law of Diffuse Reflection for Semi-infinite Isotropic Scattering Atmospheres

Let us consider that there is no diffuse downward and upward radiation at the top ($\tau = 0$) and bottom ($\tau = \tau_1$) of a semi-infinite atmosphere so that

$$I(0, -\mu_i) = 0, \qquad I(\tau_1, +\mu_i) = 0. \qquad (6.2.23)$$

Inserting the second boundary condition into the solution for the isotropic radiative transfer equation in the nth approximation denoted in Eq. (6.2.19), we obtain

$$I(\tau_1, +\mu_i) = 0 = \sum_{j=-n}^{-1} \frac{L_j}{1 - \mu_i k_j} e^{k_j \tau_1}, \qquad i = 1, \dots, n. \qquad (6.2.24)$$

In order to satisfy the boundary condition for a semi-infinite atmosphere, we must have $L_j = 0$ ($j = -n, \ldots, -1$). Thus,

$$I(\tau, \mu_i) = \sum_{j=1}^{n} \frac{L_j}{1 + \mu_i k_j} e^{-k_j \tau} + \frac{\tilde{\omega} F_{\odot} H(\mu_0) H(-\mu_0)}{4\pi(1 + \mu_i/\mu_0)} e^{-\tau/\mu_0}. \quad (6.2.25)$$

For convenience of analysis, we define

$$S(\mu) = \sum_{j=1}^{n} \frac{L_j}{1 - k_j \mu} + \frac{\tilde{\omega} F_{\odot} H(\mu_0) H(-\mu_0)}{4\pi(1 - \mu/\mu_0)}. \quad (6.2.26)$$

Thus,

$$S(\mu_i) = I(0, -\mu_i) = 0, \qquad i = 1, \ldots, n, \quad (6.2.27)$$

and the reflected intensity

$$I(0, \mu) = S(-\mu). \quad (6.2.28)$$

Moreover, we consider the function

$$(1 - \mu/\mu_0) \prod_{j=1}^{n} (1 - k_j \mu) S(\mu), \quad (6.2.29)$$

which is a polynomial of degree n in μ and vanishes for $\mu = \mu_i, i = 1, \ldots, n$. Hence, this function must be equal to $\prod_{j=1}^{n} (\mu - \mu_j)$ apart from a constant value. We may write

$$(1 - \mu/\mu_0) \prod_{j=1}^{n} (1 - k_j \mu) S(\mu) = \text{const} \cdot \frac{(-1)^n}{\mu_1 \ldots \mu_n} \prod_{j=1}^{n} (\mu - \mu_j). \quad (6.2.30)$$

Upon employing the definition of the H function defined in Eq. (6.2.17), we obtain

$$S(\mu) = \text{const} \cdot H(-\mu)/(1 - \mu/\mu_0). \quad (6.2.31)$$

To obtain the value of the constant, we note that

$$\lim_{\mu \to \mu_0} (1 - \mu/\mu_0) S(\mu) = \text{const} \cdot H(-\mu_0). \quad (6.2.32)$$

But from Eq. (6.2.26), we have

$$\lim_{\mu \to \mu_0} (1 - \mu/\mu_0) S(\mu) = \frac{1}{4\pi} \tilde{\omega} F_{\odot} H(\mu_0) H(-\mu_0). \quad (6.2.33)$$

Comparing Eq. (6.2.32) and Eq. (6.2.33), we find that

$$\text{const} = \frac{1}{4\pi} \tilde{\omega} F_{\odot} H(\mu_0), \quad (6.2.34)$$

and

$$S(\mu) = \frac{\tilde{\omega} F_{\odot} H(\mu_0) H(-\mu)}{4\pi(1 - \mu/\mu_0)}. \quad (6.2.35)$$

The reflected intensity for a semi-infinite, isotropic scattering atmosphere is then given by

$$I(0, \mu) = S(-\mu) = \frac{1}{4\pi} \tilde{\omega} F_\odot \frac{\mu_0}{\mu + \mu_0} H(\mu_0) H(\mu). \tag{6.2.36}$$

Thus, the diffuse reflection can be expressed in terms of the H function. This simple expression has been used to interpret the absorption line formation in cloudy atmospheres of other planets. Exercise 6.1 illustrates the applicability of the law of diffuse reflection.

6.2.3   General Solution for Anisotropic Scattering

To solve the general radiative transfer equation defined in Eq. (6.1.5), we first seek the solution for the homogeneous part of the differential equation and then add a particular solution for the inhomogeneous part. After some mathematical manipulation, the equation can be presented in the form

$$I^m(\tau, \mu_i) = \sum_j L_j^m \varphi_j^m(\mu_i) e^{-k_j^m \tau} + Z^m(\mu_i) e^{-\tau/\mu_0}, \tag{6.2.37}$$

where the eigenfunction derived from the associated homogeneous system is

$$\varphi_j^m(\mu_i) = \frac{1}{1 + \mu_i k_j^m} \sum_{\ell=0}^{N} \tilde{\omega}_\ell^m \xi_\ell^m P_\ell^m(\mu_i), \tag{6.2.38}$$

and the Z function is

$$Z^m(\mu_i) = \frac{1}{4\pi} \tilde{\omega} F_\odot P_m^m(-\mu_0) \frac{H^m(\mu_0) H^m(-\mu_0)}{1 + \mu_i/\mu_0} \sum_{\ell=0}^{N} \tilde{\omega}_\ell^m \xi_\ell^m \left(\frac{1}{\mu_0}\right) P_\ell^m(\mu_i). \tag{6.2.39}$$

The ξ function has the recursion form

$$\xi_{\ell+1}^m = -\frac{2\ell + 1 - \tilde{\omega}_\ell}{k(\ell - m + 1)} \xi_\ell^m - \frac{\ell + m}{\ell - m + 1} \xi_{\ell-1}^m. \tag{6.2.40}$$

Finally, the eigenvalues k_j^m can be determined from the characteristic equation described by

$$1 = \frac{\tilde{\omega}}{2} \sum_j \frac{a_j}{1 + \mu_j k} \left[\sum_{\lambda=m}^{N} \tilde{\omega}_\lambda^m \xi_\lambda^m(k) P_\lambda^m(\mu_j) P_m^m(\mu_j) \right]. \tag{6.2.41}$$

Equation (6.2.41) is of order n in k^2 and admits, in general, $2n$ distinct nonvanishing eigenvalues that must occur in pairs. For strong anisotropic scattering having a sharp phase function, a number of eigenvalues are normally contained in the interval $(0, \mu_n^{-1})$ in which the eigenvalue pattern is highly unsymmetric and differs substantially from that displayed in Fig. 6.2.

The unknown coefficients L_j^m must be determined from the radiation boundary conditions. For simple boundary conditions given by Eq. (6.2.23), and in view of the intensity expansion in Eq. (6.1.4), we have

$$\left.\begin{array}{l} I^m(0, -\mu_i) = 0 \\ I^m(\tau_1, \mu_i) = 0 \end{array}\right\} \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad m = 0, \ldots, N. \quad (6.2.42)$$

We may then determine L_j^m m times independently with the final result given by Eq. (6.1.4). At this point, the analytic solution for Eq. (6.1.5) is complete.

The solution expressed in Eq. (6.2.37) is valid only for nonconservative scattering because when $\tilde{\omega} = 1$, $k^2 = 0$ will satisfy the characteristic equation for $m = 0$, and $\xi_\ell^0(k)$ becomes indefinite. Thus, a different solution must be derived. Since there is no absorption for conservative scattering, the flux of radiation normal to the plane of stratification is constant. It can be shown that the transfer equation admits a solution of the form, for $m = 0$,

$$I^0(\tau, \mu_i) = \sum_{j=-(n-1)}^{n-1} L_j^0 \varphi_j^0(\mu_i) e^{-k_j^0 \tau} + [(1 - \tilde{\omega}_1/3)\tau + \mu_i] L_{-n}^0 + L_n^0 + Z^0(\mu_i)^{-\tau/\mu_0}.$$

$$(6.2.43)$$

A mathematical procedure has been developed to compute the eigenvalues k_j^m from a recurrence characteristic equation (Chandrasekhar, 1950). The eigenvectors $\varphi_j^m(\mu_i)$ may be expressed in terms of known functions, which contain the eigenvalues, while the particular solution is related to the known H function. The characteristic equation for the eigenvalues derived by Chandrasekhar is, however, mathematically as well as numerically ambiguous. The method is unstable for highly peaked phase functions, as pointed out by Liou (1973a), who discovered that the solution of the characteristic equation may be formulated as an algebraic eigenvalue problem. Further, Asano (1975) has shown that the degree of the characteristic equation for the eigenvalues can be reduced by a factor of 2 because the solution for the eigenvalues may be obtained by solving a characteristic polynomial of degree n for k^2. Both of these authors have expanded the matrix in polynomial form to solve the characteristic equation for the eigenvalues corresponding to the associated homogeneous system of the differential equations. However, the expansion in polynomial form is not a stable numerical scheme for obtaining eigenvalues. To solve the algebraic eigenvalue problem, a well-developed numerical subroutine found in the *IMSL User's Manual* (1987) can be used to compute the eigenvalues and eigenvectors of a real general matrix in connection with the discrete-ordinates method. Stamnes and Dale (1981) have shown that azimuthally dependent scattered intensities can be computed accurately using numerical methods.

In the discrete-ordinates method for radiative transfer, analytical solutions for diffuse intensity are explicitly given for any optical depth. Thus, the internal radiation field can be evaluated without additional computational effort. Moreover, analytic two- and four-stream approximations can be developed from this method for flux calculations. In the following, we present a matrix formulation of the discrete-ordinates

method. We shall do so by considering the azimuth-independent component in the diffuse intensity component. On replacing the integral with a summation and omitting the emission term, Eq. (6.1.12) may be written in the form

$$\mu_i \frac{dI(\tau, \mu_i)}{d\tau} = I(\tau, \mu_i) - \frac{\tilde{\omega}}{2} \sum_{j=-n}^{n} I(\tau, \mu_j) P(\mu_i, \mu_j) a_j$$

$$- \frac{\tilde{\omega}}{4\pi} F_{\odot} P(\mu_i, -\mu_0) e^{-\tau/\mu_0}, \qquad i = -n, \ldots, n, \quad (6.2.44)$$

where we may select the quadrature weights and points that satisfy $a_{-j} = a_j (\Sigma_j a_j = 2)$ and $\mu_{-j} = -\mu_j$. To simplify this equation, we may define

$$c_{i,j} = \frac{\tilde{\omega}}{2} a_j P(\mu_i, \mu_j) = \frac{\tilde{\omega}}{2} a_j \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell(\mu_i) P_\ell(\mu_j), \qquad j = -n, \ldots -0, \ldots, n,$$

$$(6.2.45)$$

and

$$I(\tau, -\mu_0) = e^{-\tau/\mu_0} F_{\odot}/2\pi, \qquad (6.2.46)$$

where we set $a_{-0} = 1$ and the notation -0 is used to be consistent with the definition $\mu_{-0} = -\mu_0$. On the basis of the definition of Legendre polynomials, we have

$$c_{i,-j} = c_{-i,j}, \qquad c_{-i,-j} = c_{i,j}, \qquad i \neq -0. \qquad (6.2.47)$$

Moreover, we define

$$b_{i,j} = \begin{cases} c_{i,j}/\mu_i, & i \neq j \\ (c_{i,j} - 1)/\mu_i, & i = j. \end{cases} \qquad (6.2.48)$$

It follows that $b_{i,j} = -b_{-i,-j}$, and $b_{i,-j} = -b_{-i,j}$. Using the preceding definitions, Eq. (6.2.44) becomes

$$\frac{dI(\tau, \mu_i)}{d\tau} = \sum_j b_{i,j} I(\tau, \mu_j). \qquad (6.2.49)$$

We may separate the upward and downward intensities in the forms

$$\frac{dI(\tau, \mu_i)}{d\tau} = \sum_{j=1}^{n} b_{i,j} I(\tau, \mu_i) + \sum_{j=0}^{n} b_{i,-j} I(\tau, -\mu_j), \qquad (6.2.50a)$$

$$\frac{dI(\tau, -\mu_i)}{d\tau} = \sum_{j=1}^{n} b_{-i,j} I(\tau, \mu_j) + \sum_{j=0}^{n} b_{-i,-j} I(\tau, -\mu_j). \qquad (6.2.50b)$$

In terms of a matrix representation for the homogeneous part, we write

$$\frac{d}{d\tau} \begin{bmatrix} \mathbf{I}^+ \\ \mathbf{I}^- \end{bmatrix} = \begin{bmatrix} \mathbf{b}^+ & \mathbf{b}^- \\ -\mathbf{b}^- & -\mathbf{b}^+ \end{bmatrix} \begin{bmatrix} \mathbf{I}^+ \\ \mathbf{I}^- \end{bmatrix}, \qquad (6.2.51)$$

where

$$\mathbf{I}^{\pm} = \begin{bmatrix} I(\tau, \pm\mu_1) \\ I(\tau, \pm\mu_2) \\ \vdots \\ I(\tau, \pm\mu_n) \end{bmatrix}, \tag{6.2.52}$$

and $\mathbf{b}^{\pm}$ denotes the elements associated with $b_{i,j}$ and $b_{i,-j}$. Since Eq. (6.2.51) is a first-order differential equation, we may seek a solution in the form

$$\mathbf{I}^{\pm} = \varphi^{\pm} e^{-k\tau}. \tag{6.2.53}$$

Substituting Eq. (6.2.53) into Eq. (6.2.51) leads to

$$\begin{bmatrix} \mathbf{b}^{+} & \mathbf{b}^{-} \\ -\mathbf{b}^{-} & -\mathbf{b}^{+} \end{bmatrix} \begin{bmatrix} \varphi^{+} \\ \varphi^{-} \end{bmatrix} = -k \begin{bmatrix} \varphi^{+} \\ \varphi^{-} \end{bmatrix}. \tag{6.2.54}$$

Equation (6.2.54) may be solved as a standard eigenvalue problem. In the discrete-ordinates method for radiative transfer, the eigenvalues associated with the differential equations are all real and occur in pairs $(\pm k)$, as pointed out previously. This property can also be understood from the symmetry of the $\mathbf{b}$ matrix. Thus, the rank of the matrix may be reduced by a factor of 2. To accomplish this reduction, we rewrite Eq. (6.2.54) in the forms

$$\mathbf{b}^{+}\varphi^{+} + \mathbf{b}^{-}\varphi^{-} = -k\varphi^{+}, \tag{6.2.55a}$$

$$\mathbf{b}^{-}\varphi^{+} + \mathbf{b}^{+}\varphi^{-} = k\varphi^{-}. \tag{6.2.55b}$$

Adding and subtracting these two equations yields

$$(\mathbf{b}^{+} - \mathbf{b}^{-})(\mathbf{b}^{+} + \mathbf{b}^{-})(\varphi^{+} + \varphi^{-}) = k^2(\varphi^{+} + \varphi^{-}). \tag{6.2.56}$$

It follows that the eigenvectors of the original system, $\varphi^{\pm}$, can now be obtained from the reduced system, $(\varphi^{+} + \varphi^{-})$, in terms of the eigenvalue k^2.

As discussed by Chandrasekhar (1950), the Gaussian quadrature formula for the complete angular range, $-1 < \mu < 1$, is efficient and accurate for the discretization of the basic radiative transfer equation. However, the Gaussian quadrature can also be applied separately to the half-ranges, $-1 < \mu < 0$ and $0 < \mu < 1$, which are referred to as *double-Gauss quadrature*. The division appears to offer numerical advantages when upward and downward radiation streams are treated separately.

6.2.4   Application to Nonhomogeneous Atmospheres

One of the fundamental difficulties in radiative transfer involves accounting for the nonhomogeneous nature of the atmosphere, which cannot be represented by a single single-scattering albedo $\tilde{\omega}$ and a phase function P. The radiative transfer equation

for diffuse intensity must be modified to include variations in $\tilde{\omega}$ and P with optical depth.

The discrete-ordinates method for radiative transfer can be applied to nonhomogeneous atmospheres by numerical means (Liou, 1975). In the following analysis, consider the azimuthally independent transfer equation for diffuse radiation as follows:

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{\tilde{\omega}(\tau)}{2} \int_{-1}^{1} I(\tau, \mu')P(\tau; \mu, \mu') \, d\mu'$$

$$- \frac{\tilde{\omega}(\tau)}{4\pi} P(\tau; \mu, -\mu_0) F_\odot e^{-\tau/\mu_0}. \tag{6.2.57}$$

Since $\tilde{\omega}$ and P are functions of optical depth, analytic solutions for this equation are generally not possible. We may, however, devise a numerical procedure to compute the diffuse intensity in nonhomogeneous atmospheres.

As illustrated in Fig. 6.3, the atmosphere may be divided into N homogeneous layers, each of which is characterized by a single-scattering albedo, a phase function, and an extinction coefficient (or optical depth). The solution for the azimuthally independent diffuse intensity ($m = 0$), as given in Eq. (6.2.37), may be written for each individual layer ℓ in the form

$$I^{(\ell)}(\tau, \mu_i) = \sum_j L_j^{(\ell)} \varphi_j^{(\ell)}(\mu_i) e^{-k_j^{(\ell)} \tau} + Z^{(\ell)}(\mu_i) e^{-\tau/\mu_0}, \qquad \ell = 1, 2, \ldots, N. \tag{6.2.58}$$

At the top of the atmosphere (TOA) ($\tau = 0$), there is no downward diffuse flux, so that

$$I^{(1)}(0, -\mu_i) = 0. \tag{6.2.59}$$

Within the atmosphere, the upward and downward intensities must be continuous at the interface of each predivided layer. Thus, we have

$$I^{(\ell)}(\tau_\ell, \mu_i) = I^{(\ell+1)}(\tau_\ell, \mu_i), \tag{6.2.60}$$

where τ_ℓ denotes the optical depth from TOA to the bottom of the ℓ layer. At the bottom of the atmosphere, surface reflectance must be accounted for. The reflection of sunlight at the surface depends significantly on the latter's optical property with respect to the incident wavelength. For example, reflectance from vegetation and soil is highly wavelength-dependent. Also, surface reflectance patterns are generally anisotropic. For the computation of solar fluxes in the atmosphere, however, it is appropriate to use a Lambertian surface. Let the surface albedo be r_s. Then the upward diffuse intensities that are reflected from the surface may be expressed by

$$I^{(N)}(\tau_N, +\mu_i) = \frac{r_s}{\pi} \left[F^\downarrow(\tau_n) + \mu_0 F_\odot e^{-\tau_N/\mu_0} \right], \tag{6.2.61}$$
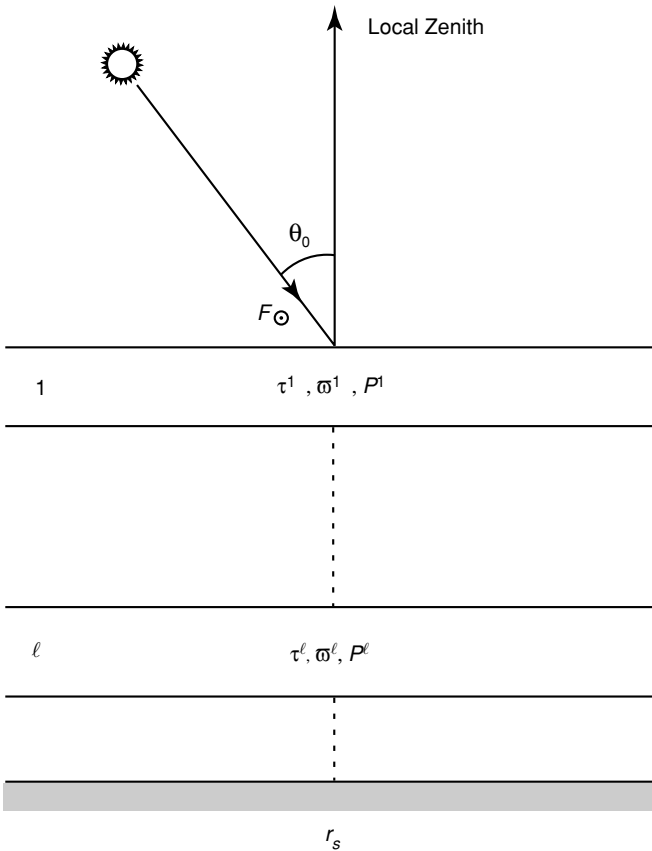
Figure 6.3   A nonhomogeneous atmosphere is divided into ℓ homogeneous layers with respect to the single-scattering albedo $\bar{\omega}$, phase function P, and optical depth τ. $F_{\odot}$ denotes the solar flux at the top of the atmosphere, θ_0 is the solar zenith angle, and r_s is the surface albedo.

where the downward diffuse flux reaching the surface is

$$F^{\downarrow}(\tau_N) = 2\pi \sum_{i=1}^{n} I^{(N)}(\tau_N, -\mu_i)\, a_i \mu_i. \tag{6.2.62}$$

Matching the required boundary and continuity conditions for the diffuse intensities, we obtain the following set of equations for the determination of the unknown coefficients:

$$\sum_{j} L_j^{(1)} \varphi_j^{(1)}(-\mu_i) = -Z^{(1)}(-\mu_i), \qquad i = 1, \dots, n, \tag{6.2.63a}$$

$$\sum_j \left[L_j^{(\ell)} \gamma_j^{(\ell)}(\mu_i) + L_j^{(\ell+1)} \delta_j^{(\ell+1)}(\mu_i) \right] = -{}^{(\ell)}\eta^{(\ell+1)}(\mu_i),$$

$$i = -n, \ldots, -1, 1, \ldots, n, \qquad \ell = 1, 2, \ldots, N-1, \quad \text{(6.2.63b)}$$

$$\sum_j L_j^{(N)} \beta_j^{(N)}(+\mu_i) = -\varepsilon^{(N)}(+\mu_i), \qquad i = 1, \ldots, n, \quad \text{(6.2.63c)}$$

where

$$\gamma_j^{(\ell)}(\mu_i) = \varphi_j^{(\ell)}(\mu_i) e^{-k_j^{(\ell)} \tau_\ell}, \tag{6.2.64a}$$

$$\delta_j^{(\ell+1)}(\mu_i) = -\varphi_j^{(\ell+1)}(\mu_i) e^{-k_j^{(\ell+1)} \tau_\ell}, \tag{6.2.64b}$$

$${}^{(\ell)}\eta^{(\ell+1)}(\mu_i) = \left[Z^{(\ell)}(\mu_i) - Z^{(\ell+1)}(\mu_i) \right] e^{-\tau_\ell/\mu_0}, \tag{6.2.64c}$$

$$\beta_j^{(N)}(+\mu_i) = \left[\varphi_j^{(N)}(+\mu_i) - 2r_s \sum_{i=1}^n \varphi_j^{(N)}(-\mu_i) a_i \mu_i \right] e^{-k_j^{(N)} \tau_N}, \tag{6.2.64d}$$

$$\varepsilon^{(N)}(+\mu_i) = \left[Z^{(N)}(+\mu_i) - 2r_s \sum_{i=1}^n Z^{(N)}(-\mu_i) a_i \mu_i - \frac{r_s}{\pi} \mu_0 F_\odot \right] e^{-\tau_N/\mu_0}. \tag{6.2.64e}$$

Thus, we have $N \times 2n$ equations for the determination of $N \times 2n$ unknown coefficients, $L_j^{(\ell)}$. Equations (6.2.63a)–(6.2.63e) may be expressed in terms of a matrix representation in the form

$$\varphi \mathbf{L} = \chi. \tag{6.2.65a}$$

where

$$\mathbf{L} = \begin{bmatrix} L_{-n}^{(1)} \\ \vdots \\ L_n^{(1)} \\ L_{-n}^{(2)} \\ \vdots \\ L_n^{(2)} \\ \vdots \\ L_{-n}^{(N)} \\ \vdots \\ L_n^{(N)} \end{bmatrix}, \qquad -\chi = \begin{bmatrix} Z^{(1)}(-\mu_n) \\ \vdots \\ Z^{(1)}(-\mu_1) \\ {}^{(1)}\eta^{(2)}(-\mu_n) \\ \vdots \\ {}^{(1)}\eta^{(2)}(+\mu_n) \\ \vdots \\ \varepsilon^{(N)}(+\mu_1) \\ \vdots \\ \varepsilon^{(N)}(+\mu_n) \end{bmatrix}, \tag{6.2.65b}$$

and

$$
\varphi = \begin{bmatrix}
\varphi_{-n}^{(1)}(-\mu_n) & \cdots & \varphi_n^{(1)}(-\mu_n) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
\varphi_{-n}^{(1)}(-\mu_1) & \cdots & \varphi_n^{(1)}(-\mu_1) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\gamma_{-n}^{(1)}(-\mu_n) & \cdots & \gamma_n^{(1)}(-\mu_n) & \delta_{-n}^{(2)}(-\mu_n) & \cdots & \delta_n^{(2)}(-\mu_n) & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
\gamma_{-n}^{(1)}(\mu_n) & \cdots & \gamma_n^{(1)}(\mu_n) & \delta_{-n}^{(2)}(\mu_n) & \cdots & \delta_n^{(2)}(\mu_n) & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \beta_{-n}^{(N)}(\mu_1) & \cdots & \beta_n^{(N)}(\mu_1) \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \beta_{-n}^{(N)}(\mu_n) & \cdots & \beta_n^{(N)}(\mu_n)
\end{bmatrix}.
$$

$$(6.2.65c)$$

Azimuthally dependent components can be formulated in a likely manner to obtain the angular intensity pattern. At this point, however, the discrete-ordinates method for radiative transfer has not been applied to cases involving polarization.

6.3   Principles of Invariance

6.3.1   Definitions of Scattering Parameters

The principles-of-invariance method for radiative transfer seeks certain physical statements and mathematical formulations regarding the fields of reflection and transmission of light beams. In this method, the radiation field is not derived directly from the transfer equation, as in the case of the discrete-ordinates method.

To introduce the principles of invariance and other multiple-scattering problems, it is necessary to define and clarify a number of parameters that have been used in the literature. We find it convenient to express the solutions to multiple-scattering problems in terms of the *reflection function R* and *transmission function T* in the forms

$$
I_r(0; \ \mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\mu, \phi; \mu', \phi') I_\odot(-\mu', \phi') \mu' d\mu' d\phi', \quad (6.3.1a)
$$

$$
I_t(\tau_1; \ -\mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T(\mu, \phi; \mu', \phi') I_\odot(-\mu', \phi') \mu' d\mu' d\phi', \quad (6.3.1b)
$$

where $I_\odot(-\mu, \phi)$ represents the intensity of sunlight incident on the top of the scattering layer. It suffices for most practical problems to approximate the intensity as monodirectional in the form

$$
I_\odot(-\mu, \phi) = \delta(\mu - \mu_0)\delta(\phi - \phi_0)F_\odot, \quad (6.3.2)
$$

where δ is the Dirac delta function, and $F_\odot$ denotes the solar flux density in the direction of the incident beam. Thus, from Eqs. (6.3.1a) and (6.3.1b), we have the definitions of the reflection and transmission functions in the forms

$$R(\mu, \phi; \mu_0, \phi_0) = \pi I_r(0; \mu, \phi)/(\mu_0 F_\odot), \qquad (6.3.3a)$$

$$T(\mu, \phi; \mu_0, \phi_0) = \pi I_t(\tau_1; -\mu, \phi)/(\mu_0 F_\odot). \qquad (6.3.3b)$$

Note here that $I_t(\tau_1; -\mu, \phi)$ represents the diffusely transmitted intensity, which does not include the directly transmitted solar intensity $F_\odot e^{-\tau_1/\mu_0}$. The direct component represents the attenuation of the incident solar beam that penetrates to the level τ_1. In the case where polarization is considered, in which the four Stokes parameters are required, R and T are composed of four rows and four columns and are referred to as *reflection* and *transmission matrices.* The reflection and transmission functions also have been referred to as reflection and transmission coefficients by Ambartzumian (1958) and Sobolev (1975). In satellite remote sensing, a parameter called *bidirectional reflectance,* which is analogous to the reflection function, is frequently used (Section 7.3.1).

On the basis of Eqs. (6.3.3a) and (6.3.3b), we may define the *reflection r* (also called the *local* or *planetary albedo*) and *transmission* (diffuse) t associated with reflected (upward) and transmitted (downward) flux densities in the forms

$$r(\mu_0) = \frac{F_{dif}^\uparrow(0)}{\mu_0 F_\odot} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\mu, \phi; \mu_0, \phi_0)\, \mu\, d\mu\, d\phi, \qquad (6.3.4a)$$

$$t(\mu_0) = \frac{F_{dif}^\downarrow(\tau_1)}{\mu_0 F_\odot} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T(\mu, \phi; \mu_0, \phi_0)\, \mu\, d\mu\, d\phi. \qquad (6.3.4b)$$

Note that the direct transmission is simply $e^{-\tau_1/\mu_0}$. In a similar manner, *absorption* of the atmosphere, bounded by the optical depths of 0 and τ_1, can be obtained from the net flux density divergence that includes the direct transmission component at levels of 0 and τ_1 and normalizes by $\mu_0 F_\odot$. To find the flux of energy reflected by the planet, we consider on the sphere a ring with radius a' and width da', where a' is the projected distance from the center of the disk, as shown in Fig. 6.4. Hence the flux of energy reflected by this ring is given by $r(\mu_0)F_\odot 2\pi a'\, da'$. But $a' = a \sin \theta_0$, and $da' = a \cos \theta_0\, d\theta_0$. Thus, we may write this flux of energy as $2\pi a^2 F_\odot r(\mu_0)\mu_0\, d\mu_0$. The flux of energy reflected by the entire planet is then given by

$$f^\uparrow(0) = 2\pi a^2 F_\odot \int_0^1 r(\mu_0)\, \mu_0\, d\mu_0. \qquad (6.3.5a)$$

The *spherical* (or *global*) *albedo,* which represents the ratio of the flux of energy reflected by the entire planet to the incident energy, is then given by

$$\bar{r} = \frac{f^\uparrow(0)}{\pi a^2 F_\odot} = 2 \int_0^1 r(\mu_0)\, \mu_0\, d\mu_0. \qquad (6.3.5b)$$
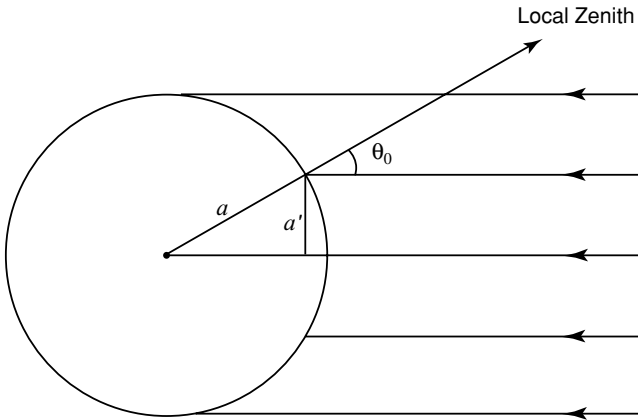
Figure 6.4   Geometry for the definition of the spherical albedo, where a is the radius of the earth and θ_0 is the solar zenith angle.

Likewise, the *global diffuse transmission* is

$$\bar{t} = \frac{f^{\downarrow}(\tau_1)}{\pi a^2 F_\odot} = 2 \int_0^1 t(\mu_0)\, \mu_0\, d\mu_0 \qquad (6.3.5c)$$

and the global direct transmission is simply $2\int_0^1 e^{-\tau_1/\mu_0} \mu_0\, d\mu_0$.

Chandrasekhar (1950) expressed the resulting laws of diffuse reflection and transmission for a finite atmosphere with an optical depth τ_1 in terms of the *scattering function S* and the *transmission function T_c* (diffuse), which differ from the parameters defined in Eqs. (6.3.1a) and (6.3.1b), in the forms

$$I_r(0; \mu, \phi) = \frac{1}{4\pi\mu} \int_0^{2\pi} \int_0^1 S(\mu, \phi; \mu', \phi') I_\odot(-\mu', \phi')\, d\mu' d\phi', \quad (6.3.6a)$$

$$I_t(\tau_1; -\mu, \phi) = \frac{1}{4\pi\mu} \int_0^{2\pi} \int_0^1 T_c(\mu, \phi; \mu', \phi') I_\odot(-\mu', \phi')\, d\mu' d\phi',$$
$$(6.3.6b)$$

where T_c is used to differentiate from T defined previously. On substituting Eq. (6.3.2) into Eqs. (6.3.6a) and (6.3.6b), we obtain the definitions of Chandrasekhar's scattering and transmission functions as follows:

$$S(\mu, \phi; \mu_0, \phi_0) = 4\mu\mu_0\pi\, I_r(0; \mu, \phi)/(\mu_0 F_\odot), \qquad (6.3.7a)$$

$$T_c(\mu, \phi; \mu_0, \phi_0) = 4\mu\mu_0\pi\, I_t(\tau_1; -\mu, \phi)/(\mu_0 F_\odot). \qquad (6.3.7b)$$

The introduction of the factor μ into the intensity parameters gives the required symmetry of S and T_c in the pair of variables (μ, ϕ) and (μ_0, ϕ_0) such that

$$S(\mu, \phi; \mu_0, \phi_0) = S(\mu_0, \phi_0; \mu, \phi), \qquad (6.3.8a)$$

$$T_c(\mu, \phi; \mu_0, \phi_0) = T_c(\mu_0, \phi_0; \mu, \phi). \qquad (6.3.8b)$$

6.3.2 Principles of Invariance for Semi-infinite Atmospheres

Consider a flux of parallel solar radiation $F_\odot$ in a direction defined by $(-\mu_0, \phi_0)$ ($-\mu_0$ denotes that the light beam is downward), incident on the outer boundary of a semiinfinite, plane-parallel atmosphere. The principles of invariance originally introduced by Ambartzumian (1942, 1958) stated that the diffuse reflected intensity from such an atmosphere cannot be changed if a plane layer of finite optical depth having the same optical properties as those of the original atmosphere is added. Let the optical depth of the added layer be $\Delta\tau$, which is so small that $(\Delta\tau)^2$ can be neglected when it is compared with $\Delta\tau$ itself. For simplicity in presenting the principles of invariance, we shall neglect the azimuthal dependence of the diffuse reflected intensity and define the *reflection function* in terms of the diffuse reflected intensity at the top of a semi-infinite atmosphere $I(0, \mu)$ in the form [see Eq. 6.3.3a]

$$R(\mu, \mu_0) = \pi I(0, \mu)/(\mu_0 F_\odot). \tag{6.3.9}$$

In reference to Fig. 6.5, the reduction or increase of the reflection function, after the addition of an infinitesimal layer, can be defined by the following principles:

1. The differential attenuation of the reflection function in passing through $\Delta\tau$ downward, based on the first term on the right-hand side of Eq. (3.4.2) ($\Delta\tau = \beta_e \Delta z$), is given by

$$\Delta R_1' = -R(\mu, \mu_0)\Delta\tau/\mu_0. \tag{6.3.10}$$

The reflection function at $\tau = 0$ is now $(R + \Delta R_1')$, which is again attenuated in passing through $\Delta\tau$ upward. Thus,

$$\Delta R_1'' = -\left[R(\mu, \mu_0) + \Delta R_1'\right]\Delta\tau/\mu. \tag{6.3.11}$$
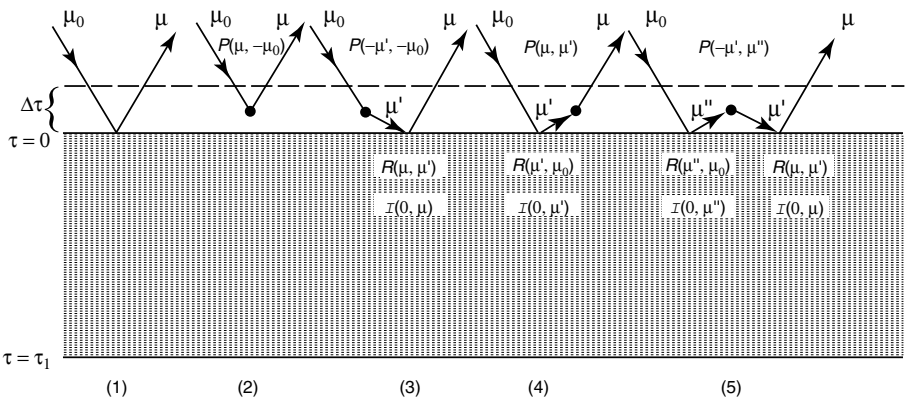
Figure 6.5 The principles of invariance for a semi-infinite, plane-parallel atmosphere. The μ and $-\mu$ denote the upward and downward directions, respectively. The black dots show that scattering events take place in which the phase function is required. The directional representation in the argument is such that the emergent angle is stated first and is then followed by the incident angle. A similar rule governs the argument of the reflection function.

The total attenuation is therefore

$$\Delta R_1 = \Delta R_1' + \Delta R_1'' = -R(\mu, \mu_0)\left[\Delta\tau\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) - \frac{\Delta\tau^2}{\mu\mu_0}\right]$$

$$\approx -R(\mu, \mu_0)\Delta\tau\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right). \qquad (6.3.12)$$

2. However, $\Delta\tau$ may scatter directly in the direction μ a part of the solar flux $F_\odot$ that is incident on it. Based on the second term on the right-hand side of Eq. (3.4.2), we find the additional reflection as follows:

$$\Delta R_2 = \frac{\pi}{\mu_0 F_\odot}\frac{\tilde{\omega}}{4\pi}F_\odot P(\mu, -\mu_0)\Delta\tau/\mu = \frac{\tilde{\omega}}{4}P(\mu, -\mu_0)\Delta\tau/(\mu\mu_0). \quad (6.3.13)$$

3. In addition, $\Delta\tau$ may scatter a part of the solar flux in the direction μ' onto the boundary $\tau = 0$. The diffuse light beam then undergoes reflection from this surface, and this additional reflection, analogous to the third term on the right-hand side of Eq. (3.4.2), is given by

$$\Delta R_3 = \frac{\pi}{\mu_0 F_\odot}\frac{\tilde{\omega}}{4\pi}\int_0^{2\pi} d\phi' \int_0^1 I(0, \mu)P(-\mu', -\mu_0)\, d\mu'\frac{\Delta\tau}{\mu'}$$

$$= \frac{\tilde{\omega}}{2}\frac{\Delta\tau}{\mu_0}\int_0^1 R(\mu, \mu')P(-\mu', -\mu_0)\, d\mu'. \qquad (6.3.14)$$

4. Moreover, $\Delta\tau$, after attenuating a fraction of the light beam diffusely reflected from the boundary $\tau = 0$ in the direction μ', may scatter a part of it in the direction μ. This incremental reflection is given by

$$\Delta R_4 = \frac{\pi}{\mu_0 F_\odot}\frac{\tilde{\omega}}{4\pi}\int_0^{2\pi} d\phi' \int_0^1 P(\mu, \mu')I(0, \mu')\, d\mu'\frac{\Delta\tau}{\mu}$$

$$= \frac{\tilde{\omega}}{2}\frac{\Delta\tau}{\mu}\int_0^1 P(\mu, \mu')R(\mu', \mu_0)\, d\mu'. \qquad (6.3.15)$$

5. Finally, the unscattered component of the solar flux $F_\odot$, which is reflected from the boundary $\tau = 0$ in the direction μ'', is scattered by $\Delta\tau$ back to $\tau = 0$ in the different direction μ', and again is reflected from the surface $\tau = 0$ in the direction μ. This additional contribution may be expressed by

$$\Delta R_5 = \frac{\pi}{\mu_0 F_\odot}\frac{\tilde{\omega}}{4\pi}\int_0^{2\pi} d\phi' \int_0^1 \frac{I(0, \mu)}{F_\odot}\, d\mu'$$

$$\times \left[\int_0^{2\pi} d\phi'' \int_0^1 P(-\mu', \mu'')\, I(0, \mu'')\, d\mu''\right]\frac{\Delta\tau}{\mu'}$$

$$= \tilde{\omega}\Delta\tau \int_0^1 R(\mu, \mu')\, d\mu'\left[\int_0^1 P(-\mu', \mu'')R(\mu'', \mu_0)\, d\mu''\right]. \quad (6.3.16)$$

On the basis of the principles of invariance stated above, we must have

$$\Delta R_1 + \Delta R_2 + \Delta R_3 + \Delta R_4 + \Delta R_5 = 0. \tag{6.3.17}$$

It follows that

$$R(\mu, \mu_0) \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) = \frac{\tilde{\omega}}{4\mu\mu_0} \left\{ P(\mu, -\mu_0) + 2\mu \int_0^1 R(\mu, \mu') P(-\mu', -\mu_0)\, d\mu' \right.$$

$$+ 2\mu_0 \int_0^1 P(\mu, \mu') R(\mu', \mu_0)\, d\mu' + 4\mu\mu_0 \int_0^1 R(\mu, \mu')\, d\mu'$$

$$\left. \times \left[\int_0^1 P(-\mu', \mu'') R(\mu'', \mu_0)\, d\mu'' \right] \right\}. \tag{6.3.18}$$

For a simple case of isotropic scattering, Eq. (6.3.18) becomes

$$R(\mu, \mu_0)(\mu + \mu_0) = \frac{\tilde{\omega}}{4} \left[1 + 2\mu \int_0^1 R(\mu, \mu')\, d\mu' + 2\mu_0 \int_0^1 R(\mu', \mu_0)\, d\mu' \right.$$

$$\left. + 4\mu\mu_0 \int_0^1 R(\mu', \mu_0)\, d\mu' \int_0^1 R(\mu, \mu'')\, d\mu'' \right]$$

$$= \frac{\tilde{\omega}}{4} \left[1 + 2\mu \int_0^1 R(\mu, \mu')\, d\mu' \right] \left[1 + 2\mu_0 \int_0^1 R(\mu', \mu_0)\, d\mu' \right]. \tag{6.3.19}$$

Inspection of Eq. (6.3.19) reveals that if it is satisfied by the function $R(\mu, \mu_0)$, it must also be satisfied by the function $R(\mu_0, \mu)$. And since this equation can have only one solution, we must have

$$R(\mu, \mu_0) \equiv R(\mu_0, \mu). \tag{6.3.20}$$

With this relationship, which is stated here without a rigorous mathematical proof, we may define

$$H(\mu) = 1 + 2\mu \int_0^1 R(\mu, \mu')\, d\mu', \tag{6.3.21}$$

such that

$$R(\mu, \mu_0) = \frac{\tilde{\omega}}{4} \frac{H(\mu)H(\mu_0)}{\mu + \mu_0}. \tag{6.3.22}$$

In reference to Section 6.2.2, we find that this expression is exactly the same as that in Eq. (6.2.36). It is indeed the exact solution for the semi-infinite isotropic scattering atmosphere. To examine the H function, we insert Eq. (6.3.22) into Eq. (6.3.21) to obtain

$$H(\mu) = 1 + \frac{\tilde{\omega}}{2} \mu H(\mu) \int_0^1 \frac{H(\mu')\, d\mu'}{\mu + \mu'}. \tag{6.3.23}$$

It is now clear that the solution of Eq. (6.3.19) is reduced to solving the H function. To do so, we may select an approximate value and then carry out appropriate iterations. We first seek the mean value of H in the form

$$H_0 = \int_0^1 H(\mu)\, d\mu. \tag{6.3.24}$$

From Eq. (6.3.23), we have

$$\int_0^1 H(\mu)\, d\mu = 1 + \frac{\tilde{\omega}}{2} \int_0^1 \int_0^1 \frac{H(\mu)H(\mu')\mu}{\mu + \mu'}\, d\mu\, d\mu'. \tag{6.3.25}$$

On interchanging μ with μ', Eq. (6.3.25) remains the same. Thus, we may write

$$
\begin{aligned}
\int_0^1 H(\mu)\, d\mu &= 1 + \frac{\tilde{\omega}}{4} \int_0^1 \int_0^1 \frac{H(\mu)H(\mu')\mu}{\mu + \mu'}\, d\mu\, d\mu' \\
&\quad + \frac{\tilde{\omega}}{4} \int_0^1 \int_0^1 \frac{H(\mu)H(\mu')\mu'}{\mu + \mu'}\, d\mu\, d\mu' \\
&= 1 + \frac{\tilde{\omega}}{4} \int_0^1 H(\mu)\, d\mu \int_0^1 H(\mu')\, d\mu'.
\end{aligned}
\tag{6.3.26}
$$

Consequently, we obtain

$$H_0 = 1 + \frac{\tilde{\omega}}{4} H_0^2. \tag{6.3.27a}$$

This gives the solution of H_0 in the form

$$H_0 \equiv \int_0^1 H(\mu)\, d\mu = \frac{2}{\tilde{\omega}}\left(1 - \sqrt{1 - \tilde{\omega}}\right), \tag{6.3.27b}$$

where the positive root is found to be unrealistic because the albedo value becomes greater than unity (Exercise 6.2). To find $H(\mu)$ in Eq. (6.3.23), we may insert this zero-order approximation into the right-hand side of this equation to obtain a first approximation. The iterative procedure can be continued until a desirable degree of accuracy is achieved.

6.3.3   Principles of Invariance for Finite Atmospheres

In the last section, we described the principles of invariance for a semi-infinite atmosphere in which only the reflection function is involved. We now introduce the general principles of invariance for a finite atmosphere developed by Chandrasekhar (1950). Consistent with our previous discussions, we shall neglect the azimuthal dependence of the scattering parameters and use the reflection and transmission functions defined in Eqs. (6.3.3a) and (6.3.3b), instead of Chandrasekahr's scattering and transmission functions defined in Eqs. (6.3.7a) and (6.3.7b). We note, however, the relationships

$$S(\mu, \mu_0) = 4\mu\mu_0 R(\mu, \mu_0), \tag{6.3.28a}$$

$$T_c(\mu, \mu_0) = 4\mu\mu_0 T(\mu, \mu_0). \tag{6.3.28b}$$
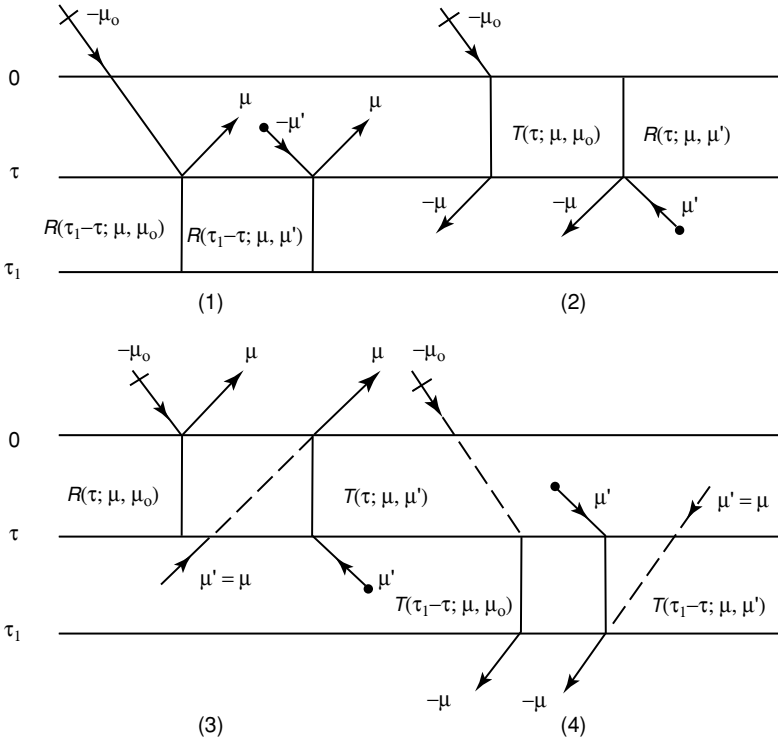
Figure 6.6 The four principles of invariance for a finite atmosphere (see text for explanation).

In reference to Fig. 6.6, we find the following four principles governing the reflection and transmission of a light beam:

1. The reflected (upward) intensity at level τ is caused by the reflection of the attenuated incident solar flux density $F_\odot e^{-\tau/\mu_0}$, and the downward diffuse intensity incident on the surface τ from the finite optical depth $(\tau_1 - \tau)$ below [see Eqs. (6.3.1a) and (6.3.1b)]. Thus, we have

$$I(\tau, \mu) = (\mu_0 F_\odot/\pi)e^{-\tau/\mu_0} R(\tau_1 - \tau; \mu, \mu_0)$$
$$+ 2 \int_0^1 R(\tau_1 - \tau; \mu, \mu')I(\tau, -\mu') \, \mu'd\mu'. \qquad (6.3.29)$$

2. The diffusely transmitted (downward) intensity at level τ is due to the transmission of the incident solar flux density by the optical depth τ above, and the reflection of the upward diffuse intensity incident on the surface τ from below. Thus, we find

$$I(\tau, -\mu) = (\mu_0 F_\odot/\pi)T(\tau; \mu, \mu_0) + 2 \int_0^1 R(\tau; \mu, \mu')I(\tau, \mu') \, \mu'd\mu'. \qquad (6.3.30)$$

3. The reflected (upward) intensity at the top of the finite atmosphere ($\tau = 0$) is a result of the reflection by the optical depth τ of the atmosphere plus the transmission of the upward diffuse and direct intensities incident on the surface τ from below. Thus, we have

$$I(0, \mu) = (\mu_0 F_\odot/\pi)R(\tau; \mu, \mu_0) + 2\int_0^1 T(\tau; \mu, \mu')I(\tau, \mu')\,\mu'd\mu' + e^{-\tau/\mu}I(\tau, \mu).$$

(6.3.31)

4. The diffusely transmitted (downward) intensity at the bottom of the finite atmosphere ($\tau = \tau_1$) is equivalent to the transmission of the attenuated incident solar flux density plus the transmission of the downward diffuse and direct intensities incident on the surface τ from above. Thus, we find

$$I(\tau_1, -\mu) = (\mu_0 F_\odot/\pi)e^{-\tau/\mu_0}T(\tau_1 - \tau; \mu, \mu_0)$$

$$+ 2\int_0^1 T(\tau_1 - \tau; \mu, \mu')I(\tau_1, -\mu')\,\mu'd\mu' + e^{-(\tau_1-\tau)/\mu}I(\tau, -\mu).$$

(6.3.32)

In order to obtain the reflection and transmission functions of a finite atmosphere with an optical depth of τ_1, we first differentiate Eqs. (6.3.29)–(6.3.32) with respect to τ and evaluate the values at $\tau = 0$ and τ_1, where the boundary conditions stated in Eq. (6.2.23) can be applied. After differentiation with respect to τ, we set $\tau = 0$ and $\tau = \tau_1$ for principles 1 and 4 and for principles 2 and 3, respectively, to obtain the equations

$$\left.\frac{dI(\tau, \mu)}{d\tau}\right|_{\tau=0} = (\mu_0 F_\odot/\pi)\left[-\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} - \frac{1}{\mu_0}R(\tau_1; \mu, \mu_0)\right]$$

$$+ 2\int_0^1 \mu'd\mu' R(\tau_1; \mu, \mu')\left.\frac{dI(\tau, -\mu')}{d\tau}\right|_{\tau=0}, \qquad (6.3.33)$$

$$\left.\frac{dI(\tau, -\mu)}{d\tau}\right|_{\tau=\tau_1} = (\mu_0 F_\odot/\pi)\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1}$$

$$+ 2\int_0^1 \mu'd\mu' R(\tau_1; \mu, \mu')\left.\frac{dI(\tau, \mu')}{d\tau}\right|_{\tau=\tau_1}, \qquad (6.3.34)$$

$$\left.\frac{dI(0, \mu)}{d\tau}\right|_{\tau=\tau_1} = 0 = (\mu_0 F_\odot/\pi)\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1}$$

$$+ 2\int_0^1 \mu'd\mu'\, T(\tau_1; \mu, \mu')\left.\frac{dI(\tau, \mu')}{d\tau}\right|_{\tau=\tau_1} + e^{-\tau_1/\mu}\left.\frac{dI(\tau, \mu)}{d\tau}\right|_{\tau=\tau_1},$$

(6.3.35)

$$\left.\frac{dI(\tau_1, -\mu)}{d\tau}\right|_{\tau=0} = 0 = (\mu_0 F_{\odot}/\pi)\left[-\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} - \frac{1}{\mu_0}T(\tau_1; \mu, \mu_0)\right]$$

$$+ 2\int_0^1 \mu' d\mu' T(\tau_1; \mu, \mu')\left.\frac{dI(\tau, -\mu')}{d\tau}\right|_{\tau=0}$$

$$+ e^{-\tau_1/\mu}\left.\frac{dI(\tau, -\mu)}{d\tau}\right|_{\tau=0}. \tag{6.3.36}$$

To eliminate the derivatives of the intensity, we utilize the azimuthally independent transfer equation [Eq. (6.1.12)] to obtain

$$\left.\mu\frac{dI(\tau, \mu)}{d\tau}\right|_{\tau=0} = (\mu_0 F_{\odot}/\pi)R(\tau_1; \mu, \mu_0) - \frac{\tilde{\omega}}{2}\int_0^1 P(\mu, \mu'')I(0, \mu'')\,d\mu''$$

$$- \frac{\tilde{\omega}}{4\pi}F_{\odot}P(\mu, -\mu_0), \tag{6.3.37}$$

$$\left.-\mu\frac{dI(\tau, -\mu)}{d\tau}\right|_{\tau=0} = 0 - \frac{\tilde{\omega}}{2}\int_0^1 P(-\mu, \mu'')I(0, \mu'')\,d\mu'' - \frac{\tilde{\omega}}{4\pi}F_{\odot}P(-\mu, -\mu_0), \tag{6.3.38}$$

$$\left.\mu\frac{dI(\tau, \mu)}{d\tau}\right|_{\tau=\tau_1} = 0 - \frac{\tilde{\omega}}{2}\int_0^1 P(\mu, -\mu'')I(\tau_1, -\mu'')\,d\mu''$$

$$- \frac{\tilde{\omega}}{4\pi}F_{\odot}P(\mu, -\mu_0)e^{-\tau_1/\mu_0}, \tag{6.3.39}$$

$$\left.-\mu\frac{dI(\tau, -\mu)}{d\tau}\right|_{\tau=\tau_1} = (\mu_0 F_{\odot}/\pi)T(\tau_1; \mu, \mu_0) - \frac{\tilde{\omega}}{2}\int_0^1 P(-\mu, -\mu'')I(\tau_1, -\mu'')\,d\mu''$$

$$- \frac{\tilde{\omega}}{4\pi}F_{\odot}P(-\mu, -\mu_0)e^{-\tau_1/\mu_0}. \tag{6.3.40}$$

In these four equations, we note that $\mu \geq 0$. We also note that $\pi I(0, \mu) = \mu_0 F_{\odot} \cdot R(\tau_1; \mu, \mu_0)$, and $\pi I(\tau_1, -\mu) = \mu_0 F_{\odot} T(\tau_1; \mu, \mu_0)$. Upon substituting Eqs. (6.3.37) and (6.3.38), (6.3.39) and (6.3.40), (6.3.39) and (6.3.38), and (6.3.38) into Eqs. (6.3.33)–(6.3.36), respectively, and rearranging the terms, we obtain

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} = -\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)R(\tau_1; \mu, \mu_0) + \frac{\tilde{\omega}}{4\mu\mu_0}P(\mu, -\mu_0)$$

$$+ \frac{\tilde{\omega}}{2\mu}\int_0^1 P(\mu, \mu'')R(\tau_1; \mu'', \mu_0)\,d\mu''$$

$$+ \frac{\tilde{\omega}}{2\mu_0}\int_0^1 R(\tau_1; \mu, \mu')P(-\mu', -\mu_0)\,d\mu'$$

$$+ \tilde{\omega}\int_0^1 R(\tau_1; \mu, \mu')\,d\mu'\left[\int_0^1 P(-\mu', \mu'')R(\tau_1; \mu'', \mu_0)\,d\mu''\right], \tag{6.3.41}$$

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} = -\frac{1}{\mu} T(\tau_1; \mu, \mu_0) + \frac{\tilde{\omega}}{4\mu\mu_0} e^{-\tau_1/\mu_0} P(\mu, -\mu_0)$$

$$+ \frac{\tilde{\omega}}{2\mu} \int_0^1 P(-\mu, -\mu'') T(\tau_1; \mu'', \mu_0) \, d\mu''$$

$$+ \frac{\tilde{\omega}}{2\mu_0} e^{-\tau_1/\mu_0} \int_0^1 R(\tau_1; \mu, \mu') P(\mu', -\mu_0) \, d\mu'$$

$$+ \tilde{\omega} \int_0^1 R(\tau_1; \mu, \mu') \, d\mu' \left[\int_0^1 P(\mu', -\mu'') T(\tau_1; \mu'', \mu_0) \, d\mu'' \right],$$

(6.3.42)

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} = \frac{\tilde{\omega}}{4\mu\mu_0} \exp\left[-\tau \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) \right] P(\mu, -\mu_0)$$

$$+ \frac{\tilde{\omega}}{2\mu} e^{-\tau_1/\mu} \int_0^1 P(\mu, -\mu'') T(\tau_1; \mu'', \mu_0) \, d\mu''$$

$$+ \frac{\tilde{\omega}}{2\mu_0} e^{-\tau_1/\mu_0} \int_0^1 T(\tau_1; \mu, \mu') P(\mu', -\mu_0) \, d\mu'$$

$$+ \tilde{\omega} \int_0^1 T(\tau_1; \mu, \mu') \, d\mu' \left[\int_0^1 P(\mu', -\mu'') T(\tau_1; \mu'', \mu_0) \, d\mu'' \right],$$

(6.3.43)

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} = -\frac{1}{\mu_0} T(\tau_1; \mu, \mu_0) + \frac{\tilde{\omega}}{4\mu\mu_0} e^{-\tau_1/\mu} P(-\mu, -\mu_0)$$

$$+ \frac{\tilde{\omega}}{2\mu} e^{-\tau_1/\mu} \int_0^1 P(-\mu, \mu'') R(\tau_1; \mu'', \mu_0) \, d\mu''$$

$$+ \frac{\tilde{\omega}}{2\mu_0} \int_0^1 T(\tau_1; \mu, \mu') P(-\mu', -\mu_0) \, d\mu'$$

$$+ \tilde{\omega} \int_0^1 T(\tau_1; \mu, \mu') \, d\mu' \left[\int_0^1 P(-\mu', \mu'') R(\tau_1; \mu'', \mu_0) \, d\mu'' \right].$$

(6.3.44)

Equations (6.3.41)–(6.3.44) represent four nonlinear integral equations that govern the complete radiation field at $\tau = 0$ and $\tau = \tau_1$ in plane-parallel atmospheres. In the preceding analysis, we neglected the azimuthal dependence on the reflection, transmission, and phase-function terms in the derivation of the four integral equations. However, it is simple to include ϕ and ϕ' terms in these four integral equations. Further, we note that as $\tau_1 \to \infty$, $\partial R/\partial \tau_1 \to 0$, and Eq. (6.3.41) reduces to Eq. (6.3.18) that was derived for a semi-infinite atmosphere. Equations (6.3.41) and (6.3.44) may be obtained by adding a thin layer ($\Delta \tau \ll 1$) to the top of a finite atmosphere following the principles outlined in Section 6.3.2. In addition, by adding a thin layer to the bottom

of a finite atmosphere, Eqs. (6.3.42) and (6.3.43) may be derived. The addition of thin layers to a finite atmosphere is referred to as *invariant imbedding* (Bellman *et al.,* 1963), which is equivalent to the principles of invariance (Exercise 6.3).

6.3.4 The *X* and *Y* Functions

In Section 6.3.2, we showed that the reflection function of a semi-infinite atmosphere for isotropic scattering is given by the *H* function. In the following, we wish to demonstrate that the reflection and transmission functions of a finite atmosphere for isotropic scattering are governed by the *X* and *Y* functions. In the case of isotropic scattering, Eqs. (6.3.41)–(6.3.44) become

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) R(\tau_1; \mu, \mu_0)$$
$$= \frac{\tilde{\omega}}{4\mu\mu_0} \left[1 + 2\mu_0 \int_0^1 R(\tau_1; \mu'', \mu_0) \, d\mu'' + 2\mu \int_0^1 R(\tau_1; \mu, \mu') \, d\mu' \right.$$
$$\left. + 4\mu\mu_0 \int_0^1 R(\tau_1; \mu, \mu') \int_0^1 R(\tau_1; \mu'', \mu_0) \, d\mu'' \right], \qquad (6.3.45)$$

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} = \frac{\tilde{\omega}}{4\mu\mu_0} \left[\exp\left\{ -\tau_1 \left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) \right\} + 2\mu_0 e^{-\tau_1/\mu} \int_0^1 T(\tau_1; \mu'', \mu_0) d\mu'' \right.$$
$$+ 2\mu e^{-\tau_1/\mu_0} \int_0^1 T(\tau_1; \mu, \mu') \, d\mu'$$
$$\left. + 4\mu\mu_0 \int_0^1 T(\tau_1; \mu, \mu') \, d\mu' \int_0^1 T(\tau_1; \mu'', \mu_0) \, d\mu'' \right], \qquad (6.3.46)$$

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \frac{1}{\mu} T(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0} \left[e^{-\tau_1/\mu_0} + 2\mu_0 \int_0^1 T(\tau_1; \mu'', \mu_0) \, d\mu'' \right.$$
$$+ 2\mu e^{-\tau_1/\mu_0} \int_0^1 R(\tau_1; \mu, \mu') \, d\mu'$$
$$\left. + 4\mu\mu_0 \int_0^1 R(\tau_1; \mu, \mu') \, d\mu' \int_0^1 T(\tau_1; \mu'', \mu_0) \, d\mu'' \right], \quad (6.3.47)$$

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \frac{1}{\mu_0} T(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0} \left[e^{-\tau_1/\mu_0} + 2\mu \int_0^1 T(\tau_1; \mu, \mu') \, d\mu' \right.$$
$$+ 2\mu_0 e^{-\tau_1/\mu_0} \int_0^1 R(\tau_1; \mu'', \mu_0) \, d\mu''$$
$$\left. + 4\mu\mu_0 \int_0^1 T(\tau_1; \mu, \mu') \, d\mu' \int_0^1 R(\tau_1; \mu'', \mu_0) \, d\mu'' \right]. \qquad (6.3.48)$$

From Eqs. (6.3.45)–(6.3.48), the integral terms may be expressed in terms of Chandrasekhar's X and Y functions in the forms

$$X(\mu) = 1 + 2\mu \int_0^1 R(\tau_1; \mu, \mu')\, d\mu', \tag{6.3.49}$$

$$Y(\mu) = e^{-\tau_1/\mu} + 2\mu \int_0^1 T(\tau_1; \mu, \mu')\, d\mu'. \tag{6.3.50}$$

It follows that Eqs. (6.3.45)–(6.3.48) may be rewritten as follows:

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) R(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0} X(\mu)X(\mu_0), \tag{6.3.51}$$

$$\frac{\partial R(\tau_1; \mu, \mu_0)}{\partial \tau_1} = \frac{\tilde{\omega}}{4\mu\mu_0} Y(\mu)Y(\mu_0), \tag{6.3.52}$$

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \frac{1}{\mu} T(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0} X(\mu)Y(\mu_0), \tag{6.3.53}$$

$$\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} + \frac{1}{\mu_0} T(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0} X(\mu_0)Y(\mu). \tag{6.3.54}$$

By eliminating $\partial R/\partial \tau_1$ from Eqs. (6.3.51) and (6.3.52), we obtain

$$\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) R(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0}[X(\mu)X(\mu_0) - Y(\mu)Y(\mu_0)], \tag{6.3.55}$$

and by eliminating $\partial T/\partial \tau_1$ from Eqs. (6.3.53) and (6.3.54), we have

$$\left(\frac{1}{\mu} - \frac{1}{\mu_0}\right) T(\tau_1; \mu, \mu_0) = \frac{\tilde{\omega}}{4\mu\mu_0}[X(\mu)Y(\mu_0) - X(\mu_0)Y(\mu)]. \tag{6.3.56}$$

Inserting Eqs. (6.3.55) and (6.3.56) into Eqs. (6.3.49) and (6.3.50), we find

$$X(\mu) = 1 + \mu \int_0^1 \frac{\Psi(\mu')}{\mu + \mu'}[X(\mu)X(\mu') - Y(\mu)Y(\mu')]\, d\mu', \tag{6.3.57}$$

$$Y(\mu) = e^{-\tau_1/\mu} + \mu \int_0^1 \frac{\Psi(\mu')}{\mu' - \mu}[X(\mu)Y(\mu') - X(\mu')Y(\mu)]\, d\mu', \tag{6.3.58}$$

where the characteristic function $\Psi(\mu') = \tilde{\omega}/2$. Thus, the exact solutions of the reflection and transmission functions are now governed by the X and Y functions, solutions of the two nonlinear integral equations. For a semi-infinite atmosphere $Y(\mu) = 0$ and the X function defined in Eqs. (6.3.49) and (6.3.57) is equivalent to the H function introduced in Eqs. (6.3.21) and (6.3.23). The characteristic function $\Psi(\mu')$ differs from problem to problem and has a simple algebraic form for the Rayleigh scattering phase function. For a more general case involving Lorenz–Mie scattering phase functions, however, the analytic characteristic functions $\Psi(\mu')$ appear to be extremely

complicated and have not been derived for practical applications. The iteration procedure may be used to solve the preceding nonlinear integral equations for the X and Y functions. Extensive tables of these two functions for conservative and non-conservative isotropic scattering, as well as anisotropic phase functions, with as many as three terms, have been constructed.

6.3.5 Inclusion of Surface Reflection

For planetary applications, surface reflection plays an important role in reflected and transmitted sunlight. In this section, we introduce the inclusion of surface reflection in the scattered intensity and flux density equations. The ground is considered to reflect according to Lambert's law, with a reflectivity (or surface albedo) of r_s. Under this condition, the diffuse upward intensity is

$$I(\tau_1; \mu, \phi) = I_s = \text{const.} \tag{6.3.59}$$

Let $I^*(0; \mu, \phi)$ represent the reflected intensity including the contribution of surface reflection, and in reference to Fig. 6.7a, we find

$$I^*(0; \mu, \phi) = I(0; \mu, \phi) + \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T(\mu, \phi; \mu', \phi') I_s \mu' d\mu' d\phi' + I_s e^{-\tau_1/\mu}. \tag{6.3.60}$$

The last two terms represent, respectively, the diffuse and direct transmission of the upward isotropic intensity I_s.

Equation (6.3.60) can be rewritten in terms of the reflection function and the direct and diffuse transmission defined in Section 6.3.1 in the form

$$I^*(0; \mu, \phi) = \mu_0 F_\odot R(\mu, \phi; \mu_0, \phi_0) + I_s \gamma(\mu), \tag{6.3.61}$$

where

$$\gamma(\mu) = e^{-\tau_1/\mu} + t(\mu), \tag{6.3.62}$$

and the diffuse transmission $t(\mu)$ is defined in Eq. (6.3.4b). The principle of reciprocity involving $T(\mu, \phi; \mu', \phi') = T(\mu', \phi'; \mu, \phi)$ is used to obtain $t(\mu)$.

The upward isotropic intensity from the surface also will be reflected by the atmosphere and will contribute to the downward intensity by an additional amount (see Fig. 6.7b)

$$I_s^R(-\mu) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\mu, \phi; \mu', \phi') I_s \mu' d\mu' d\phi' = I_s r(\mu), \tag{6.3.63}$$

where again the principle of reciprocity, $R(\mu, \phi; \mu', \phi') = R(\mu', \phi'; \mu, \phi)$, is used. Thus, the total transmitted intensity, including the ground contribution, is given by

$$I^*(\tau_1; -\mu, \phi) = I(\tau_1; -\mu, \phi) + I_s^R(-\mu)$$

$$= \mu_0 F_\odot T(\mu, \phi; \mu_0, \phi_0) + I_s r(\mu). \tag{6.3.64}$$
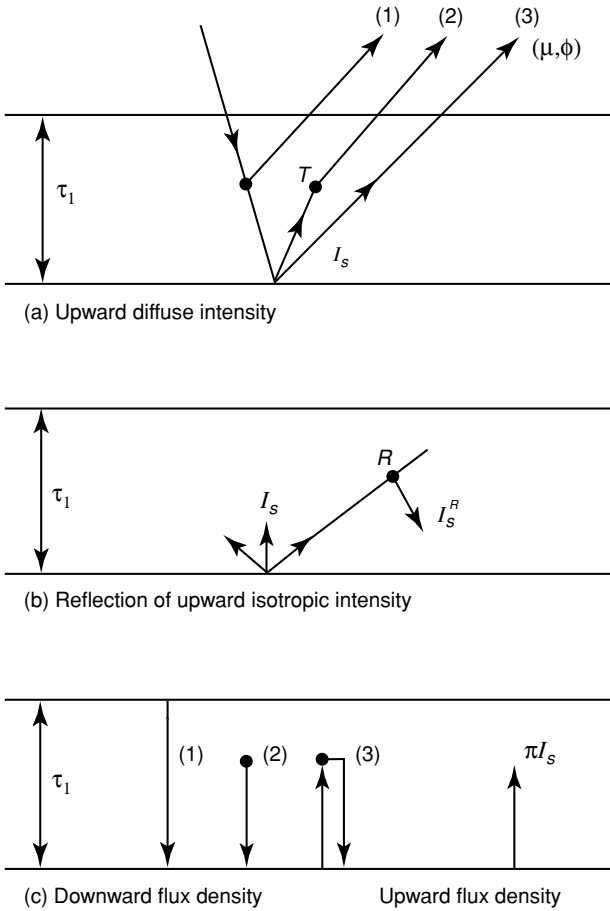
Figure 6.7   Scattering configuration for the inclusion of surface reflection: (a) upward diffuse intensity; (b) reflection of upward isotropic intensity; and (c) downward flux density and upward flux density.

We now require an equation to determine I_s. Since the upward flux density must be equal to the downward flux density multiplied by the surface albedo r_s, we have

$$\pi I_s = r_s \times \text{downward flux density.} \qquad (6.3.65)$$

The downward flux density includes three components, as shown in Fig. 6.7c:

1. Direct transmission component:

$$\mu_0 F_\odot e^{-\tau_1/\mu_0}.$$

2. Diffuse transmission component:

$$\int_0^{2\pi} \int_0^1 I(\tau_1; -\mu, \phi)\mu \, d\mu \, d\phi = \int_0^{2\pi} \int_0^1 (\mu_0 F_\odot/\pi) T(\mu, \phi; \mu_0, \phi_0)\mu \, d\mu \, d\phi$$

$$= \mu_0 F_\odot t(\mu_0).$$

3. The component of I_s reflected by the atmosphere:

$$\int_0^{2\pi} \int_0^1 I_s^R(-\mu)\mu \, d\mu \, d\phi = \pi I_s \bar{r}.$$

From Eq. (6.3.65), we have the following equality at $\tau = \tau_1$:

$$\pi I_s = r_s \left[\mu_0 F_\odot e^{-\tau_1/\mu_0} + \mu_0 F_\odot t(\mu_0) + \pi I_s \bar{r} \right]. \tag{6.3.66}$$

We then rearrange these terms to yield

$$I_s = \frac{r_s}{1 - r_s \bar{r}} \frac{\mu_0 F_\odot}{\pi} \gamma(\mu_0). \tag{6.3.67}$$

It follows from Eqs. (6.3.61) and (6.3.64) that the reflected and transmitted intensities, including the ground reflection, are, respectively, given by

$$I^*(0; \mu, \phi) = I(0; \mu, \phi) + \frac{r_s}{1 - r_s \bar{r}} \frac{\mu_0 F_\odot}{\pi} \gamma(\mu)\gamma(\mu_0), \tag{6.3.68a}$$

$$I^*(\tau_1; \mu, \phi) = I(\tau_1; -\mu, \phi) + \frac{r_s}{1 - r_s \bar{r}} \frac{\mu_0 F_\odot}{\pi} \gamma(\mu_0) r(\mu). \tag{6.3.68b}$$

To obtain the reflected and transmitted flux densities, we may perform an integration of the intensity over the solid angle according to Eqs. (6.1.7a) and (6.1.7b) to yield

$$F^*(0) = F(0) + \frac{r_s}{1 - r_s \bar{r}} \mu_0 F_\odot \gamma(\mu_0) \bar{\gamma}, \tag{6.3.69a}$$

$$F^*(\tau_1) = F(\tau_1) + \frac{r_s}{1 - r_s \bar{r}} \mu_0 F_\odot \gamma(\mu_0) \bar{r}, \tag{6.3.69b}$$

where

$$\bar{\gamma} = \bar{t} + 2 \int_0^1 e^{-\tau_1/\mu_0} \mu_0 \, d\mu_0, \tag{6.3.70}$$

and $\bar{t}$ and $\bar{r}$ are defined in Eqs. (6.3.5b) and (6.3.5c). Further, by dividing $\mu_0 F_\odot$ and adding $e^{-\tau_1/\mu_0}$ to both sides in Eq. (6.3.69b), the preceding two equations become

$$r^*(\mu_0) = r(\mu_0) + f(\mu_0)\bar{\gamma}, \tag{6.3.71a}$$

$$\gamma^*(\mu_0) = \gamma(\mu_0) + f(\mu_0)\bar{r}, \tag{6.3.71b}$$

where

$$f(\mu_0) = \frac{r_s}{1 - r_s \bar{r}} \gamma(\mu_0). \tag{6.3.72}$$

Exercise 6.4 requires the derivation of Eqs. (6.3.71a) and (6.3.71b) by means of the ray-tracing technique.

6.4 Adding Method for Radiative Transfer

The principle of the adding method for radiative transfer was stated by Stokes (1862) in a problem dealing with reflection and transmission by glass plates. Peebles and Plesset (1951) developed the adding theory for application to gamma-ray transfer. van de Hulst (1980) presented a set of adding equations for multiple scattering that is now commonly used. Hansen (1971) applied the adding/doubling method to the interpretation of intensity and polarization of sunlight reflected from clouds. Takano and Liou (1989b) modified the adding method for radiative transfer by including polarization for application to randomly and horizontally oriented ice crystals. The exact adding/doubling method and its approximations appear to be a powerful tool for multiple scattering calculations, particularly with reference to remote-sensing applications from the ground, the air, and space.

6.4.1 Definitions of Physical Parameters

To introduce the adding principle for radiative transfer, we shall first define the reflection function R and the transmission function T. Consider a light beam incident from above, as shown in Fig. 6.8. The reflected and transmitted intensities of this beam may be expressed in terms of the incident intensity in the forms

$$I_{\text{out,top}}(\mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R(\mu, \phi; \mu', \phi') I_{\text{in,top}}(\mu', \phi') \mu' \, d\mu' \, d\phi', \quad (6.4.1a)$$

$$I_{\text{out,bottom}}(\mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T(\mu, \phi; \mu', \phi') I_{\text{in,top}}(\mu'\phi') \mu' \, d\mu' \, d\phi'. \quad (6.4.1b)$$
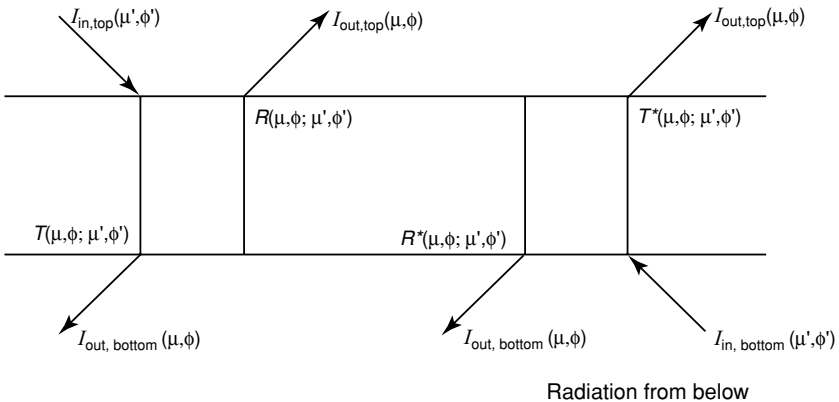
Figure 6.8 Configurations for radiation incident from above and below, and the definitions of the reflection and transmission functions for the adding method.

Likewise, if the light beam is incident from below (Fig. 6.8) we may write

$$I_{\text{out,bottom}}(\mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 R^*(\mu, \phi; \mu', \phi') I_{\text{in,bottom}}(\mu', \phi') \, \mu' \, d\mu' \, d\phi', \quad (6.4.2a)$$

$$I_{\text{out,top}}(\mu, \phi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 T^*(\mu, \phi; \mu', \phi') I_{\text{in,bottom}}(\mu', \phi') \, \mu' \, d\mu' \, d\phi'. \quad (6.4.2b)$$

Equations (6.4.2a) and (6.4.2b) define R^* and T^*, respectively, where the superscript* denotes that the light beam comes from below.

Consider now the transfer of monochromatic solar radiation. The incident solar intensity, in the present notation, may be written in the form

$$I_{\text{in,top}}(-\mu_0, \phi_0) = \delta(\mu' - \mu_0)\delta(\phi' - \phi_0)F_\odot, \quad (6.4.3)$$

where δ is the Dirac delta function. Using Eq. (6.4.3), the reflection and transmission functions defined in Eqs. (6.4.1a) and (6.4.1b) are given by

$$R(\mu, \phi; \mu_0, \phi_0) = \pi I_{\text{out,top}}(\mu, \phi)/\mu_0 F_\odot, \quad (6.4.4a)$$

$$T(\mu, \phi; \mu_0, \phi_0) = \pi I_{\text{out,bottom}}(\mu, \phi)/\mu_0 F_\odot. \quad (6.4.4b)$$

Under the single-scattering approximation and neglecting the emission contribution, the source function defined in Eq. (6.1.1) may be written in the form [see also Eq. (3.4.10)]

$$J(\tau; \mu, \phi) = \frac{\tilde{\omega}}{4\pi} F_\odot P(\mu, \phi; -\mu_0, \phi_0)e^{-\tau/\mu_0}. \quad (6.4.5)$$

Assuming that there are no diffuse intensities from the top and bottom of the layer with an optical depth $\Delta\tau$, then the radiation boundary conditions are as follows:

$$I_{\text{in,top}}(\mu, \phi) = 0,$$

$$I_{\text{in,bottom}}(\mu, \phi) = 0. \quad (6.4.6)$$

Subject to these boundary conditions, the reflected and transmitted diffuse intensities due to single scattering can be derived directly from the basic radiative transfer equation. Thus, the solutions for the reflection and transmission functions for an optical depth $\Delta\tau$ are given by

$$R(\mu, \phi; \mu_0, \phi_0) = \frac{\tilde{\omega}}{4(\mu + \mu_0)} P(\mu, \phi; -\mu_0, \phi_0) \left\{ 1 - \exp\left[-\Delta\tau \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) \right] \right\}, \quad (6.4.7a)$$

$$T(\mu, \phi; \mu_0, \phi_0) = \begin{cases} \dfrac{\tilde{\omega}}{4(\mu - \mu_0)} P(-\mu, \phi; -\mu_0, \phi_0)(e^{-\Delta\tau/\mu} - e^{-\Delta\tau/\mu_0}), & \mu \neq \mu_0 \\[2ex] \dfrac{\tilde{\omega}\Delta\tau}{4\mu_0^2} P(-\mu, \phi; -\mu_0, \phi_0)e^{-\Delta\tau/\mu_0}, & \mu = \mu_0. \end{cases} \quad (6.4.7b)$$

Consider a layer in which $\Delta\tau$ is very small (e.g., $\Delta\tau \approx 10^{-8}$). Equations (6.4.7a) and (6.4.7b) may then be further simplified to give

$$R(\mu, \phi; \mu_0, \phi_0) = \frac{\tilde{\omega}\Delta\tau}{4\mu\mu_0} P(\mu, \phi; -\mu_0, \phi_0), \qquad (6.4.8a)$$

$$T(\mu, \phi; \mu_0, \phi_0) = \frac{\tilde{\omega}\Delta\tau}{4\mu\mu_0} P(-\mu, \phi; -\mu_0, \phi_0). \qquad (6.4.8b)$$

Equations (6.4.7a) and (6.4.8a) were presented in the first subsection of Section 3.4.2. For a thin homogeneous layer, the reflection and transmission functions are the same regardless of whether the light beam comes from above or below. Thus, $R^* = R$ and $T^* = T$. However, when we proceed with the adding of layers, the reflection and transmission functions for combined layers will depend on the direction of the incoming light beam.

6.4.2 Adding Equations

In reference to Fig. 6.9, consider two layers, one on top of the other. Let the reflection and total (direct and diffuse) transmission functions be denoted by R_1 and $\tilde{T}_1$ for the first layer and R_2 and $\tilde{T}_2$ for the second layer. We further define $\tilde{D}$ and U for the combined total transmission and reflection functions between layers 1 and 2. In principle, the light beam may undergo an infinite number of scattering events. Accounting for multiple reflections of the light beam in the two layers, as shown in Fig. 6.9, the combined reflection and transmission functions are given by

$$\begin{aligned}
R_{12} &= R_1 + \tilde{T}_1^* R_2 \tilde{T}_1 + \tilde{T}_1^* R_2 R_1^* R_2 \tilde{T}_1 + \tilde{T}_1^* R_2 R_1^* R_2 R_1^* R_2 \tilde{T}_1 + \dots \\
&= R_1 + \tilde{T}_1^* R_2 \left[1 + R_1^* R_2 + \left(R_1^* R_2 \right)^2 + \dots \right] \tilde{T}_1 \\
&= R_1 + \tilde{T}_1^* R_2 \left(1 - R_1^* R_2 \right)^{-1} \tilde{T}_1, \qquad (6.4.9a)
\end{aligned}$$

$$\begin{aligned}
\tilde{T}_{12} &= \tilde{T}_2 \tilde{T}_1 + \tilde{T}_2 R_1^* R_2 \tilde{T}_1 + \tilde{T}_2 R_1^* R_2 R_1^* R_2 \tilde{T}_1 + \dots \\
&= \tilde{T}_2 \left[1 + R_1^* R_2 + \left(R_1^* R_2 \right)^2 + \dots \right] \tilde{T}_1 \\
&= \tilde{T}_2 \left(1 - R_1^* R_2 \right)^{-1} \tilde{T}_1. \qquad (6.4.9b)
\end{aligned}$$

Likewise, the expressions U and $\tilde{D}$ may be written as

$$\begin{aligned}
U &= R_2 \tilde{T}_1 + R_2 R_1^* R_2 \tilde{T}_1 + R_2 R_1^* R_2 R_1^* R_2 \tilde{T}_1 + \dots \\
&= R_2 \left[1 + R_1^* R_2 + \left(R_1^* R_2 \right)^2 + \dots \right] \tilde{T}_1 \\
&= R_2 \left(1 - R_1^* R_2 \right)^{-1} \tilde{T}_1, \qquad (6.4.10a)
\end{aligned}$$
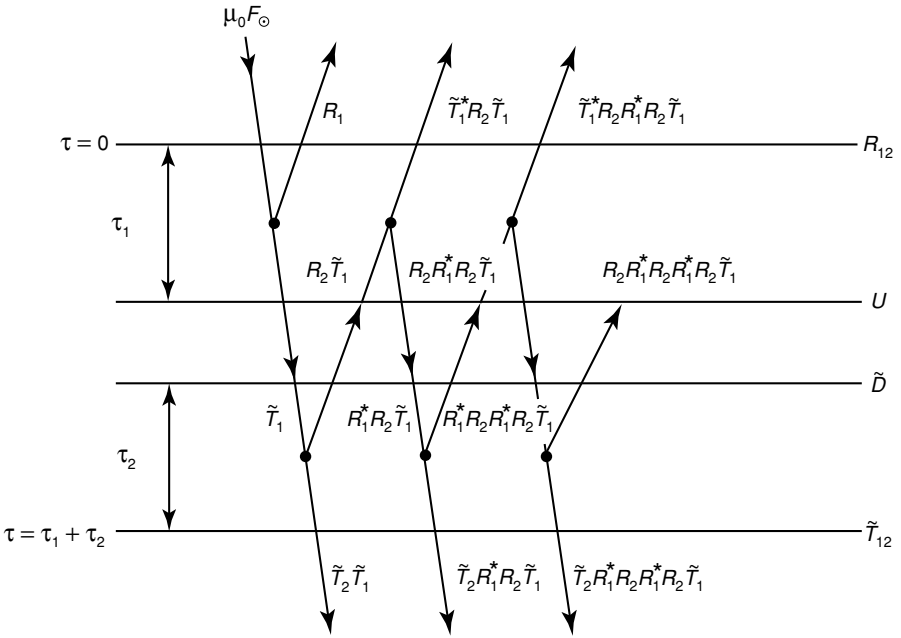
Figure 6.9 Configuration of the adding method. The two layers of optical depths τ_1 and τ_2 are rendered, for convenient illustration, as though they were physically separated (see text for the definitions of R, R^*, $\widetilde{T}$, and $\widetilde{T}^*$).

$$\widetilde{D} = \widetilde{T}_1 + R_1^* R_2 \widetilde{T}_1 + R_1^* R_2 R_1^* R_2 \widetilde{T}_1 + \ldots$$

$$= \left[1 + R_1^* R_2 + \left(R_1^* R_2 \right)^2 + \ldots \right] \widetilde{T}_1$$

$$= \left(1 - R_1^* R_2 \right)^{-1} \widetilde{T}_1. \tag{6.4.10b}$$

In Eqs. (6.4.9a) and (6.4.9b), the infinite series can be replaced by a single inverse function. We may define an operator in the form

$$S = R_1^* R_2 \left(1 - R_1^* R_2 \right)^{-1}, \tag{6.4.11a}$$

so that

$$\left(1 - R_1^* R_2 \right)^{-1} = 1 + S. \tag{6.4.11b}$$

From the preceding adding equations, we have

$$R_{12} = R_1 + \widetilde{T}_1^* U, \tag{6.4.12a}$$

$$\widetilde{T}_{12} = \widetilde{T}_2 \widetilde{D}, \tag{6.4.12b}$$

$$U = R_2 \widetilde{D}. \tag{6.4.12c}$$

We shall now separate the diffuse and direct components of the total transmission function defined by

$$\widetilde{T} = T + e^{-\tau/\mu'}, \tag{6.4.13}$$

where $\mu' = \mu_0$ when transmission is associated with the incident solar beam, and $\mu' = \mu$ when it is associated with the emergent light beam in the direction μ. Using Eq. (6.4.13), we may separate the direct and diffuse components in Eqs. (6.4.10b) and (6.4.12b) to obtain

$$\widetilde{D} = D + e^{-\tau_1/\mu_0} = (1 + S)\left(T_1 + e^{-\tau_1/\mu_0}\right)$$

$$= (1 + S)T_1 + Se^{-\tau_1/\mu_0} + e^{-\tau_1/\mu_0}, \tag{6.4.14a}$$

$$\widetilde{T}_{12} = \left(T_2 + e^{-\tau_2/\mu}\right)\left(D + e^{-\tau_1/\mu_0}\right)$$

$$= e^{-\tau_2/\mu}D + T_2 e^{-\tau_1/\mu_0} + T_2 D + \exp\left[-\left(\frac{\tau_1}{\mu_0} + \frac{\tau_2}{\mu}\right)\right]\delta(\mu - \mu_0), \tag{6.4.14b}$$

where D, T_1, and T_2 denote the diffuse components only, and a delta function has been added to the exponential term to signify that the direct transmission function is a function of μ_0 only.

On the basis of the preceding analysis, a set of iterative equations for the computation of diffuse transmission and reflection involving the two layers may be written in the form

$$Q = R_1^* R_2, \tag{6.4.15a}$$

$$S = Q(1 - Q)^{-1}, \tag{6.4.15b}$$

$$D = T_1 + ST_1 + Se^{-\tau_1/\mu_0}, \tag{6.4.15c}$$

$$U = R_2 D + R_2 e^{-\tau_1/\mu_0}, \tag{6.4.15d}$$

$$R_{12} = R_1 + e^{-\tau_1/\mu}U + T_1^* U, \tag{6.4.15e}$$

$$T_{12} = e^{-\tau_2/\mu}D + T_2 e^{-\tau_1/\mu_0} + T_2 D. \tag{6.4.15f}$$

The direct transmission function for the combined layer is given by $\exp[-(\tau_1 + \tau_2)/\mu_0]$. In these equations, the product of the two functions implies an integration over the appropriate solid angle so that all possible multiple-scattering contributions can be accounted for, as in the following example:

$$R_1^* R_2 = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 R_1^*(\mu, \phi; \mu', \phi')R_2(\mu', \phi'; \mu_0, \phi_0)\mu'\, d\mu'\, d\phi'. \tag{6.4.16}$$

In numerical computations, we may set $\tau_1 = \tau_2$. This is referred to as the *doubling method*. We may start with an optical depth $\Delta\tau$ on the order of 10^{-8} and use Eqs. (6.4.8a) and (6.4.8b) to compute the reflection and transmission functions. Equations (6.4.15a)–(6.4.15f) are subsequently employed to compute the reflection and transmission functions for an optical depth of $2\Delta\tau$. For the initial layers, $R_{1,2}^* = R_{1,2}$ and

$T_{1,2}^* = T_{1,2}$. Using the adding equations, the computations may be repeated until a desirable optical depth is achieved.

For a light beam incident from below, R_{12}^* and T_{12}^* may be computed from a scheme analogous to Eq. (6.4.15). Let the incident direction be μ'; then the adding equations may be written in the forms (Exercise 6.5)

$$Q = R_2 R_1^*, \tag{6.4.17a}$$

$$S = Q(1 - Q)^{-1}, \tag{6.4.17b}$$

$$U = T_2^* + ST_2^* + Se^{-\tau_2/\mu'}, \tag{6.4.17c}$$

$$D = R_1^* U + R_1^* e^{-\tau_2/\mu'}, \tag{6.4.17d}$$

$$R_{12}^* = R_2^* + e^{-\tau_2/\mu} D + T_2 D, \tag{6.4.17e}$$

$$T_{12}^* = e^{-\tau_1/\mu} U + T_1^* e^{-\tau_2/\mu'} + T_1^* U. \tag{6.4.17f}$$

When polarization and azimuth dependence are neglected, the transmission function is the same regardless of whether the light beam is from above or below so that $T^*(\mu, \mu') = T(\mu', \mu)$. This relation can be derived based on the Helmholtz principle of reciprocity in which the light beam may reverse its direction (Hovenier, 1969).

For practical applications, we may begin with the computations of reflection and transmission functions given in Eqs. (6.4.8a) and (6.4.8b). The phase function must be expressed as a function of the incoming and outgoing directions via Eq. (6.1.3a) in the form

$$P(\mu, \phi; \mu', \phi') = P^0(\mu, \mu') + 2 \sum_{m=1}^{N} P^m(\mu, \mu') \cos m(\phi' - \phi), \tag{6.4.18}$$

where $P^m(\mu, \mu')(m = 0, 1, \ldots N)$ denotes the Fourier expansion coefficients. The number of terms required in the expansion depends on the sharpness of the forward diffraction peak in phase function (see Fig. 3.13).

The preceding adding equations for radiative transfer have been written in scalar forms involving diffuse intensity. However, these equations can be applied to the case that takes polarization into account in which the light beam is characterized by the Stokes parameters and the phase function is replaced by the phase matrix. The phase matrix must be expressed with respect to the local meridian plane in a manner defined in Section 6.6. Finally, it should be noted that the numerical techniques referred to as *matrix formulation*, *matrix operator*, or *star product* are esssentially the same as the adding method, so far as the principle and actual computations are concerned.

6.4.3   Equivalence of the Adding Method and the Principles of Invariance

In reference to the principles of invariance for finite atmospheres defined in Eqs. (6.3.29)–(6.3.32) of Section 6.3.3, we replace τ by τ_1 and $\tau_1 - \tau$ by τ_2 and define the

dimensionless upward and downward internal intensities as follows:

$$U(\mu, \mu_0) = \frac{\pi I(\tau_1, \mu)}{\mu_0 F_\odot}, \tag{6.4.19a}$$

$$D(\mu, \mu_0) = \frac{\pi I(\tau_1, -\mu)}{\mu_0 F_\odot}. \tag{6.4.19b}$$

The four principles of invariance may be expressed in terms of the reflection and transmission functions as follows:

$$U(\mu, \mu_0) = R_2(\mu, \mu_0)e^{-\tau_1/\mu_0}$$
$$+ 2\int_0^1 R_2(\mu, \mu')D(\mu', \mu_0)\mu'd\mu', \tag{6.4.20}$$

$$D(\mu, \mu_0) = T_1(\mu, \mu_0) + 2\int_0^1 R_1(\mu, \mu'')U(\mu'', \mu_0)\mu''d\mu'', \tag{6.4.21a}$$

$$R_{12}(\mu, \mu_0) = R_1(\mu, \mu_0) + e^{-\tau_1/\mu_0}U(\mu, \mu_0)$$
$$+ 2\int_0^1 T_1^*(\mu, \mu')U(\mu', \mu_0)\mu'd\mu', \tag{6.4.22}$$

$$T_{12}(\mu, \mu_0') = T_2(\mu, \mu_0)e^{-\tau_1/\mu_0} + e^{-\tau_2/\mu}D(\mu, \mu_0)$$
$$+ 2\int_0^1 T_2(\mu, \mu')D(\mu', \mu_0)\mu'd\mu'. \tag{6.4.23}$$

The geometric configuration involving the basic variables is illustrated in Fig. 6.10. Although the preceding equations are written for azimuthally independent cases, these equations may be modified for general radiative transfer involving azimuthal terms
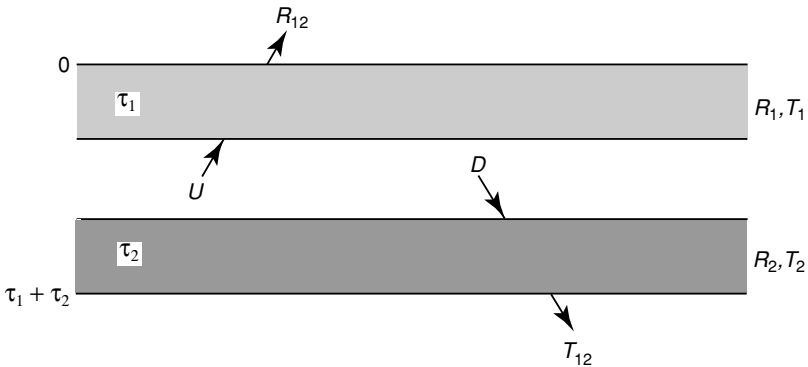
Figure 6.10   Geometric configuration for the reflection and transmission functions defined in Eqs. (6.4.19)–(6.4.23) based on the principles of invariance for a finite atmosphere. For illustration purposes, we have defined $\tau = \tau_1$ and $\tau_1 - \tau = \tau_2$ in Eqs. (6.3.29)–(6.3.32).

and polarization effects by replacing μ with (μ, ϕ) and the diffuse intensity with the Stokes parameters. Substituting Eq. (6.4.20) into Eq. (6.4.21a) leads to

$$D(\mu, \mu_0) = T_1(\mu, \mu_0) + S_{12}(\mu, \mu_0)e^{-\tau_1/\mu_0}$$

$$+ 2 \int_0^1 S_{12}(\mu, \mu'')D(\mu'', \mu_0)\mu''d\mu'', \qquad (6.4.21b)$$

where

$$S_{12}(\mu, \mu'') = 2 \int_0^1 R_1^*(\mu, \mu')R_2(\mu', \mu'')\mu'd\mu'. \qquad (6.4.24)$$

In Eqs. (6.4.22) and (6.4.24), the superscript * denotes that radiation comes from below. Equation (6.4.21a) can be rewritten as follows:

$$2 \int_0^1 \left[\delta(\mu - \mu'') - S_{12}(\mu, \mu'')\right]D(\mu'', \mu_0)\mu''d\mu''$$

$$= T_1(\mu, \mu_0) + S_{12}(\mu, \mu_0)e^{-\tau_1/\mu_0}. \qquad (6.4.21c)$$

In terms of the integral operator defined in Eq. (6.4.16), and noting again that $(1 - R_1^*R_2)^{-1} = 1 + S$ and $S_{12} = R_1^*R_2$, Eqs. (6.4.21c), (6.4.20), (6.4.22) and (6.4.23) can then be expressed by

$$D = T_1 + ST_1 + Se^{-\tau_1/\mu_0}, \qquad (6.4.25a)$$

$$U = R_2D + R_2e^{-\tau_1/\mu_0}, \qquad (6.4.25b)$$

$$R_{12} = R_1 + e^{-\tau_1/\mu}U + T_1^*U, \qquad (6.4.25c)$$

$$T_{12} = e^{-\tau_2/\mu}D + T_2e^{-\tau_1/\mu_0} + T_2D. \qquad (6.4.25d)$$

Equations (6.4.25a)–(6.4.25d) are exactly the same as Eqs. (6.4.15c)–(6.4.15f). We have, therefore, proved that the adding method for radiative transfer is "equivalent" to the principles of invariance for the case involving radiation from above. The principles of invariance can also be formulated for the case involving radiation from below, and the resulting equations would be "equivalent" to the adding equations presented in Eqs. (6.4.17c)–(6.4.17f).

6.4.4   Extension to Nonhomogeneous Atmospheres for Internal Fields

As demonstrated in the preceding sections, the adding principle is an efficient method for determining the radiation fields at the top and bottom of the atmosphere. To obtain the internal radiation fields, we may employ the following numerical procedures:

1. We may divide the atmosphere into N homogeneous layers, each of which is characterized by a single-scattering albedo, phase function, and optical depth. Let R_ℓ and $T_\ell (\ell = 1, 2, \ldots, N)$ denote the reflection and transmission functions for each homogeneous layer. Since homogeneous layers are considered, we have $R_\ell^* = R_\ell$ and $T_\ell^* = T_\ell . R_\ell$ and T_ℓ may be obtained from the doubling method described previously.
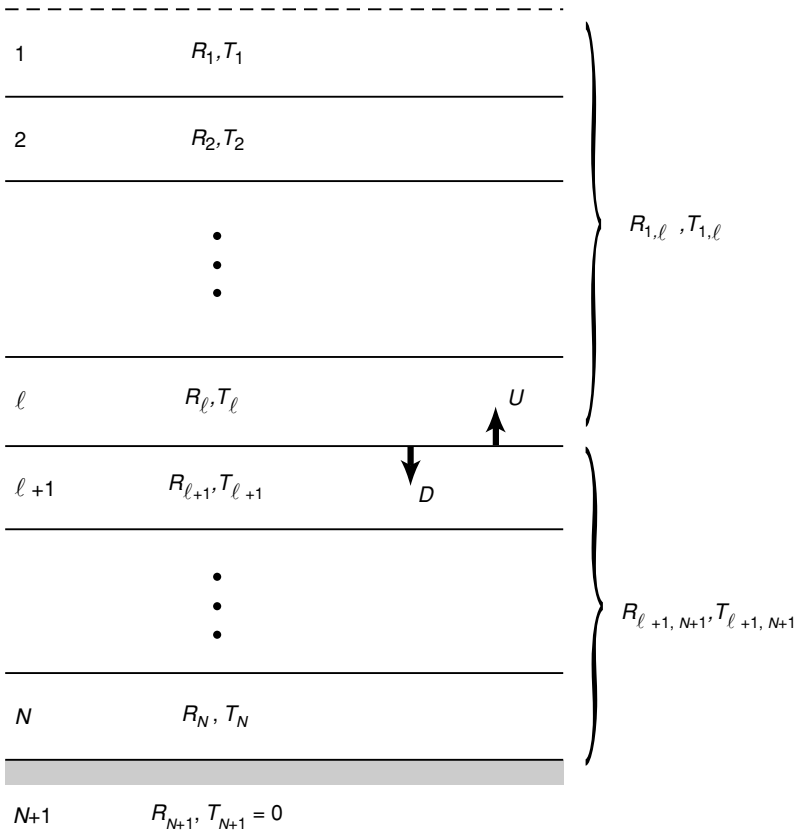
Figure 6.11 Illustrating diagram for the computation of internal intensities based on the adding principle for radiative transfer. The symbols U and D are defined in Eqs. (6.4.26a) and (6.4.26b).

The surface is considered to be a layer whose reflection function is R_{N+1} and whose transmission function is $T_{N+1} = 0$. If the surface is approximated as a Lambertian reflector, then R_{N+1} is the surface albedo r_s.

2. As shown in Fig. 6.11, the layers are added one at a time from TOA downward to obtain $R_{1,\ell}$ and $T_{1,\ell}$ for ℓ from 2 to $(N + 1)$, and $R^*_{1,\ell}$, and $T^*_{1,\ell}$ for ℓ from 2 to N. For example, $R_{1,\ell}$ is the reflection function for composite layers 1 through ℓ, with the lower part of the atmosphere and surface absent.

3. The layers are then added one at a time from the surface upward to obtain $R_{\ell+1,N+1}$ and $T_{\ell+1,N+1}$ for ℓ from $(N - 1)$ to 1.

4. We now consider the composite layers $(1, \ell)$ and $(\ell + 1, N + 1)$ and use the adding equations for the internal intensities noted in Eqs. (6.4.15c,d) to obtain

$$D = T_{1,\ell} + ST_{1,\ell} + S \exp(-\tau_{1,\ell}/\mu_0), \tag{6.4.26a}$$

$$U = R_{\ell+1,N+1}D + R_{\ell+1,N+1} \exp(-\tau_{1,\ell}/\mu_0), \tag{6.4.26b}$$

where

$$S = Q(1 - Q)^{-1}, \tag{6.4.27a}$$

$$Q = R^*_{1,\ell} R_{\ell+1,N+1}, \tag{6.4.27b}$$

and $\tau_{1,\ell}$ is the optical depth from TOA to the bottom of the layer.

To obtain the upward and downward fluxes at the interface between the ℓ and $\ell + 1$ layers, angular integrations can be performed. It is necessary to consider only the azimuth-independent condition, so that

$$F^\uparrow = \mu_0 F_\odot \left(2 \int_0^1 U(\mu, \mu_0)\mu \, d\mu \right), \tag{6.4.28a}$$

$$F^\downarrow_{\text{dif}} = \mu_0 F_\odot \left(2 \int_0^1 D(\mu, \mu_0)\mu \, d\mu \right). \tag{6.4.28b}$$

The downward direct solar flux is

$$F^\downarrow_{\text{dir}} = \mu_0 F_\odot \exp(-\tau_{1,\ell}/\mu_0). \tag{6.4.29}$$

Thus, the net flux is

$$F = \left(F^\downarrow_{\text{dif}} + F^\downarrow_{\text{dir}} \right) - F^\uparrow. \tag{6.4.30}$$

The heating rate can then be evaluated from the divergence of the net fluxes, as discussed in Section 3.5.

6.4.5   Similarity between the Adding and Discrete-Ordinates Methods

The adding equation can also be derived from an approach referred to as the *interaction principle,* which is based on a conservation relationship between the radiation emerging from a layer and the radiation incident to the boundary and emitted within the layer. For the present discussion, the emission contribution will be omitted.

Consider the geometric configuration illustrated in Fig. 6.9. Let the incoming and outgoing diffuse intensities at the top be $I^-(0)$ and $I^+(0)$, respectively. Similarly, the incoming and outgoing diffuse intensities at the level $\tau = \tau_1$ are denoted by $I^+(\tau_1)$ and $I^-(\tau_1)$, respectively. The $+$ or $-$ sign signifies that the intensity is upward or downward, respectively. The interaction principle states that the outgoing intensities $I^+(0)$ and $I^-(\tau_1)$ are related to the incoming intensities $I^-(0)$ and $I^+(\tau_1)$ via a linear conservation principle such that

$$I^+(0) = R_1 I^-(0) + \widetilde{T}^*_1 I^+(\tau_1), \tag{6.4.31a}$$

$$I^-(\tau_1) = \widetilde{T}_1 I^-(0) + R^*_1 I^+(\tau_1), \tag{6.4.31b}$$

where the notations R, R^*, $\widetilde{T}$, and $\widetilde{T}^*$ have been defined in Section 6.4.2. In matrix form, we write

$$
\begin{bmatrix} I^+(0) \\ I^-(\tau_1) \end{bmatrix} = \begin{bmatrix} R_1 & \widetilde{T}_1^* \\ \widetilde{T}_1 & R_1^* \end{bmatrix} \begin{bmatrix} I^-(0) \\ I^+(\tau_1) \end{bmatrix}.
\tag{6.4.32a}
$$

Application of the interaction principle to the layer bounded by τ_1 and $\tau = \tau_1 + \tau_2$, as defined in Fig. 6.9, leads to

$$
\begin{bmatrix} I^+(\tau_1) \\ I^-(\tau) \end{bmatrix} = \begin{bmatrix} R_2 & \widetilde{T}_2^* \\ \widetilde{T}_2 & R_2^* \end{bmatrix} \begin{bmatrix} I^-(\tau_1) \\ I^+(\tau) \end{bmatrix}.
\tag{6.4.32b}
$$

On eliminating the upward and downward diffuse intensities at the τ_1 level, $I^\pm(\tau_1)$, we can prove that

$$
\begin{bmatrix} I^+(0) \\ I^-(\tau) \end{bmatrix} = \begin{bmatrix} R_1 + \widetilde{T}_1^* R_2 \left(1 - R_1^* R_2\right)^{-1} \widetilde{T}_1 & \widetilde{T}_1^* \left(1 - R_2 R_1^*\right)^{-1} \widetilde{T}_2^* \\ \widetilde{T}_2 \left(1 - R_1^* R_2\right)^{-1} \widetilde{T}_1 & R_2^* + \widetilde{T}_2 R_1^* \left(1 - R_2 R_1^*\right)^{-1} \widetilde{T}_2^* \end{bmatrix} \begin{bmatrix} I^-(0) \\ I^+(\tau) \end{bmatrix}.
\tag{6.4.33a}
$$

The interaction principle is general and can be applied to the combined layer, $\tau = \tau_1 + \tau_2$, so that

$$
\begin{bmatrix} I^+(0) \\ I^-(\tau) \end{bmatrix} = \begin{bmatrix} R_{12} & \widetilde{T}_{12}^* \\ \widetilde{T}_{12} & R_{12}^* \end{bmatrix} \begin{bmatrix} I^-(0) \\ I^+(\tau) \end{bmatrix}.
\tag{6.4.33b}
$$

It follows that Eqs. (6.4.33a) and (6.4.33b) must be identical, leading to the following relationships:

$$
R_{12} = R_1 + \widetilde{T}_1^* R_2 \left(1 - R_1^* R_2\right)^{-1} \widetilde{T}_1,
\tag{6.4.34a}
$$

$$
\widetilde{T}_{12} = \widetilde{T}_2 \left(1 - R_1^* R_2\right)^{-1} \widetilde{T}_1,
\tag{6.4.34b}
$$

$$
R_{12}^* = R_2^* + \widetilde{T}_2 R_1^* \left(1 - R_2 R_1^*\right)^{-1} \widetilde{T}_2^*,
\tag{6.4.35a}
$$

$$
\widetilde{T}_{12}^* = \widetilde{T}_1^* \left(1 - R_2 R_1^*\right)^{-1} \widetilde{T}_2^*.
\tag{6.4.35b}
$$

The first two equations are exactly the same as those derived from the geometric ray-tracing technique for the light beam incident from above, denoted in Eqs. (6.4.9a) and (6.4.9b). We may also apply this technique to the light beam incident from below to obtain Eqs. (6.4.17e) and (6.4.17f) (Exercise 6.5). Equations (6.4.34a) and (6.4.34b) contain a term $(1 - R_1^* R_2)^{-1}$, which is associated with the inversion of a matrix. Moreover, the term $R_1^* R_2$, as defined in Eq. (6.4.16), can be expressed by summation according to Gaussian quadratures in the form

$$
R_1^* R_2(\mu_i, \mu_0) = \frac{1}{2} \sum_{j=-n}^{n} R_1^*(\mu_i, \mu_j) R_2(\mu_j, \mu_0) \mu_j a_j, \quad i = -n, n.
\tag{6.4.36}
$$

The discrete-ordinates method for radiative transfer presented in Section 6.2 directly solves the basic integrodifferential equation by means of discretization of the

integral in terms of Gauss's formula. In line with the present discussion, we shall consider a homogeneous layer with an optical depth τ_1. We may begin with the upward and downward components of the diffuse intensity, as presented in Eq. (6.2.51), in matrix form as follows:

$$\frac{d}{d\tau}\begin{bmatrix} I^+(\tau) \\ I^-(\tau) \end{bmatrix} = \begin{bmatrix} b^+ & b^- \\ -b^- & -b^+ \end{bmatrix}\begin{bmatrix} I^+(\tau) \\ I^-(\tau) \end{bmatrix}. \qquad (6.4.37)$$

The solutions of the reflected intensity at the top, $I^+(0)$, and the transmitted intensity at the bottom, $I^-(\tau_1)$, require the imposition of boundary conditions. For the sake of discussion, we may employ the vacuum boundary conditions such that there are no upward and downward diffuse intensities at the bottom and top layers, respectively. In this case, we have $I^+(\tau_1) = I^-(0) = 0$. The discrete-ordinates method, as presented in Eq. (6.4.37), consists of a differential operator and, hence, must be solved mathematically. The adding method that is derived from the interaction principle, as presented in Eqs. (6.4.33a) and (6.4.33b), however, entails calculations of the reflection and transmission functions in a direct way. Nevertheless, the two methods are similar to the extent that both are related to a 2×2 matrix, representing upward and downward radiation components.

Moreover, in the discrete-ordinates method, the solution of the diffuse intensity can be written as

$$I(\tau, \mu_i) = \sum_{j=-n}^{n} L_j \varphi_j(\mu_i) e^{-k_j \tau} + Z(\mu_i) e^{-\tau/\mu_0}, i = -n, n. \qquad (6.4.38)$$

Equation (6.4.38) is similar to Eq. (6.2.58), except without the layer index. The unknown coefficients L_j must be determined from boundary conditions. We may follow the approach presented in Section 6.2.4 and write

$$\mathbf{L} = \varphi^{-1} \chi, \qquad (6.4.39)$$

where $\mathbf{L}$ is a $2n$ column vector, φ is a $2n \times 2n$ matrix involving eigenvectors and eigenvalues, and χ is a $2n$ column vector associated with the direct solar flux. When $n = 1$ and 2, the system of equations reduces to 2 and 4, referred to as the *two-stream* and *four-stream* approximations, respectively (Sections 6.5.2 and 6.5.4). Once L_j are determined, diffuse intensity can be computed at the top and bottom of a layer, as well as within the layer. For comparison with the adding method, we shall use the reflection function R_{12} defined by $\pi I^+(0, \mu_i)/\mu_0 F_\odot$. Thus, we have

$$R_{12} = (\pi/\mu_0 F_\odot)\left[\sum_{j=-n}^{n} L_j \varphi_j(\mu_i) + Z(\mu_i) \right], \qquad (6.4.40a)$$

where L_j values are computed from the inversion of a matrix given by Eq. (6.4.39). On the other hand, using Eq. (6.4.36), the reflection function derived from the adding

method can be expressed by

$$R_{12} = R_1 + \widetilde{T}_1^* R_2 \left[1 - \frac{1}{2} \sum_{j=-n}^{n} R_1^*(\mu_i, \mu_j) R_2(\mu_j, \mu_0) \mu_j a_j \right]^{-1} \widetilde{T}_1. \quad (6.4.40b)$$

It is not possible to match each term in Eqs. (6.4.40a) and (6.4.40b). However, to the extent that both involve summation over discrete streams and inversion of matrices, it appears appropriate to conclude that the adding and discrete-ordinates methods are similar in terms of numerical computation.

6.5   Approximations for Radiative Transfer

6.5.1   Successive-Orders-of-Scattering Approximation

In Section 6.4.1, we showed that under the single-scattering approximation and subject to the condition that there are no diffuse intensities from the top and bottom of the layer with an optical depth τ_1, the reflection and transmission functions are directly proportional to the phase function. Rewriting Eqs. (6.4.7a) and (6.4.7b) in terms of reflected and transmitted intensities, we have

$$I_1(0; \mu, \phi) = \frac{\tilde{\omega} \mu_0 F_\odot}{4\pi(\mu + \mu_0)} P(\mu, \phi; -\mu_0, \phi_0) \left\{ 1 - \exp\left[-\tau_1 \left(\frac{1}{\mu} + \frac{1}{\mu_0} \right) \right] \right\},$$
$$(6.5.1a)$$

$$I_1(\tau_1; -\mu, \phi) = \begin{cases} \dfrac{\tilde{\omega} \mu_0 F_\odot}{4\pi(\mu - \mu_0)} P(-\mu, \phi; -\mu_0, \phi_0)(e^{-\tau_1/\mu} - e^{-\tau_1/\mu_0}), & \mu \neq \mu_0 \\[3mm] \dfrac{\tilde{\omega} \tau_1 F_\odot}{4\pi \mu_0} P(-\mu_0, \phi_0; -\mu_0, \phi_0) e^{-\tau_1/\mu_0}, & \mu = \mu_0. \end{cases}$$
$$(6.5.1b)$$

These two equations are fundamental in building the successive-orders-of-scattering (SOS) approximation, in which the intensity is computed individually for photons scattered once, twice, three times, and so forth. The total intensity is then obtained as the sum over all orders. Hence, for reflected and diffuse transmitted intensities, we may write, respectively,

$$I(\tau; \mu, \phi) = \sum_{n=1}^{\infty} I_n(\tau; \mu, \phi), \quad (6.5.2a)$$

$$I(\tau; -\mu, \phi) = \sum_{n=1}^{\infty} I_n(\tau; -\mu, \phi), \quad (6.5.2b)$$

where n denotes the order of scattering.

Subject to the boundary conditions denoted in Eq. (6.4.6), the formal solution of the equations of transfer for upward and downward intensities is given by

$$I(\tau; \mu, \phi) = \int_{\tau}^{\tau_1} J(\tau'; \mu, \phi) \exp[-(\tau' - \tau)/\mu] \frac{d\tau'}{\mu}, \qquad (6.5.3a)$$

$$I(\tau; -\mu, \phi) = \int_{0}^{\tau} J(\tau'; -\mu, \phi) \exp[-(\tau - \tau')/\mu] \frac{d\tau'}{\mu}, \qquad (6.5.3b)$$

where τ_1 is the total optical depth. The source function for the first-order scattering is given by Eq. (6.4.5). Inserting this expression into the above formal solutions of the equation of transfer and integrating over the appropriate optical depths, the intensity due to photons scattered once may be obtained. It follows that the source functions and intensities may be derived successively by means of the following recursion relationships:

$$J_{n+1}(\tau; \mu, \phi) = \frac{\tilde{\omega}}{4\pi} \int_{0}^{2\pi} \int_{-1}^{1} P(\mu, \phi; \mu', \phi') I_n(\tau; \mu', \phi') \, d\mu' \, d\phi', \qquad (6.5.4a)$$

$$I_n(\tau; \mu, \phi) = \int_{\tau}^{\tau_1} J_n(\tau'; \mu, \phi) \exp[-(\tau' - \tau)/\mu] \frac{d\tau'}{\mu}, \qquad n \geq 1, \quad (6.5.4b)$$

$$I_n(\tau; -\mu, \phi) = \int_{0}^{\tau} J_n(\tau'; -\mu, \phi) \exp[-(\tau - \tau')/\mu] \frac{d\tau'}{\mu}, \qquad n \geq 1, \quad (6.5.4c)$$

where the zero-order intensity is given by the Dirac δ function [see also Eq. (6.4.3)]

$$I_{\odot}(\tau; \mu', \phi') = F_{\odot} e^{-\tau/\mu_0} \delta(\mu' - \mu_0) \delta(\phi' - \phi_0). \qquad (6.5.5)$$

Note that J_1 can be determined from Eq. (6.5.4a) and is given in Eq. (6.4.5). I_1 can be evaluated from Eqs. (6.5.4b) and (6.5.4c) and is given in Eqs. (6.5.1a) and (6.5.1b). Numerical techniques may be devised to carry out the integrations for a finite interval of τ in Eq. (6.5.4) to obtain the intensity distribution. Exercise 6.7 requires the derivation of the reflected intensity due to second-order scattering ($n = 2$).

The SOS method just outlined is an integral solution approach that can be directly applied to specific geometry without the requirement of solving the basic radiative transfer equation in differential form. In addition, the inhomogeneous structure of the medium can be incorporated in the calculation in a straightforward manner in terms of integration along the line path. However, it is well known, based on the experience of plane-parallel radiative-transfer calculations, that this method requires considerable computational effort to converge the intensity solution, especially for an optically thick medium.

6.5.2   Two-Stream and Eddington's Approximations

In Sections 3.4.2 and 4.6.3, we introduced the two-stream approximation for radiative transfer and pointed out the potential applicability of this approximation to the

parameterization of radiative transfer for use in climate models. In this section, we present the details of this method.

In order to solve Eq. (6.1.6) analytically, the integral must be replaced by a summation over a finite number of quadrature points. In numerical integrations, Gauss's formula has been found to be superior to other formulas for quadratures in the interval $(-1, 1)$. For any function $f(\mu)$, Gauss's formula is expressed by

$$\int_{-1}^{1} f(\mu)d\mu \approx \sum_{j=-n}^{n} a_j f(\mu_j), \tag{6.5.6a}$$

where the weights are

$$a_j = \frac{1}{P'_{2n}(\mu_j)} \int_{-1}^{1} \frac{P_{2n}(\mu)}{\mu - \mu_j} d\mu, \tag{6.5.6b}$$

μ_j are the zeros of the even-order Legendre polynomials $P_{2n}(\mu)$, and the prime denotes the differentiation with respect to μ_j. Also,

$$a_{-j} = a_j, \qquad u_{-j} = -\mu_j, \qquad \sum_{j=-n}^{n} a_j = 2. \tag{6.5.6c}$$

Table 6.1 lists the Gaussian points and weights for the first four approximations.

Employing Gauss's formula, Eq. (6.1.6) can be written as

$$\mu_i \frac{dI(\tau, \mu_i)}{d\tau} = I(\tau, \mu_i) - \frac{\tilde{\omega}}{2} \sum_{\ell=0}^{N} \tilde{\omega}_\ell P_\ell(\mu_i) \sum_{j=-n}^{n} a_j P_\ell(\mu_j)I(\tau, \mu_j)$$

$$- \frac{\tilde{\omega}}{4\pi} F_\odot \left[\sum_{\ell=0}^{N} (-1)^\ell \tilde{\omega}_\ell P_\ell(\mu_i) P_\ell(\mu_0) \right] e^{-\tau/\mu_0}, \quad i = -n, n, \tag{6.5.7}$$

where $\mu_i(-n, n)$ represent the directions of the radiation streams. In the two-stream approximation, we take two radiation streams, i.e., $j = -1$ and 1, and $N = 1$. Note

Table 6.1

Gaussian Points and Weights

n	$2n$	$\pm\mu_n$	a_n
1	2	$\mu_1 = 0.5773503$	$a_1 = 1$
2	4	$\mu_1 = 0.3399810$	$a_1 = 0.6521452$
		$\mu_2 = 0.8611363$	$a_2 = 0.3478548$
3	6	$\mu_1 = 0.2386192$	$a_1 = 0.4679139$
		$\mu_2 = 0.6612094$	$a_2 = 0.3607616$
		$\mu_3 = 0.9324695$	$a_3 = 0.1713245$
4	8	$\mu_1 = 0.1834346$	$a_1 = 0.3626838$
		$\mu_2 = 0.5255324$	$a_2 = 0.3137066$
		$\mu_3 = 0.7966665$	$a_3 = 0.2223810$
		$\mu_4 = 0.9602899$	$a_4 = 0.1012285$

that $\mu_1 = 1/\sqrt{3}$ and $a_1 = a_{-1} = 1$. After rearranging terms and denoting $I^\uparrow = I(\tau, \mu_1)$ and $I^\downarrow = I(\tau, -\mu_1)$, we obtain two simultaneous equations as follows:

$$\mu_1 \frac{dI^\uparrow}{d\tau} = I^\uparrow - \tilde{\omega}(1-b)I^\uparrow - \tilde{\omega}bI^\downarrow - S^- e^{-\tau/\mu_0}, \tag{6.5.8a}$$

$$-\mu_1 \frac{dI^\downarrow}{d\tau} = I^\downarrow - \tilde{\omega}(1-b)I^\downarrow - \tilde{\omega}bI^\uparrow - S^+ e^{-\tau/\mu_0}, \tag{6.5.8b}$$

where

$$g = \frac{\tilde{\omega}_1}{3} = \frac{1}{2}\int_{-1}^{1} P(\cos\Theta)\cos\Theta \, d\cos\Theta = \langle\cos\Theta\rangle, \tag{6.5.9a}$$

$$b = \frac{1-g}{2} = \frac{1}{2}\int_{-1}^{1} P(\cos\Theta)\frac{1-\cos\Theta}{2} d\cos\Theta, \tag{6.5.9b}$$

$$S^\pm = \frac{F_\odot\tilde{\omega}}{4\pi}(1 \pm 3g\mu_1\mu_0). \tag{6.5.9c}$$

The asymmetry factor g was introduced in Section 3.4.1. It is the first moment of the phase function. Note that the zero moment of the phase function is simply equal to $\tilde{\omega}_0(=1)$. For isotropic scattering, g is zero as it is for Rayleigh scattering. The asymmetry factor increases as the diffraction peak of the phase function becomes increasingly sharp. Conceivably, the asymmetry factor may be negative if the phase function is peaked in backward directions (90–180°). For Lorenz–Mie particles, whose phase function has a general sharp peak at a 0° scattering angle, the asymmetry factor denotes the relative strength of forward scattering. The parameters b and $(1-b)$ can be interpreted as the integrated fractions of energy backscattered and forward scattered, respectively. Thus, the multiple scattering contribution in the context of the two-stream approximation is represented by the upward and downward intensities weighed by the appropriate fraction of the forward or backward phase function, as illustrated in Eq. (6.5.8). The upward intensity is strengthened by its coupling with the forward fraction (0–90°) of the phase function plus the downward intensity that appears in backward fractions (90–180°) of the phase function. A similar argument applies to the downward intensity.

Equations (6.5.8a) and (6.5.8b) represent two first-order inhomogeneous differential equations. For the solutions, we first let $M = I^\uparrow + I^\downarrow$ and $N = I^\uparrow - I^\downarrow$ and note that $(1-2b) = g$. Subsequently, by subtracting and adding, these two equations become

$$\mu_1 \frac{dM}{d\tau} = (1 - \tilde{\omega}g)N - (S^- - S^+)e^{-\tau/\mu_0}, \tag{6.5.10a}$$

$$\mu_1 \frac{dN}{d\tau} = (1 - \tilde{\omega})M - (S^- + S^+)e^{-\tau/\mu_0}. \tag{6.5.10b}$$

Further, by differentiating both equations with respect to τ, we obtain

$$\mu_1 \frac{d^2 M}{d\tau^2} = (1 - \tilde{\omega}g)\frac{dN}{d\tau} + \frac{(S^- - S^+)}{\mu_0}e^{-\tau/\mu_0}, \tag{6.5.11a}$$

$$\mu_1 \frac{d^2 N}{d\tau^2} = (1 - \tilde{\omega})\frac{dM}{d\tau} + \frac{(S^- + S^+)}{\mu_0}e^{-\tau/\mu_0}. \tag{6.5.11b}$$

Upon inserting Eqs. (6.5.10b) and (6.5.10a) into Eqs. (6.5.11a) and (6.5.11b), respectively, we find

$$\frac{d^2 M}{d\tau^2} = k^2 M + Z_1 e^{-\tau/\mu_0}, \tag{6.5.12a}$$

$$\frac{d^2 N}{d\tau^2} = k^2 N + Z_2 e^{-\tau/\mu_0}, \tag{6.5.12b}$$

where the eigenvalue is given by

$$k^2 = (1 - \tilde{\omega})(1 - \tilde{\omega}g)/\mu_1^2, \tag{6.5.13}$$

and

$$Z_1 = -\frac{(1 - \tilde{\omega}g)(S^- + S^+)}{\mu_1^2} + \frac{S^- - S^+}{\mu_1\mu_0}, \tag{6.5.14a}$$

$$Z_2 = -\frac{(1 - \tilde{\omega})(S^- - S^+)}{\mu_1^2} + \frac{S^- + S^+}{\mu_1\mu_0}. \tag{6.5.14b}$$

Equations (6.5.12a,b) represent a set of second-order differential equations, which can be solved by first seeking the homogeneous part and then adding a particular solution. In seeking the homogeneous solution, the homogeneous parts of Eq. (6.5.10) need to be satisfied so that there are only two unknown constants involved. Straightforward analyses yield the solutions

$$I^\uparrow = I(\tau, \mu_1) = Kve^{k\tau} + Hue^{-k\tau} + \varepsilon e^{-\tau/\mu_0}, \tag{6.5.15a}$$

$$I^\downarrow = I(\tau, -\mu_1) = Kue^{k\tau} + Hve^{-k\tau} + \gamma e^{-\tau/\mu_0}, \tag{6.5.15b}$$

where

$$v = (1 + a)/2, \qquad u = (1 - a)/2, \tag{6.5.16a}$$

$$a^2 = (1 - \tilde{\omega})/(1 - \tilde{\omega}g), \tag{6.5.16b}$$

$$\varepsilon = (\alpha + \beta)/2, \qquad \gamma = (\alpha - \beta)/2, \tag{6.5.16c}$$

$$\alpha = Z_1\mu_0^2/(1 - \mu_0^2 k^2), \qquad \beta = Z_2\mu_0^2/(1 - \mu_0^2 k^2). \tag{6.5.16d}$$

The constants K and H are to be determined from the diffuse radiation boundary conditions at the top and bottom of the scattering layer. Assuming no diffuse components from the top and bottom, we have

$$K = -(\varepsilon v e^{-\tau_1/\mu_0} - \gamma u e^{-k\tau_1})/(v^2 e^{k\tau_1} - u^2 e^{-k\tau_1}), \tag{6.5.17a}$$

$$H = -(\varepsilon u e^{-\tau_1/\mu_0} - \gamma v e^{-k\tau_1})/(v^2 e^{k\tau_1} - u^2 e^{-k\tau_1}). \tag{6.5.17b}$$

Once the upward and downward intensities have been evaluated, the upward and downward diffuse flux densities are simply given by

$$F^\uparrow(\tau) = 2\pi \mu_1 I^\uparrow, \qquad F^\downarrow(\tau) = 2\pi \mu_1 I^\downarrow. \tag{6.5.18}$$

The preceding analyses constitute the two-stream approximation for radiative transfer.

The form of Eq. (6.5.8), without the direct solar source term, was first presented by Schuster (1905). The terms $\pm k$ in Eq. (6.5.13) are the eigenvalues for the solution of the differential equations, and u and v represent the eigenfunctions, which are defined by the similarity parameter a in Eq. (6.5.16b) (see Section 6.5.3 for discussion on the similarity principle in radiative transfer). For conservative scattering, $\tilde{\omega} = 1$. Simpler solutions can be derived from Eqs. (6.5.8a) and (6.5.8b) with one of the eigenvalues, $k = 0$ (Exercise 6.8). In practice, however, we may set $\tilde{\omega} = 0.999999$ in Eqs. (6.5.16a)–(6.5.16d) and obtain the results for conservative scattering.

To introduce *Eddington's approximation* for radiative transfer, we begin with the general approach of decomposing the equation of radiative transfer using the property of Legendre polynomials. In line with the Legendre polynomial expansion for the phase function denoted in Eq. (6.1.11), the scattered intensity may be expanded in terms of Legendre polynomials such that

$$I(\tau, \mu) = \sum_{\ell=0}^{N} I_\ell(\tau) P_\ell(\mu). \tag{6.5.19}$$

Using the orthogonal and recurrence properties of Legendre polynomials and omitting the emission term, Eq. (6.1.12) may be decomposed in N harmonics in the form

$$\frac{\ell}{2\ell-1} \frac{dI_{\ell-1}}{d\tau} + \frac{\ell+1}{2\ell+3} \frac{dI_{\ell+1}}{d\tau}$$
$$= I_\ell \left(1 - \frac{\tilde{\omega}\tilde{\omega}_\ell}{2\ell+1}\right) - \frac{\tilde{\omega}}{4\pi}\tilde{\omega}_\ell P_\ell(-\mu_0)F_\odot e^{-\tau/\mu_0},$$
$$\ell = 0, 1, 2, \ldots, N. \tag{6.5.20}$$

The method of solving the basic radiative transfer equation using the aforementioned procedure is referred to as the *spherical harmonics method* (Kourganoff, 1952).

Eddington's approximation uses an approach similar to that of the two-stream approximation and was originally used for studies of radiative equilibrium in stellar atmospheres (Eddington, 1916). Letting $N = 1$, the phase function and intensity expressions may be written as follows:

$$P(\mu, \mu') = 1 + 3g\mu\mu',$$
$$I(\tau, \mu) = I_0(\tau) + I_1(\tau)\mu, \qquad -1 \le \mu \le 1. \tag{6.5.21}$$

Consequently, Eq. (6.5.20) reduces to a set of two simultaneous equations in the form

$$\frac{dI_1}{d\tau} = 3(1 - \tilde{\omega})I_0 - \frac{3\tilde{\omega}}{4\pi} F_\odot e^{-\tau/\mu_0}, \qquad (6.5.22a)$$

$$\frac{dI_0}{d\tau} = (1 - \tilde{\omega}g)I_1 + \frac{3\tilde{\omega}}{4\pi} g\mu_0 F_\odot e^{-\tau/\mu_0}. \qquad (6.5.22b)$$

Differentiating Eq. (6.5.22b) with respect to τ and substituting the expression for $dI_1/d\tau$ from Eq. (6.5.22a) leads to

$$\frac{d^2 I_0}{d\tau^2} = k^2 I_0 - \chi e^{-\tau/\mu_0}, \qquad (6.5.23)$$

where $\chi = 3\tilde{\omega}F_\odot(1 + g - \tilde{\omega}g)/4\pi$, and the eigenvalues k^2 are exactly the same as they are for the two-stream approximation defined in Eq. (6.5.13). Equation (6.5.23) represents a well-known diffusion equation for radiative transfer. This diffusion approximation is particularly applicable to the radiation field in the deep domain of an optically thick layer (see Subsection 3.4.2.2).

Straightforward analysis yields the following solution for the diffusion equation:

$$I_0 = K e^{k\tau} + H e^{-k\tau} + \Psi e^{-\tau/\mu_0}, \qquad (6.5.24a)$$

where

$$\Psi = \frac{3\tilde{\omega}}{4\pi} F_\odot [1 + g(1 - \tilde{\omega})] / (k^2 - 1/\mu_0^2).$$

Following a similar procedure, the solution for the second harmonic, I_1, is given by

$$I_1 = aK e^{k\tau} - aH e^{-k\tau} - \xi e^{-\tau/\mu_0}, \qquad (6.5.24b)$$

where $a^2 = 3(1 - \tilde{\omega})/(1 - \tilde{\omega}g)$, defined in the two-stream approximation [Eq. (6.5.16b)], and

$$\xi = \frac{3\tilde{\omega}}{4\pi} \frac{F_\odot}{\mu_0} [1 + 3g(1 - \tilde{\omega})\mu_0^2] / (k^2 - 1/\mu_0^2). \qquad (6.5.24c)$$

The integration constants, K and H, are to be determined from proper boundary conditions. Finally, the upward and downward fluxes are given by

$$\left.\begin{array}{c} F^\uparrow(\tau) \\ F^\downarrow(\tau) \end{array}\right\} = 2\pi \int_0^{\pm 1} (I_0 + \mu I_1)\mu \, d\mu = \begin{cases} \pi \left(I_0 + \frac{2}{3}I_1\right) \\ \pi \left(I_0 - \frac{2}{3}I_1\right). \end{cases} \qquad (6.5.25)$$

In the following, we shall present the general two-stream equations. Using the radiative transfer equations denoted in Eq. (6.1.6) and the upward and downward diffuse fluxes defined in Eqs. (6.1.7a,b) and omitting the emission contribution, we may write the following equations:

$$\frac{1}{2\pi} \frac{dF^\uparrow(\tau)}{d\tau} = \int_0^1 I(\tau, \mu) \, d\mu - \frac{\tilde{\omega}}{2} \int_0^1 \int_{-1}^1 I(\tau, \mu)P(\mu, \mu') \, d\mu' \, d\mu$$

$$- \frac{\tilde{\omega}}{4\pi} F_\odot e^{-\tau/\mu_0} \int_0^1 P(\mu, -\mu_0) \, d\mu, \qquad (6.5.26a)$$

$$\frac{1}{2\pi}\frac{dF^{\downarrow}(\tau)}{d\tau} = -\int_0^1 I(\tau, -\mu)\,d\mu + \frac{\tilde{\omega}}{2}\int_0^1\int_{-1}^1 I(\tau, \mu')P(-\mu, \mu')\,d\mu'\,d\mu$$

$$+ \frac{\tilde{\omega}}{4\pi}F_{\odot}e^{-\tau/\mu_0}\int_0^1 P(-\mu, -\mu_0)\,d\mu. \tag{6.5.26b}$$

Thus, the generalized two-stream approximation may be expressed by

$$\frac{dF^{\uparrow}(\tau)}{d\tau} = \gamma_1 F^{\uparrow}(\tau) - \gamma_2 F^{\downarrow}(\tau) - \gamma_3\tilde{\omega}F_{\odot}e^{-\tau/\mu_0}, \tag{6.5.27a}$$

$$\frac{dF^{\downarrow}(\tau)}{d\tau} = \gamma_2 F^{\uparrow}(\tau) - \gamma_1 F^{\downarrow}(\tau) + (1 - \gamma_3)\tilde{\omega}F_{\odot}e^{-\tau/\mu_0}. \tag{6.5.27b}$$

The differential changes in the upward and downward diffuse fluxes are directly related to the upward and downward diffuse fluxes, as well as the downward direct flux. The coefficients γ_i ($i = 1, 2, 3$) depend on the manner in which the intensity and phase function are approximated in Eq. (6.5.26). In the two-stream approximation, there are only upward and downward intensities in the directions μ_1 and $-\mu_1$ given by the Gauss quadrature formula, while the phase function is expanded in two terms in Legendre polynomials. In Eddington's approximation, both the intensity and phase functions are expanded in two polynomial terms. The coefficients γ_i can be directly derived from Eqs. (6.5.8a,b) and (6.5.22a,b) and are given in Table 6.2.

In Eqs. (6.5.26a,b), we let the last integral involving the phase function be

$$q = \frac{1}{2}\int_0^1 P(\mu, -\mu_0)\,d\mu. \tag{6.5.28a}$$

Since the phase function is normalized to unity, we have

$$\frac{1}{2}\int_0^1 P(-\mu, -\mu_0)\,d\mu = 1 - q. \tag{6.5.28b}$$

Equations (6.5.28a) and (6.5.28b) can be evaluated exactly by numerical means. We may take $\gamma_3 = q$ in the two-stream approximation. This constitutes the modified two-stream approximation proposed by Liou (1973b) and Meador and Weaver (1980). The two-stream approximation yields negative albedo values for a thin atmosphere when $\gamma_3 < 0$ (i.e., $g > \mu_1/\mu_0$). This also occurs in Eddington's approximation when $g > 0.67/\mu_0$. These negative albedo values can be avoided by using q, the full phase function integration for the direct solar beam, denoted in Eq. (6.5.28b). The accuracy of the two-stream approximation has been discussed in Liou (1973a). The overall accuracy of the two-stream and Eddington's approximations can be improved by

Table 6.2

Coefficients in Two-Stream Approximations

Method	γ_1	γ_2	γ_3
Two-stream	$[1 - \tilde{\omega}(1 + g)/2]/\mu_1$	$\tilde{\omega}(1 - g)/2\mu_1$	$(1 - 3g\mu_1\mu_0)/2$
Eddington's	$[7 - (4 + 3g)\tilde{\omega}]/4$	$-[1 - (4 - 3g)\tilde{\omega}]/4$	$(2 - 3g\mu_0)/4$

incorporating the δ-function adjustment for forward scattering and will be discussed in Section 6.5.3.

The solutions for the equations of the generalized two-stream approximation expressed in Eqs. (6.5.27a) and (6.5.27b) are as follows:

$$F^\uparrow = vKe^{k\tau} + uHe^{-k\tau} + \varepsilon e^{-\tau/\mu_0}, \tag{6.5.29a}$$

$$F^\downarrow = uKe^{k\tau} + vHe^{-k\tau} + \gamma e^{-\tau/\mu_0}, \tag{6.5.29b}$$

where K and H are unknown coefficients to be determined from the boundary conditions, and

$$k^2 = \gamma_1^2 - \gamma_2^2, \tag{6.5.30a}$$

$$v = \frac{1}{2}[1 + (\gamma_1 - \gamma_2)/k], \qquad u = \frac{1}{2}[1 - (\gamma_1 - \gamma_2)/k], \tag{6.5.30b}$$

$$\varepsilon = [\gamma_3(1/\mu_0 - \gamma_1) - \gamma_2(1 - \gamma_3)]\,\mu_0^2\tilde{\omega}F_\odot, \tag{6.5.30c}$$

$$\gamma = -[(1 - \gamma_3)(1/\mu_0 + \gamma_1) + \gamma_2\gamma_3]\mu_0^2\tilde{\omega}F_\odot. \tag{6.5.30d}$$

6.5.3   Delta-Function Adjustment and Similarity Principle

The two-stream and Eddington methods for radiative transfer are good approximations for optically thick layers, but they produce inaccurate results for thin layers and when significant absorption is involved. The basic problem is that scattering by atmospheric particulates is highly peaked in the forward direction. This is especially evident for cloud particles, for which the forward-scattered energy within $\sim5°$ scattering angles produced by diffraction is five to six orders of magnitude greater than it is in the side and backward directions (see, e.g., Figs 5.15 and 5.23). The highly peaked diffraction pattern is typical for atmospheric particulates. It is clear that a two-term expansion in the phase function is far from adequate.

To incorporate the forward peak contribution in multiple scattering, we may consider an adjusted absorption and scattering atmosphere, such that the fraction of scattered energy residing in the forward peak, f, is removed from the scattering parameters: optical depth, τ; single-scattering albedo, $\tilde{\omega}$; and asymmetry factor g. We use primes to represent the adjusted radiative parameters, as shown in Fig. 6.12. The optical (extinction) depth is the sum of the scattering (τ_s) and absorption (τ_a) optical depths. The forward peak is produced by diffraction without the contribution of absorption. Thus, the adjusted scattering and absorption optical depths must be

$$\tau_s' = (1 - f)\tau_s, \tag{6.5.31a}$$

$$\tau_a' = \tau_a. \tag{6.5.31b}$$

The total adjusted optical depth is

$$\tau' = \tau_s' + \tau_a' = (1 - f)\tau_s + \tau_a = \tau(1 - f\tilde{\omega}), \tag{6.5.32}$$
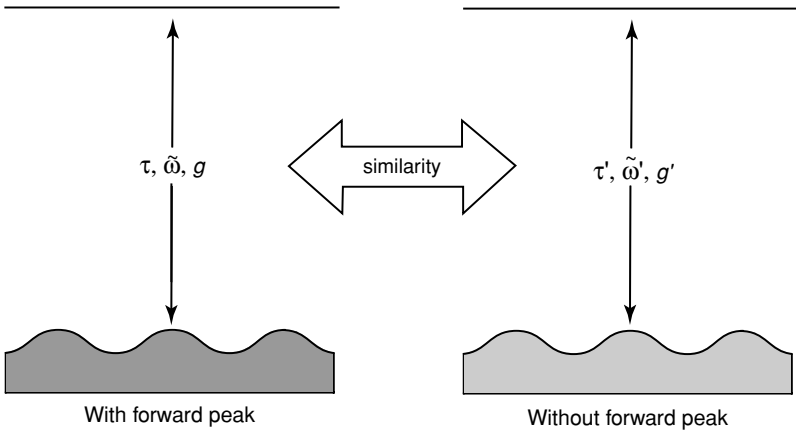
Figure 6.12 Similarity principle for radiative transfer. The prime system represents adjusted radiative parameters such that the forward diffraction peak in scattering processes is removed.

and the adjusted single-scattering albedo is then

$$\tilde{\omega}' = \frac{\tau_s'}{\tau'} = \frac{(1-f)\tau_s}{(1-f\tilde{\omega})\tau} = \frac{(1-f)\tilde{\omega}}{1-f\tilde{\omega}}. \tag{6.5.33}$$

Moreover, we multiply the asymmetry factor by the scattering optical depth to get the similarity equation

$$\tau_s'g' = \tau_s g - \tau_s f, \qquad \text{or} \qquad g' = \frac{g-f}{1-f}, \tag{6.5.34}$$

where we note that the asymmetry factor in the forward peak is equal to unity. In the diffusion domain, the solution for diffuse intensity is given by exponential functions with eigenvalues defined in Eq. (6.5.13). We may set the intensity solution in the adjusted atmosphere so that it is equivalent to that in the real atmosphere in the form

$$k\tau = k'\tau'. \tag{6.5.35}$$

From Eqs. (6.5.32)–(6.5.35), the similarity relations for radiative transfer can be expressed in the forms

$$\frac{\tau}{\tau'} = \frac{k'}{k} = \frac{1-\tilde{\omega}'}{1-\tilde{\omega}} = \frac{\tilde{\omega}'(1-g')}{\tilde{\omega}(1-g)}. \tag{6.5.36a}$$

Using the expression for the eigenvalue defined in Eq. (6.5.13), we also find the relation for the similarity parameter defined in Eq. (6.5.16b) as follows:

$$a = \left(\frac{1-\tilde{\omega}}{1-\tilde{\omega}g}\right)^{1/2} = \left(\frac{1-\tilde{\omega}'}{1-\tilde{\omega}'g'}\right)^{1/2}. \tag{6.5.36b}$$

The similarity principle can also be derived from the basic radiative transfer equation. We may begin with the following equation in the form

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{\tilde{\omega}}{2} \int_{-1}^{1} I(\tau, \mu')P(\mu, \mu')\, d\mu'. \qquad (6.5.37)$$

From Eq. (6.1.11), the phase function in the limit of the two-stream and Eddington's approximations is given by $P(\mu, \mu') = 1 + 3g\mu\mu'$. However, the phase functions involving cloud and aerosol particles are highly peaked in the forward direction, and two-term expansions do not adequately account for the strong forward scattering. Let the fraction of the energy scattered in the forward direction ($\Theta = 0°$) be denoted by f. The normalized phase function may be expressed in terms of this value as follows:

$$P(\mu, \mu') = 2f\delta(\mu - \mu') + (1 - f)(1 + 3g'\mu\mu'), \qquad (6.5.38)$$

where $\mu = \mu'$ when $\Theta = 0$, δ is the delta function, and g' denotes the adjusted asymmetry factor. The phase function so defined is normalized to unity, and the asymmetry factor is given by

$$g = f + (1 - f)g'. \qquad (6.5.39)$$

The second moment of the phase function is $\tilde{\omega}_2/5 = f$. Thus, the scaled asymmetry factor can be expressed by

$$g' = \frac{g - \tilde{\omega}_2/5}{1 - \tilde{\omega}_2/5}. \qquad (6.5.40)$$

Substituting Eq. (6.5.38) into Eq. (6.5.37) leads to

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu)(1 - \tilde{\omega}f) - \frac{\tilde{\omega}(1 - f)}{2} \int_{-1}^{1} (1 + 3g'\mu\mu')I(\tau, \mu')\, d\mu'. \quad (6.5.41)$$

Consequently, if we redefine the optical depth, single-scattering albedo, and phase function such that

$$\tau' = (1 - \tilde{\omega}f)\tau, \qquad (6.5.42a)$$

$$\tilde{\omega}' = \frac{(1 - f)\tilde{\omega}}{1 - \tilde{\omega}f}, \qquad (6.5.42b)$$

$$P'(\mu, \mu') = 1 + 3g'\mu\mu', \qquad (6.5.42c)$$

Eq. (6.5.41) then becomes

$$\mu \frac{dI(\tau', \mu)}{d\tau'} = I(\tau', \mu) - \frac{\tilde{\omega}'}{2} \int_{-1}^{1} I(\tau', \mu')P'(\mu, \mu')\, d\mu'. \qquad (6.5.43)$$

Equation (6.5.43) is exactly the same as Eq. (6.5.37), except that g, τ, and $\tilde{\omega}$ have been replaced by g', τ', and $\tilde{\omega}'$. By redefining the asymmetry factor, optical depth, and single-scattering albedo, the forward-scattering nature of the phase function is

approximately accounted for in the basic radiative transfer equation. In essence, we have incorporated the second moment of the phase function expansion in the formulation of the radiative transfer equation. The "equivalence" between Eqs. (6.5.43) and (6.5.37) represents the similarity principle stated previously.

The phase functions for aerosol and cloud particles require involved scattering calculations. For many applications to radiative transfer in planetary atmospheres, an analytic expression for the phase function in terms of the asymmetry factor has been proposed (Henyey and Greenstein, 1941):

$$P_{HG}(\cos \Theta) = (1 - g^2)/(1 + g^2 - 2g \cos \Theta)^{3/2}$$

$$= \sum_{\ell=0}^{N} (2\ell + 1)g^\ell P_\ell(\cos \Theta). \qquad (6.5.44)$$

This is referred to as the *Henyey–Greenstein phase function,* which is adequate for scattering patterns that are not strongly peaked in the forward direction. Using this expression, the second moment of the phase function is given by $\tilde{\omega}_2/5 = f = g^2$. Thus, in the limit of the Henyey–Greenstein approximation, the forward fraction of the scattered light is now expressed in terms of the asymmetry factor (Joseph *et al.*, 1976). Subsequently, the adjusted asymmetry factor, optical depth, and single-scattering albedo can now be expressed by

$$g' = \frac{g}{1 + g}, \quad \tau' = (1 - \tilde{\omega}g^2)\tau, \quad \tilde{\omega}' = \frac{(1 - g^2)\tilde{\omega}}{1 - \tilde{\omega}g^2}. \qquad (6.5.45)$$

The similarity principle for radiative transfer was first stated by Sobolev (1975) for isotropic scattering. The general similarity relationships have been presented by van de Hulst (1980). The two-stream approximations are popular because they enable the derivation of the analytic solutions for upward and downward fluxes, leading to efficient numerical computations. The incorporation of the delta-function adjustment to account for the strong forward scattering of large size parameters in the context of two-stream approximations has led to a significant improvement in the accuracy of radiative flux calculations. As pointed out previously, the δ adjustment provides a third term closure through the second moment of the phase function expansion. The δ-two-stream and δ-Eddington approximations have the same accuracy. Relative errors of 15–20% could result for some values of optical depths, solar zenith angles, and single-scattering albedos.

6.5.4 Four-Stream Approximation

The four-stream approximation, as first derived in Liou (1974), is based on the general solution for the discrete-ordinates method for radiative transfer. In order to be able to understand the merit of the four-stream approximation, it is necessary to have some background in solving a set of differential equations based on Chandrasekhar's (1950) formulations. In particular, it is noted that the search for eigenvalues from the recurrence equation developed in the solution is both mathematically ambiguous

and numerically troublesome (Section 6.2.3). More recently, a systematic development of the solution for this approximation has been presented by Liou *et al.* (1988). Specifically, this solution involves the computation of solar radiative fluxes using a relatively simple, convenient, and accurate method. Knowledge of the discrete-ordinates method for radiative transfer is desirable but not necessary. In addition, a wide range of accuracy checks for this approximation has been developed, including the δ-adjustment to account for the forward diffraction peak based on the generalized similarity principle for radiative transfer. In the following, we present the key equations for the four-stream approximation for readers who may wish to use this method for numerical calculations.

Consider two radiative streams in the upper and lower hemispheres (i.e., let $n = 2$). At the same time, expand the scattering phase function into four terms (i.e., $N = 3$) in line with the four radiative streams. On the basis of Eqs. (6.2.50a) and (6.2.50b), four first-order differential equations can then be written explicitly in matrix form:

$$
\frac{d}{d\tau}
\begin{bmatrix} I_2 \\ I_1 \\ I_{-1} \\ I_{-2} \end{bmatrix}
= -
\begin{bmatrix}
b_{2,-2} & b_{2,-1} & b_{2,1} & b_{2,2} \\
b_{1,-2} & b_{1,-1} & b_{1,1} & b_{1,2} \\
-b_{1,2} & -b_{1,1} & -b_{1,-1} & -b_{1,-2} \\
-b_{2,2} & -b_{2,1} & -b_{2,-1} & -b_{2,-2}
\end{bmatrix}
\begin{bmatrix} I_2 \\ I_1 \\ I_{-1} \\ I_{-2} \end{bmatrix}
-
\begin{bmatrix} b_{2,-0} \\ b_{1,-0} \\ b_{-1,-0} \\ b_{-2,-0} \end{bmatrix}
I_\odot,
\tag{6.5.46}
$$

where the terms $b_{i,j} (i = \pm1, 2; j = -0, \pm1, 2)$ are defined in Eq. (6.2.48). The four-by-four matrix represents the contribution of multiple scattering. Thus, the derivative of the diffuse intensity at a specific quadrature angle is the weighted sum of the multiple-scattered intensity from all four quadrature angles. The last term represents the contribution of the unscattered component of the direct solar flux at position τ.

After a lengthy and laborious derivation, the solution for Eq. (6.5.46) is given by

$$
\begin{bmatrix} I_2 \\ I_1 \\ I_{-1} \\ I_{-2} \end{bmatrix}
=
\begin{bmatrix}
\varphi_2^+ e2 & \varphi_1^+ e1 & \varphi_1^- e3 & \varphi_2^- e4 \\
\Phi_2^+ e2 & \Phi_1^+ e1 & \Phi_1^- e3 & \Phi_2^- e4 \\
\Phi_2^- e2 & \Phi_1^- e1 & \Phi_1^+ e3 & \Phi_2^+ e4 \\
\varphi_2^- e2 & \varphi_1^- e1 & \varphi_1^+ e3 & \varphi_2^+ e4
\end{bmatrix}
\begin{bmatrix} G_2 \\ G_1 \\ G_{-1} \\ G_{-2} \end{bmatrix}
+
\begin{bmatrix} Z_2^+ \\ Z_1^+ \\ Z_1^- \\ Z_2^- \end{bmatrix}
e^{-f_0\tau}, \quad (6.5.47)
$$

where $f_0(\text{solar}) = 1/\mu_0$, $f_0(\text{thermal}) = -1/\tau_1 \ln(B_1/B_0)$, B_0 and B_1 are Planck functions evaluated at the top and bottom of the layer (Section 4.6.3), respectively, and the other terms except $G_{\pm1,2}$ are defined as follows:

$$
e1 = e^{-k_1\tau}, \qquad e2 = e^{-k_2\tau}, \tag{6.5.48a}
$$

$$
e3 = e^{-k_1(\tau_1-\tau)}, \qquad e4 = e^{-k_2(\tau_1-\tau)}, \tag{6.5.48b}
$$

$$
Z_{1,2}^{\pm} = \frac{1}{2}(\eta_{1,2} \pm \eta_{1,2}'), \tag{6.5.48c}
$$

$$\varphi_{1,2}^{\pm} = \frac{1}{2}\left(1 \pm \frac{b_{11}^{-} - A_{1,2}b_{21}^{-}}{a^{-}}k_{1,2}\right), \qquad (6.5.48d)$$

$$\Phi_{1,2}^{\pm} = \frac{1}{2}\left(A_{1,2} \pm \frac{A_{1,2}b_{22}^{-} - b_{12}^{-}}{a^{-}}k_{1,2}\right), \qquad (6.5.48e)$$

where the parameters $k_{1,2}$, $A_{1,2}$, $\eta_{1,2}$, $\eta_{1,2}'$, b_{11}^{-}, b_{21}^{-}, b_{22}^{-}, b_{12}^{-}, a^{-}, and other related terms can be computed in successive order from the following equations:

$$b_i(\text{solar}) = \frac{\tilde{\omega}}{4\pi}F_{\odot}\sum_{\ell=0}^{3}\tilde{\omega}_\ell P_\ell(\mu_i)P_\ell(-\mu_0)/\mu_i,$$

$$b_i(\text{thermal}) = (1 - \tilde{\omega})B_0/\mu_i, \qquad i = -2, -1, 1, 2,$$

$$c_{i,j} = \frac{\tilde{\omega}}{2}a_j\sum_{\ell=0}^{3}\tilde{\omega}_\ell P_\ell(\mu_i)P_\ell(\mu_j), \qquad i, j = -2, -1, 1, 2,$$

$$b_{i,j} = \begin{cases} c_{i,j}/\mu_i, & i \neq j \\ (c_{i,j} - 1)/\mu_i, & i = j, \end{cases}$$

$$b_{22}^{\pm} = b_{2,2} \pm b_{2,-2}, \qquad b_{21}^{\pm} = b_{2,1} \pm b_{2,-1},$$

$$b_{12}^{\pm} = b_{1,2} \pm b_{1,-2}, \qquad b_{11}^{\pm} = b_{1,1} \pm b_{1,-1},$$

$$b_2^{\pm} = b_2 \pm b_{-2}, \qquad b_1^{\pm} = b_1 \pm b_{-1},$$

$$a_{22} = b_{22}^{+}b_{22}^{-} + b_{12}^{+}b_{21}^{-}, \qquad a_{21} = b_{22}^{-}b_{21}^{+} + b_{21}^{-}b_{11}^{+},$$

$$a_{12} = b_{12}^{-}b_{22}^{+} + b_{11}^{-}b_{12}^{+}, \qquad a_{11} = b_{12}^{-}b_{21}^{+} + b_{11}^{-}b_{11}^{+}, \qquad (6.5.49)$$

$$d_2 = b_{22}^{-}b_2^{-} + b_{21}^{-}b_1^{-} + b_2^{+}f_0, \qquad d_1 = b_{12}^{-}b_2^{-} + b_{11}^{-}b_1^{-} + b_1^{+}f_0,$$

$$a_{22}' = b_{22}^{-}b_{22}^{+} + b_{12}^{-}b_{21}^{+}, \qquad a_{21}' = b_{22}^{+}b_{21}^{-} + b_{21}^{+}b_{11}^{-},$$

$$a_{12}' = b_{12}^{+}b_{22}^{-} + b_{11}^{+}b_{12}^{-}, \qquad a_{11}' = b_{12}^{+}b_{21}^{-} + b_{11}^{+}b_{11}^{-},$$

$$d_2' = b_{22}^{+}b_2^{+} + b_{21}^{+}b_1^{+} + b_2^{-}f_0, \qquad d_1' = b_{12}^{+}b_2^{+} + b_{11}^{+}b_1^{+} + b_1^{-}f_0,$$

$$b = a_{22} + a_{11}, \qquad c = a_{21}a_{12} - a_{11}a_{22},$$

$$a^{-} = b_{22}^{-}b_{11}^{-} - b_{12}^{-}b_{21}^{-}, \qquad A_{1,2} = (k_{1,2}^{2} - a_{22})/a_{21},$$

$$k_1 = \left[(b + \sqrt{b^2 + 4c})/2\right]^{1/2}, \qquad k_2 = \left[(b - \sqrt{b^2 + 4c})/2\right]^{1/2},$$

$$\eta_1 = (d_1 f_0^2 + a_{12}d_2 - a_{22}d_1)/f', \qquad \eta_2 = (d_2 f_0^2 + a_{21}d_1 - a_{11}d_2)/f',$$

$$\eta_1' = (d_1' f_0^2 + a_{12}'d_2' - a_{22}'d_1')/f', \qquad \eta_2' = (d_2' f_0^2 + a_{21}'d_1' - a_{11}'d_2')/f',$$

$$f' = f_0^4 - bf_0^2 - c.$$

The coefficients $G_i(i = \pm 1, 2)$ are to be determined from radiation boundary conditions. If we consider a homogeneous cloud layer characterized by an optical depth τ_1 and assume that there is no diffuse radiation from the top and bottom of this layer, then the boundary conditions are

$$\left. \begin{array}{l} I_{-1,-2}(\tau = 0) = 0 \\ I_{1,2}(\tau = \tau_1) = 0 \end{array} \right\}. \qquad (6.5.50)$$

The boundary conditions can be modified to include nonzero diffuse radiation; G_i can be obtained by an inversion of a four-by-four matrix in Eq. (6.5.47). We note that the only difference in the four-stream formulation between infrared and solar wavelengths is the definition of f_0 and $b_i(i = \pm 1, 2)$. Finally, the upward and total (diffuse plus direct) downward fluxes at a given level τ are given by

$$F^{\uparrow}(\tau) = 2\pi(a_1\mu_1 I_1 + a_2\mu_2 I_2), \tag{6.5.51a}$$

$$F^{\downarrow}(\tau) = 2\pi(a_1\mu_1 I_{-1} + a_2\mu_2 I_{-2}) + \mu_0 F_{\odot}e^{-\tau/\mu_0}. \tag{6.5.51b}$$

We may also apply the four-stream solutions to nonhomogeneous atmospheres in the manner presented in Section 6.2.4.

The regular Gauss quadratures and weights in the four-stream approximation are $\mu_1 = 0.3399810, \mu_2 = 0.8611363, a_1 = 0.6521452$, and $a_2 = 0.3478548$, as defined in Table 6.1. When the isotropic surface reflection is included in this approximation or when it is applied to thermal infrared radiative transfer involving isothermal emission, double Gauss quadratures and weights ($\mu_1 = 0.2113248, \mu_2 = 0.7886752$, and $a_1 = a_2 = 0.5$) offer some advantage in flux calculations because $\sum_i a_i\mu_i = 0.5$ in this case. In the case of conservative scattering, $\tilde{\omega} = 1, \varphi_2^{\pm} = \Phi_2^{\pm} = 0.5$, the 4×4 matrix becomes 0 in Eq. (6.5.47). The solution for this equation does not exist. We may undertake direct formulation and solution from Eq. (6.4.46) by setting $\tilde{\omega} = 1$. However, we may also use $\tilde{\omega} = 0.999999$ in numerical calculations and obtain the results for conservative scattering. In the case $\tilde{\omega} = 0$, the multiple-scattering term vanishes.

We may incorporate a δ-function adjustment to account for the forward diffraction peak in the context of the four-stream approximation. In reference to Eq. (6.5.44), the normalized phase function expansion can be expressed by incorporating the δ-forward adjustment in the form

$$P_{\delta}(\cos\Theta) = 2f\delta(\cos\Theta - 1) + (1 - f)\sum_{\ell=0}^{N}\tilde{\omega}'_{\ell}P_{\ell}(\cos\Theta), \tag{6.5.52a}$$

where $\tilde{\omega}'_{\ell}$ is the adjusted coefficient in the phase function expansion. The forward peak coefficient f in the four-stream approximation can be evaluated by demanding that the next highest order coefficient in the prime expansion, $\tilde{\omega}'_4$, vanish. Setting $P(\cos\Theta) = P_{\delta}(\cos\Theta)$ and utilizing the orthogonal property of Legendre polynomials, we find

$$\tilde{\omega}'_{\ell} = [\tilde{\omega}_{\ell} - f(2\ell + 1)]/(1 - f). \tag{6.5.52b}$$

Letting $\tilde{\omega}'_4 = 0$, we obtain $f = \tilde{\omega}_4/9$. Based on Eq. (6.5.52b), $\tilde{\omega}'_{\ell}(\ell = 0, 1, 2, 3)$ can be evaluated from the expansion coefficients of the phase function, $\tilde{\omega}_{\ell}(\ell = 0, 1, 2, 3, 4)$.

The adjusted phase function from Eq. (6.5.52a) is then given by

$$P'(\cos\Theta) = \sum_{\ell=0}^{N}\tilde{\omega}'_{\ell}P_{\ell}(\cos\Theta). \tag{6.5.52c}$$

This equation, together with Eqs. (6.5.42a,b), constitutes the generalized similarity principle for radiative transfer. That is, the removal of the forward diffraction peak in scattering processes using adjusted single-scattering parameters is "equivalent" to actual scattering processes.

6.6 Radiative Transfer Including Polarization

6.6.1 Representation of a Light Beam

Electromagnetic waves are characterized by certain polarization configurations that are described by the vibration of the electric vector and by the phase difference between the two components of this vector. These components are commonly denoted by E_l and E_r, the electric fields parallel (l) and perpendicular (r) to a reference plane (defined in Section 5.2.3). This reference plane is commonly described as the plane containing the incident and scattered directions and is referred to as the scattering plane.

Electric fields are complex, oscillating functions and may be expressed by

$$E_l = a_l \exp[-i(\xi + \delta_l)], \tag{6.6.1a}$$

$$E_r = a_r \exp[-i(\xi + \delta_r)], \tag{6.6.1b}$$

where a_l and a_r are amplitudes, δ_l and δ_r are phases, $\xi = kz - \omega t$, $k = 2\pi/\lambda$, ω is the circular frequency, and $i = \sqrt{-1}$. From these two equations, we can show that the electric fields are defined by the equation of an ellipse (Exercise 6.12).

An electromagnetic wave can be represented by the amplitudes of its two electric components and their phase difference. Based on the Stokes parameters defined in Section 5.2.4, we have

$$I = E_l E_l^* + E_r E_r^* = a_l^2 + a_r^2, \tag{6.6.2a}$$

$$Q = E_l E_l^* - E_r E_r^* = a_l^2 - a_r^2, \tag{6.6.2b}$$

$$U = E_l E_r^* + E_r E_l^* = 2a_l a_r \cos \delta, \tag{6.6.2c}$$

$$V = i(E_r E_l^* - E_l E_r^*) = 2a_l a_r \sin \delta, \tag{6.6.2d}$$

where the superscript * denotes the complex conjugate and the phase difference $\delta = \delta_r - \delta_l$.

The Stokes parameters can be expressed in terms of the geometry defining an ellipse. Let β denote an angle whose tangent is the ratio of the axes of the ellipse traced by the endpoint of the electric vector, as displayed in Fig. 6.13. If the semimajor and -minor axes of the ellipse are given by a and b, respectively, then $\tan \beta = \pm b/a$. Also, let χ be the orientation angle between the major axis of the ellipse and the l direction. When the plane waves are time harmonics, we may express the electric field vectors along the l and r directions in terms of amplitude and phase using the
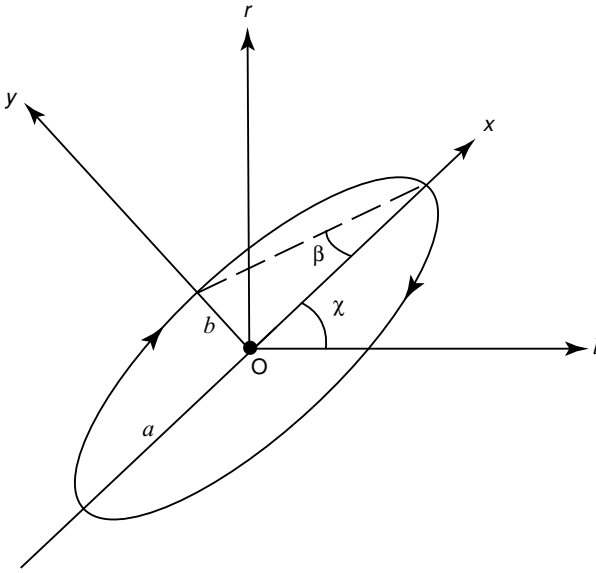
Figure 6.13   Geometric representation of elliptical polarization of a light beam in which the direction of propagation is into the paper, a and b are the lengths of the semimajor and –minor axes, respectively; χ is the orientation angle between the Ol and Ox axes; and β is the ellipticity angle whose tangent is the ratio of the ellipse traced by the endpoint of the electric vector, i.e., $\tan \beta = \pm b/a$, where $+$ and $-$ stand for the right- and left-handed polarization, respectively.

cosine representation in the forms

$$E_l = a_l \cos(\xi + \delta_l), \tag{6.6.3a}$$

$$E_r = a_r \cos(\xi + \delta_r). \tag{6.6.3b}$$

Let x and y denote the directions along the major and minor axes, respectively. Then the electric fields in the x–y plane may be written

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \cos \chi & \sin \chi \\ -\sin \chi & \cos \chi \end{pmatrix} \begin{pmatrix} E_l \\ E_r \end{pmatrix}, \tag{6.6.3c}$$

where E_x and E_y may also be expressed in terms of amplitudes (a, b) and an arbitrary phase δ_0 using cosine and sine representations such that they satisfy the elliptical equation in the forms

$$E_x = a \cos(\xi + \delta_0), \tag{6.6.3d}$$

$$E_y = \pm b \cos(\xi + \delta_0). \tag{6.6.3e}$$

After eliminating the propagation constant and all the phases using Eqs. (6.6.3a)–(6.6.3e), the Stokes parameters can be written in terms of the total intensity, and the
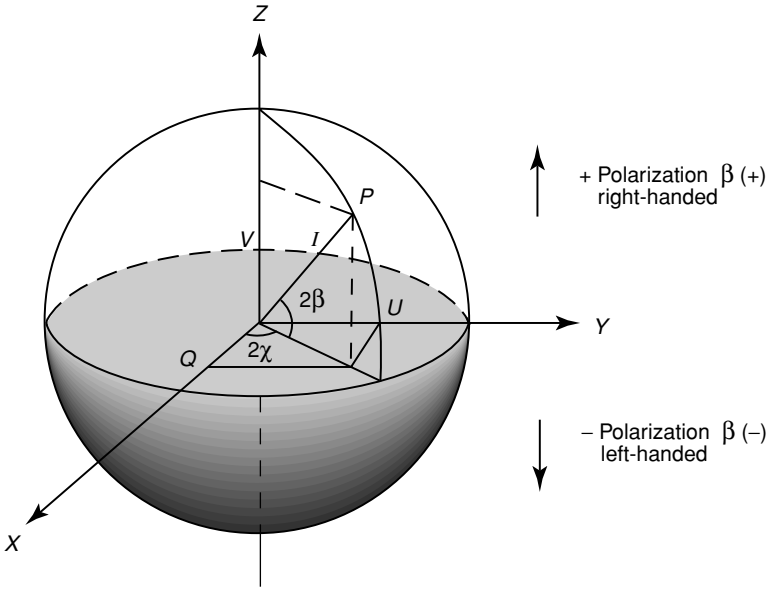
Figure 6.14   Polarization representation of the Stokes parameters (I, Q, U, V) on a Poincaré sphere. The angles β and χ are defined in Fig. 6.13. When β is positive (negative), the polarization configuration is referred to as right- (left-) handed polarization.

ellipticity and orientation angles in the forms (Exercise 6.13)

$$I = I_l + I_r, \tag{6.6.4a}$$

$$Q = I_l - I_r = I \cos 2\beta \cos 2\chi, \tag{6.6.4b}$$

$$U = I \cos 2\beta \sin 2\chi, \tag{6.6.4c}$$

$$V = I \sin 2\beta. \tag{6.6.4d}$$

It is noted that I and V are independent of the orientation angle χ. Equation (6.6.4) may be represented in Cartesian coordinates on a sphere called the *Poincaré sphere,* shown in Fig. 6.14. The radius of the sphere is given by I, and the zenithal and azimuthal angles are given by $\pi/2 - 2\beta$ and 2χ, respectively. Thus, Q, U, and V denote the lengths in the x, y, and z directions, respectively. On this sphere, the northern and southern hemispheres represent right-handed and left-handed elliptic polarizations, respectively. The north and south poles denote right-handed and left-handed circular polarizations, respectively, and points on the equatorial plane represent linear polarization. For a simple wave, we have $I^2 = Q^2 + U^2 + V^2$.

In representing the wave vibration using Eq. (6.6.2) we have assumed a constant amplitude and phase. However, the actual light beam consists of many simple waves in very rapid succession. Within a very short duration (on the order of, say, 1 second), millions of simple waves are collected by a detector. Consequently, measurable

intensities are associated with the superimposition of many millions of simple waves with independent phases. Let the operator $\langle\ \rangle$ denote the time average for a time interval (t_1, t_2). Then the Stokes parameters of the entire beam of light for this time interval may be expressed by

$$
\begin{aligned}
I &= \langle a_l^2 \rangle + \langle a_r^2 \rangle = I_l + I_r, \\
Q &= \langle a_l^2 \rangle - \langle a_r^2 \rangle = I_l - I_r, \\
U &= \langle 2a_l a_r \cos\delta \rangle, \\
V &= \langle 2a_l a_r \sin\delta \rangle.
\end{aligned}
\tag{6.6.5}
$$

In this case, we can prove that $I^2 \geq Q^2 + U^2 + V^2$. A light beam is generally characterized by partial elliptical polarization. If the phase differences between the two electric components are $0°$ or an integer order of $180°$ (i.e., $\beta = 0$), the light beam is linearly polarized ($V = 0$). If, on the other hand, the amplitudes of the two electric components are the same and their phase differences are an odd integer order of $90°$, then the light beam is circularly polarized ($Q = U = 0$). When the ellipticity angle β is positive or negative, the circular polarization is said to be right- or left-handed. The degree of polarization of a light beam is defined by

$$
PO = (Q^2 + U^2 + V^2)^{1/2}/I.
\tag{6.6.6a}
$$

If the ellipticity is not considered, we may define the degree of linear polarization in the form [see also Eq. (3.3.22)]

$$
LP = -Q/I = -(I_l - I_r)/(I_l + I_r).
\tag{6.6.6b}
$$

From the measurement perspective, we may represent the Stokes parameters in terms of detectable variables. Referring to Fig. 6.15, we introduce a retardation ε in the r direction with respect to the l direction and consider the component of the electric field vector in the direction making an angle Ψ with the positive l direction.
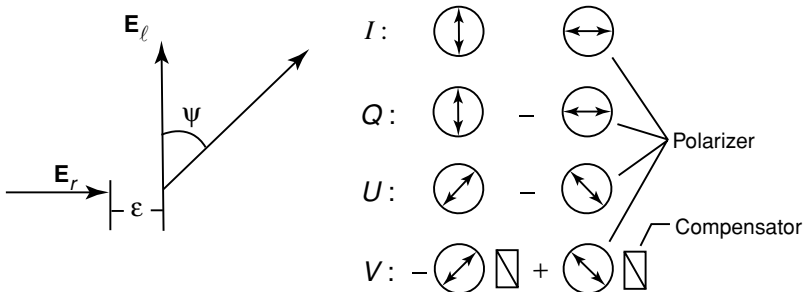
Figure 6.15   Representation of the electric field in terms of the retardation ε and the polarization angle Ψ.

Thus, for a simple wave at time t, the electric field may be represented in the form

$$E(t; \Psi, \varepsilon) = E_l \cos \Psi + E_r e^{-i\varepsilon} \sin \Psi$$
$$= a_l \cos \Psi e^{-i\xi} + a_r e^{-i(\delta+\varepsilon)-i\xi} \sin \Psi. \qquad (6.6.7)$$

The average intensity measured at a time interval (t_1, t_2) is then given by

$$I(\Psi, \varepsilon) = \langle E(t; \Psi, \varepsilon) E^*(t; \Psi, \varepsilon) \rangle$$
$$= \langle a_l^2 \rangle \cos^2 \Psi + \langle a_r^2 \rangle \sin^2 \Psi + \frac{1}{2} \langle 2a_l a_r \cos \delta \rangle \sin 2\Psi \cos \varepsilon$$
$$- \frac{1}{2} \langle 2a_l a_r \sin \delta \rangle \sin 2\Psi \sin \varepsilon. \qquad (6.6.8a)$$

By making use of Eq. (6.6.5) and noting that $I_l \cos^2 \Psi + I_r \sin^2 \Psi = (I + Q \cos 2\Psi)/2$, we obtain

$$I(\Psi, \varepsilon) = \frac{1}{2}[I + Q \cos 2\Psi + (U \cos \varepsilon - V \sin \varepsilon) \sin 2\Psi]. \qquad (6.6.8b)$$

On the basis of Eq. (6.6.8b), the Stokes parameters may be expressed by the retardation and polarization angles as follows:

$$\begin{aligned}
I &= I(0°, 0) + I(90°, 0), \\
Q &= I(0°, 0) - I(90°, 0), \\
U &= I(45°, 0) - I(135°, 0), \\
V &= -[I(45°, \pi/2) - I(135°, \pi/2)].
\end{aligned} \qquad (6.6.9)$$

Thus, the Stokes parameters of a light beam can be measured by a combination of a number of polarizers and a compensator (e.g., a quarter-wave plate) as illustrated in Fig. 6.15.

We may now define natural light. It is the light whose intensity remains unchanged and is unaffected by the retardation of one of the orthogonal components relative to the other when resolved in any direction in the transverse plane. That is to say, for natural light it is required that $I(\Psi, \varepsilon) = I/2$. The intensity is then independent of Ψ and ε. Thus, the necessary and sufficient condition that light be natural is $Q = U = V = 0$. Under this condition, the percentage of the degree of polarization defined in Eq. (6.6.6a) for natural light is zero. As a consequence, natural light is also referred to as *unpolarized light*. Light emitted from the sun is unpolarized. However, after interacting with molecules and particles through scattering events, the unpolarized sunlight generally becomes partially polarized. Natural light characterized by $Q = U = V = 0$ can be shown to be equivalent to a mixture of any two independent oppositely polarized streams of half the intensity (Exercise 6.15).

In the atmosphere, light is generally partially polarized, and its Stokes parameters (I, Q, U, V) may be decomposed into two independent groups characterized by

natural light and elliptically polarized light as follows:

$$
\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} I - (Q^2 + U^2 + V^2)^{1/2} \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} (Q^2 + U^2 + V^2)^{1/2} \\ Q \\ U \\ V \end{bmatrix}. \tag{6.6.10}
$$

Moreover, from Eq. (6.6.4), the plane of polarization can be determined by $\tan 2\chi = U/Q$, and the ellipticity by $\sin 2\beta = V/(Q^2 + U^2 + V^2)^{1/2}$.

6.6.2 Formulation

On inserting Eq. (6.6.3c) into the Stokes parameters defined on the x–y coordinate and after some straightforward analysis, we have

$$
\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix}_{x-y} = \mathbf{L}(\chi) \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix}, \tag{6.6.11}
$$

where the transformation matrix for the Stokes parameters is given by

$$
\mathbf{L}(\chi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\chi & \sin 2\chi & 0 \\ 0 & -\sin 2\chi & \cos 2\chi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.6.12}
$$

From Eqs. (6.6.11) and (6.6.12), it is clear that I and V are invariant in the transformation process. We also note that $\mathbf{L}(\chi_1)\mathbf{L}(\chi_2) = \mathbf{L}(\chi_1 + \chi_2)$, and the inverse matrix $\mathbf{L}^{-1}(\chi) = \mathbf{L}(-\chi)$ (Exercise 6.16).

Having the transformation matrix defined, we can now formulate the transfer equation to include polarization. In Section 6.1, the transfer equation for plane-parallel atmospheres was presented for diffuse intensity, without taking into account the effect of polarization. In order to describe the radiation field completely at each point in space, we replace the scalar intensity I by the vector intensity $\mathbf{I} = (I, Q, U, V)$. The four Stokes parameters give, respectively, the intensity, the degree of polarization, the plane of polarization, and the ellipticity of the light waves, as functions of the incoming and outgoing directions.

The transfer equation given in Eq. (1.4.22) may now be written in vector form as follows:

$$
\mu \frac{d\mathbf{I}(\tau; \mu, \phi)}{d\tau} = \mathbf{I}(\tau; \mu, \phi) - \mathbf{J}(\tau; \mu, \phi), \tag{6.6.13}
$$

where the source function is a vector consisting of four elements. To obtain the expression for the vector source function, consider a differential increment $d\mathbf{J}(\tau; \mu, \phi; \mu', \phi')$, produced by multiple scattering involving a pencil of radiation of solid angle $d\Omega'$ in
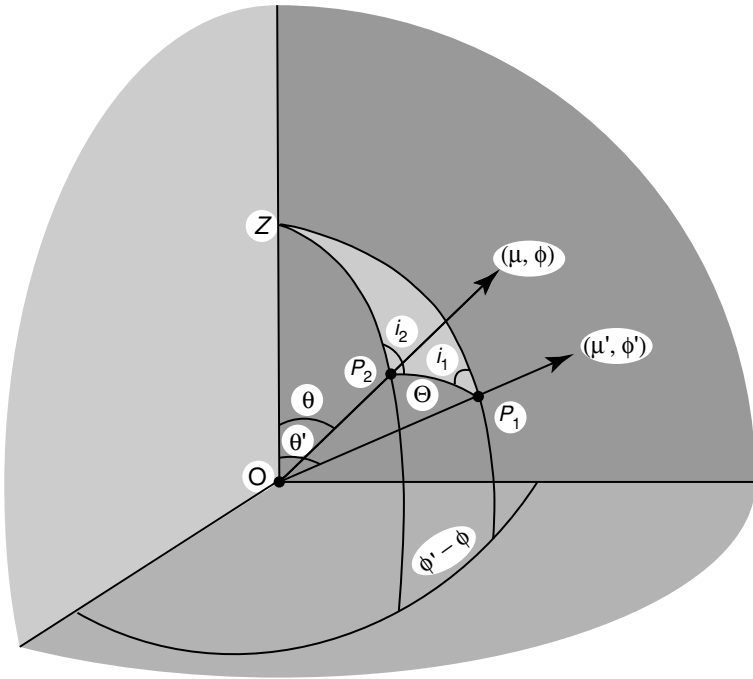
Figure 6.16 Scattering plane OP_1P_2 with respect to the meridian planes OP_1Z and OP_2Z (see text for the definitions of the angles).

the direction (μ', ϕ'). The diffuse intensity vector $\mathbf{I}(\tau; \mu', \phi')$, which generates the source term, is in reference to the meridian plane OP_1Z, as shown in Fig. 6.16. However, the phase matrix derived from the scattering theory [e.g., see Eq. (5.2.113)] is in reference to the plane of scattering OP_1P_2 that contains the incident and scattered beams. Thus, we must first transform $\mathbf{I}(\tau; \mu', \phi')$ to the plane of scattering in order to obtain the proper source function. We may first transform $\mathbf{I}(\tau; \mu', \phi')$ to the plane of scattering by applying the transformation matrix $\mathbf{L}(-i_1)$, where i_1 denotes the angle between the meridian plane OP_1Z and the plane of scattering OP_1P_2, and the minus sign signifies that the rotation of the plane is counterclockwise. Thus, the contribution to the source function with reference to the plane of scattering at P_2 is given by

$$\tilde{\omega}\mathbf{P}(\Theta)\mathbf{L}(-i_1)\mathbf{I}(\tau; \mu', \phi')\,d\Omega'/4\pi. \tag{6.6.14}$$

To transform this vector to the scattering direction (μ, ϕ), i.e., the meridian plane OP_2Z, we must again apply the transformation matrix $\mathbf{L}(\pi - i_2)$ through the angle $(\pi - i_2)$ clockwise, where i_2 denotes the angle between the meridian plane OP_2Z and the plane of scattering OP_1P_2. It follows that the desired differential source function due to the diffuse component is

$$d\mathbf{J}(\tau; \mu, \phi; \mu', \phi') = \tilde{\omega}\mathbf{L}(\pi - i_2)\mathbf{P}(\Theta)\mathbf{L}(-i_1)\mathbf{I}(\tau; \mu', \phi')\,d\Omega'/4\pi. \tag{6.6.15}$$

Thus, by performing the integration over all directions (μ', ϕ'), we obtain the source

function vector for multiple scattering as follows:

$$\mathbf{J}(\tau; \mu, \phi; \mu', \phi') = \frac{\tilde{\omega}}{4\pi} \int_0^{2\pi} \int_{-1}^1 \mathbf{Z}(\mu, \phi; \mu', \phi') \mathbf{I}(\tau; \mu', \phi') d\mu' d\phi', \quad (6.6.16)$$

where the phase matrix is defined by

$$\mathbf{Z}(\mu, \phi; \mu', \phi') = \mathbf{L}(\pi - i_2)\mathbf{P}(\Theta)\mathbf{L}(-i_1). \quad (6.6.17)$$

Note that we differentiate $\mathbf{Z}$ and $\mathbf{P}$ by using the terms phase matrix and scattering phase matrix, respectively. From the spherical trigonometry, as illustrated in Appendix C, the angles i_1 and i_2 can be expressed by

$$\cos i_1 = \frac{-\mu + \mu' \cos \Theta}{\pm(1 - \cos^2 \Theta)^{1/2}(1 - \mu'^2)^{1/2}}, \quad (6.6.18)$$

$$\cos i_2 = \frac{-\mu' + \mu \cos \Theta}{\pm(1 - \cos^2 \Theta)^{1/2}(1 - \mu^2)^{1/2}}, \quad (6.6.19)$$

where the plus sign is to be used when $\pi < \phi - \phi' < 2\pi$ and the minus sign is to be used when $0 < \phi - \phi' < \pi$. Also note that $\cos \Theta$ has been defined in Eq. (3.4.7).

Following the same procedures, the direct component of the source function associated with the point source $\mathbf{I}_\odot(-\mu, \phi) = \delta(\mu - \mu_0)\delta(\phi - \phi_0)\mathbf{F}_\odot$ is given by

$$\mathbf{J}(\tau; \mu, \phi) = \frac{\tilde{\omega}}{4\pi} \mathbf{Z}(\mu, \phi; -\mu_0, \phi_0)\mathbf{F}_\odot e^{-\tau/\mu_0}. \quad (6.6.20)$$

Thus, the equation of transfer of sunlight including polarization can be written as follows:

$$\mu \frac{d\mathbf{I}(\tau; \mu, \phi)}{d\tau} = \mathbf{I}(\tau; \mu, \phi) - \frac{\tilde{\omega}}{4\pi} \int_0^{2\pi} \int_{-1}^1 \mathbf{Z}(\mu, \phi; \mu', \phi')\mathbf{I}(\tau; \mu', \phi') d\mu' d\phi'$$

$$- \frac{\tilde{\omega}}{4\pi} \mathbf{Z}(\mu, \phi; -\mu_0, \phi_0)\mathbf{F}_\odot e^{-\tau/\mu_0}. \quad (6.6.21)$$

Comparing Eq. (6.6.21) with Eq. (6.1.1) without the emission component, we see that the scalar intensity is now replaced by a vector intensity consisting of four elements. In the preceding formulation, we have assumed that an optical depth can be defined for the medium. This applies to spherical particles that are randomly located with the scattering phase matrix $\mathbf{P}$ given by Eq. (5.2.113), and to nonspherical particles that are randomly oriented in space with the scattering phase matrix $\mathbf{P}$ given by Eq. (5.4.31). In these cases, we may replace the phase function with the four-by-four scattering phase matrix to account for the full polarization effect and employ the adding method presented in Section 6.4 to proceed with numerical calculations.

The phase matrix $\mathbf{Z}$ obeys a number of unique properties associated with the symmetry principle of light beams. For the preceding cases, a light beam can reverse its direction with final results being the same such that

$$\mathbf{Z}(-\mu, -\phi; -\mu', -\phi') = \mathbf{Z}(\mu, \phi; \mu', \phi'). \quad (6.6.22)$$

Other relationships have been developed by Hovenier (1969). These relationships can be employed to optimize numerical computations involving multiple scattering

processes in planetary atmospheres. Finally, it should be pointed out that a new formulation is required for nonspherical particles having a specific orientation (Section 6.7.1).

6.7   Advanced Topics in Radiative Transfer

6.7.1   Horizontally Oriented Ice Particles

The scattering of a light beam by a nonspherical particle depends on the directions of the incoming and outgoing radiation, and on the orientation of the particle with respect to the incoming beam. To formulate the transfer of solar radiation in a medium composed of horizontally oriented nonspherical particles, we begin by assuming that such a medium is plane-parallel so that the intensity varies only in the z direction. In reference to Fig. 6.17, we select a fixed coordinate system xyz such that the z axis is in the zenith direction. Also, we let $x'y'z'$ represent a coordinate system in reference to the incoming light beam, which is placed on the z' axis. Angles ϕ', ϕ, γ', and γ are azimuthal angles corresponding to θ', θ, α', and α, denoted in the figure, and Θ is the scattering angle. The scattering parameters for a nonspherical particle, including the phase function, and the extinction and scattering cross sections may be expressed with respect to either of these two coordinate systems. Thus, we may write
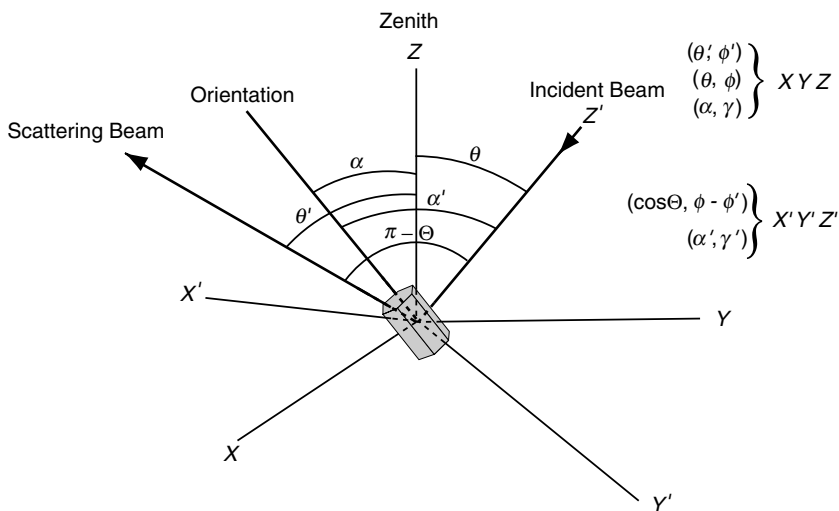
Figure 6.17   Geometry of single scattering involving a nonspherical particle. The coordinate system (x', y', z') is in reference to the incident light beam, while (x, y, z) is fixed in space. The angles (θ, θ') are zenith angles associated with the incident and scattered beams with respect to the (x, y, z) coordinates, while (ϕ, ϕ') are corresponding azimuthal angles. The angles (α, α') are zenith angles with respect to the orientation of the particle, while (γ, γ') are corresponding azimuthal angles. These angles can be transferred to the (x', y', z') coordinates in terms of a set of angles, $(\Theta, \phi - \phi')$ and (α', γ'), for the analysis of radiative transfer involving horizontally oriented ice crystals.

symbolically

$$P(\alpha, \gamma; \mu', \phi'; \mu, \phi) = P(\alpha', \gamma'; \cos \Theta, \phi - \phi'),$$

$$\sigma_{e,s}(\alpha, \gamma; \mu', \phi') = \sigma_{e,s}(\alpha', \gamma'). \tag{6.7.1}$$

Here we note that the phase function depends on the directions of the incident and scattered beams as well as on the orientation of the nonspherical particle. The extinction and scattering cross sections, however, depend only on the direction of the incident beam and the orientation of the particle.

For a sample of nonspherical particles randomly oriented in space, average single-scattering properties may be expressed in the forms

$$P(\cos \Theta, \phi - \phi') = \frac{1}{2\pi \sigma_s} \int_0^{2\pi} \int_0^{\pi/2} P(\alpha', \gamma'; \cos \Theta, \phi - \phi') \sigma_s(\alpha', \gamma') \sin \alpha' \, d\alpha' d\gamma',$$

$$\sigma_{e,s} = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\pi/2} \sigma_{e,s}(\alpha', \gamma') \sin \alpha' d\alpha' d\gamma'. \tag{6.7.2}$$

It is clear that the extinction and scattering cross sections for randomly oriented nonspherical particles are independent of direction. Moreover, since $\cos \Theta$ can be expressed in terms of μ, ϕ, and μ', ϕ', as defined in Eq. (3.4.7), the source function in this case has the same form as that defined in Eq. (3.4.6). If all of the nonspherical particles have rotational symmetry (e.g., circular cylinders), then the phase function is independent of the azimuthal angle $\phi - \phi'$. Consequently, multiple scattering of the diffuse intensity in randomly oriented, symmetrical nonspherical particles can be formulated by following conventional procedures such as those presented for the adding method.

The spatial orientation of hexagonal and irregular ice crystals in cirrus clouds is a significant factor in the discussion of the transfer of radiation in the atmosphere. The fact that numerous halos and arcs have been observed demonstrates that specific orientation of ice particles must exist in some cirrus. Based on laboratory experiments, cylinders with a diameter-to-length ratio of less than 1 tend to fall with their long axes horizontally oriented. Observations of columnar and plate crystals in cirrus clouds have shown that these particles fall with their major axes parallel to the ground. The orientation of ice particles in cirrus clouds has been observed by numerous lidar measurements based on the depolarization technique in the backscattering direction (see Section 7.6.2). The depolarization ratio of the backscattered return from horizontally oriented plates is close to zero, but this ratio increases significantly as the lidar scans a few degrees off the vertical. Specific orientation occurs when the ice particles have relatively large sizes and defined shapes, such as columns and plates. However, if the ice crystals are irregular, such as aggregates, preferred orientation is unlikely to occur. Furthermore, smaller ice crystals in cirrus clouds where substantial turbulence occurs tend to orient in three-dimensional space. Finally, it has been noted that ice particle orientation and alignment are closely modulated by the electric field in clouds.

In the case of horizontally oriented ice crystals, their single-scattering parameters are dependent on the direction of the incident light beam. Thus, the conventional formulation for the multiple-scattering problem requires modification. The basic

equation for the transfer of solar radiation in an optically anisotropic medium has been discussed and formulated in the first edition of this text. More recently, Takano and Liou (1989b) have used realistic scattering parameters and the Stokes vector for horizontally oriented ice crystals in association with the adding method for radiative transfer. Takano and Liou (1993) have further presented the theoretical formulation and numerical calculations involving the transfer of polarized thermal infrared radiation in optically anisotropic media with a specific application to horizontally oriented ice particles. In the following, we present a unified theoretical formulation that is applicable to both solar and thermal infrared radiative transfer, including polarization for horizontally oriented ice crystals.

In the case of ice particles randomly oriented in a horizontal plane, we have $\alpha = \pi/2$ from Eq. (6.7.1). Thus, the phase function and cross sections are dependent only on the incident angle and may be symbolically written in the forms

$$P(\mu', \phi'; \mu, \phi) = \frac{2}{\pi} \int_0^{\pi/2} P(\pi/2, \gamma; \mu', \phi'; \mu, \phi) \, d\gamma, \qquad (6.7.3a)$$

$$\sigma_{e,s}(\mu) = \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{\pi/2} \sigma_{e,s}(\pi/2, \gamma; \mu, \phi) \, d\gamma \, d\phi. \qquad (6.7.3b)$$

With the preceding understanding of the incident direction of a light beam with respect to particle geometry, we may define a differential normal optical depth such that $d\tilde{\tau} = -\tilde{\beta}_e \, dz$, where the vertical extinction coefficient $\tilde{\beta}_e = \beta_e(\mu = 1)$ and z is the distance. The general equation governing the transfer of the Stokes vector may be expressed in the form

$$\mu \frac{d\mathbf{I}(\tilde{\tau}; \mu, \phi)}{d\tilde{\tau}} = \mathbf{k}(\mu)\mathbf{I}(\tilde{\tau}; \mu, \phi) - \mathbf{J}(\tilde{\tau}; \mu, \phi), \qquad (6.7.4a)$$

where $\mathbf{I} = (I, Q, U, V)$ and the actual extinction coefficient normalized by the vertical extinction coefficient is defined by

$$\mathbf{k}(\mu) = \beta_e(\mu)/\tilde{\beta}_e. \qquad (6.7.4b)$$

For horizontally oriented particles the extinction coefficient is dependent on both the energy characteristics of the incident beam and its state of polarization, referred to as *dichroism* of the scattering medium. This generally occurs when the light beam passes through a cloud of aligned nonspherical particles associated with an electric and/or magnetic field, as noted previously. Because of dichroism, the extinction coefficients corresponding to the Stokes vector are represented by the 4×4 extinction matrix. For nonspherical particles randomly oriented in a plane, the extinction matrix may be written in the form (Martin, 1974; Mishchenko, 1991)

$$\boldsymbol{\beta}_e = \begin{bmatrix} \beta_e & \beta_{pol} & 0 & 0 \\ \beta_{pol} & \beta_e & 0 & 0 \\ 0 & 0 & \beta_e & \beta_{cpol} \\ 0 & 0 & -\beta_{cpol} & \beta_e \end{bmatrix}, \qquad (6.7.5)$$

where β_{pol} and β_{cpol} are polarized and cross-polarized components, respectively, of

the extinction coefficients with respect to the incident Stokes vector. For all practical purposes, we may use the scalar β_e for applications to ice-crystal cases.

The source function in the basic radiative transfer equation may be written as follows:

$$\mathbf{J}(\tilde{\tau}; \mu, \phi) = \frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} \mathbf{k}(\mu')\tilde{\boldsymbol{\omega}}(\mu')\mathbf{Z}(\mu, \phi; \mu', \phi')\mathbf{I}(\tilde{\tau}, \mu', \phi') \, d\mu' d\phi'$$

$$+ \frac{1}{4\pi}\mathbf{k}(-\mu_0)\tilde{\boldsymbol{\omega}}(-\mu_0)\mathbf{Z}(\mu, \phi; -\mu_0, \phi_0)\mathbf{F}_\odot \exp[-\mathbf{k}(-\mu_0)\tilde{\tau}/\mu_0]$$

$$+ \mathbf{k}(\mu)[1 - \tilde{\boldsymbol{\omega}}(\mu)]B(T)\mathbf{I}_e, \tag{6.7.6a}$$

where the single-scattering albedo is defined by

$$\tilde{\boldsymbol{\omega}}(\mu) = \boldsymbol{\beta}_s(\mu)/\boldsymbol{\beta}_e(\mu), \tag{6.7.6b}$$

where $\boldsymbol{\beta}_s$ is the scattering coefficient matrix, which has a form similar to the extinction coefficient matrix, $\mathbf{F}_\odot$ represents the Stokes vector for the incident solar irradiance, $B(T)$ is the Planck intensity at temperature T, and $\mathbf{I}_e = (I, Q_e, 0, 0)$, with $-Q_e$ the linear polarization component associated with emission. The phase matrix,

$$\mathbf{Z}(\mu, \phi; \mu', \phi') = \mathbf{L}(\pi - i_2)\mathbf{P}(\mu, \phi; \mu', \phi')\mathbf{L}(i_1), \tag{6.7.7}$$

is duplicated here for the continuity of discussion [see also Eq. (6.6.17)], and the transformation matrix is given in Eq. (6.6.12). In general, the scattering phase matrix $\mathbf{P}$ consists of 16 elements as defined in Eq. (5.4.30). In Eq. (6.7.6a), the second and third terms on the right-hand side represent the contributions from direct solar radiation and thermal emission from a medium having a temperature T that is azimuthally independent. Also note that $\mathbf{k}\tilde{\omega} = \boldsymbol{\beta}_s/\tilde{\boldsymbol{\beta}}_e$. For wavelengths shorter than about 3.7 μm, thermal emission within the earth–atmosphere system can be neglected in comparison to radiation from the sun. For wavelengths longer than 5 μm, the reverse is true. Between 3.7 and 5 μm, the relative importance of thermal emission and solar reflection for a cloud layer depends largely on the position of the sun and the cloud temperature. We have stated these important constraints in various parts of this text. If the particles are randomly oriented in space in such a manner that each one of them has a plane of symmetry and the law of reciprocity may be applied, the scattering phase matrix $\mathbf{P}$ consists of only six independent elements as shown in Eq. (5.4.32). In this case, $\mathbf{k}(\mu) = 1$ and $\beta_s, \beta_e, \tilde{\omega}$ are independent of μ.

We may approach the radiative transfer problem involving ice particles randomly oriented in a horizontal plane using the adding method introduced in Section 6.4.2. The phase function and single-scattering parameters are now dependent on the direction of the incident beam. From Eqs. (6.7.4a) and (6.7.6a), we may omit the multiple-scattering term and define the reflection and transmission functions based on single-scattering and optically thin approximations, as shown in Eqs. (6.4.8a,b). However, in the case of solar radiation, we need to use the normal optical depth $\Delta\tilde{\tau}$ and the single-scattering albedo, which is a function of the cosine of the solar zenith angle, i.e., $\tilde{\omega}(\mu_0)$. We must also distinguish between the reflection and transmission functions for radiation from above and below, since the phase functions for horizontally oriented

ice particles differ in these two configurations. The adding equations will be the same, except that the optical depth is replaced by $k(\mu_0)\tilde{\tau}_{a,b}$, where k is the scalar ratio of the vertical extinction coefficient to the actual extinction coefficient [see Eq. (6.7.4b)].

Figure 6.18 shows an example of the bidirectional reflectances for horizontally oriented (2D) and randomly oriented (3D) columns using a representative cirrostratus ice crystal size distribution with an optical depth of 1 in the plane defined by the zenith (θ) and relative azimuthal ($\phi - \phi_0$) angles. The wavelength and the cosine of the solar zenith angle used are 0.63 μm and 0.5, respectively. Note that the domain for the reflectances is from $60°$ to $180°$ scattering angles, which does not cover the commonly observed halos located at $22°$ and $46°$. In the 3D case, the maximum at $\theta \approx 80°$ and close to the principal plane $\phi - \phi_0 = 0°$, is related to the limb brightening. Otherwise, the reflectance variations are relatively small in the linear scale. In the case of 2D columns, we see numerous reflection maxima. The chief ones are: (1) the subsun located at $\theta \approx 60°$ in the principal plane produced by external reflections; (2) the lower tangent arc located at $\theta \approx 80°$ in the principal plane generated by two refractions; and (3) the antisolar peak located at $\theta \approx 60°$ and $\phi - \phi_0 = 180°$ caused by internal reflections. Much larger anisotropy occurs in this case as compared with the 3D case. In realistic cirrus clouds, we would anticipate that some of the large and defined ice particles are horizontally oriented. In Section 7.3.5, we demonstrate that the reflected polarization of sunlight contains information regarding ice crystal shape and orientation.

6.7.2   Three-Dimensional Nonhomogeneous Clouds

From satellite cloud pictures, as well as our day-to-day experience, we see that a portion of the clouds and cloud systems that cover the earth either are finite in extent or occur in the form of cloud bands. This is especially evident in the tropics and the midlatitudes in the summertime. One generally common feature is the presence of cumulus clouds whose horizontal dimensions are on the same order as their vertical dimensions. Satellite mapping of the optical depth in the midlatitude and tropical regions has illustrated that cirrus clouds are frequently finite in nature and display substantial horizontal variabilities. Vertical inhomogeneity of the ice crystal size distribution and ice water content is also demonstrated in the replicator sounding observations (see Fig. 5.3) and the time series of backscattering coefficients derived from lidar returns. Thus, the potential effects of cloud geometry and inhomogeneity on the transfer of radiation must be studied to understand their impact on the radiative properties of the atmosphere, as well as to accurately interpret radiometric measurements from the ground, the air, and space.

Following the discussion presented in Section 1.4.5, the basic steady-state radiative transfer diffuse intensity, I, can be expressed in the form

$$-\frac{1}{\beta_e(\mathbf{s})}(\mathbf{\Omega} \cdot \mathbf{\nabla})I(\mathbf{s}, \mathbf{\Omega}) = I(\mathbf{s}, \mathbf{\Omega}) - J(\mathbf{s}, \mathbf{\Omega}), \qquad (6.7.8)$$
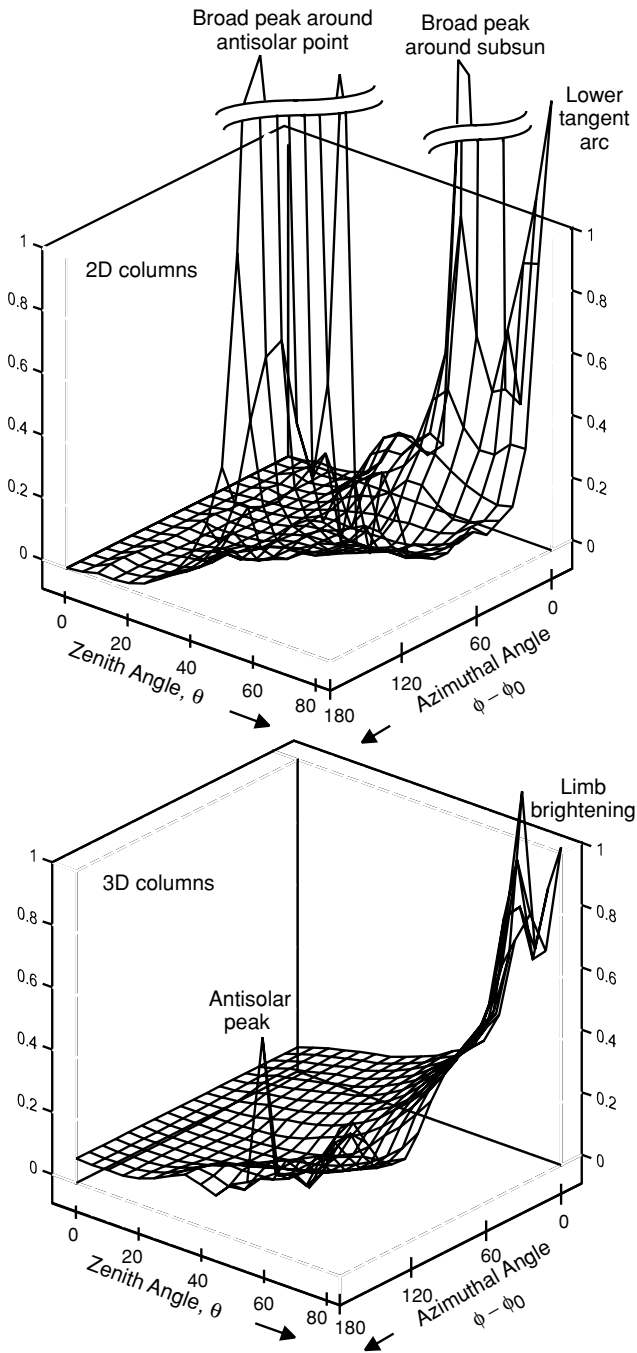
Figure 6.18 Bidirectional reflectances at a wavelength of 0.63 μm for ice columns randomly oriented in space (3D) and oriented parallel to the ground (2D columns) in the plane of zenith and azimuthal angles. The cosine of the solar zenith angle and the optical depth used are 0.5 and 1, respectively. For the 2D column case, three peak features are marked in the diagram.

where **s** is the position vector; $\mathbf{\Omega}$ is a unit vector representing the angular direction of scattering through the position vector; β_e is the extinction coefficient for cloud particles, which is a function of the position vector; and the source function, which is produced by the single scattering of the direct solar irradiance, multiple scattering of the diffuse intensity, and emission of the cloud, can be written as follows:

$$J(\mathbf{s}, \mathbf{\Omega}) = \frac{\tilde{\omega}(\mathbf{s})}{4\pi} \int_{4\pi} I(\mathbf{s}, \mathbf{\Omega}')P(\mathbf{s}; \mathbf{\Omega}, \mathbf{\Omega}')\, d\mathbf{\Omega}'$$

$$+ \frac{\tilde{\omega}(\mathbf{s})}{4\pi} P(\mathbf{s}; \mathbf{\Omega}, \mathbf{\Omega}_0) F_\odot \exp\left[-\int_0^{s_0} \beta_e(\mathbf{s}')\, d\mathbf{s}' \right] + [1 - \tilde{\omega}(\mathbf{s})]B[T(\mathbf{s})],$$

$$(6.7.9)$$

where $\tilde{\omega} = \beta_s/\beta_e$ is the single-scatterng albedo with β_s the scattering coefficient; the phase function P is defined by the position of the light beam and the incoming and outgoing solid angles $\mathbf{\Omega}'(\mathbf{\Omega}_0)$ and $\mathbf{\Omega}$, respectively; $F_\odot$ is the incident solar irradiance; s_0 is defined as a spatial coordinate in the direction of the incident solar radiation; and $B(T)$ is the Planck function of temperature T. Solutions for I in Eqs. (6.7.8) and (6.7.9) in multidimensional space must be carried out numerically subject to the coordinate system imposed.

We shall consider the Cartesian coordinate system in which three-dimensional flux densities may be defined by

$$F_{\pm x_i}(x, y, z) = \int_{2\pi} I(x, y, z; \mathbf{\Omega})\mathbf{\Omega}_{x_i}\, d\mathbf{\Omega}, \qquad (6.7.10)$$

where $x_1 = x$, $x_2 = y$, $x_3 = z$, $\mathbf{\Omega}_x = (1 - \mu^2)^{1/2} \cos\phi$, $\mathbf{\Omega}_y = (1 - \mu^2)^{1/2} \sin\phi$, and $\mathbf{\Omega}_z = \mu$. There are six flux components corresponding to three coordinates. Angular integrations over the upward and downward hemispheres are the same as in the plane–parallel case, i.e., $(2\pi, 0)$ for ϕ and $(\pm 1, 0)$ for μ. The angular integrations in the x-direction are $(\pi/2, -\pi/2)$ and $(-1, 1)$, and $(3\pi/2, \pi/2)$ and $(-1, 1)$. In the y-direction, the angular integrations are $(\pi, 0)$ and $(-1, 1)$, and $(2\pi, \pi)$ and $(-1, 1)$. The local rate of change of temperature is produced by the 3D radiative flux divergence in the form

$$\frac{\partial T}{\partial t}(x, y, z) = -\frac{1}{\rho C_p}\mathbf{\nabla} \cdot \mathbf{F}, \qquad (6.7.11a)$$

where

$$\mathbf{F} = \mathbf{i}F_x + \mathbf{j}F_y + \mathbf{k}F_z, \qquad (6.7.11b)$$

with $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ the unit vectors and F_x, F_y, and F_z the net flux densities in the x, y, and z directions, respectively.

A final note is in order. The conventional definition of the absorbed flux in a cloud layer from both the measurement and calculation perspectives is given by the divergence of net fluxes at the cloud top (z_t) and bottom (z_b) in the form

$$F_{\text{abs}} = \left[F^\downarrow(z_t) - F^\uparrow(z_t)\right] - \left[F^\downarrow(z_b) - F^\uparrow(z_b)\right]. \qquad (6.7.12a)$$

Dividing by the downward solar flux at the cloud top, $F^\downarrow(z_t)$, we obtain the relative percentage as follows:

$$a = 1 - r - t, \tag{6.7.12b}$$

with the absorptance $a = F_{abs}/F^\downarrow(z_t)$, the reflectance $r = F^\uparrow(z_t)/F^\downarrow(z_t)$, and the net transmittance $t = [F^\downarrow(z_b) - F^\uparrow(z_b)]/F^\downarrow(z_t)$, where the upward flux at the cloud base is related to the contribution from the surface and the atmosphere below the cloud. The preceding definition of cloud absorption is correct if the horizontal extent of the cloud system is sufficiently large, such as those associated with large-scale frontal activities. However, for a cloud whose horizontal scale is comparable to or less than its vertical scale, such as tropical cumulus towers, the absorption definition based on the fluxes at the cloud top and bottom requires modification.

From a theoretical perspective, and if Cartesian coordinates are imposed, the side fluxes must be included in the discussion of cloud absorption. The problem is further complicated because the flux is a function of position (x, y, z) and appropriate domain averages are needed to obtain absorption in the cloud as a unit. Let the areas corresponding to the x, y, and z directions be A_x, A_y, and A_z, respectively. Based on the energy conservation principle, we must have

$$\int_{A_z} F_\odot \mu_0 \, dxdy + \int_{A_x} F_\odot \left(1 - \mu_0^2\right)^{1/2} \cos\phi_0 \, dydz + \int_{A_y} F_\odot \left(1 - \mu_0^2\right)^{1/2} \sin\phi_0 \, dxdz$$

$$= \int_{A_z} \left(F_z^\uparrow + F_z^\downarrow\right) dxdy + \int_{A_x} \left(F_x^\rightarrow + F_x^\leftarrow\right) dydz$$

$$+ \int_{A_y} \left(F_y^\rightarrow + F_y^\leftarrow\right) dxdz + \int_V f_a \, dxdydz, \tag{6.7.13}$$

where V is the volume and f_a is the absorbed flux per volume. Averaging over the respective areas and volume, we obtain

$$F_\odot\left[\mu_0 A_z + \left(1 - \mu_0^2\right)^{1/2} \cos\phi_0 A_x + \left(1 - \mu_0^2\right)^{1/2} \sin\phi_0 A_y\right]$$

$$= \left(\bar{F}_z^\uparrow + \bar{F}_z^\downarrow\right) A_z + \left(\bar{F}_x^\rightarrow + \bar{F}_x^\leftarrow\right) A_x + \left(\bar{F}_y^\rightarrow + \bar{F}_y^\leftarrow\right) A_y + \bar{f}_a V, \tag{6.7.14}$$

where the flux notations are self-explanatory.

6.7.2.1   Monte Carlo Method

The Monte Carlo method involves releasing photons from a source and tracing them through a medium that is divided into a suitable number of cubic cells. The absorption and scattering of photons can be considered stochastic processes in which the scattering phase function may be thought of as a transformation probability function that redistributes the photons in different directions. The single-scattering properties within the predivided cubic cells are prescribed. The Monte Carlo program then computes the free path length of a single photon from its initial entry point through these cells.

Consider a nonhomgeneous finite cloud and let the incident energy be $h\tilde{\nu}$. Let l denote the path length that a photon travels in the cloud before the first scattering, and let RN be a random number in the interval $(0, 1)$ generated by computer. The mean path length l_0 between scattering must be inversely proportional to the scattering coefficient β_s (in units of per length) associated with the cloud particles. Using the Poisson distribution, we have

$$\text{RN} = e^{-l/l_0}, \quad l = l_0 \ln(1/\text{RN}). \tag{6.7.15a}$$

When the photon encounters scattering, its new direction in terms of the scattering and azimuthal angles may be determined by

$$\left. \begin{array}{l} \int_0^\Theta P(\cos\Theta) \sin\Theta\, d\Theta = \text{RN} \cdot \int_0^\pi P(\cos\Theta) \sin\Theta\, d\Theta \\ \phi = 2\pi \cdot \text{RN} \end{array} \right\}, \tag{6.7.15b}$$

where P is the phase function and it converts the incident direction (μ', ϕ') to the scattered direction (μ, ϕ) via the definition of the scattering angle. With the new direction, a new path length for the next scattering is then determined from Eq. (6.7.15a) and this procedure continues until the photon is either absorbed in the cloud or escapes through a cloud boundary.

Analogous to the definition of the scattering path length, l_0, we may define the absorption and extinction path lengths as n_0 and m_0, respectively. Thus, we have $m_0/n_0 = 1 - \tilde{\omega}$, and the mean number of scattering events, which is an integer, is given by

$$N_0 \geq \frac{n_0}{m_0} = \frac{1}{1 - \tilde{\omega}}. \tag{6.7.15c}$$

In the extreme case involving conservative scattering, $\tilde{\omega} = 1$, there will be infinite numbers of scattering events and the computation continues until the photon departs the cloud. In the other extreme case when $\tilde{\omega} = 0$, scattering does not occur and the photon is absorbed immediately. The number of scattering events before absorption is also determined from the Poisson distribution and is given by

$$N = N_0 \ln(1/\text{RN}). \tag{6.7.15d}$$

The Monte Carlo program contains codes that sample the relevant behavior of each photon as a function of depth in the cloud. It will keep a record of the number and direction of photons that cross a predivided plane in a cloud. The intensity may be collected within preset solid angles from which the flux density can be evaluated.

In principle, the Monte Carlo method may be applied to the transfer of radiation in a medium with any geometric configuration (see, e.g., Marchuk *et al.*, 1980). The calculation involves technical input/output operations. The results computed from the Monte Carlo method are subject to statistical fluctuations, which decrease in magnitude as the square root of the number of photons used in the calculation increases. Hence, enormous amounts of computer time may be required in order to achieve reliable accuracy. The Monte Carlo method has long been employed to

simulate the transfer of solar radiation in planetary atmospheres (Plass and Kattawar, 1968). Applications to three-dimensional (3D) cloud problems have also been made by many researchers (e.g., Cahalan *et al.,* 1994).

6.7.2.2   SUCCESSIVE-ORDERS-OF-SCATTERING (SOS) APPROACH

In Section 6.5.1, we pointed out that the SOS approach can be directly applied to specific geometry without the requirement of solving the basic radiative transfer equation in differential form. In addition, the nonhomogeneous structure of a medium can be incorporated in the calculation in a straightforward manner in terms of integration along the line path. To begin this method, we perform the line integration along the spatial coordinates in Eq. (6.7.8) to obtain

$$I(s, \Omega) = I(0, \Omega) \exp[-\tau(s)_\Omega] + \int_0^s \beta_e(s') J(s', \Omega) \exp\{-[\tau(s) - \tau(s')]_\Omega\} \, ds',$$
(6.7.16a)

where $I(0, \Omega)$ is the incident diffuse intensity at position $s = 0$. The effective optical depth $\tau(s)$, or $\tau(s')$, or $\tau(s_0)$ is defined by

$$\tau(s)_\Omega = \int_0^s \beta_e(s') \, ds'.$$
(6.7.16b)

For simplicity of presentation, we shall assume that there is no diffuse downward, upward, or inward intensity at the top, base, and sides of a finite cloud layer so that $I(0, \Omega) = 0$. Moreover, using the index n to denote each order-of-scattering event, we can write

$$I_n(s, \Omega) = \int_0^s \beta_e(s') J_n(s', \Omega) \exp\{-[\tau(s) - \tau(s')]_\Omega\} \, ds', \quad n \geq 1. \quad (6.7.17)$$

Based on the SOS principle, the source function defined in Eq. (6.7.9) can be decomposed into the forms

$$J_1(s, \Omega) = \frac{\tilde{\omega}(s)}{4\pi} P(s; \Omega, -\Omega_0) F_\odot \exp[-\tau(s)_{\Omega_0}],$$
(6.7.18a)

$$J_n(s, \Omega) = \frac{\tilde{\omega}(s)}{4\pi} \int_{4\pi} I_{n-1}(s, \Omega') P(s; \Omega, \Omega') \, d\Omega', \quad n \geq 2. \quad (6.7.18b)$$

Equations (6.7.17) and (6.7.18) are iterative equations in which the intensity and source function for each order of scattering can be computed successively, beginning with $n = 1$. The total diffuse intensity is then

$$I(s, \Omega) = \sum_{n=1}^M I_n(s, \Omega),$$
(6.7.19)

where the index M represents the order-of-scattering events such that $|I_M - I_{M-1}|/ I_{M-1} < \varepsilon$, a prescribed small number depending on the accuracy requirement.

The SOS method has been employed by a number of researchers for applications to radiative transfer concerning specific geometry. Liou *et al.* (1990a) used the method

to study the transmission of thermal infrared radiation for a target–detector system. Weinman (1976) applied it to compute the backscattering return in a collimated lidar system. Herman *et al.* (1994) utilized the general principle of the SOS approach to calculate the radiation field in spherical atmospheres. However, the method has not been widely applied to radiative transfer in plane-parallel atmospheres, primarily because of the substantial computer time requirement to achieve the solution convergence for optically thick media. Liou and Rao (1996) employed the SOS method to investigate the effects of cloud geometry and nonhomogeneity on the reflection and transmission of sunlight with verifications based on the plane-parallel adding method and the Monte Carlo method for 3D cloud fields.

For 3D nonhomogeneous radiative transfer problems, we may define normalized scattered intensities in terms of the conventional reflection function (bidirectional reflectance) and transmission function in the forms

$$R(x, y, 0; \mu, \mu_0, \Delta\phi) = \pi I(x, y, 0; \mu, \mu_0, \Delta\phi)/\mu_0 F_\odot, \qquad (6.7.20a)$$

$$T(x, y, z_c; \mu, \mu_0, \Delta\phi) = \pi I(x, y, z_c; \mu, \mu_0, \Delta\phi)/\mu_0 F_\odot, \qquad (6.7.20b)$$

where z_c is the cloud-base height and the cloud-top height is set at $z = 0$. Normalized side intensities can also be defined in a similar manner. For fluxes, we define the reflectance and transmittance in the forms

$$r(x, y, 0; \mu_0) = F_z^\uparrow(x, y, 0; \mu_0)/\mu_0 F_\odot, \qquad (6.7.21a)$$

$$t(x, y, z_c; \mu_0) = F_z^\downarrow(x, y, z_c; \mu_0)/\mu_0 F_\odot. \qquad (6.7.21b)$$

Figure 6.19 shows the difference patterns for reflection and transmission functions (upper two diagrams) and reflectance and transmittance (lower two diagrams) defined in Eqs. (6.7.20)–(6.7.21) as functions of latitude and longitude for a visible wavelength. The 3D extinction coeffcient field for a finite cirrus cloud field with a horizontal dimension of 20 km × 30 km and a vertical thickness of 3.5 km was constructed from the optical depth retrieved from AVHRR and the ice crystal size distribution determined from point vertical replicator sounding. The single-scattering albedo and phase function were assumed to be the same for the cloud domain. The differences are between results from 3D nonhomogeneous and pixel-by-pixel plane-parallel models. For the reflection and transmission functions, the emergent angles used were 30° and 150°, respectively, and the results are presented in the principal plane ($\Delta\phi = 0$). The relatively large positive differences are associated with the low sun angle ($\theta_0 = 71°$), at which a significant amount of solar flux is available to two cloud sides in addition to the cloud top, as well as with the specific emergent angles used in the calculation. For other incoming and outgoing directions, differences can be either positive or negative, revealing the complexity of the intensity field associated with 3D nonhomogeneous clouds. Because of the larger horizontal dimension (compared to the vertical) used in the calculation, the absolute differences of reflectance and transmittance associated with fluxes are relatively small.
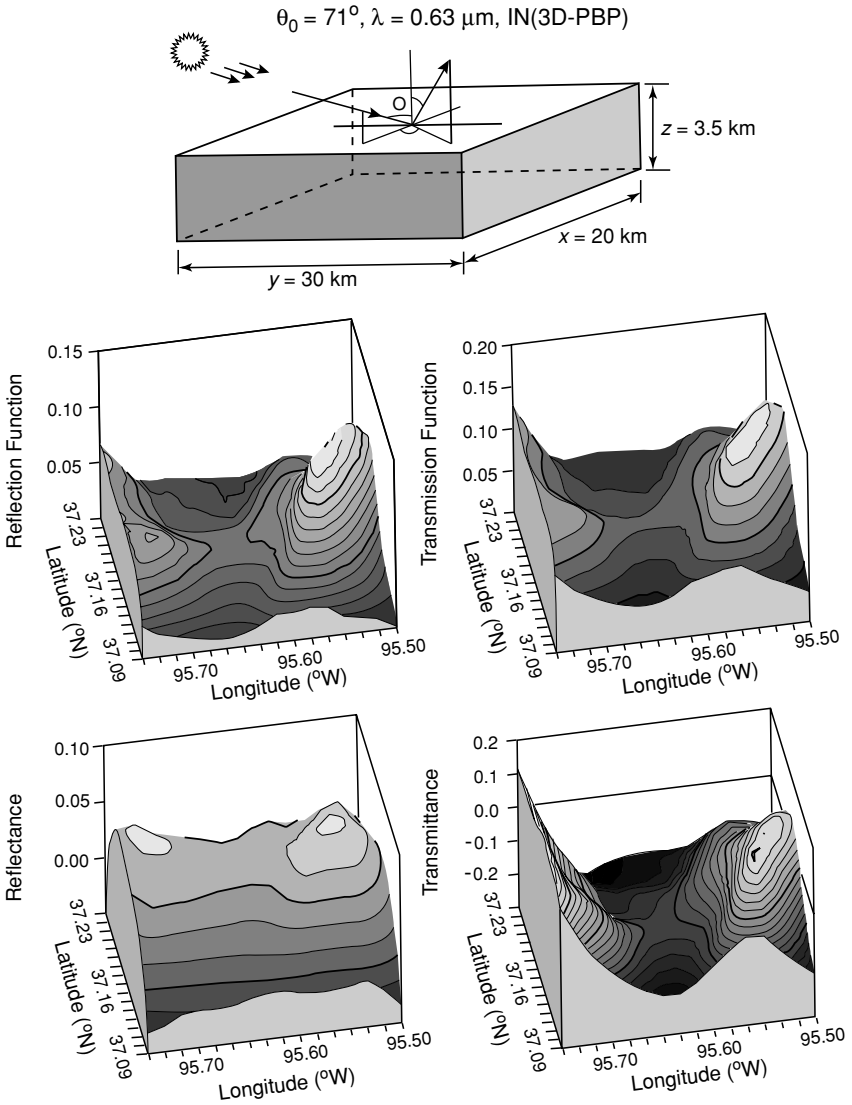
Figure 6.19   Differences of the reflection function, transmission function, reflectance, and transmittance distributions computed from a 3D inhomogeneous (3DIH) model with those computed from a pixel-by-pixel (PBP) plane-parallel model for the 0.63-μm wavelength. The geometrical parameters are given in the figure. The 3D extinction coefficient field was constructed on the basis of the optical depth retrieved from satellite radiometers and the point vertical ice crystal profile determined from a balloon sounding (after Liou and Rao, 1996).

6.7.2.3 DELTA FOUR-TERM (DIFFUSION) APPROXIMATION

In Section 6.5.2, we presented the two-stream approximation that is particularly useful for applications to broadband flux calculations. In the following, we describe a similar approximation for 3D radiative transfer. Referring to Eqs. (6.7.8) and (6.7.9), the phase function and diffuse intensity may be expressed in terms of the spherical harmonics expansion as follows:

$$P(\mathbf{s}, \mathbf{\Omega}, \mathbf{\Omega}') = \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} \tilde{\omega}_\ell Y_\ell^m(\mathbf{\Omega}) Y_\ell^{m*}(\mathbf{\Omega}'), \qquad (6.7.22a)$$

$$I(\mathbf{s}, \mathbf{\Omega}) = \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} I_\ell^m(\mathbf{s}) Y_\ell^m(\mathbf{\Omega}), \qquad (6.7.22b)$$

where $\tilde{\omega}_\ell$ are certain coefficients, and N denotes the number of terms in the spherical harmonics expansion defined by

$$Y_\ell^m(\theta, \phi) = (-1)^{(m+|m|)/2} \left(\frac{(\ell - |m|)!}{(\ell + |m|)!} \right) P_\ell^{|m|}(\cos\theta) e^{im\phi}, \qquad (6.7.23a)$$

where P_ℓ^m is the associated Legendre polynomial defined in Appendix E, $|m|$ is the absolute value of m, and $i = \sqrt{-1}$. The complex conjugates of the spherical harmonics are given by

$$Y_\ell^{m*}(\theta, \phi) = Y_\ell^{-m}(\theta, \phi)/(-1)^m. \qquad (6.7.23b)$$

The spherical harmonics are normalized such that

$$\frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^1 Y_\ell^m(\mu, \phi) Y_\alpha^{\beta*}(\mu, \phi) \, d\mu \, d\phi = \delta_\ell^\alpha \delta_m^\beta / (2\ell + 1), \qquad (6.7.23c)$$

where δ_ℓ^α and δ_m^β are Kronecker delta functions.

To decompose Eq. (6.7.9) in accordance with spherical harmonics, we may insert Eqs. (6.7.22a,b) into Eqs. (6.7.8) and (6.7.9) to obtain

$$\frac{1}{\beta_e(\mathbf{s})}(\mathbf{\Omega} \cdot \nabla) \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} I_\ell^m(\mathbf{s}) Y_\ell^m(\mathbf{\Omega}) = -\sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} \gamma_\ell I_\ell^m(\mathbf{s}) Y_\ell^m(\mathbf{\Omega})$$

$$+ \frac{\tilde{\omega}(\mathbf{s})}{4\pi} \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} \tilde{\omega}_\ell Y_\ell^m(\mathbf{\Omega}) Y_\ell^{m*}(\mathbf{\Omega}_0) F_\odot e^{-\tau_s}$$

$$+ [1 - \tilde{\omega}(\mathbf{s})] B[T(\mathbf{s})], \qquad (6.7.24)$$

where $\gamma_\ell = 1 - \tilde{\omega}\tilde{\omega}_\ell/(2\ell + 1)$. Subsequently, we perform the following successive integrations:

$$\int_{4\pi} \text{Eq. (6.7.24)} \times Y_\alpha^{\beta*}(\mathbf{\Omega}) \, d\mathbf{\Omega}, \qquad \alpha = 0, 1, \ldots, N; \beta = -\alpha, \ldots, \alpha. \qquad (6.7.25)$$

We find that

$$-\frac{1}{\beta_e(\mathbf{s})} \sum_{\ell=0}^{N} \sum_{m=-\ell}^{\ell} \int_{4\pi} (\mathbf{\Omega} \cdot \mathbf{\nabla}) Y_\ell^m(\mathbf{\Omega}) Y_\alpha^{\beta*}(\mathbf{\Omega}) I_\ell^m(\mathbf{s}) \, d\Omega$$

$$= -\gamma_\alpha I_\alpha^\beta(\mathbf{s}) \frac{4\pi}{2\alpha+1} + \frac{\tilde{\omega}(\mathbf{s})\tilde{\omega}_\ell}{2\alpha+1} Y_\alpha^{\beta*}(\mathbf{\Omega}_0) F_\odot e^{-\tau_s} + 4[1 - \tilde{\omega}(\mathbf{s})]\pi B[T(\mathbf{s})]. \quad (6.7.26)$$

The left-hand side of this equation may be decomposed by using the recursion relationships in each coordinate system (Ou and Liou, 1982; Evans, 1993).

For application to the finite homogeneous cloud problem, we may make a first-order approximation (i.e., $N = 1$). Using Cartesian coordinates and the definition of spherical harmonics and their recursion relationships, we obtain the following four partial differential equations:

$$\frac{\partial I_1^0}{\partial z} + \frac{1}{\sqrt{2}} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) I_1^{-1} - \frac{1}{\sqrt{2}} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) I_1^1$$

$$= -\beta_e I_0^0 (1 - \tilde{\omega}) + \frac{\tilde{\omega}}{4\pi} \beta_e F_\odot e^{-\tau_s}, \quad (6.7.27a)$$

$$\beta_e(1 - \tilde{\omega}g) I_1^{-1} = -\frac{1}{3\sqrt{2}} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) I_0^0$$

$$+ \frac{3}{\sqrt{2}} \frac{\tilde{\omega}g}{4\pi} (1 - \mu_0^2)^{1/2} (\cos\phi_0 + i\sin\phi_0) F_\odot e^{-\tau_s}, \quad (6.7.27b)$$

$$\beta_e(1 - \tilde{\omega}g) I_1^0 = -\frac{1}{3} \frac{\partial I_0^0}{\partial z} + \frac{3\tilde{\omega}g}{4\pi} \mu_0 F_\odot e^{-\tau_s}, \quad (6.7.27c)$$

$$\beta_e(1 - \tilde{\omega}g) I_1^1 = \frac{1}{3\sqrt{2}} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) I_0^0$$

$$- \frac{1}{\sqrt{2}} \frac{3\tilde{\omega}g}{4\pi} (1 - \mu_0^2)^{1/2} (\cos\phi_0 - i\sin\phi_0) F_\odot e^{-\tau_s}. \quad (6.7.27d)$$

On substituting Eqs. (6.7.27b–d) into Eq. (6.7.27a), the following 3D nonhomogeneous diffusion equation may be derived:

$$\mathbf{\nabla} \cdot \left(\mathbf{\nabla} I_0^0 / \beta_t \right) - 3\alpha_t I_0^0 = -F_t + \mathbf{\Omega}_0 \cdot \mathbf{\nabla}(F_t g / \beta_t), \quad (6.7.28a)$$

where

$$\beta_t = \beta_e(1 - \tilde{\omega}g), \quad (6.7.28b)$$

$$\alpha_t = \beta_e(1 - \tilde{\omega}), \quad (6.7.28c)$$

$$F_t = \begin{cases} 3\beta_e F_\odot e^{-\tau_s}/4\pi & \text{solar} \\ 3\beta_e(1 - \tilde{\omega})B(T) & \text{IR}. \end{cases} \quad (6.7.28d)$$

In these equations, all the variables are functions of the coordinates (x, y, z). In the case when the single-scattering parameters β_e, $\tilde{\omega}$, and g are independent of (x, y, z), a general diffusion equation in radiative transfer can be obtained (Exercise 6.17) from which an analytic solution for the diffuse intensity can be derived (Liou, 1992).

Equation (6.7.28a) represents a second-order nonhomogeneous partial differential equation, which must be solved numerically. From the spherical harmonics expansion for diffuse intensity denoted in Eq. (6.7.22b), we have

$$I(x, y, z; \Omega) = I_0^0 + I_1^{-1} Y_1^{-1}(\Omega) + I_1^0 Y_1^0(\Omega) + I_1^1 Y_1^1(\Omega). \qquad (6.7.29a)$$

Substituting Eqs. (6.7.27b–d) into Eq. (6.7.29a), we obtain

$$I(x, y, z; \Omega) = I_0^0 - \frac{3}{2h} \sum_{j=1}^{3} \frac{\partial I_0^0}{\partial x_j} \Omega_{xj} + \frac{9q}{2h}(\Omega \cdot \Omega_0) e^{-\tau_s}, \qquad (6.7.29b)$$

where $x_1 = x$, $x_2 = y$, $x_3 = z$, $h = 3(1 - \tilde{\omega}g)/2$, $q = \tilde{\omega}g F_\odot/12\pi$, $\Omega_z = \mu$, $\Omega_x = (1 - \mu^2)^{1/2} \cos\phi$, and $\Omega_y = (1 - \mu^2)^{1/2} \sin\phi$. For thermal infrared, the last term vanishes. To increase computational accuracy, we may apply the similarity principle for radiative transfer introduced in Section 6.5.3 for each computational point such that $\beta'_e = \beta_e(1 - \tilde{\omega}f)$, $\tilde{\omega}' = \tilde{\omega}(1 - f)(1 - \tilde{\omega}f)$, and $g' = (g - f)/(1 - f)$. The fractional energy in the diffraction peak of the phase function f can be taken to be $\tilde{\omega}_2/5$, where $\tilde{\omega}_2$ is the second moment in the phase function expansion.

Contrail cirrus are a typical example of finite clouds. Observations from lidar backscattering and depolarization demonstrate that these clouds are also highly non-homogeneous (see Section 8.3.3 for a discussion of contrail cirrus). At this point, a 3D extinction coefficient field for cirrus derived directly from observations is not available. Figure 6.20 illustrates a hypothetical extinction coefficient field constructed in the x–z plane. For radiative transfer calculations, the extinction coefficients are assumed to be the same in the y-direction. The solar and emergent angles used are $10°$ and $40°$, respectively, with a relative azimuthal angle of $140°$. The reflectance pattern is presented in the x–y plane at the cloud top. Maximum values close to the edge are shown because of the position of the sun and larger extinction coefficients. Similar patterns are displayed along the y direction. By using a mean extinction coefficient in the x–z plane, the reflectance pattern now corresponds to a homogeneous cloud. Except near the left edge, associated with the finite geometry, the pattern is uniform. This example demonstrates the significance and intricacy of the finite and nonhomogeneous cloud structure with respect to its radiative properties (see also Fig. 6.19).

6.7.3   Spherical Atmospheres

When calculating the transfer of solar radiation involving low sun, such as in twilight, or when the limb extinction technique is used to infer ozone, aerosols, and trace gases (Section 7.2.3), the effect of spherical geometry must be accounted for. In reference to Fig. 6.21, the spatial operator in conventional spherical coordinates (the SO system)

Contrail Extinction Coefficient Field (km^{-1})
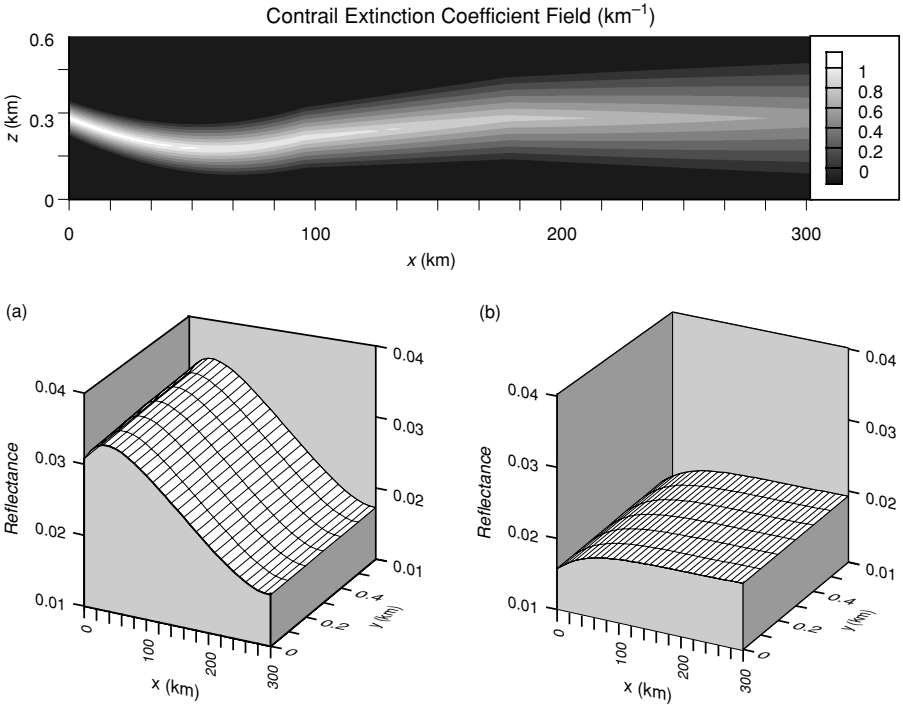
Figure 6.20   The top panel illustrates an extinction coefficient field mimicking a contrail in the x–z plane. The lower panel displays the bidirectional reflectances for (a) inhomogeneous and (b) homogeneous (mean extinction value) contrail fields in the x–y plane. The solar and emergent angles and the azimuthal difference used in the calculations are $10°$, $40°$, and $140°$, respectively.

may be written in the form

$$\mathbf{\Omega} \cdot \mathbf{\nabla} = \Omega_r \frac{\partial}{\partial r} + \Omega_{Z_0} \frac{\partial}{r \partial Z_0} + \Omega_{A_0} \frac{\partial}{r \sin Z_0 \partial A_0}, \qquad (6.7.30)$$

where the directional cosines may be derived by a transformation from those in polar coordinates and are given by

$$\begin{bmatrix} \Omega_r \\ \Omega_{Z_0} \\ \Omega_{A_0} \end{bmatrix} = \begin{bmatrix} \sin Z_0 \cos A_0 & \sin Z_0 \sin A_0 & \cos Z_0 \\ \cos Z_0 \cos A_0 & \cos Z_0 \sin A_0 & -\sin Z_0 \\ -\sin A_0 & \cos A_0 & 0 \end{bmatrix} \begin{bmatrix} \sin Z \cos A \\ \sin Z \sin A \\ \cos Z \end{bmatrix}, \quad (6.7.31)$$

where Z_0 and Z are the zenith angles and $A_0 - A$ denotes the azimuthal difference in the SO system.

However, the system of equations for radiative transfer in plane-parallel atmospheres that has been developed is in reference to the local zenith (see Fig. 6.1). Thus, we must transform the SO system of coordinates to those in the PO system with respect to the local zenith, as defined in Fig. 6.21. In this system, the scattered intensity
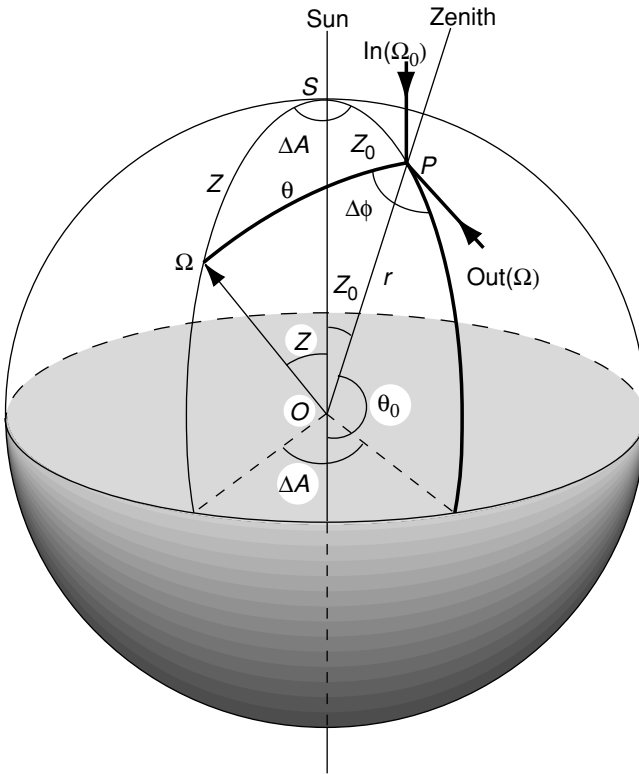
Figure 6.21 Spherical coordinate systems with respect to the sun and the center of the earth (*SO*), and with respect to the zenith and the center of the earth (*PO*), which can be compared with the plane-parallel system. The former involves the zenith angles Z_0 and Z and the azimuthal angle difference $\triangle A$, while the latter involves the zenith angles $\theta_0 (= \pi - Z_0)$ and θ and the azimuthal angle difference $\triangle \phi$. The point P is the location where scattering takes place, O is the center of the sphere, r (*OP*) is the radius, θ_0 is the solar zenith angle associated with the solid angle Ω_0, and θ is the emergent angle associated with the solid angle Ω.

is a function of four variables: θ, θ_0, $\Delta\phi$, and r. The spatial operator can, therefore, be written as follows:

$$\mathbf{\Omega} \cdot \mathbf{\nabla} = \Omega_r \frac{\partial I}{\partial r} + \Omega_\theta \frac{\partial I}{\partial \theta} + \Omega_{\theta_0} \frac{\partial I}{\partial \theta_0} + \Omega_{\Delta\phi} \frac{\partial I}{\partial \Delta\phi}. \tag{6.7.32}$$

Determination of the directional consines in this coordinate system is quite involved. After considering a number of geometric relationships for the angles on the sphere as shown in Fig. 6.21, we find (Exercise 6.18)

$$\Omega_r = \frac{dr}{ds} = \cos\theta = \mu, \tag{6.7.33a}$$

$$\Omega_\theta = \frac{d\theta}{ds} = -\frac{\sin\theta}{r}, \tag{6.7.33b}$$

$$\Omega_{\theta_0} = \frac{d\theta_0}{ds} = -\frac{\sin\theta\cos\Delta\phi}{r}, \tag{6.7.33c}$$

$$\Omega_{\Delta\phi} = \frac{d\Delta\phi}{ds} = \frac{\sin\theta\cos\theta_0\sin\Delta\phi}{r\sin\theta_0}, \tag{6.7.33d}$$

where ds is an arbitrary differential distance in space with respect to the local zenith direction r. It follows that the radiative transfer equation for spherical atmospheres with reference to the local zenith may be written in the form

$$\mu\frac{\partial I}{\partial r} + (1-\mu^2)\frac{\partial I}{r\partial\mu} + (1-\mu^2)^{1/2}(1-\mu_0)^{1/2}\cos\Delta\phi\frac{\partial I}{r\partial\mu_0}$$

$$+ \frac{(1-\mu^2)^{1/2}\mu_0\sin\Delta\phi}{(1-\mu_0^2)^{1/2}}\frac{\partial I}{r\partial\Delta\phi} = -\beta_e(r)[I(r;\mu,\mu_0,\Delta\phi) - J(r;\mu,\mu_0,\Delta\phi)],$$

$$\tag{6.7.34}$$

where the source function is defined in Eq. (6.7.9) and we note that the extinction coefficient β_e and the single-scattering albedo $\tilde{\omega}$ are functions of position only and are not dependent on the incoming and outgoing directions.

We may decompose Eq. (6.7.34) in a manner similar to that developed in Section 6.7.2. For this purpose, we may expand the diffuse intensity in the form

$$I(r;\mu,\phi;\mu_0,\phi_0) = \sum_{m=0}^{N} I_m(r;\mu,\mu_0)\cos m\Delta\phi. \tag{6.7.35a}$$

The expansion of the phase function was introduced in Section 6.1 and is given by

$$P(r;\mu,\phi;\mu',\phi') = \sum_{m=0}^{N}\sum_{\ell=m}^{N}\tilde{\omega}_\ell^m(r)P_\ell^m(\mu)P_\ell^m(\mu')\cos m\Delta\phi. \tag{6.7.35b}$$

Performing the integration $\int_0^{2\pi}$ [Eq. (6.7.34)] $d\Delta\phi$ leads to

$$\mu\frac{\partial I_0}{\partial r} + \frac{(1-\mu^2)}{r}\frac{\partial I_0}{\partial\mu} = -\beta_e(I_0 - J_0)$$

$$- \frac{1}{2}\left[\frac{(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}}{r}\frac{\partial I_1}{\partial\mu_0} - \frac{(1-\mu^2)^{1/2}\mu_0}{(1-\mu_0^2)^{1/2}r}I_1\right], \tag{6.7.36}$$

where the azimuthally averaged source function is given by

$$J_0 = \int_0^{2\pi} J\,d\Delta\phi = \frac{\tilde{\omega}}{2}\int_{-1}^{1} P_0(\mu,\mu')I_0(r;\mu,\mu')\,d\mu'$$

$$+ \frac{\tilde{\omega}}{4\pi}P_0(\mu,-\mu_0)F_\odot\exp[-\tau\mathrm{Ch}(r,\mu_0)] + (1-\tilde{\omega})B(T), \tag{6.7.37}$$

where Ch denotes the Chapman function (see Exercise 3.3), and the single-scattering properties $(\beta_e, \tilde{\omega}, \text{and } P_0)$ as well as the temperature field can be functions of position.

Equation (6.7.36) contains two variables: I_0 and I_1. The solution requires one additional equation for I_1. Consequently, the intensity will be a function of $\Delta\phi$, the azimuthal difference of the incoming and outgoing beams, even in the case of isotropic scattering, i.e., $P(\mu, \phi; \mu', \phi') = 1$.

It appears that Lenoble and Sekera (1961) were the first to present the basic radiative transfer equation in spherical atmospheres with reference to the local zenith. Chandrasekhar (1950) briefly discussed radiative transfer problems with spherical symmetry, stating "And when, further, no radiation from the outside is incident, the intensity and the source function will be functions only of the distance r and the inclination θ to the radius vector." This is the situation when the source function is given by the emission in Eq. (6.7.37) without the direct solar radiation term. In this case the last two terms in Eq. (6.7.36) can be omitted, leading to Eq. (133) in Chandrasekhar's Section 14. Dahlback and Stamnes (1991) also presented correct equations for radiative transfer in spherical atmospheres with applications to the calculation of atmospheric photodissociation and heating rates in middle atmospheres in which the I_1 terms were neglected. The importance of azimuthal dependence in the case of spherical atmospheres, even with isotropic scattering, has been pointed out in Herman *et al.* (1995).

To complete the solution for Eq. (6.7.36), we may carry out an approach similar to that presented in Section 6.7.2 by truncating the intensity expansion to the I_1 term. By performing the integration $\int_0^{2\pi}$ [Eq. (6.7.34)] $\cos\Delta\phi\,d\Delta\phi$ and by setting $I_2 = 0$, we obtain

$$\mu\frac{\partial I_1}{\partial r} + \frac{(1-\mu^2)}{r}\frac{\partial I_1}{\partial\mu} = -\beta_e(I_1 - J_1) - \frac{1}{2}\frac{(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}}{r}\frac{\partial I_0}{\partial\mu_0}, \quad (6.7.38)$$

where

$$J_1 = \int_0^{2\pi} J\cos\Delta\phi\,d\Delta\phi = \frac{\tilde\omega}{2}\int_{-1}^1 P_1(\mu, \mu')I_1(r; \mu, \mu')\,d\mu'$$

$$+ \frac{\tilde\omega}{4\pi}P_1(\mu, -\mu_0)F_\odot\exp[-\tau\mathrm{Ch}(r, \mu_0)], \quad (6.7.39)$$

where P_1 is the first Fourier component of the phase function ($m = 1$). In principle, Eqs. (6.7.36) and (6.7.38) can be used to solve for I_0 and I_1 simultaneously. The theoretical subject of radiative transfer in spherical atmospheres appears not to have been fully explored at this point.

Exercises

6.1 A satellite radiometer measures the solar radiation reflected from a semi-infinite, isotropic-scattering atmosphere composed of particulates and gases near the vicinity of an absorption line whose line shape is given by the Lorentz profile and whose absorption coefficient can be written as

$$k_\nu = \frac{S}{\pi}\frac{\alpha}{(\nu - \nu_0)^2 + \alpha^2}.$$

Assuming that the particulates are nonabsorbing and that the scattering optical depth is equal to the gaseous absorption optical depth at the line center, calculate the reflected intensity as a function of wavenumber ν using the two-stream approximation. Do the problem by formulating (a) the single-scattering albedo as a function of ν; and (b) the reflected intensity in terms of the two-stream approximation.

6.2   For a semi-infinite, isotropic-scattering atmosphere, show that the planetary albedo

$$r(\mu_0) = 1 - H(\mu_0)\sqrt{1 - \tilde{\omega}}$$

and the spherical albedo

$$\bar{r} = 1 - 2\sqrt{1 - \tilde{\omega}} \int_0^1 H(\mu_0)\mu_0 \, d\mu_0.$$

Using the first approximation for the H function and assuming single-scattering albedos of 0.4 and 0.8, compute the planetary albedo for μ_0 of 1 and 0.5 and the spherical albedo.

6.3   An optically thin layer $\Delta\tau$ is added to a finite atmosphere with an optical depth of τ_1, and all the possible transmissions of the incident beam due to the addition of the thin layer are displayed in Fig. 6.22. Formulate Eq. (6.3.44) using the principles of invariance discussed in Section 6.3.2. The method is also referred to as *invariant imbedding*. In this diagram, the dotted lines represent direct transmission.
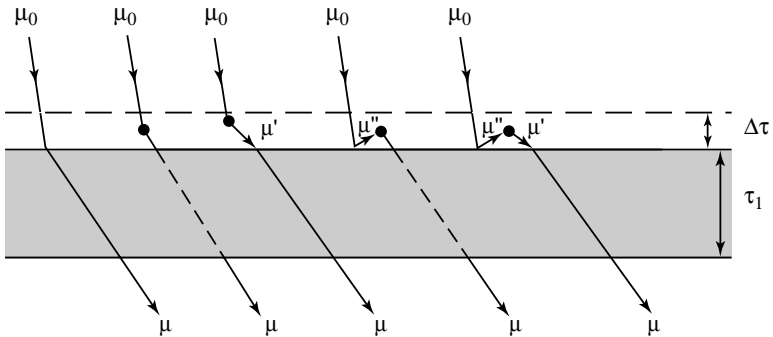
Figure 6.22

6.4   Consider a cloud layer having a total global transmission of $\bar{\gamma}$ and a global reflection (spherical albedo) of $\bar{r}$ above a Lambertian surface with an albedo of r_s. Assuming no atmosphere between the cloud and surface, derive Eqs. (6.3.69a) and (6.3.69b) by means of geometric ray-tracing for multiple reflections between the cloud and the surface.

6.5 Based on the geometric configuration shown in Fig. 6.9, derive the adding equations [Eqs. (6.4.17a)–(6.4.17f)] for a light beam incident from below the layer defined by μ'.

6.6 Neglecting the ground reflection effect, use the single-scattering approximation to compute and plot the reflected intensity (reflection) at the top of nonabsorbing molecular atmospheres whose optical depths are assumed to be 0.1 and 1 for $\mu_0 = 0.8$.

6.7 Derive an analytical expression for the diffuse reflection at the top of the atmosphere using the second-order scattering approximation (neglect the surface reflection). Carry out the analysis for $\mu \neq \mu'$ and $\mu = \mu'$.

6.8 Derive the two-stream solution for conservative scattering and calculate the reflection and transmission, assuming an asymmetry factor of 0.75 for optical depths of 0.25, 1, 4, and 16. Plot the results as functions of the cosine of the solar zenith angle μ_0.

6.9 Derive the one-dimensional diffuse equation for radiative transfer denoted in Eq. (6.5.23) from Eqs. (6.5.22a) and (6.5.22b), and solve for I_1 using the radiation boundary condition defined in Eq. (6.2.23).

6.10 By employing the orthogonality property of the Legendre polynomials and the recurrence formula

$$\mu P_\ell(\mu) = \frac{\ell + 1}{2\ell + 1} P_{\ell+1} + \frac{\ell}{2\ell + 1} P_{\ell-1},$$

show that Eq. (6.1.6) without the emission term can be reduced to the set of first-order differential equations given in Eq. (6.5.20), based on the intensity expansion given in Eq. (6.5.19). For simple cases of $\ell = 0,1$, Eq. (6.5.20) leads to Eqs. (6.5.22a,b).

6.11 Formulate the transfer of thermal infrared radiation in a scattering atmosphere having an isothermal temperature T in local thermodynamic equilibrium, assuming the intensity is azimuthally independent. By means of the discrete-ordinates method for radiative transfer and assuming isotropic scattering, show that the scattered intensity is given by

$$I(\tau, \mu_i) = \sum_{\alpha=-n}^{n} \frac{L_\alpha}{1 + \mu_i k_\alpha} e^{-k_\alpha \tau} + B_\nu(T),$$

where L_α are unknown constants of proportionality, μ_i denote the discrete streams, k_α are the eigenvalues, and B_ν represents the Planck function.

6.12 From Eqs. (6.6.1a) and (6.6.1b), show that an electromagnetic wave is defined by the equation of an ellipse. The wave is said to be *elliptically polarized*. If the phase difference between the two electric vectors $\delta = \delta_r - \delta_l$ is of the order of π, show that the vibration of the wave is governed by two lines, referred to as *linearly polarized*. If, on the other hand, it is of the order of $\pi/2$ and if the

amplitudes $a_l = a_r$, show that the vibration of the wave is governed by a circle where the waves are referred to as *circularly polarized*.

6.13 From Eqs. (6.6.3a)–(6.6.3e) derive the Stokes parameters in terms of the ellipticity and orientation angles given by Eqs. (6.6.4a)–(6.6.4d).

6.14 Any time-average quantity may be represented by the summation of individual components, e.g., $\langle x \rangle = \sum_{n=1}^{N} t_n x_n$. Using this principle, show that $I^2 \geq Q^2 + U^2 + V^2$ based on the relationships given in Eq. (6.6.5). In doing this exercise, let $N = 2$ for simplicity.

6.15 (a) Show that elliptically polarized light can be decomposed into a circularly polarized part and a linearly polarized part. Then rotate the linearly polarized beam through the angle χ and show that χ, which makes the intensity maximum (or minimum) in the direction y (see Fig. 6.13), is given by $\tan 2\chi = U/Q$.

   (b) Assuming a light beam with 50% linear polarization in the r direction and another independent light beam also with 50% right-handed circular polarization: (1) What would be the Stokes parameters for the mixture and the resulting total intensity and percentage polarization? (2) What would be the measured intensity if a polarizer having a plane of polarization along the r direction is used? And (3) sketch a diagram to denote the resultant polarization.

   (c) With reference to (b), decompose the partially polarized light beam into natural light and 100% elliptically polarized light and compute the plane of polarization χ and ellipticity angle β for the polarized component.

   (d) Natural light is equivalent to any two independent oppositely polarized beams of half the intensity. By virtue of this principle, evaluate the Stokes parameters for these two polarized beams based on the results obtained from (c).

   (e) Upon combining the polarized beams derived from (c) and (d), what would be the Stokes parameters corresponding to two independent polarized beams?

6.16 Derive Eq. (6.6.11) from Eq. (6.6.3c) and prove that $\mathbf{L}(\chi_1)\mathbf{L}(\chi_2) = \mathbf{L}(\chi_1 + \chi_2)$ and $\mathbf{L}^{-1}(\chi) = \mathbf{L}(-\chi)$.

6.17 Assuming that the single-scattering parameters β_e, $\tilde{\omega}$, and g are independent of the coordinate system in Eqs. (6.7.27a)–(6.7.27d), derive the general diffusion equation for radiative transfer. Compare the final result with Eq. (6.5.23).

6.18 On the basis of the geometric configuration shown in Fig. 6.21 and the trigonometric relationships among various angles depicted in this figure, derive Eqs. (6.7.33a)–(6.7.33d).

6.19 Prove that

$$\frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} P(\mu, \phi; \mu', \phi')\mu \, d\mu d\phi = \mu'g,$$

where g is the asymmetry factor defined in Eq. (6.5.9a). In carrying out this exercise, expand the phase function in the form presented in Eq. (6.1.3a) and use the orthogonal property of the Legendre polynomial.

Suggested Reading

Ambartzumian, V. A. (1958). *Theoretical Astrophysics*. Pergamon Press, New York. Chapters 33 and 34 contain fundamental discussions of the principles of invariance for a semi-infinite, isotropic atmosphere.

Chandrasekhar, S. (1950). *Radiative Transfer*. Dover, New York. Chapters 1, 3, 6, and 7 contain basic materials for the discrete-ordinates method for radiative transfer and the principles of invariance.

Kourganoff, V. (1952). *Basic Methods in Transfer Problems*. Oxford University Press, London. Chapter 3 contains some elementary discussions of the discrete-ordinates and spherical harmonics methods for radiative transfer.

Lenoble, J. (1993). *Atmospheric Radiative Transfer*. A. Deepak Publishing, Hampton, Virginia. Chapters 13 and 14 consist of useful discussions of radiative transfer methodologies and polarization.

Liou, K. N. (1992). *Radiative and Cloud Processes in the Atmosphere. Theory, Observation, and Modeling*. Oxford University Press, New York. Chapter 3 presents various radiative transfer methods and approximations with a specific emphasis on flux calculations.

Sobolev, V. V. (1975). *Light Scattering in Planetary Atmospheres*. Pergamon Press, New York. Chapters 2 and 3 contain discussions of the principles of invariance.

Thomas, G. E., and Stamnes, K. (1999). *Radiative Transfer in the Atmosphere and Ocean*. Cambridge University Press, Cambridge, U.K. Chapters 7 and 8 discuss useful approximations and numerical solutions for radiative transfer.

van de Hulst, H. C. (1980). *Multiple Light Scattering. Tables, Formulas, and Applications*, Vol. 1 and 2. Academic Press, New York. Chapter 4 in Vol. 1 contains a general review of various methods for radiative transfer, including the adding/doubling method. Chapter 13 in Vol. 2 provides some useful numerical values that can be used to check various radiative transfer calculations.

Chapter 7 | Application of Radiative Transfer Principles to Remote Sensing

7.1 Introduction

Remote sensing is differentiated from *in situ* measurements whereby specific observations are made within the medium. Remote sensing involves the interpretation and inversion of radiometric measurements of electromagnetic radiation measured some distance away, where the radiation is characterized by a specific wavelength that is sensitive to some physical aspect of the medium. The interpretation and inversion require the use of fundamental light scattering and radiative transfer theories. The physical principle of remote sensing can be understood from the simple configuration illustrated in Fig. 7.1. Basically, an electromagnetic signal is recorded by a detector after it interacts with a target containing molecules, particulates, and/or surfaces. If T and S denote the target and signal, respectively, then we may write symbolically

$$S = F(T),$$

where F represents a function, not necessarily linear. This function governs the processes involving the transfer of electromagnetic radiation. The inverse of the preceding relation gives

$$T = F^{-1}(S),$$

where F^{-1} represents a function of F. Generally, if a functional relationship can be established, one may carry out the inversion directly. However, because of the complexity of F, its inverse cannot be performed in a straightforward manner. In those cases where a specific functional form does not exist, one may search the desired parameters of the target that would best fit the measured signals.

The fundamental obstacle in all remote sensing inversion problems is the *uniqueness* of the solution. The nonuniqueness arises because the medium under investigation may be composed of a number of unknown parameters, a combination of which may lead to the same radiation signature. In addition to this physical problem, there are also mathematical problems associated with the existence and stability of the solution, and the manner in which the solution is constructed.

There are two basic types of remote sensing: *active* and *passive*. Active remote sensing employs a radiation source generated by artificial means such as the lasers
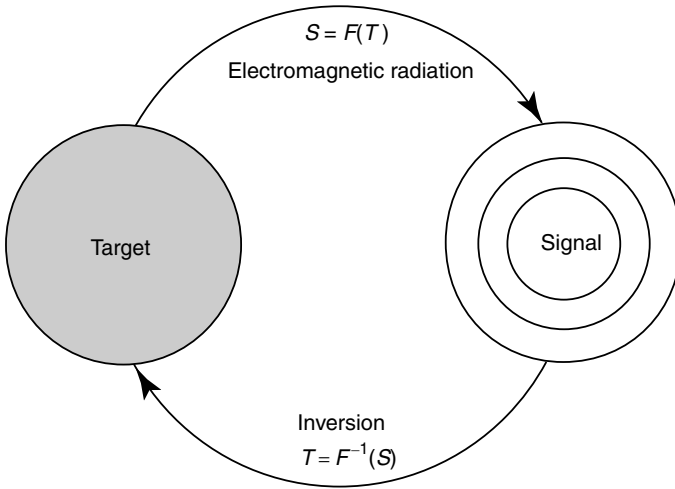
Figure 7.1 Principle of remote sensing. The symbols T and S denote target and signal, respectively, F represents a function which is generally nonlinear and cannot be represented by analytic equations, and F^{-1} is the inverse of this complicated function.

used in lidar or the microwaves used in radar. The radiant energy source corresponding to a specific wavelength is sent to the atmosphere. Some of the energy is scattered back to the detector and recorded. From the recorded scattered energy, one analyzes the composition and structure of the atmosphere with which the radiant energy has interacted. Active remote sensing is normally concerned with backscattering; i.e., the transmitter and detector are collocated. In recent years, significant developments in lidar systems based on differential absorption and depolarization techniques and in millimeter (mm)-wave radar have been made in conjunction with the detection of aerosols, clouds, water vapor, and trace gases. Deployment of lidar and mm-wave radar on space platforms is being continuously explored in ongoing research and development programs.

Passive remote sensing utilizes the natural radiation sources of the sun or the earth–atmosphere system. For example, spectral solar radiation interacts with a cloud and leaves a scattered signature, which may be used for its identification. Similarly, spectral thermal infrared or microwave radiation emitted from the earth–atmosphere system may be utilized to understand its thermodynamic state and composition. In reference to Fig. 1.1 regarding the electromagnetic spectrum, all wavelengths are possible from the emitting medium. However, for atmospheric applications, the ultraviolet (UV), visible, infrared (IR), and microwave spectra are most important.

Passive remote sensing allows the global inference of atmospheric and surface temperatures, composition profiles, surface properties, and radiative budget components from orbiting meteorological satellites. The first meteorological satellite experiment was an array of hemispheric sensors flown on the Explorer 7 satellite launched in 1959 to measure the radiation balance of the earth–atmosphere system. Shortly after,

a five-channel scanning radiometer was carried on board TIROS 2. The scanning radiometer has provided the general characteristics of sounding instrumentation for research and operational satellites over the past four decades. In April 1969, two spectrometers providing spatial infrared measurements for the determination of the vertical profiles of temperature, water vapor, and ozone were flown on the Nimbus 3 satellite. On board was another instrument for measuring reflected UV radiation, which allowed the determination of the global ozone concentration. Also on board was an infrared radiation interferometer spectrometer (IRIS) for measurement of the emitted infrared spectrum. The concept of using the detailed emission spectrum for remote sensing and climate studies has been revisited only recently. The launch of Nimbus 5 in December 1972 marked the first use of microwave techniques for the remote sensing of atmospheric temperature and total water content. These are the key milestones of atmospheric sounding from orbiting meteorological satellites.

In this chapter, we first discuss the information content of transmitted and reflected sunlight. Following a description of the principle of temperature and gaseous profile determination from thermal infrared emission, the use of microwave emissions for atmospheric studies is introduced. These sections emphasize the application of radiative transfer principles to the remote sensing of atmospheric temperature, aerosols, ozone, water vapor, clouds, and precipitation from satellites. The principle of active remote sensing utilizing backscattered energy is subsequently presented with an emphasis on the propagation of laser energy in the atmosphere. Determination of the radiation budget of the earth–atmosphere system from satellites for use in radiation and climate studies will be discussed in Chapter 8.

7.2   Remote Sensing Using Transmitted Sunlight

The remote sensing of aerosols and ozone from ground-based radiometers utilizes the direct beam of solar radiation transmitted through the cloudless atmosphere. Specific wavelengths in the UV and visible spectra are normally selected to optimize the effects of aerosols and ozone for retrieval purposes. Considering the geometry illustrated in Fig. 7.2 and following the Beer–Bouguer–Lambert law presented in Section 1.4.2, the solar intensity of a given wavelength measured at the ground at a given time of the year can be written in the form

$$\hat{I}(\lambda) = \left(\frac{r_0}{r}\right)^2 I_\odot(\lambda) \exp[-\tau(\lambda)m(\theta_0)], \qquad (7.2.1)$$

where r and r_0 are the actual and mean sun-earth distances, respectively (Section 2.2.1); $I_\odot$ denotes the known solar intensity at the top of the atmosphere corresponding to r_0; the air mass factor $m(\theta_0) = 1/\cos\theta_0$, where θ_0 is the solar zenith angle involving the latitude, the solar declination angle, and the hour angle defined in Eq. (2.2.1); and the total optical depth is the sum of the individual values contributed from aerosols (A), Rayleigh molecules (R), ozone (3) and nitrogen dioxide (2), respectively, as follows:

$$\tau(\lambda) = \tau_A(\lambda) + \tau_R(\lambda) + \tau_3(\lambda) + \tau_2(\lambda). \qquad (7.2.2)$$
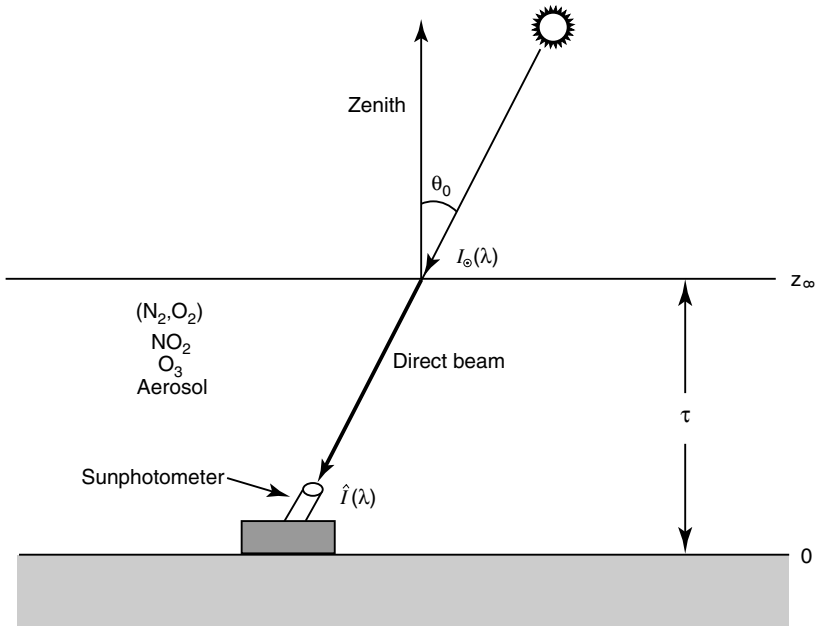
Figure 7.2   Geometry of a ground-based radiometer for the measurement of a direct solar beam. The solar zenith angle is defined by θ_0, the optical depth is denoted by τ, and the solar intensity and measured intensity are given by $I_\odot$ and $\hat{I}$, respectively. A sunphotometer can have a number of different filters to record direct solar radiation in component wavelengths.

The optical depths of Rayleigh molecules and NO_2 for a given wavelength (Section 3.3.1) are generally assumed to be known quantities in the determination of aerosol and ozone properties.

7.2.1   Determination of Aerosol Optical Depth and Size Distribution

In Section 5.1, we reviewed the origin of atmospheric aerosols. Aerosols not only scatter, but can also absorb incoming sunlight. Various kinds of aerosols can have a significant effect on the radiative balance of the earth–atmosphere system and hence on the earth's climate because of their global presence. Increasing concerns that aerosols in general and man-made pollution in particular have caused worldwide weather disturbances have prompted intensified observation of the concentrations and physical and chemical characteristics of aerosols from the ground and space. Determination of the optical properties of aerosols and their size distribution around the globe has been a significant contemporary research effort of late.

Observational methods to determine the dust loading of the atmosphere were developed during the 1920s by Linke and Ångström. In essence, aerosol total optical depth, sometimes also referred to as *turbidity,* is derived from direct spectral solar intensity measured on the ground. Wavelengths in the visible spectrum are normally

employed because absorption due to water vapor can be neglected and the effect of ozone absorption is small.

From Eqs. (7.2.1) and (7.2.2), aerosol optical depth is given by

$$\tau_A(\lambda) = \frac{1}{m(\theta_0)}[\ln I^*(\lambda) - \ln \hat{I}(\lambda)] - [\tau_R(\lambda) + \tau_3(\lambda) + \tau_2(\lambda)], \quad (7.2.3a)$$

where we define $I^* = I_\odot(r_0/r)^2$. This term can be evaluated from the known solar spectral data or can be directly evaluated from the Langley plot (Section 2.3.2). In the determination of aerosol optical depth, the optical depths associated with NO_2 and O_3 molecules are usually computed from the following parameterizations:

$$\tau_R(\lambda) = (a + bH)\lambda^{-(c+d\lambda+e/\lambda)} p/p_s,$$
$$\tau_2(\lambda) = k_2(\lambda) C(NO_2), \quad (7.2.3b)$$
$$\tau_3(\lambda) = k_3(\lambda) C(O_3),$$

where the empirical coefficients for Rayleigh scattering are $a = 0.00864$, $b = 6.5 \times 10^{-6}$, $c = 3.916$, $d = 0.074$, and $e = 0.050$; H is the radiometer height in kilometers; p is atmospheric pressure at the radiometer; p_s is 1013.25 mb; k_2 and k_3 are absorption coefficients for NO_2 and O_3, respectively; and C denotes the concentration notation and $C(NO_2) \cong 4 \times 10^{15}$ cm^{-2}. Table 7.1 lists the pertinent values of six typical wavelengths used for aerosol measurements (Russell *et al.*, 1993). Ozone concentration is determined using other methods (see Section 7.2.2). Further, some diffuse light that enters the radiometer can be taken into account by empirical adjustments. Thus, once these optical depths have been evaluated and the position of the sun is known, a measurement of the direct solar intensity can be used to infer the aerosol optical depth. In the following, we wish to show that in addition to the optical depth, the aerosol size distribution can also be retrieved. We shall first demonstrate the principle of using two wavelengths for this purpose.

The aerosol optical depth corresponding to the entire atmospheric column can be expressed in terms of the extinction coefficient in the form

$$\tau_A(\lambda) = \int_0^{z_\infty} \beta_e(\lambda, z) \, dz. \quad (7.2.4)$$

Let the height-dependent aerosol size distribution be defined by $n(z, a)(cm^{-3}\mu m^{-1})$.

Table 7.1

Useful Scattering or Absorption Coefficients for a Number of Solar Wavelengths for Molecules, O_3, and NO_2

$\lambda(\mu m)$	0.382	0.451	0.526	0.778	0.861	1.060
k_R	0.4407	0.2198	0.1175	0.0240	0.0159	0.00069
k_3 (atm − cm^{-1})	—	0.004	0.061	0.009	—	—
k_2 (cm^2)	5.39×10^{-19}	4.66×10^{-19}	1.74×10^{-19}	—	—	—

Then the extinction coefficient (cm^{-1}) is given in the form [Eq. (5.2.115)]

$$\beta_e(\lambda, z) = \int_{a_1}^{a_2} \sigma_e(a, \lambda) n(z, a) \, da, \qquad (7.2.5)$$

where σ_e represents the extinction cross section (cm^2) for an individual particle. The size distribution of aerosols in the atmosphere has been a subject of extensive research in the past four decades, as pointed out in Section 5.1. For the purpose of discussing retrieval, the aerosol size distribution can best be defined by the Junge distribution displayed in Fig. 5.1 in the form

$$n(z, a) = C(z, a) a^{-(\nu^*+1)}, \qquad (7.2.6)$$

where C is a scaling factor directly proportional to the aerosol concentration and is, therefore, a function of height z in the atmosphere, and ν^* represents a shaping constant which is normally found to lie in the range $2 \leq \nu^* < 4$. As shown in Fig. 5.1, the size distribution of aerosols typically covers the range from about 0.01 to about 10 μm.

Using the Junge size distribution, the aerosol optical depth is given by

$$\tau_A(\lambda) = k\lambda^{(-\nu^*+2)}, \qquad (7.2.7)$$

where k is a certain constant (see Exercise 7.1). When $\nu^* = 3.3$, k is known as the *Ångström turbidity coefficient*. If the aerosol optical depth is measured at two wavelengths, then we have

$$\hat{z} = \frac{\tau_A(\lambda_1)}{\tau_A(\lambda_2)} = \left(\frac{\lambda_1}{\lambda_2}\right)^{(-\nu^*+2)} = y^{(-\nu^*+2)}. \qquad (7.2.8a)$$

Thus, the shaping factor can be inferred from

$$\nu^* = 2 - \ln \hat{z} / \ln y. \qquad (7.2.8b)$$

We may also retrieve the aerosol size distribution directly if a number of aerosol optical depths are measured with a multispectral instrument known as a *sunphotometer* (Fig. 7.2). A sunphotometer tracks the sun and measures the intensity of a solar beam in several spectral channels. For the retrieval of aerosol size distribution, we shall use Eqs. (7.2.4)–(7.2.6) and define the column aerosol size distribution in the form

$$n_c(a) = \int_0^{z_\infty} n(a, z) \, dz = f(a) h(a), \qquad (7.2.9.a)$$

where we have separated $n_c(a)$ into slowly and rapidly varying functions represented by $f(a)$ and $h(a) = a^{-(\nu^*+1)}$, respectively (King *et al.,* 1978). Thus, we have

$$\tau_A(\lambda) = \int_{a_1}^{a_2} f(a)[h(a)\pi a^2 Q_e(m, a/\lambda)] \, da. \qquad (7.2.9.b)$$

In Eq. (7.2.9b), we have expressed, based on the Lorenz–Mie theory, the extinction cross section in terms of the extinction efficiency Q_e, which is a function of the

aerosol radius, wavelength, and complex refractive index. In the retrieval, the refractive index for aerosols must first be assigned. The objective of the retrieval is now to determine $f(a)$.

To simplify the introduction of the inversion method we let $g = \tau_A(\lambda)$ and $K_\lambda(a) = \pi a^2 Q_e(m, a/\lambda)h(a)$. Thus, we have

$$g_\lambda = \int_{a_1}^{a_2} f(a)K_\lambda(a)\,da. \tag{7.2.10}$$

This is the well-known Fredholm equation of the first kind in which $K_\lambda(a)$, the weighting function, is the kernel, and $f(a)$ is the function to be recovered from a set of g_λ. In practice, because only finite values of g_λ are available, the solution for $f(a)$ is mathematically ill-conditioned, even if the weighting function and the measured data are without errors.

Weighting functions based on Lorenz–Mie calculations involving seven sunphotometer wavelengths for a typical Junge aerosol size distribution are illustrated in Fig. 7.3. The peaks of the weighting functions shift from larger to smaller radius
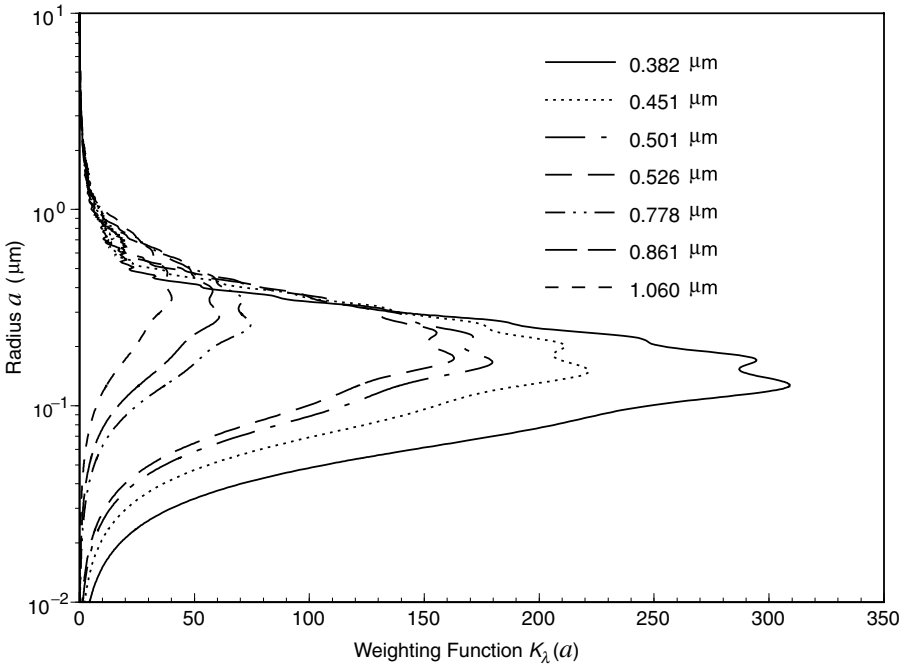
Figure 7.3   Weighting function as a function of aerosol radius for seven sunphotometer wavelengths. The calculations use the Junge size distribution depicted in Fig. 5.1 with $\nu^* = 3$ and an index of refraction of 1.45. The second peak above the maximum for each weighting function is associated with the extinction efficiency curve computed from the Lorenz–Mie spherical particles (see Fig. 5.7 for $m_i = 0$).

corresponding to the increase in wavelength. The weighting functions cover a range of aerosol radii from about 0.01 to 1 μm. In the following, we introduce the direct linear inversion method, which has been frequently used for the retrieval of aerosol size distribution.

7.2.1.1   DIRECT LINEAR INVERSION

Consider the following Fredholm equation of the first kind:

$$g_i = \int_{a_1}^{a_2} f(a)K_i(a)\,da, \qquad i = 1, 2, \ldots, M, \tag{7.2.11}$$

where the index i (replacing λ) denotes a discrete set of M observations. Thus, a set of M integral equations is to be used to recover the unknown function $f(a)$ given the known kernels $K_i(a)$. It is clear that the solution is an ill-posed problem, because the unknown profile is a continuous function of particle radius and because only a finite number of observations are available.

It is convenient to express $f(a)$ in a series such that

$$f(a) = \sum_{j=1}^{N} f_j W_j(a), \tag{7.2.12}$$

where f_j are unknown coefficients and W_j are the known representation functions, which can be orthogonal functions, such as polynomials or Fourier series. Substituting Eq. (7.2.12) into Eq. (7.2.11) we obtain

$$g_i = \sum_{j=1}^{N} A_{ij} f_j, \qquad i = 1, 2, \ldots, M, \tag{7.2.13a}$$

where

$$A_{ij} = \int_{a_1}^{a_2} W_j(a)K_i(a)\,da. \tag{7.2.13b}$$

To find $f_j(j = 1, \ldots, N)$, we require the observations $g_i(i = 1, \ldots, M)$ and the general condition that $M \geq N$.

For convenience of analysis, we can use vectors and matrices to represent the fundamental parameters. The observations and unknown coefficients are denoted by the following column vectors:

$$\mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_M \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{bmatrix}. \tag{7.2.14a}$$

A matrix consisting of M rows and N columns is defined by

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{bmatrix}. \tag{7.2.14b}$$

$\mathbf{A}$ is said to be an $(M \times N)$ matrix and can also be written as $\|A_{ij}\|$. Thus, Eq. (7.2.13a) can be expressed in matrix form:

$$\mathbf{g} = \mathbf{A}\mathbf{f}. \tag{7.2.15}$$

The inverse of a matrix is denoted by $\mathbf{A}^{-1}$. In general, the inverse exists only when $\mathbf{A}$ is a square $(N \times N)$ and the determinant of the array set $(\mathbf{A})$ is not zero (nonsingular). Inversion of a matrix usually requires a computer. If $M = N$, then we can obtain $\mathbf{f}$ from direct inversion as follows:

$$\mathbf{f} = \mathbf{A}^{-1}\mathbf{g}. \tag{7.2.16}$$

Exercise 7.2 requires the direct inversion of a known mathematical function.

If $M > N$, i.e., the number of observations is more than the terms for the representation function, Eq. (7.2.15) cannot be inverted directly because there are more equations than unknowns, leading to the problem that the system is overdetermined. In practical terms, the matrix $\mathbf{A}$ cannot be inverted. The conventional approach to such a problem is to use the method of least squares. The difference between the left- and right-hand sides of Eq. (7.2.13a) can be written in the form

$$\varepsilon_i = g_i - \sum_{j=1}^{N} A_{ij} f_j, \qquad i = 1, 2, \ldots, M. \tag{7.2.17}$$

The least squares solution minimizes the quantity

$$\sum_{i=1}^{M} \varepsilon_i{}^2 = \sum_{i=1}^{M} \left(\sum_{j=1}^{N} A_{ij} f_j - g_i \right)^2 \tag{7.2.18a}$$

by setting the partial derivatives of this term with respect to $f_k (k = 1, 2, \ldots, N)$ equal to zero so that

$$\frac{\partial}{\partial f_k} \left[\sum_i \left(\sum_{j=1}^{N} A_{ij} f_j - g_i \right)^2 \right] = 0. \tag{7.2.18b}$$

Performing the operation leads to

$$\sum_i \left(\sum_{j=1}^{N} A_{ij} f_j - g_i \right) A_{ik} = 0. \tag{7.2.18c}$$

In matrix form, we have

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{f} = \mathbf{A}^{\mathrm{T}}\mathbf{g}, \tag{7.2.19}$$

where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$, i.e., an $(N \times M)$ matrix. It follows that

$$\mathbf{f} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{g}. \tag{7.2.20}$$

This is the least squares solution which requires the inverse of a symmetric and square matrix. Many studies have pointed out that the solution derived from Eq. (7.2.20) is unstable because it is under constraint.

7.2.1.2 CONSTRAINED LINEAR INVERSION

In addition to the instability due to the governing mathematical function itself, instability can also be traced to the following sources of error: (a) errors produced by the numerical quadrature used for the calculation of A_{ij} in Eq. (7.2.13b); and (b) numerical round-off errors. Further, radiometers have inherent instrument noise and thus the observed intensities probably generate errors in a random fashion. All of the preceding errors make direct inversion from the integral transfer equation impractical. In practice, the true g_i are never known and the measured data can be expressed in terms of its deviation from the true value as follows:

$$\hat{g}_i = g_i + \varepsilon_i. \tag{7.2.21}$$

From Eq. (7.2.17), it is clear that to within the measurement error, the solution f_j is not unique. The ambiguity in the solution can be removed by imposing an additional condition that would allow one of the possible sets of f_i to be selected.

Consider now a function that utilizes a least squares method with quadratic constraints in the form

$$\sum_i \varepsilon_i^2 + \gamma \sum_{j=1}^{N}(f_j - \bar{f})^2, \tag{7.2.22}$$

where γ is an arbitrary smoothing coefficient that determines how strongly the solution f_j is constrained to be near the mean $\bar{f}$, i.e., the constraint is given by the variance of f_j. We may select a solution such that the measurement error is minimized while the solution is constrained to be close to the mean $\bar{f}$ such that

$$\frac{\partial}{\partial f_k}\left[\sum_i \left(\sum_{j=1}^{N} A_{ij}f_j - \hat{g}_i \right)^2 + \gamma \sum_{j=1}^{N}(f_j - \bar{f})^2 \right] = 0. \tag{7.2.23}$$

This leads to

$$\sum_i \left(\sum_{j=1}^{N} A_{ij}f_j - \hat{g}_i \right) A_{ik} + \gamma(f_k - \bar{f}) = 0. \tag{7.2.24}$$

In matrix form, we have

$$\mathbf{A}^T\mathbf{A}\mathbf{f} - \mathbf{A}^T\hat{\mathbf{g}} + \gamma\mathbf{H}\mathbf{f} = 0, \tag{7.2.25}$$

where $\mathbf{H}$ is an $(N \times N)$ matrix given by

$$\mathbf{H} = \begin{bmatrix} 1 - N^{-1} & -N^{-1} & \cdots & -N^{-1} \\ -N^{-1} & 1 - N^{-1} & \cdots & -N^{-1} \\ \vdots & \vdots & & \vdots \\ -N^{-1} & -N^{-1} & \cdots & 1 - N^{-1} \end{bmatrix}. \tag{7.2.26}$$

Note that $\bar{f} = N^{-1} \sum_{k=1}^{N} f_k$. The solution is given by

$$\mathbf{f} = (\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{H})^{-1}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{g}}. \tag{7.2.27}$$

This is the equation for constrained linear inversion derived by Phillips (1962) and Twomey (1963). The quadrature constraint for smoothing can also be imposed on the first differences, i.e., $\sum(f_{j-1} - f_j)^2$, or the second differences, i.e., $\sum(f_{j-1} - 2f_j + f_{j+1})^2$, and so on (Exercise 7.3). In Exercise 7.4, interested readers may wish to carry out a numerical exercise for the retrieval of aerosol size distribution based on the direct inversion technique.

If there is a considerable amount of background data available from direct methods, it can be advantageous to construct an appropriate set of base functions to approximate the unknown function f. Let the mean of all past data be $\bar{\mathbf{f}}$. Then we may find a constrained solution that would minimize the mean square departure from this mean. From Eq. (7.2.25), we can write

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{f} - \mathbf{A}^{\mathrm{T}}\hat{\mathbf{g}} + \gamma(\mathbf{f} - \bar{\mathbf{f}}) = \mathbf{0}. \tag{7.2.28}$$

The solution is then

$$\mathbf{f} = (\mathbf{A}^{\mathrm{T}}\mathbf{A} + \gamma\mathbf{1})^{-1}(\mathbf{A}^{\mathrm{T}}\hat{\mathbf{g}} + \gamma\bar{\mathbf{f}}), \tag{7.2.29}$$

where $\mathbf{1}$ is an $(N \times N)$ identity matrix. This inversion will yield an improved measure of the solution if there is a reasonable basis for selecting $\bar{\mathbf{f}}$.

7.2.2   Determination of Total Ozone Concentration

A classic example of utilizing measured transmitted sunlight as a means of inferring composition information is the method proposed by Dobson (1957) for the estimate of total ozone concentration from a ground-based instrument. Basically, this method uses the Beer–Bouguer–Lambert law defined by Eq. (7.2.1) for the transfer of UV radiation to retrieve ozone concentration.

For application to ozone retrieval, it is convenient to rewrite Eqs. (7.2.1) and (7.2.2) in the form

$$\hat{I}(\lambda) = I_{\odot}(\lambda) \exp\left[-\sum_i \tau_i(\lambda)m_i(\theta_0) \right], \tag{7.2.30}$$

so that

$$\log_{10} \frac{\hat{I}(\lambda)}{I_{\odot}(\lambda)} = -[\tau_3^*(\lambda)\sec Z + \tau_A^*(\lambda)m(\theta_0) + \tau_R^*(\lambda)m(\theta_0)], \tag{7.2.31}$$

where all the values with * should be multiplied by $\log_{10} e$. The solar zenith angle Z is in reference to the height of about 20 km corresponding to the maximum ozone concentration. The factor $(r_0/r)^2$ can be removed by a subtraction operation described later. Further, we define the total ozone concentration in the vertical column as

$$\Omega = \int_0^{z_\infty} \rho_3(z)\,dz.$$

Let the absorption coefficient be denoted by $k(\lambda)$. We may select a pair of wavelengths (λ_1, λ_2) in the Hartley–Huggins ozone absorption bands described in Section 3.2 such that $k(\lambda_1) > k(\lambda_2)$. Thus,

$$\log_{10} \frac{\hat{I}(\lambda_i)}{I_\odot(\lambda_i)} = -k^*(\lambda_i)\Omega \sec Z - \tau_A^*(\lambda_i)m - \tau_R^*(\lambda_i)m, \qquad i = 1, 2. \quad (7.2.32)$$

Subtraction and rearrangement operations lead to

$$N = \log_{10} \frac{\hat{I}(\lambda_1)}{\hat{I}(\lambda_2)} - \log_{10} \frac{I_\odot(\lambda_1)}{I_\odot(\lambda_2)} = -\Omega \sec Z \Delta k - m \Delta \tau_A - m \Delta \tau_R, \quad (7.2.33)$$

where $\Delta k = k^*(\lambda_1) - k^*(\lambda_2)$ and $\Delta \tau_{A,R} = \tau_{A,R}^*(\lambda_1) - \tau_{A,R}^*(\lambda_2)$. Two pairs of wavelengths have been selected to minimize the aerosol effect because $\Delta \tau_A$ is the most uncertain term due to aerosol scattering. In the standard procedure developed by the World Meteorological Organization (WMO), these pairs are (0.3055, 0.3254 μm) and (0.3176, 0.3398 μm). An instrument using these pairs is referred to as a *Dobson spectrometer* (Dobson, 1957). Applying these two pairs to Eq. (7.2.33), we have

$$\Omega = \frac{N^{(1)} - N^{(2)}}{a \sec Z} - b, \qquad (7.2.34a)$$

where the parameterized coefficients a and b are determined from known ozone absorption coefficients and the Rayleigh scattering theory in the forms

$$a = \Delta k^{(2)} - \Delta k^{(1)} \cong 1.388\,(\text{atm}^{-1}\text{cm}^{-1}),$$

$$b \cong 0.009\,p_s. \qquad (7.2.34b)$$

The superscripts (1) and (2) denote the two pairs of wavelengths and p_s is the surface pressure in units of atmospheres.

Total ozone concentration has traditionally been measured in milli atm-cm, called Dobson units (DU). A DU is a vertical thickness of atmosphere in thousandths of a centimeter that is occupied by O_3 when concentrated into a uniform layer of pure gas at the standard temperature and pressure. The total column ozone concentration normally ranges from 200 to 400 DU.

Total ozone has been measured by the Dobson spectrometer at some 80 ground stations around the world. As discussed in Section 3.2.2, observations by Farman *et al.* (1985) using this instrument first detected the springtime (November of the Northern Hemisphere) lowering of ozone, beginning in about 1976.
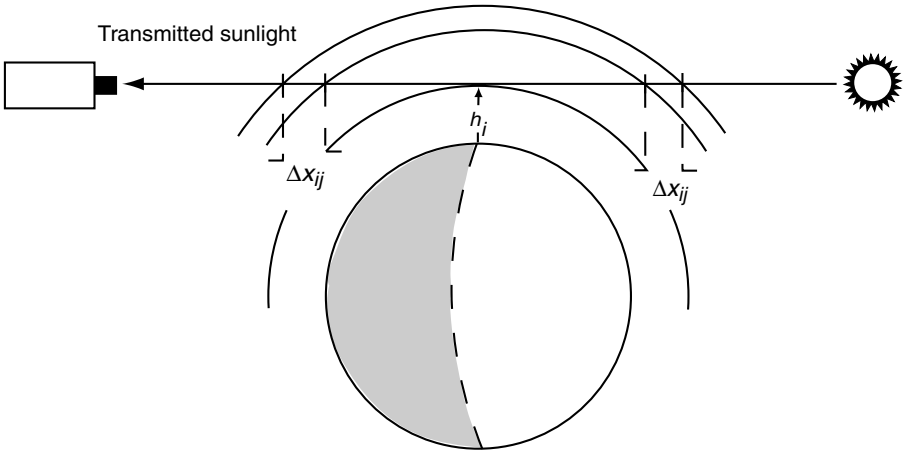
Figure 7.4 The geometry for limb extinction measurements, which are made from the satellite by scanning the sun's disk as the sun rises and sets relative to the motion of the satellite. In the retrieval of the atmospheric gaseous and aerosol profiles, the atmosphere is divided into many layers, in which h_i is the tangent height of the ith layer and Δx_{ij} corresponds to the horizontal distance.

7.2.3   Limb Extinction Technique

Analogous to ground-based sunphotometer measurements of transmitted sunlight, measurements can also be made from space by scanning the sun's disk as the sun rises and sets relative to the motion of the spacecraft, referred to as *solar occultation*. In reference to the geometry depicted in Fig. 7.4, let the intensity measured at the center of the scan be denoted by $I_\odot$, and the intensity determined from a scan at a lower altitude h_i be $I(h_i)$. The transmittance in this case is given by

$$T(\lambda) = I(\lambda, h_i)/I_\odot(\lambda) = \exp[-\tau(\lambda, h_i)], \qquad (7.2.35)$$

where the optical depth along the horizontal tangent path, shown in Fig. 7.4, is defined by the tangent height h_i as follows:

$$\tau(\lambda, h_i) = -\ln\left[\frac{I(\lambda, h_i)}{I_\odot(\lambda)}\right] = \int_{-\infty}^{\infty} \beta_e(\lambda, x)\, dx, \qquad (7.2.36a)$$

where the extinction coefficient is generally contributed by the extinction (scattering and absorption) of aerosols, Rayleigh molecules, O_3, and NO_2, as noted in Eqs. (7.2.2) and (7.2.4), in the form

$$\beta_e(\lambda, x) = \beta_{e,A}(\lambda, x) + \beta_{e,R}(\lambda, x) + \beta_{e,3}(\lambda, x) + \beta_{e,2}(\lambda, x). \qquad (7.2.36b)$$

The basic principle of the solar occultation technique is to take measurements at several wavelengths, similar to the sunphotometer, and at as many altitudes as possible

to obtain the profiles of aerosols and other trace gases. We may divide the atmosphere into an appropriate number of layers (e.g., 80) so that Eq. (7.2.36a) can be expressed in finite difference form such that

$$\tau(\lambda, h_i) = 2 \sum_j \beta_e(\lambda, z_j) \Delta \mathrm{x}_{ij}, \tag{7.2.36c}$$

where Δx_{ij} is the path length in the jth layer represented by z_j associated with the direct solar beam passing through the tangent height h_i. The geometrical factor Δx_{ij} can be devised as the weighting function for inverting the extinction coefficient profile from the optical depth measurements [see Eq. (7.2.13a)].

The limb extinction technique is specifically useful for the determination of aerosols and other minor gases in the stratosphere. It was explored by the Stratospheric Aerosol Measurement (SAM) experiment aboard the Nimbus 7 satellite in 1978. Subsequently, the Stratospheric Aerosol and Gas Experiment I (SAGE I) and SAGE II were conducted in 1979 and 1984, respectively. SAGE II contains seven channels centered at 0.385, 0.448, 0.453, 0.525, 0.6, 0.94, and 1.02 μm. The 0.94 and 0.6 μm channels were used to infer the water vapor and ozone amounts, while the difference between 0.448 and 0.453 μm was used to determine the NO_2 concentration. In this manner, aerosol extinction coefficient profiles can be derived from the seven measurements, following the constrained linear inversion method introduced in Subsection 7.2.1.2. Pioneering work in the solar occultation technique for stratospheric aerosol studies has been reported by McCormick *et al.* (1979).

The solar occultation technique is extremely sensitive to the presence of high-level clouds. As a by-product, the data from SAGE have been used to derive the frequency of occurrence of cirrus clouds based on the extinction values measured (Woodbury and McCormick, 1986). The cirrus results derived from SAGE have complemented those from other satellite remote sensing methods, particularly in view of the fact that it is difficult to determine thin cirrus with optical depths less than about 0.5 based on reflected sunlight and/or emitted infrared radiation from nadir-looking radiometers (see Sections 7.3.5 and 7.4.5 for discussions of the remote sensing of clouds).

7.3   Remote Sensing Using Reflected Sunlight

7.3.1   Satellite–Sun Geometry and Theoretical Foundation

In reference to Fig. 6.1, the sunlight (denoted as In) reflected from a position-vector on the earth and the atmosphere is detected by a satellite (denoted as Out). The position of the sun is defined by the solar zenith angle θ_0, while the satellite position is defined by the emergent zenith angle θ. The relative positions of the sun and the satellite are given by the azimuthal angle difference $\Delta\phi$. From spherical geometry (Appendix C and Fig. 6.1), the angle between the incoming and outgoing light

beams counterclockwise from the incoming beam, known as the scattering angle Θ, is defined by

$$\cos \Theta = \cos \theta \cos \theta_0 + \sin \theta \sin \theta_0 \cos \Delta\phi$$

$$= \mu\mu_0 + (1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2} \cos \Delta\phi, \qquad (7.3.1)$$

where we set $\mu = \cos \theta$ and $\mu_0 = \cos \theta_0$.

Any radiometer on board a satellite will have a finite field-of-view on a horizontal plane that collects radiation from the earth and the atmosphere, referred to as *resolution,* which is dependent on the specific instrument designed and its scan angle. The orbit of a satellite is determined by the earth's gravitational attraction force and the centrifugal acceleration of the satellite and is governed by Kepler's laws discussed in Section 2.2. Nearly all the present meteorological satellites are in one of two orbits: sunsynchronous or geostationary.

The earth makes one complete resolution about the sun (2π radian) in one tropical year (365.2422 days). Thus, the right ascension of the sun changes at an average rate of about $1°\text{day}^{-1}$. If the inclination of the satellite is correctly chosen, the right ascension of its ascending node can be made to precess at the same rate. An orbit that is synchronized with the sun is called a *sunsynchronous orbit* or *polar orbit*. The point at which a satellite crosses the earth's equatorial plane from south to north is known as the ascending node; the point passed as it crosses the plane from north to south is known as the descending node. The point directly beneath the satellite is called the subsatellite point or nadir. For a satellite with a height of about 870 km, its inclination angle needs to be about $99°$ for its orbit to be sunsynchronous. This is the nominal height of the National Oceanic and Atmospheric Administration (NOAA) polar-orbiting weather satellites, which orbit the earth in a period of about 100 minutes and cross the equator at the same local standard time (LST) every day, as illustrated in Fig. 7.5a. Morning satellites ascend (or descend) between 06 and 12 h LST and descend (or ascend) between 18 and 24 h LST. Afternoon satellites ascend (or descend) between 12 and 18 h LST, and descend (or ascend) between 00 and 06 h LST.

The space–time coverage of a polar orbiting satellite depends on its orbit and the scanning geometry of its instruments. The poles or near poles are observed on every orbit and the instruments view every point on the earth at least twice per day. Most meteorological satellite instruments are designed such that the area viewed on one orbit touches or overlaps the area viewed on previous and successive orbits. The data is produced in the form of scan lines, each divided into elements or samples known as *pixels* or *scan spots*. Each pixel has a unique time associated with it and is defined by its latitude and longitude. The size of a pixel is controlled by the field-of-view of the instrument and its scan pattern.

If a satellite is moved farther from the earth, it would experience a weaker gravitational field and the centrifugal acceleration required to keep the satellite in orbit would be smaller. The rotation period of the satellite would also become longer. It is possible to select a distance at which the rotation period is exactly equal to the 1 day

(a) Polar
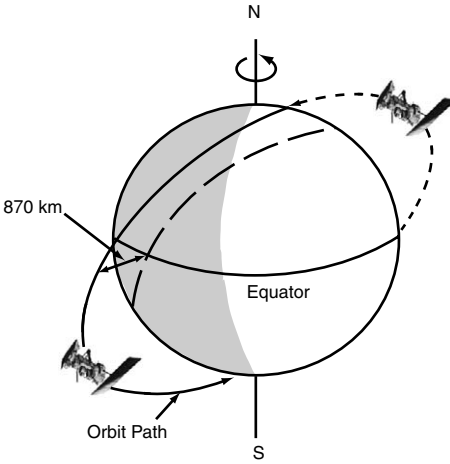
(b) Geostationary
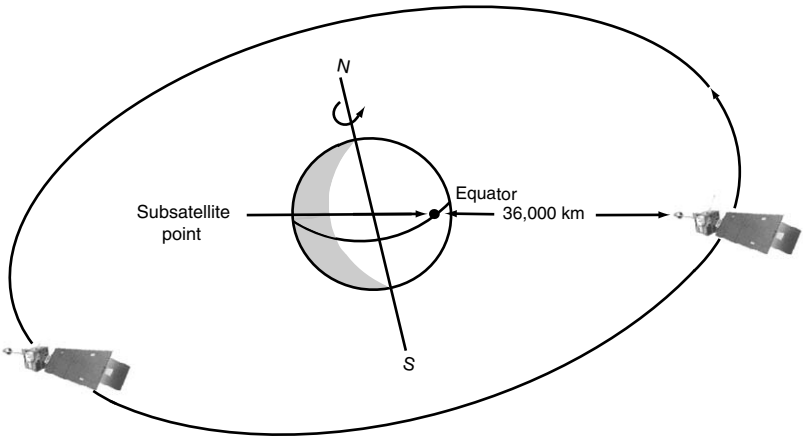
Figure 7.5 (a) Polar and (b) geostationary orbits for NOAA satellites. The polar orbit rotates $1°$ per day to make it synchronous with the sun. The geostationary satellite stays continuously above one spot on Earth.

rotation period of the earth in such a manner that the satellite moves in a counterclockwise fashion. This orbit is called a *geosynchronous* or *geostationary orbit*. To achieve this distance, the satellite must be about 36,000 km above the earth's surface (see Exercise 2.4). Geostationary satellites remain essentially stationary above a point on the equator and are classified by the longitude of their subsatellite points (Fig. 7.5b). They view a fixed area comprising about 42% of the globe at the same time so that

any point in this area can be observed as frequently as their instruments will allow. However, because each point has a fixed geometric relationship to the satellite, it is viewed at only one set of zenith and azimuthal angles. Presently, five geostationary satellites orbit the earth to gather weather data, including two operated by the United States, one by the European Space Agency (METEOSAT), one by Japan (GMS), and one by India (INSAT). The current U.S. system consists of two Geostationary Operational Environmental Satellites (GOES).

In Section 1.4.4, we introduced the basic equation of radiative transfer for plane-parallel atmospheres. In conjunction with satellite applications, we shall rewrite this equation for diffuse intensity as

$$\mu\frac{dI(\tau;\mu,\phi)}{d\tau} = I(\tau;\mu,\phi) - J(\tau;\mu,\phi), \tag{7.3.2a}$$

where the source function, representing the scattering contributions by the direct solar beam and by multiple scattering of the diffuse beam, is given by (Section 6.1)

$$J(\tau;\mu,\phi) = \frac{\tilde{\omega}}{4\pi}\int_0^{2\pi}\int_{-1}^{1} I(\tau;\mu',\phi')P(\mu,\phi;\mu',\phi')\,d\mu'd\phi'$$

$$+ \frac{\tilde{\omega}}{4\pi}F_\odot P(\mu,\phi;-\mu_0,\phi_0)e^{-\tau/\mu_0}, \tag{7.3.2b}$$

where P is the phase function, which represents the angular distribution of scattered energy as a function of direction and is an important parameter associated with satellite remote sensing of aerosols and clouds, $F_\odot$ is the direct solar irradiance at the top of the atmosphere (TOA), and the single-scattering albedo $\tilde{\omega}$ is defined as the ratio of the scattering cross section σ_s to the extinction cross section σ_e. The phase function, single-scattering albedo, and extinction cross section are the three fundamental parameters in radiative transfer.

To seek a solution for Eq. (7.3.2a), we shall separate the atmosphere and the surface and assume that there are no diffuse intensities from the top and bottom of the atmosphere. Under this condition, the reflected intensity at TOA can be expressed in terms of the integral equation as follows:

$$I(0;\mu,\phi) = \int_0^\tau J(\tau';\mu,\phi)e^{-\tau'/\mu}\frac{d\tau'}{\mu}. \tag{7.3.3}$$

Moreover, we may seek a solution by using the single-scattering approximation to obtain the nondimensional bidirectional reflectance introduced in Section 3.4.2 or *bidirectional reflection distribution function* (BRDF) defined by

$$R(\mu,\phi;\mu_0,\phi_0) = \frac{\pi I(0;\mu,\phi)}{\mu_0 F_\odot}$$

$$= \frac{\tilde{\omega}}{4(\mu+\mu_0)}P(\mu,\phi;-\mu_0,\phi_0)\left\{1 - \exp\left[-\tau\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)\right]\right\}. \tag{7.3.4a}$$

If we consider an atmosphere such that its optical depth is extremely small, then Eq. (7.3.4a) can be further simplified in the form

$$R(\mu, \phi; \mu_0, \phi_0) = \frac{\tilde{\omega}\tau}{4\mu\mu_0} P(\mu, \phi; -\mu_0, \phi_0). \qquad (7.3.4b)$$

Under the optically thin and single-scattering approximations, it is clear that the BRDF at TOA is directly proportional to the phase function and optical depth.

Consider now an underlying surface and account for multiple reflections between the atmosphere and the surface. For simplicity of presentation, let the surface be Lambertian with an albedo r_s (isotropic reflection, although it is not required in the formulation). The BRDF for the combined atmosphere and surface system is the summation of all the light beams emergent at TOA, as shown in Fig. 7.6. Thus, we have

$$\begin{aligned} R &= R_a + \tilde{T}_a r_s \tilde{T}_a^* + \tilde{T}_a r_s R_a^* r_s \tilde{T}_a^* + \cdots \\ &= R_a + \tilde{T}_a r_s \tilde{T}_a^* / (1 - R_a^* r_s), \end{aligned} \qquad (7.3.5)$$

where R_a denotes the BRDF from the atmosphere only and the second term represents the surface contribution. In this equation, R_a^* is the BRDF of the atmosphere for radiation from below, and for practical purposes, $R_a^* \approx R_a$. The term $\tilde{T}_a(\mu_0)$ is the total transmission function (direct plus diffuse) from the sun to the surface, while $\tilde{T}_a^*(\mu)$ is the total transmission function from the surface to the satellite. These terms
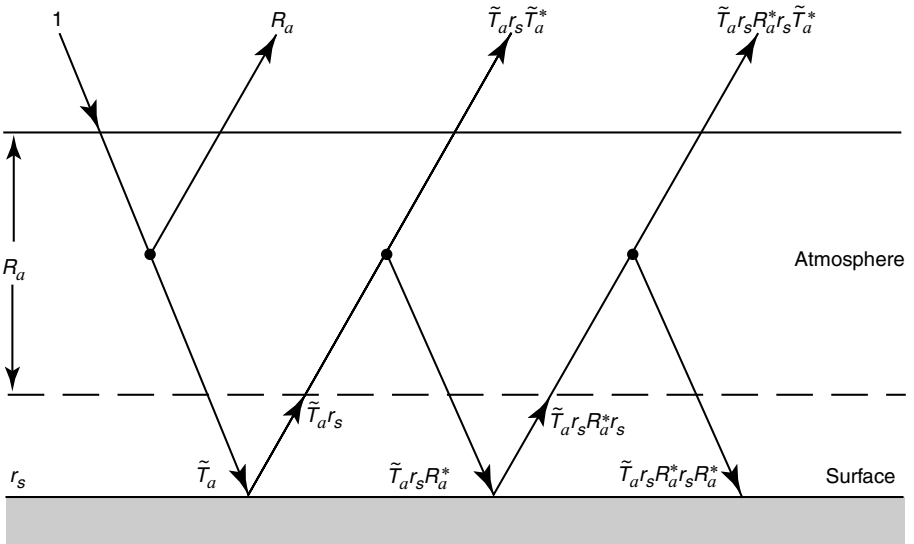
Figure 7.6 Contribution of the bidirectional reflectance from the atmosphere–surface system. The term R_a is the bidirectional reflectance from the atmosphere only, and r_s is the surface albedo, which can be directionally dependent. See also Fig. 6.9.

can be computed from the theory of radiative transfer such as the adding principle introduced in Section 6.4. Within the limits of single-scattering and optically thin approximations, however, we have the diffuse transmission function

$$T_{\text{dif}}(\mu, \phi; \mu_0, \phi_0) = \frac{\tilde{\omega}}{4(\mu - \mu_0)} P(-\mu, \phi; -\mu_0, \phi_0)\left(e^{-\tau/\mu} - e^{-\tau/\mu_0}\right)$$

$$\cong \frac{\tilde{\omega}\tau}{4\mu\mu_0} P(-\mu, \phi; -\mu_0, \phi_0), \quad \mu \neq \mu_0. \tag{7.3.6}$$

The direct transmission function is simply $T_{\text{dir}} = \exp(-\tau/\mu_0)$. Equation (7.3.5) serves as the fundamental radiative transfer equation for the determination of aerosols and ozone in the atmosphere on the one hand, and for the inference of surface properties on the other.

7.3.2   Satellite Remote Sensing of Ozone

The basic principle involved in the estimation of ozone concentration utilizing reflected sunlight is to select a pair of wavelengths in the Hartley–Huggins ozone absorption band. The selection process is similar to that for the Dobson ozone spectrometer described in Section 7.2.2. Wavelengths near the long-wavelength end of the band at which absorption is relatively weak are chosen so that most of the photons reaching the satellite instrument have passed through the ozone layer and backscattered from within the troposphere. The two wavelengths are separated by about 200 Å so that the scattering effect is about the same at each wavelength, while absorption for one of these wavelengths is stronger than the other. A pair such as (3125, 3312 Å), for example, has been selected in the Nimbus 4 satellite experiment.

The backscattering radiance in the ozone band at the point of a satellite with a nadir-looking instrument depends on the attenuation of the direct solar flux through the ozone layer, the reflecting power of the atmosphere and the associated surface, and the attenuation of the diffusely reflected photons to the point of the satellite. If Z denotes the solar zenith angle at the level of maximum ozone concentration (about 20 km) at the subsatellite point, then the total attenuation path of the backscattered photons through the ozone layer is proportional to $1 + \sec Z$. Let $F_\odot$ and I be the incident solar irradiance and the measured backscattered intensity at TOA, respectively. Following the procedure described in Eq. (7.2.33), we then define

$$\hat{N}(\lambda_1, \lambda_2) = \log_{10} \frac{F_\odot(\lambda_1)}{\hat{I}(\lambda_1)} - \log_{10} \frac{F_\odot(\lambda_2)}{\hat{I}(\lambda_2)}. \tag{7.3.7}$$

Determination of the total ozone concentration can be made by comparing the observed $\hat{N}$ with values precomputed for a series of different standard ozone profiles by means of the searching method.

The computational method for the transfer of solar radiation in a scattering and absorbing atmosphere for ozone studies usually follows Eq. (7.3.5). Writing this

equation in terms of scattered intensity we have

$$\hat{I}(\Omega, \mu_0, r_s) = I(\Omega, \mu_0, 0) + \frac{r_s T(\Omega, \mu_0)}{1 - r_s \bar{r}(\Omega)}, \tag{7.3.8}$$

where $T(\Omega, \mu_0) = \mu_0 F_\odot \widetilde{T}_a(\mu = 1) \widetilde{T}_a(\mu_0)/\pi$, $\mu_0 = \cos Z$, $\bar{r}$ is a mean atmospheric reflection, and Ω denotes the total ozone concentration. The first term is the atmospheric contribution alone, while the second is produced by interaction with the underlying surface. In this equation, all the relevant parameters in the radiative terms are included in parentheses. Note that since the ozone instrument is normally looking in the nadir direction, the azimuthal dependence can be neglected.

The three basic procedures for the determination of total ozone concentration from the observed $\hat{N}$ are as follows: (a) A set of tables containing the computed quantities $I(\Omega, \mu_0, 0)$, $T(\Omega, \mu_0)$, and $\bar{r}(\Omega)$ for different values of μ_0 and Ω are prepared *a priori*. (b) The effective surface albedo is determined by utilizing the radiometric measurement at a wavelength outside the ozone absorption band, say $\lambda_3(3800 \text{ Å})$. At this wavelength, the ozone dependence drops out of all the terms in Eq. (7.3.8), and a measurement of $\hat{I}_3(\mu_0, r_s)$ permits the direct calculation of the surface albedo with the formula

$$r_s(\lambda_3) = \frac{\hat{I}_3(\mu_0, r_s) - I_3(\mu_0, 0)}{T(\mu_0) + \bar{r}[\hat{I}_3(\mu_0, r_s) - I_3(\mu_0, 0)]}. \tag{7.3.9}$$

The assumption is made that r_s is independent of wavelength so that it can be used for the pair of wavelengths (λ_1, λ_2). However, an empirical adjustment may also be performed from known surface albedo measurements. (c) With the surface albedo known, computations are then carried out to generate $N(\lambda_1, \lambda_2)$ versus total ozone concentration Ω. Best estimates of Ω from the observed intensities via $\hat{N}$ can then be made by an optimized search method. The matching and search method has been used by Mateer *et al.* (1971) to estimate total ozone concentration from the Nimbus 4 satellite measurements of backscattered intensities. Dave (1978) further discussed the effect of atmospheric aerosols on the estimate of total ozone concentration.

As a result of environmental concerns and potential health hazards, ozone concentration has been monitored by satellites since the beginning of the sounding era. The Backscatter Ultraviolet Spectrometer (BUV) on board Nimbus 4 operated for seven years, from 1970 to 1977. The Total Ozone Mapping Spectrometer (TOMS) on Nimbus 7 produced daily global maps of total ozone at 50–150 km resolution in 1978. The Solar Backscatter Ultraviolet Radiometer (SBUV) on the NOAA satellites has provided daily ozone data over the globe on a routine basis since 1980.

7.3.3 Satellite Remote Sensing of Aerosols

The retrieval of aerosol optical and microphysical properties from satellite measurements of reflected sunlight is a difficult task because of the relatively small influence of aerosols. Aerosol retrieval is further complicated by the variability of the underlying surfaces. In reference to Section 7.2.1, the basic aerosol parameters include

optical depth, size distribution, and the refractive index, which is wavelength dependent. The last parameter is related to the composition of the aerosol. Given the measured reflected intensities at some appropriate wavelengths that are sensitive to aerosol properties, we wish to determine the pertinent parameters. The simplest situation would be a black surface that makes no contribution to the scattered intensity. In this case, we can use Eq. (7.3.4b) so that the aerosol optical depth is determined by

$$\tau = \frac{\hat{R}(\mu, \phi; \mu_0, \phi_0)}{P(\mu, \phi; \mu_0, \phi_0)} \frac{4\mu\mu_0}{\tilde{\omega}}. \tag{7.3.10}$$

In addition to the measured bidirectional reflectance $\hat{R}$ and the sun–satellite geometry, we also need the information of phase function P and single-scattering albedo $\tilde{\omega}$. Because the aerosol size distribution and the refractive indices are generally unknown *a priori, P* and $\tilde{\omega}$ must be assumed or parameterized from other information. In this manner, a measured $\hat{R}$ will yield an optical depth for the atmosphere. Since the Rayleigh scattering of the atmosphere is known, aerosol optical depth can subsequently be determined. The oceans are close to black in the visible ($\sim$6% albedo on average) and near-infrared wavelengths. Thus, retrieval of aerosol optical depth from satellites has had some success over the oceans. The remote sensing of aerosols over land, however, despite its importance, is a subject still requiring research and development.

Since 1979, the Advanced Very High Resolution Radiometer (AVHRR) aboard NOAA satellites has been used to study aerosols. It consists of five nominal channels centered at 0.63, 0.86, 3.7, 10.9, and 12 μm, with a horizontal resolution of about 1 km $\times$ 1 km. The 0.63 μm channel has been extensively employed to map aerosol optical depth over the oceans. With the addition of the 0.86 μm channel, the size parameter of a Junge distribution can be estimated. The 0.86 μm channel data has also been used to provide an estimate of the combination of $\tilde{\omega}P$ occurring in Eq. (7.3.10).

A number of other instruments have also been designed to study aerosols and clouds from space. Under NASA's earth science programs, specifically the Earth Observing System (EOS) Program, a series of polar-orbiting and low-inclination satellites have been and continue to be in orbit for long-term global observations of the atmosphere, land surface, biosphere, solid earth, and oceans. Two of the EOS instruments are particularly pertinent to the study of aerosols. The Moderate–Resolution Imaging Spectrometer (MODIS) (King *et al.,* 1992) is designed to make 0.25–1 km resolution observations in 36 visible and infrared bands. Many of the visible and near-infrared channels have been used for the mapping of aerosol optical depth and size. The Multi-Angle Imaging SpectroRadiometer (MISR) provides multiangle observations. MISR produces images at nine fixed angles in the along-track direction with one at the nadir, and four both fore and aft out to $\pm 70.5°$. Four spectral bands centered at 0.443, 0.555, 0.670, and 0.805 μm are available with ground resolution between about 240 m and 2 km (Diner *et al.,* 1998). This instrument appears to be useful for studying the scattering phase function of spherical and nonspherical aerosols.

7.3.4   Satellite Remote Sensing of Land Surfaces

The land surface of the earth is highly complex involving a combination of vegetations, soils, sand, rocks, water, and ice. A precise determination of the vegetation and surface properties is critical to the present and future management of earth's resources, as well as to the understanding of the role that land surfaces play in the climate system. The remote sensing of vegetation and surface properties by satellites is a subject that is of interest to numerous disciplines, including meteorology, hydrology, geography, geology, biology, ecology, and electrical engineering. Most of the research areas involve the analysis of satellite images and are beyond the scope of this text. However, we shall discuss the fundamentals associated with the determination of surface albedo and present an important application of the AVHRR /NOAA data for the inference of vegetation.

The surface reflectivity or albedo determines the solar flux available at the surface and is the essential parameter in the discussion of climate over land. The surface albedo is dependent on the type of surface. For water surfaces, the albedo ranges from about 6 to 9%, except for cases involving the low solar angle that is associated with the high latitudes of the winter hemisphere. The albedo can range from 10 to 40% for various land surfaces. For example, deserts and sand dunes have albedos of about 30–40%, whereas those for meadows and forests are about 10%. The albedos of snow and ice are greater than 40%. Moreover, the albedos of some vegetation surfaces vary greatly with solar wavelength as noted below.

In reference to Eq. (7.3.5), if we select a wavelength in the visible spectrum (0.4–0.7 μm) such that the product of the surface albedo r_s and the atmospheric BRDF (generally less than 0.1) is much smaller than 1, then we can express the surface albedo in the form

$$r_s \cong a\hat{R} - b, \qquad (7.3.11)$$

where $a = 1/\widetilde{T}_a\widetilde{T}_a^*$ and $b = R_a/\widetilde{T}_a\widetilde{T}_a^*$. The coefficients a and b are the so-called atmospheric correction terms involving the scattering contributions of aerosols and molecules, which can be empirically determined for specific applications. Thus, a measurement of BRDF $\hat{R}$ at TOA determined by the sun–satellite geometry will provide a surface albedo value. A surface albedo so determined is averaged over the pixels of the satellite instrument and is normally assumed to be independent of the incoming and outgoing directions. In reality, the surface reflection function can be extremely complex, especially over vegetation areas.

An important characteristic of the reflection by vegetation surfaces is the sharp transition in the reflection at a wavelength of about 0.7 μm. The presence of chlorophyll in vegetation leads to strong absorption at wavelengths shorter than 0.7 μm. Figure 7.7 shows the reflectance as a function of wavelength for dry and wet bare soil and vegetation cover (Li *et al.,* 2002). In view of these spectral properties, we may select two AVHRR wavelengths ($\lambda_1 = 0.63$ μm and $\lambda_2 = 0.86$ μm) and define a term called the normalized difference vegetation index (NDVI), in the form

$$\text{NDVI} = \frac{I(\lambda_2) - I(\lambda_1)}{I(\lambda_2) + I(\lambda_1)}. \qquad (7.3.12)$$
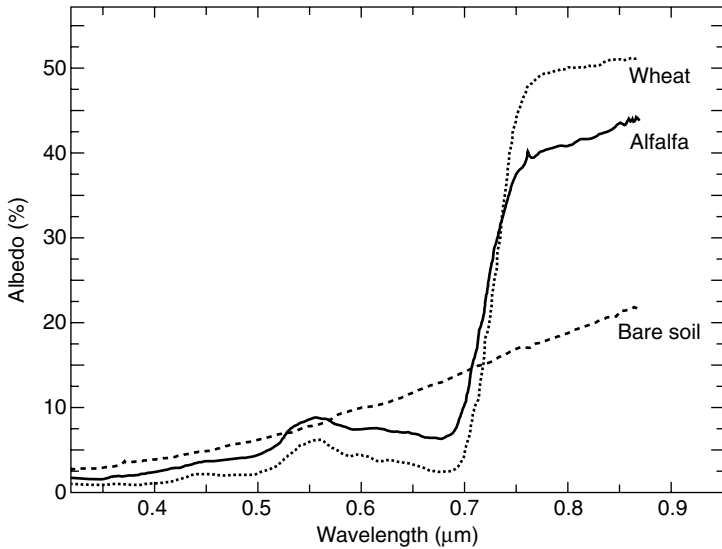
Figure 7.7   Illustration of the step function transition in the surface spectral albedo of two types of vegetation at 0.7 μm, compared to the continuous spectral albedo for soil surfaces. These results were derived from measurements made in the Southern Great Plains of Oklahoma surrounding the Central Facility of the Department of Energy's Atmospheric Radiation Measurement site (data taken from Li *et al.,* 2002).

It is an index of biospheric activity indicating the amount of live vegetation in the field-of-view of the satellite sensor. The NDVI field has been routinely processed from the NOAA /AVHRR data for research and operational purposes.

7.3.5   Cloud Optical Depth and Particle Size

Clouds cover more than 50% of the planet earth and are the most important regulators of the radiation budget of the earth–atmosphere system. The transfer of radiation through cloud layers depends on particle concentration, size distribution and shape, and cloud thickness and geometry. Information about cloud composition and structure is of vital importance to the understanding of the radiation balance and energetics of the earth–atmosphere system.

Because of the number of variables involved and the associated problem of multiple scattering by particulates, determination of the highly varying cloud variables by remote sensing is a difficult task. With the availability of high-speed computers and efficient numerical methods for the solution of the radiative transfer equation, advanced techniques for the inference of cloud composition using the reflected intensity and polarization of sunlight have been developed. Analogous to the inference of total ozone concentration, the method involved is an ad hoc direct matching between the observed and computed intensity and/or polarization. In this section, we describe the information content of the reflected intensity and polarization of sunlight from water and ice clouds.

7.3.5.1 BIDIRECTIONAL REFLECTANCE

The remote sensing of clouds from space using reflected sunlight requires a fundamental understanding of the radiative properties of water and ice clouds. Determination of these properties must begin with knowledge of the complex refractive index of water and ice as a function of wavelength. The real refractive indices of water exhibit substantial deviations from those of ice at wavelengths greater than about 10 μm. The imaginary refractive index of ice is much more complex, with values ranging from about 10^{-6} at 0.9 μm to about 0.75 at 3 μm. Ice exhibits relatively strong absorption at about 1.6 μm, where water shows a minimum. The imaginary refractive indices for ice and water vary rapidly in the solar and IR spectra. There are significant variations in the absorption properties of ice and water in the near-IR solar spectrum.

In Section 3.2.3, we discussed the absorption bands in the solar spectrum. In the near-IR region, the absorption of solar radiation is primarily due to water vapor. Absorption by water vapor is strong in the band centers and falls off sharply away from these centers. Absorption by ice and water, on the other hand, is more constant across each band, being less intense than water vapor in the band centers and more intense away from the centers. Moreover, the maximum absorption due to ice and water does not coincide exactly with the water-vapor band centers. There are regions in the near-IR where water vapor absorption dominates, whereas in other regions, ice and water absorption is more important. The reflection of sunlight from clouds depends on the optical depth and some measure of the particle size within the cloud. Figure 7.8 illustrates the spectral bidirectional reflectance and absorptance for three mean effective ice crystal sizes [see Eq. (7.3.15a) for definition]. The calculation uses a solar zenith angle of $30°$, an emergent angle of $0°$, an ice water path (IWP) of 1 g m^{-2}, and a spectral interval of 50 cm^{-1} containing 30 cumulative probability functions (g's) covering the spectrum from 0.2 to 5 μm, based on the correlated k-distribution method for infrared radiative transfer introduced in Section 4.3. From the absorption spectrum, the H_2O absorption bands located at 3.2, 2.7, 1.87, 1.38, 1.1, 0.94, 0.82, and 0.72 μm are evident, as are the 4.3 μm CO_2 and O_3 UV bands. The bidirectional reflectances have maxima in between the H_2O absorption bands. Small ice crystals reflect more solar radiation. Sufficient ice crystal size information is clearly demonstrated in the near-IR spectrum. The spectral features are similar for clouds containing water droplets.

Variation of the reflectance in the near-IR region could provide a means for the inference of some of the physical properties of clouds. This possibility has been suggested by Hansen and Pollack (1970), who have interpreted spectral near-IR reflectances of clouds measured from aircraft. Suggestions of the inference of the optical depth and mean radius from spectral reflectance measurements have also been made by Twomey and Seton (1980). The preceding suggestions are based on the principle of radiative transfer that the reflection of clouds at a wavelength with little absorption (such as in the visible spectrum) is primarily a function of their optical depth, whereas in a near-IR wavelength at which the absorption of water vapor in the clouds is small, their reflection is largely dependent on particle size.
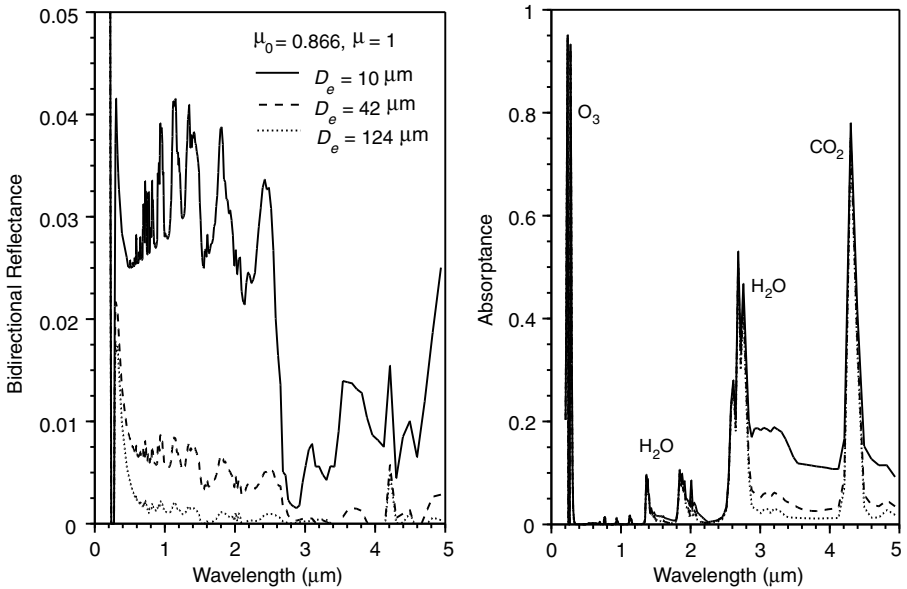
Figure 7.8   Spectral bidirectional reflectance and absorptance as functions of wavelength for cirrus clouds with mean effective ice crystal sizes of 10, 42, and 124 μm. Absorption bands due to ozone, water vapor, and carbon dioxide are identified. The spectral features are similar for water clouds.

Before we present some pertinent radiative transfer calculations and observed data, we must define the particle size that can be inferred from reflected sunlight. A brief review of the droplet size distribution in water clouds, including stratus, nimbostratus, altostratus, and various types of cumulus, has been presented by Liou (1992). A determination of the detailed droplet size spectrum from space, employing currently available instruments, has yet to be developed and verified. Nevertheless, we may define a mean effective size to represent the actual size distribution taking into account the scattering property of spherical droplets such that they scatter an amount of light proportional to their cross sectional area as follows:

$$a_e = \int a \cdot \pi a^2 n(a) \, da \bigg/ \int \pi a^2 n(a) \, da. \tag{7.3.13a}$$

The mean effective radius differs from the simple mean radius in that the droplet cross section is included as a weighting factor. As shown by many radiative transfer calculations, the reflected and transmitted sunlight is primarily dependent on this parameter regardless of the detailed distribution of droplet sizes (Hansen and Travis, 1974). In the following, we wish to relate this mean parameter to the amount of vertically integrated liquid water in the cloud, referred to as the liquid water path (LWP, g cm^{-2}), and optical depth.

From the definition of liquid water content (LWC, g cm^{-3}) for spherical droplets, we have

$$\text{LWC} = \frac{4\pi}{3}\rho_l \int a^3 n(a)\, da, \tag{7.3.13b}$$

where ρ_l is the density of water. Thus, $\text{LWP} = \text{LWC} \cdot \Delta z$. The optical depth is defined by [see also Eq. (7.2.9b)]

$$\tau = \Delta z \cdot \int Q_e \pi a^2 n(a)\, da, \tag{7.3.13c}$$

where Q_e is the efficiency factor for extinction, a function of the droplet radius, wavelength, and refractive index. For wavelengths in the visible, $Q_e \cong 2$ for cloud droplets. Thus, combining the preceding three equations, we obtain

$$a_e \cong \frac{3}{2\rho_l}\text{LWP}/\tau, \tag{7.3.13d}$$

which relates the optical depth, LWP, and droplet size. The relation is significant. Consider two clouds having the same LWP. The cloud that contains a smaller droplet would have a larger optical depth and therefore would reflect more sunlight. Anthropogenic pollution sources can affect droplet size, referred to as the *indirect aerosol effect* (see Exercise 7.5 and Section 8.4.4 for further discussion).

In conjunction with the retrieval of the optical depth and mean effective droplet radius of water clouds from the NASA EOS/Terra satellite using the MODIS instrument discussed in Section 7.3.3, King *et al.* (1997) performed radiative transfer calculations for a wide variety of solar zenith angles and observational zenith and azimuthal angles at a number of selected wavelengths in the visible and near-IR. Shown in Fig. 7.9 are bidirectional reflection function (BRDF or reflectance) correlations at 0.664/1.62 μm and 0.664/2.142 μm. These wavelengths were selected because they are outside the water vapor and oxygen absorption bands and, at the same time, have substantially different water droplet absorption properties. The minimum value of the reflectance at each wavelength corresponds to the underlying ocean reflectance in the absence of an atmosphere. The dashed curves represent the reflectances for specified values of cloud optical depth, whereas the solid lines denote those for specified values of droplet mean effective radius. As evidenced by these results at a nonabsorbing wavelength (0.664 μm), cloud optical depth is largely determined by reflectance with little dependence on droplet radius. The reflectance at 2.142/1.621 μm, in contrast, is largely sensitive to the mean effective radius, with the largest values occurring for small droplet sizes. The data points superimposed on the theoretical curves represent over 400 measurements obtained from the MODIS Airborne Simulator (MAS) instrument, which is a 50-band scanning spectrometer that was mounted in the NASA ER-2 aircraft during an experiment that was carried out over marine stratocumulus clouds in the vicinity of the Azores approximately 1000 km southwest of Lisbon on 22 June, 1992. Simultaneous determination of cloud optical depth and
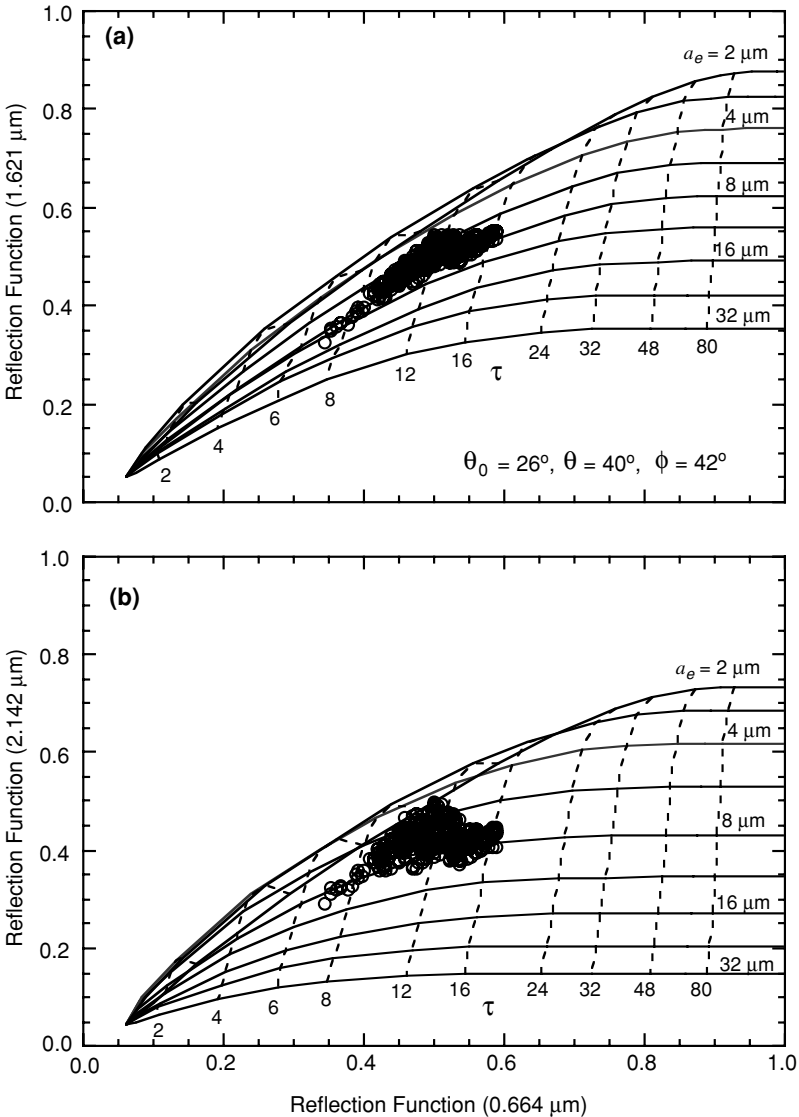
Figure 7.9   Theoretical relationship between the reflectance at 0.664 and (a) 1.621 μm and (b) 2.142 μm for various values of optical depth at 0.664 μm and effective droplet radius, for the specific geometry denoted in the diagram. Data from over 400 measurements obtained with the MAS that was aboard the NASA ER-2 aircraft are superimposed. These observations were obtained when the aircraft flew above marine stratocumulus clouds during a stratus cloud experiment on June 22,1992 (data taken from King *et al.*, 1997).

mean effective droplet radius can be performed by maximizing the probability that the measured reflectances $R_m(\mu_0, \mu, \Delta\phi)$ have the functional form $R_c(\tau, a_e; \mu_0, \mu, \Delta\phi)$, such that

$$\chi^2 = \sum_{i=1}^{3} \left[\ln R_m^i(\mu_0, \mu, \Delta\phi) - \ln R_c^i(\tau, a_e; \mu_0, \mu, \Delta\phi) \right]^2, \qquad (7.3.14)$$

where the summation extends over the three wavelengths. Validation of this retrieval algorithm is critical because of the nonuniqueness of the solutions determined from the statistical optimization method, and is a subject of continuing research.

Analogous to the retrieval of water cloud optical depth and effective radius, one can also apply the reflectance correlation technique to clouds containing ice crystals (Rolland and Liou, 1998; Ou *et al.,* 1999). Ice crystals are much more complicated than spherical water droplets with respect to their scattering and radiative properties, as presented in Sections 5.4 and 6.7.1. We have defined the mean effective ice crystal size in Eq. (5.1.1) to represent the ice crystal size distribution, and it is duplicated here for the continuity of the present discussion on remote sensing:

$$D_e = \int V n(L)\, dL \Big/ \int A\, n(L)\, dL, \qquad (7.3.15a)$$

where L is the maximum dimension of an ice crystal, V is the volume, A is the geometric projected area of an ice crystal on a surface perpendicular to the incident light beam, and $n(L)$ is the ice-crystal size distribution. In this manner, the shape of irregular ice crystals is accounted for. Further, the ice water content (IWC) for a given ice crystal size distribution is defined by

$$\text{IWC} = \int V\rho_i n(L)\, dL, \qquad (7.3.15b)$$

where ρ_i is the density of ice. The volume of a hexagonal ice crystal is given by $3\sqrt{3}LD^2/8$, with D the ice crystal width. Moreover, for randomly oriented ice crystals in the limits of geometric optics, the extinction cross section is $3D(\sqrt{3}D/4 + L)/2$ (Takano and Liou, 1989a). With these relationships and following the principle introduced for water droplets, we may also relate visible optical depth and mean effective ice crystal size as follows:

$$\tau \cong \text{IWP}(c + b/D_e), \qquad (7.3.15c)$$

where $\text{IWP} = \text{IWC} \cdot \Delta z$ denotes ice water path, and $c \cong -6.656 \times 10^{-3}$ and $b \cong 3.686$ for ice columns. Below we present the inference of the optical depth and ice crystal mean effective size based on reflected solar radiation.

Radiative transfer calculations using the MAS channels of 0.681, 1.617, and 2.139 μm were performed for the six ice-crystal size distributions with mean effective ice-crystal size ranging from 24 to 124 μm, and for two water-droplet size distributions with mean effective radii of 4 and 8 μm (Fig. 7.10). The results are displayed in two-dimensional reflectance diagrams in terms of 0.681–1.617 μm and
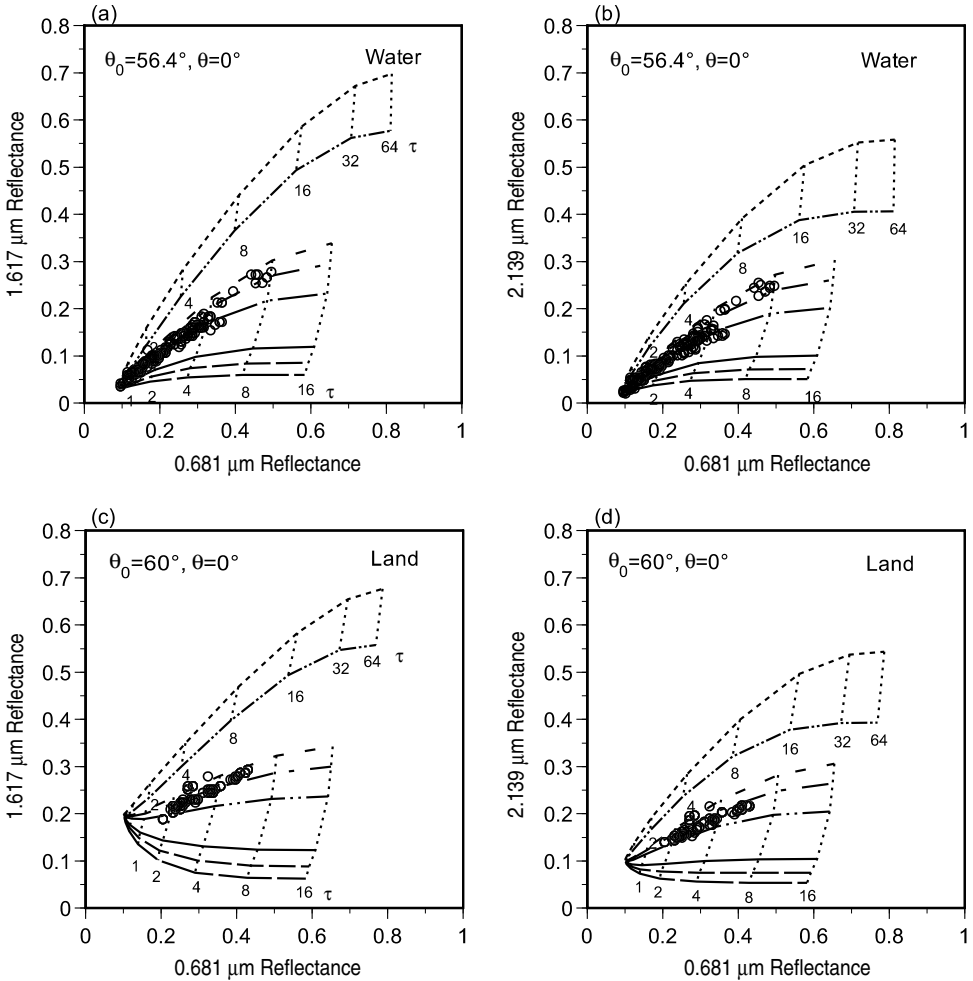
Figure 7.10   Correlation between the reflectances at $0.681/1.617$ μm and at $0.681/2.139$ μm. Radiative transfer calculations were performed for the six ice-crystal size distributions with mean effective sizes ranging from 24 to 124 μm, and for two water clouds with mean effective radii of 4 and 8 μm. The optical depth ranges from 0.5 to 64. Overlapped with the curves are the MAS data obtained from FIRE-II-IFO on 5 December 1991, over the northern Gulf of Mexico and eastern Oklahoma. Frames (a) and (b) are cases over water, and frames (c) and (d) are cases over land.

0.681–2.139 μm for optical depths ranging from 0.5 to 64. For both cirrus and water clouds, the 0.681 μm reflectance mainly depends on the optical depth, whereas the 1.617 and 2.139 μm reflectances are primarily functions of the mean effective particle size. A clear distinction is seen between the correlations for water clouds (mean effective radius < 8 μm) and for cirrus clouds (mean effective size >20 μm). It is possible that the correlations for larger water-droplet mean effective radius and smaller ice-crystal mean effective size may overlap. However, based on the statistics of aircraft observations compiled by Liou (1992), the mean droplet radius for various water clouds is within the 3.5–5.0 μm range and the spectrum of ice-crystal size distribution is generally between 20 and 2000 μm. Thus, it is expected that the probability of occurrence of the correlation overlap for water and ice clouds is very small. Even if the water droplets and ice crystals are about the same size, they can be distinguished by inferring the cloud phase and cloud temperature from separate means.

Also shown in Fig. 7.10 are the MAS data obtained from FIRE-II-IFO on 5 December 1991. The top and bottom diagrams correspond to the cases over water and land surfaces, respectively. The case over water was taken at 1636 UTC, 5 December 1991, when the ER-2 aircraft was flying over the northern part of the Gulf of Mexico near the southern coastal region of Louisiana. The case over land was taken at 1923 UTC on the same date, when ER-2 was flying over eastern Oklahoma. In the calculations, the effective surface albedos used were determined from the MAS reflectances over clear pixels. The data points indicate that the detected cirrus clouds appear to contain small ice particles with optical depths less than about 6. Larger optical depths indicate the possibility of cirrus overlying low clouds. For each data point, an optical depth and a mean effective ice-crystal size can be determined based on Eq. (7.3.14).

7.3.5.2   POLARIZATION

In Sections 5.3 and 5.4, we presented the single-scattering characteristics of spherical droplets and various types of ice crystals. Polarization of the sunlight reflected from clouds appears to show a strong imprint of the cloud thermodynamic phase (spherical water droplets and nonspherical ice crystals) as well as ice crystal shape.

Perhaps the most intriguing results in connection with the use of polarization data for the determination of particle size and optical characteristics have been found in the study of the cloud deck of Venus. Although Venus is the nearest planet, it is the most mysterious as it is surrounded by a veil of clouds. Polarization observations of Venus have been made by the French astronomer Lyot (1929) using visible light. Hansen and Hovenier (1974) have performed an extensive investigation of the particle shape, size, and refractive index of the Venus cloud deck by comparing the observed linear polarization with comprehensive multiple scattering computations, including Lorenz–Mie particles and Rayleigh molecules. Shown in Fig. 7.11 are observations and theoretical computations of the linear polarization of visible sunlight reflected by Venus. After varying the size parameter and refractive index, the best fit to the observed data is given by the dashed-dot curve. The maximum at the phase angle
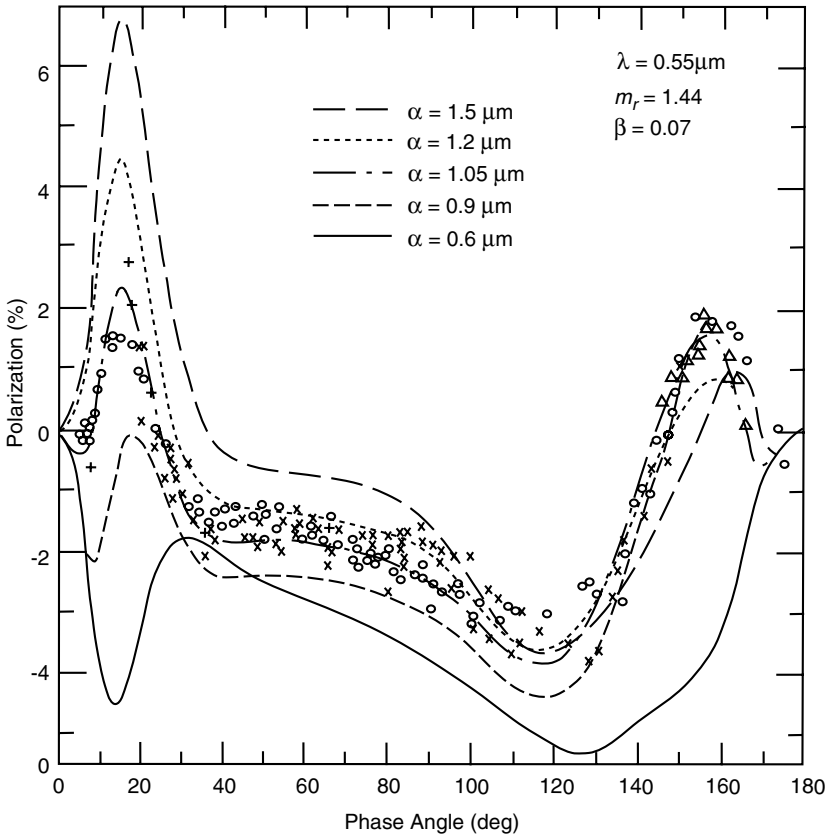
Figure 7.11   Observations ($\circ$, $\times$, $+$, Δ) of the polarization of sunlight reflected by Venus as a function of the phase angle ($180° - \Theta$) in the visual wavelength region and theoretical computations for $\lambda = 0.55\,\mu m$. α denotes the mean radius, β is an effective dispersion of the size distribution, and m_r is the refractive index (after Hansen and Hovenier, 1974).

of about $20°$ (scattering angle $160°$) is the primary rainbow, the product of light rays undergoing one internal reflection, which nonspherical particles do not show, as illustrated in Sections 5.3 and 5.4. Based on this study, it was concluded that the Venus cloud layer was composed of spherical particles having a mean radius of about $1.05\,\mu m$ and an effective dispersion of 0.07. The refractive index of the particles is about 1.44 at a wavelength of $0.55\,\mu m$ with a normal dispersion.

Photometric interpretations may also be carried out to understand the physical, optical, and chemical properties of clouds and/or hazes of other planets. Mariner spacecraft studies of Mars reveal that clouds of H_2O ice as well as possible CO_2 ice haze are present in the Martian atmosphere. Photometric and scattering techniques can provide significant data for understanding the physical and chemical composition of NH_3 clouds in the Jovian atmosphere and the nature of Saturn's rings. Particles that

occur in Mars, Jupiter, and Saturn are likely to be nonspherical. As a consequence of their nonsphericity and the associated orientation problems, reliable single-scattering information must be developed to allow multiple scattering calculations for the purposes of interpretation of observed data.

Further, Takano and Liou (1989b) and Liou and Takano (1994) have used the single-scattering results for hexagonal plates and columns, and later a combination of regular and irregular ice particles, as shown in the upper panel of Fig. 7.12, to interpret the measured linear polarization pattern of sunlight reflected from optically thick cirrus clouds presented by Coffeen (1979). More recently, polarization of sunlight reflected from cirrus clouds has been reported by Chepfer *et al.* (1998) during the European Cloud Radiation Experiment 1994 campaign based on airborne polarimeter measurements from the Polarization and Directionality of the Earth's Reflectances (POLDER) instrument. Two channels in this instrument at 0.443 and 0.864 μm were used for polarization measurements. The lower panel of Fig. 7.12 shows the observed polarization defined by $PO = (Q^2 + U^2 + V^2)^{1/2}$ for two cirrus cloud episodes. In the left and right diagrams, the scattering angle ranges are $75°$ to $165°$ and $90°$ to $180°$, respectively, with the respective solar zenith angles depicted in the diagrams. After extensive trial- and-error analyses, we find that the results employing an optical depth of about 3 match the observed data in the left diagram most closely. Radiative transfer calculations are subsequently performed using randomly oriented hollow columns, plates, and irregular ice particles. The plate case ($L/2a = 8/80$ μm) appears to fit the observations in all scattering angle ranges, as displayed in the left diagram, except in the backscattering range in which the irregular ice particle case fits better. The polarization data presented in the right lower panel illustrates a peak at about $104°$ (subsun feature) associated with the horizontal orientation of ice plates and columns. To provide adequate interpretation, all 16 phase matrix elements are needed in radiative transfer calculations. Following the approach presented in Section 6.7.1 for polarized radiative transfer in horizontally oriented ice particles, 0.1% of hollow Parry columns (horizontal random orientation in a fixed direction) was added to a combination of 60% rough-surface ice particles, 30% hollow columns, and 10% plates randomly oriented in three-dimensional space to provide the best interpretation of the observed data.

7.3.5.3   REFLECTED LINE SPECTRUM

The signature of high-level cirrus appears in the bidirectional reflectance associated with a number of water-vapor absorption bands, as shown in Fig. 7.8, particularly those containing small ice crystals. Among these, the 1.38 μm band appears to be most useful for the detection of thin cirrus, because of the combination of the moderate strength of this absorption band and the amount of solar energy residing within the band (Gao and Kaufman, 1995). The spectral lines of sunlight reflected from other planets have been observed and utilized to determine the composition of planetary atmospheres (Goody and Yung, 1989). Such observations have not been made from satellites for the earth, however. Based on the computational results from a line-by-line equivalent radiative transfer program, we study the information content of
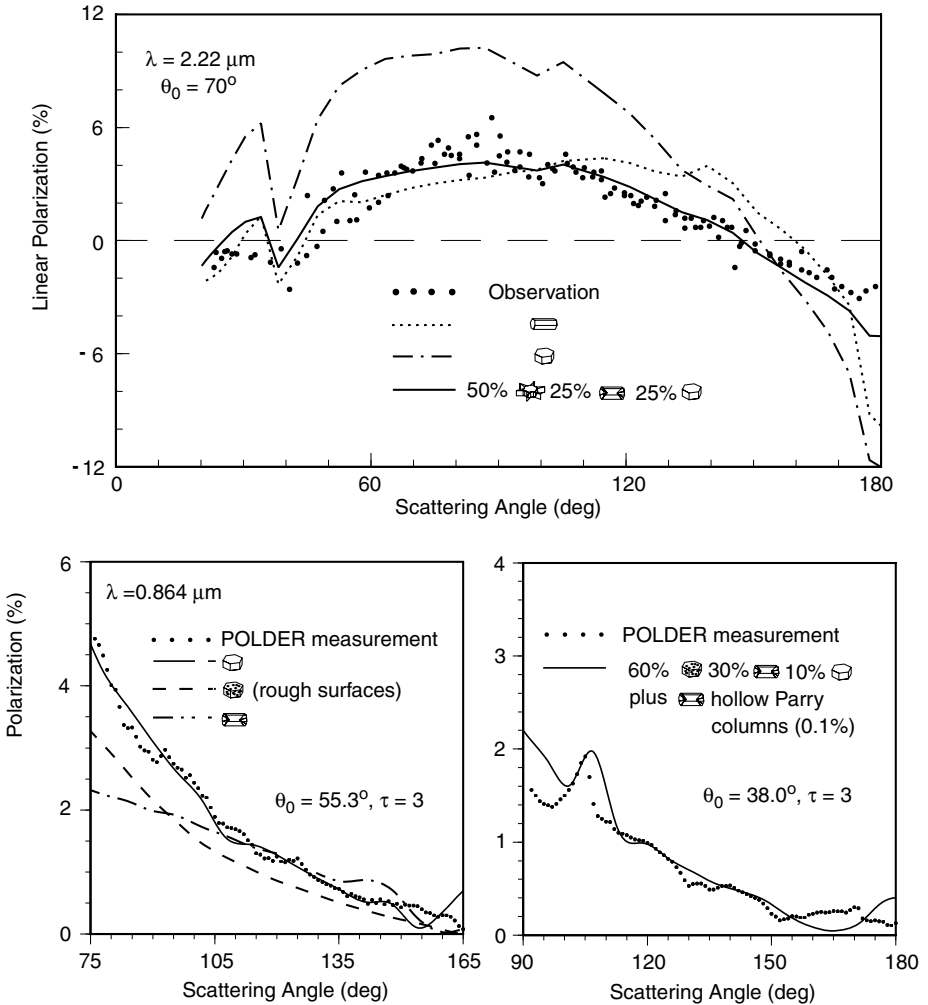
Figure 7.12   The upper panel shows linear polarization of sunlight reflected from cirrus clouds measured at 2.22 μm (Coffeen, 1979). The lower pannel displays full polarization observed from the polarimeter at 0.864 μm on board the POLDER in the scattering-angle domain (Chepfer *et al.,*1998). The theoretical results are computed for hollow columns, plates, and ice particles with rough surfaces using the best fit optical depth of 3. The lower left panel illustrates comparisons between observed values with theoretical results using 3D randomly oriented plates, rough-surface ice particles, and hollow columns. The lower right panel shows a best interpretation of the observed data, particularly for subsun peak at about the 104° scattering angle, using a combination of 0.1% hollow Parry columns and other 3D randomly oriented ice crystals indicated in the diagram.

bidirectional reflectances in the 1.38 μm H_2O band covering 6600–7500 cm^{-1}. The calculations employ two ice-crystal size distributions having mean effective ice crystal sizes of 42 and 124 μm with a shape composition of 50% aggregates/bullet rosettes, 30% hollow columns, and 20% plates.

Figure 7.13 shows the results for clear and cirrus cloudy conditions in which the midlatitude summer atmosphere water vapor profile is used. The cirrus cloud base is placed at 8 km with an optical depth of 1. The line structure of water vapor absorption
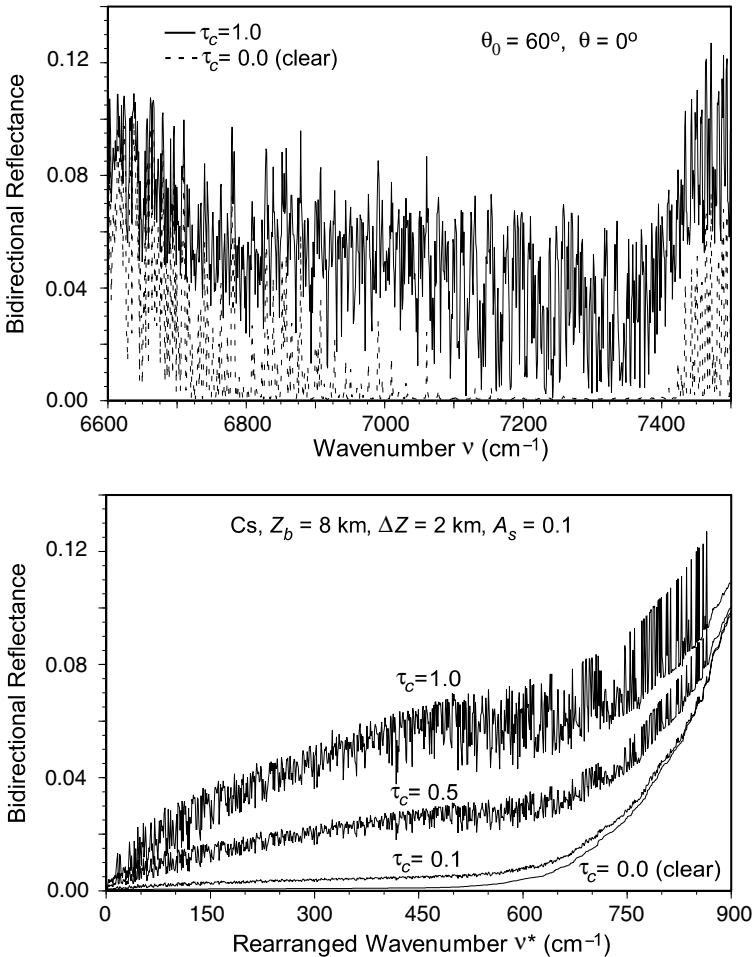
Figure 7.13   The upper diagram shows the bidirectional reflectance in the 1.38 μm water vapor band as a function of wavenumber from 6600 to 7500 cm^{-1} for clear and cirrus cloudy atmospheres. A cirrostratus located at 8 km with a thickness of 2 km in the midlatitude summer atmosphere is used in the calculation. The surface albedo and the solar and emergent angles used are 0.1 and 60° and 0°, respectively. The lower diagram presents the bidirectional reflectance in the rearranged wavenumber domain according to the order of the clear reflectance values, a monotonic increasing curve, for a number of optical depths.

exhibits significant fluctuations. At about 7100–7400 cm^{-1}, reflectances in clear conditions are extremely small because of strong water vapor absorption. Multiple scattering due to ice crystals contributes to the strength of reflectances in the line-wing regions. It is difficult to examine the full information content of cirrus clouds because of the complexity of the line structure. For this reason, we order the bidirectional line spectra for the clear condition according to their magnitudes in such a manner that a monotonic increasing function is displayed in the rearranged wavenumber domain. Subsequently, the bidirectional line spectra for cirrus cloudy conditions are also ordered in accord with this rearranged wavenumber domain. Low values represent reflectances associated with line centers, whereas high values correspond to line wings.

In the lower panel, the largest reflectance in the line wing corresponding to the surface albedo used is about 0.1 for the clear case. For the small optical depth of 0.1, reflectance increases in the center of the water vapor absorption lines produced by the scattering contribution of ice crystals. The pronounced scattering events that occur at an optical depth of 1 are due to the fact that the scattering and absorption coefficients for ice do not align with the absorption coefficients for water vapor. Sufficient information with respect to the cloud height and ice crystal size, in addition to the optical depth, appears to be present in the rearranged spectrum.

Because of the relative strength of water vapor lines in the 1.38 μm band, it appears that some aspect of the vertical profile of cirrus clouds may be inferred from the spectral bidirectional reflectance measurements. Consider a mixture of ice crystals and water vapor in a predivided layer, Δz. The single-scattering albedo for this layer may be expressed by

$$\tilde{\omega}(\nu) = \frac{\sigma_s}{\sigma_e + k(\nu)} \cong \frac{\tilde{\omega}_c}{1 + 2k(\nu)/\pi D_e^2}, \tag{7.3.16a}$$

where σ_e and σ_s are the extinction and scattering cross sections for ice particles; $k(\nu)$ is the absorption coefficient in units of cross section; and $\tilde{\omega}_c = \sigma_s/\sigma_e$ is the single-scattering albedo for ice particles where the optical theorem is used to obtain $\sigma_e \cong 2 \cdot \pi (D_e/2)^2$. We may also employ the parameterization equation developed for optical depth denoted in Eq. (7.3.15c) in the form

$$\Delta\tau(\nu) \cong [\text{IWC}(c + b/D_e) + \rho_w k(\nu)] \cdot \Delta z, \tag{7.3.16b}$$

where ρ_w denotes the number density for water vapor. In these two equations all variables are dependent on the predivided layer.

Consider further an optically thin layer such that the reflectance (reflection function) and transmittance (transmission function) are given in Eqs. (6.4.8a) and (6.4.8b). The spectral reflectance can then be written as follows:

$$R(\nu) = \frac{1}{4\mu\mu_0} \frac{P\tilde{\omega}_c(D_e)}{1 + 2k(\nu)/\pi D_e^2}[\text{IWC}(c + b/D_e) + \rho_w k(\nu)] \cdot \Delta z. \tag{7.3.17a}$$

A similar expression can also be written for the transmittance (direct + diffuse). For the retrieval of IWC and D_e from the reflectance measurements to be feasible, the term $P\tilde{\omega}_c$ must be expressed in terms of D_e. We may now consider a single two-layer

system such that the combined reflectance via Eq. (6.4.9a) is defined by

$$R_{12}(v) = R_1(v) + \frac{\widetilde{T}_1^*(v)R_2(v)\widetilde{T}_1(v)}{1 - R_1^*(v)R_2(v)}. \qquad (7.3.17b)$$

For an optically thin layer, the reflectance and transmittance for radiation from above and below may be assumed to be the same, i.e., $\widetilde{T}_1^* = \widetilde{T}_1$ and $R_1^* = R_1$. Equation (7.3.17b) contains four unknowns: $D_e(1, 2)$ and IWC(1, 2). Thus, four or more spectral reflectance measurements may be selected to perform retrievals. Direct inversion is obviously not feasible, but a numerical solution could be developed. The adding method for reflectance may be extended to a number of layers such that a vertical profile for D_e and IWC may be inferred from a set of carefully selected spectral measurements. The preceding analysis suffices to point out the rich information inherent in the water vapor line spectra reflected from cloudy atmospheres, a subject that has not been explored for the remote sensing of cloud vertical structure.

In Section 3.2.3, we discussed the characteristics of the oxygen absorption bands. The oxygen A-band is produced by the transitions of ground electronic states accompanied by vibrational–rotational transitions and is centered at 13,121 cm^{-1} (red band). The optical depths of the A-band are determined by molecular oxygen (or air) density and therefore are known quantities. The absorption of solar irradiance by the oxygen A-band is similar to but not exactly the same as the 1.38 μm water vapor bands. The A-band radiometer sees the earth's atmosphere principally in the upper troposphere and low stratosphere and, can therefore, be used to infer the position of high-level cirrus clouds, a research area of considerable interest.

7.4   Remote Sensing Using Emitted Infrared Radiation

7.4.1   Theoretical Foundation

We shall first present the theoretical foundation of infrared remote sensing in the context of satellite applications. In a nonscattering atmosphere that is in local thermodynamic equilibrium, the basic equation that governs the transfer of emitted thermal infrared (IR) radiance at a given wavenumber, v, can be described by

$$\mu \frac{dI_v(\tau, \mu)}{d\tau} = I_v(\tau, \mu) - B_v(\tau), \qquad (7.4.1)$$

where $\mu = \cos\theta$, θ is the emergent angle, B_v is the Planck function, and the optical depth is defined by

$$\tau = \int_z^{z_\infty} k_v(z')\rho_a(z')\,dz', \qquad (7.4.2)$$

with ρ_a the density of the absorbing gases and k_v the absorption coefficient. In thermal IR radiative transfer, radiance in the wavenumber domain is used instead of intensity for reflected and transmitted sunlight in the wavelength domain.

From the basic equation, the solution for upward radiance is given by the following integral equation:

$$I_\nu(\tau, \mu) = I_\nu(\tau_*)e^{-(\tau_*-\tau)/\mu} + \int_\tau^{\tau_*} B_\nu(\tau')e^{-(\tau'-\tau)/\mu}\frac{d\tau'}{\mu}, \qquad (7.4.3)$$

where τ_* is the optical depth at the surface and $I_\nu(\tau_*)$ denotes the emitted surface radiance generally assumed to be isotropic. The first term on the right-hand side represents the surface emission contribution attenuated to the level τ, while the second term denotes the emission contribution of the atmosphere between τ and τ_*.

For application to satellite remote sensing, it suffices to assume that the satellite instrument observes in a narrow cone in the local vertical so that everywhere within the cone the cosine of the emergent angle $\mu \cong 1$, referred to as the upwelling direction. The emitted radiance at the surface $I_\nu(\tau_*) = \varepsilon_\nu B_\nu(T_s)$, where ε_ν is the surface emissivity and T_s is the surface temperature. The emissivities of most of the earth's surfaces in the thermal IR region are close to 1. Thus, for all practical purposes, we may use the Planck function for the emitted radiance at the surface.

Moreover, in remote sensing the exponential terms are generally expressed in terms of the transmittance and weighting function defined in the following. The monochromatic transmittance is defined by

$$T_\nu(\tau) = e^{-\tau}, \qquad (7.4.4)$$

and the weighting function by

$$\frac{\partial T_\nu(\tau)}{\partial \tau} = -e^{-\tau}. \qquad (7.4.5)$$

It follows that at the top of the atmosphere (TOA), we have

$$I_\nu(0) = B_\nu(T_s)T_\nu(\tau_*) + \int_{\tau_*}^0 B_\nu(\tau)\frac{\partial T_\nu(\tau)}{\partial \tau}\, d\tau. \qquad (7.4.6)$$

For application to atmospheric remote sensing, the height or pressure coordinate is usually employed. Height and pressure are related via the hydrostatic equation, $dp = -\rho g dz$, where ρ is the air density and g is the gravitational acceleration. The mixing ratio for a specific gas with density ρ_a is defined as $q = \rho_a/\rho$. Thus, we can rewrite Eq. (7.4.6) in the pressure coordinates as follows:

$$I_\nu(0) = B_\nu(T_s)T_\nu(p_s) + \int_{p_s}^0 B_\nu[T(p)]\frac{\partial T_\nu(p)}{\partial p}\, dp, \qquad (7.4.7)$$

where p_s denotes the surface pressure.

An instrument can distinguish only a finite band width $\Psi(\bar{\nu}, \nu)$, where Ψ denotes the instrumental response (or slit) function and $\bar{\nu}$ is the mean wavenumber. The measured radiance from a spectrometer over a wavenumber interval (ν_1, ν_2) in the normalization form is given by

$$I_{\bar{\nu}} = \int_{\nu_1}^{\nu_2} I_\nu\Psi(\bar{\nu}, \nu)\, d\nu \bigg/ \int_{\nu_1}^{\nu_2} \Psi(\bar{\nu}, \nu)\, d\nu. \qquad (7.4.8)$$

The effective spectral interval for the response function is usually small enough that the variation of the Planck function is insignificant. We can then replace its value by $B_{\bar{\nu}}(T)$ without introducing noticeable errors. It follows that by carrying out the wavenumber integration over Eq. (7.4.7), we obtain

$$I_{\bar{\nu}}(0) = B_{\bar{\nu}}(T_s)T_{\bar{\nu}}(p_s) + \int_{p_s}^{0} B_{\bar{\nu}}[T(p)]\frac{\partial T_{\bar{\nu}}(p)}{\partial p}\, dp, \qquad (7.4.9)$$

where, when the instrumental response function is accounted for, the spectral transmittance is defined by

$$T_{\bar{\nu}}(p) = \int_{\nu_1}^{\nu_2} \Psi(\bar{\nu}, \nu)\exp\left[-\frac{1}{g}\int_0^p k_\nu(p')q(p')\, dp'\right]d\nu \Big/ \int_{\nu_1}^{\nu_2} \Psi(\bar{\nu}, \nu)\, d\nu. \qquad (7.4.10)$$

Equation (7.4.9) is fundamental to remote sensing of the atmosphere and the surface from orbiting meteorological satellites. The upwelling radiance is a result of the product of the Planck function, the spectral transmittance, and the weighting function. The temperature information is included in the Planck function, while the density profiles of relevant absorbing gases are involved in the transmittance. Observed radiances will thus be directly or indirectly associated with the temperature and gaseous profiles. Extracting the relevant information about the atmospheric state and composition from observed IR radiances is the essence of remote sensing from space.

We shall now review the potential information content in the thermal IR spectrum (Fig. 4.3). There are four regions over which water vapor, ozone, and carbon dioxide exhibit a significant absorption spectrum. Carbon dioxide absorbs IR radiation in the 15 μm band from about 600 to 800 cm^{-1}. In addition, carbon dioxide also absorbs radiation in the 4.3 μm region that overlaps with solar radiation. Absorption due to ozone is primarily confined to the 9.6 μm band. Water vapor exhibits absorption lines over the entire infrared spectrum. The most pronounced absorption occurs in the 6.3 μm vibrational–rotational band and in the pure rotational band with wavenumbers less than about 500 cm^{-1}. From about 800 to 1200 cm^{-1}, referred to as the atmospheric window, absorption due to atmospheric gases shows a minimum, except in the 9.6 μm ozone band. There are also absorption bands for various greenhouse gases that can be used for their determination by remote sensing: the CH$_4$ 7.6 μm band, the N$_2$O 7.9 μm band, and some CFC lines in the window.

7.4.2   Surface Temperature Determination

If observations are taken in the window region where the effect of the atmosphere is at a minimum, the upwelling radiance at TOA must be closely associated with emission from the surface. Replacing $\bar{\nu}$ by i for convenience of discussion, we may define a mean temperature for the atmosphere, T_a, and simplify Eq. (7.4.9) in the

form

$$I_i \cong B_i(T_s)T_i + B_i(T_a)(1 - T_i), \qquad (7.4.11)$$

where we let $T_i = T_i(p_s)$.

We now introduce the split-window technique for the determination of surface temperature that uses observations at two channels to eliminate the term involving T_a and solve for T_s. The value of T_a generally varies by less than 1 K in the window region from about 10.5 to 12.5 μm, in which variability of the surface emissivity is insignificant. The atmospheric transmittance in the window region is primarily produced by the continuous absorption of water vapor and to a good approximation is given by

$$T_i \cong e^{-k_i u} = 1 - k_i u, \qquad (7.4.12)$$

where k_i is the absorption coefficient of water vapor for a spectral band in the window and u is the water vapor path length.

Applying the window equation to two channels, we have

$$I_1 = B_1(T_{b1}) = B_1(T_s)T_1 + B_1(T_a)(1 - T_1), \qquad (7.4.13)$$

$$I_2 = B_2(T_{b2}) = B_2(T_s)T_2 + B_2(T_a)(1 - T_2), \qquad (7.4.14)$$

where we have also expressed the observed radiances I_1 and I_2 in terms of the brightness temperatures T_{b1} and T_{b2}. The objective of the split-window technique is to eliminate T_a, and to do so, we may expand the Planck function of temperature T by means of the Taylor series with respect to T_a in the form

$$B_i(T) = B_i(T_a) + \frac{\partial B_i}{\partial T}(T - T_a). \qquad (7.4.15)$$

Applying Eq. (7.4.15) to the two channels with $i = 1, 2$ and eliminating $(T - T_a)$ yields

$$B_2(T) = B_2(T_a) + \frac{\partial B_2}{\partial T} \Big/ \frac{\partial B_1}{\partial T}[B_1(T) - B_1(T_a)]. \qquad (7.4.16)$$

Further, by replacing T by T_{b2} and T_s in Eq. (7.4.16) and utilizing Eq. (7.4.14), after some analysis we obtain the following:

$$B_1(T_{b2}) = B_1(T_s)T_2 + B_1(T_a)(1 - T_2). \qquad (7.4.17)$$

By eliminating $B_1(T_a)$ from Eqs. (7.4.17) and (7.4.13), we obtain the split-window equation in the form

$$B_1(T_s) = B_1(T_{b1}) + \eta[B_1(T_{b1}) - B_1(T_{b2})], \qquad (7.4.18)$$

where

$$\eta = \frac{1 - T_1}{T_1 - T_2} \cong \frac{k_1}{k_2 - k_1}. \qquad (7.4.19)$$

In practice, the Planck function is replaced by the brightness temperature. Since a local linear relation can be established between the two, particularly in a small spectral interval in the window region, we may write

$$T_s \cong T_{b1} + \eta(T_{b1} - T_{b2}). \tag{7.4.20}$$

Sea surface temperatures (SSTs) have been routinely inferred from the operational NOAA satellites using one of the instruments aboard, the AVHRR discussed in Section 7.3.3. The 10.9 and 12.0 μm channels have been used to determine surface temperature, particularly over oceans. At night, the 3.7 μm channel has also been added to increase the accuracy of retrieval. Based on the split-window technique, surface temperature may be expressed by a general form of regression as follows:

$$SST = aT_{b1} + b(T_{b1} - T_{b2}) - c, \tag{7.4.21}$$

where a, b, and c are empirical coefficents derived from *in situ* observations obtained from drifting buoys (McClain *et al.,* 1985), and T_{b1} and T_{b2} are the brightness temperatures involving a combination of AVHRR 10.9, 12.0, and 3.7 μm channels. The satellite-derived temperatures correspond to the temperature of a surface (skin temperature), whereas the buoy measurements are associated with a layer of water some meters deep. The regression approach of the multichannel technique is used partly to adjust (or tune) the satellite skin temperatures to the *in situ* bulk temperatures and partly to account for the water vapor absorption in the window, particularly in the moist tropical region. Since 1970, global SSTs have been operationally produced and archived. An important part of SST retrieval is the detection and elimination of clouds. Several threshold methods are employed for these purposes based on the bidirectional reflectance of the solar channel in daytime and the emission characteristics of IR channels with respect to the sea surface uniformity in nighttime. The split-window technique has also been used for the determination of surface temperature over land in conjunction with land–atmosphere interaction studies. (Brutsaert *et al.,* 1993).

7.4.3   Remote Sensing of Temperature Profiles

Inference of atmospheric temperature profiles from satellite observations of thermal infrared emission was first suggested by King (1956). In his pioneering paper, King pointed out that the angular radiance (intensity) distribution is the Laplace transform of the Planck intensity distribution as a function of optical depth, and illustrated the feasibility of deriving the temperature profile from satellite intensity scan measurements.

Kaplan (1959) advanced the sounding concept by demonstrating that vertical resolution of the temperature field can be inferred from the spectral distribution of atmospheric emission. Kaplan pointed out that observations in the wings of a spectral band sense deeper into the atmosphere, whereas observations in the band center see only the very top layer of the atmosphere since the radiation mean free path is small. Thus, by properly selecting a set of different sounding wavenumbers, the observed

radiances can be used to make an interpretation leading to the vertical temperature distribution in the atmosphere. This principle is demonstrated in the following.

From Eq. (7.4.9), it is clear that the temperature of the underlying surface must be determined if surface emission represents a significant contribution to the observed radiance. For the following discussion, we may drop the surface term, i.e., assume $T_{\bar{\nu}}(p_s) = 0$, and write the forward radiative transfer equation in the form

$$I_{\bar{\nu}} = \int_{p_s}^{0} B_{\bar{\nu}}[T(p)] \frac{\partial T_{\bar{\nu}}(p)}{\partial p} \, dp. \tag{7.4.22}$$

The fundamental problem in temperature profile retrieval is how to solve for the function $B_{\bar{\nu}}[T(p)]$, given a set of observed radiances corresponding to different wavenumbers and the known weighting functions. Since the Planck function is a function of wavenumber, we must first eliminate this dependence in the retrieval. The Planck function is smooth and in a small spectral interval it can be approximated in a linear form as follows:

$$B_{\bar{\nu}}(T) = c_{\bar{\nu}} B_{\bar{\nu}_r}(T) + d_{\bar{\nu}}, \tag{7.4.23}$$

where $\bar{\nu}_r$ denotes a fixed reference wavenumber and $c_{\bar{\nu}}$ and $d_{\bar{\nu}}$ are fitting coefficients. It follows that Eq. (7.4.22) can be expressed by

$$g_{\bar{\nu}} = \int_{p_s}^{0} f(p) K_{\bar{\nu}}(p) \, dp, \tag{7.4.24a}$$

where we let

$$g_{\bar{\nu}} = (I_{\bar{\nu}} - d_{\bar{\nu}})/c_{\bar{\nu}}, \quad f(p) = B_{\bar{\nu}_r}[T(p)], \quad K_{\bar{\nu}}(p) = \partial T_{\bar{\nu}}(p)/\partial p. \tag{7.4.24b}$$

Equation (7.4.24a), similar to Eq. (7.2.10), is the well-known Fredholm equation of the first kind; $K_{\bar{\nu}}(p)$ is the weighting function; and $f(p)$ is the function to be recovered from a set of $g_{\bar{\nu}}$.

To determine atmospheric temperatures from measurements of thermal emission, the source of emission must be a relatively abundant gas of known and uniform distribution. Otherwise, its uncertainty will make the determination of temperature from the measurements ambiguous. There are two gases in the earth–atmosphere system that occur in uniform abundance at altitudes below about 100 km, and which also show emission bands in the spectral regions that are convenient for measurement. As discussed in Section 4.2, carbon dioxide, a minor constituent with a relative volume abundance of about 365 ppm at the present time, exhibits vibrational–rotational lines. In addition, oxygen, a major constituent with a relative volume abundance of 0.21, also satisfies the requirement of a uniform mixing ratio and has a microwave spin-rotational band. The microwave spectrum will be discussed in the next section.

Shown in Fig. 7.14 is a spectrum of outgoing radiance in terms of blackbody temperature in the vicinity of the 15 μm band observed by the Infrared Interfero-meter Spectrometer (IRIS) on board the Nimbus 4 satellite. The equivalent blackbody temperature generally decreases when approaching the band center. This decrease is associated with the decrease of tropospheric temperature as altitude increases. Near
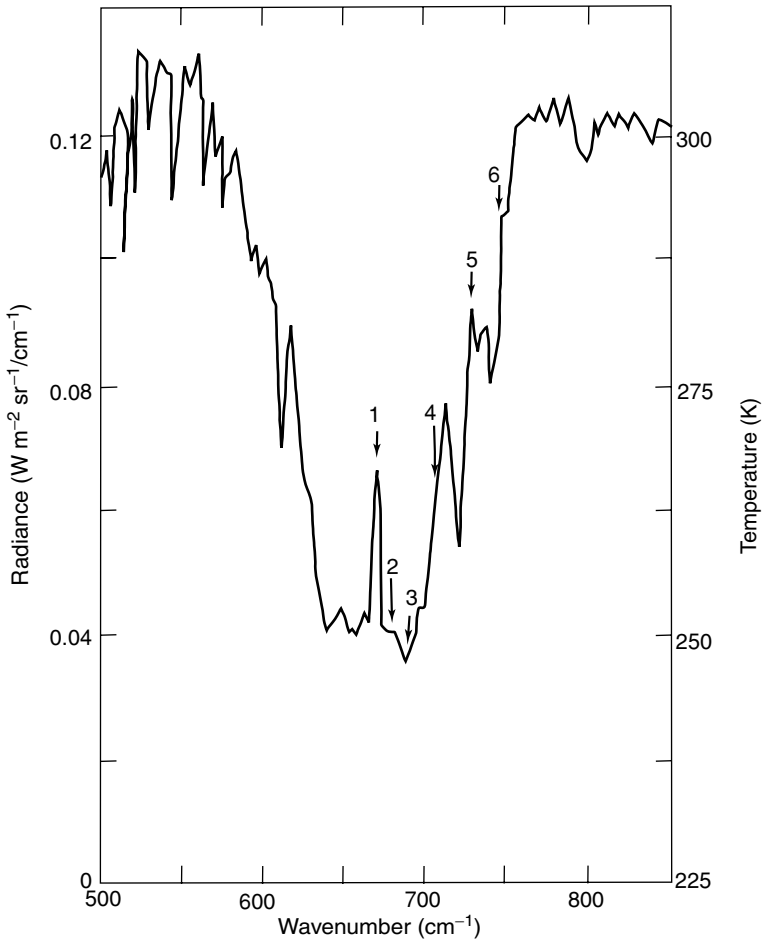
Figure 7.14 Outgoing radiance in terms of blackbody temperature in the vicinity of 15 μm CO_2 band observed by the IRIS on Nimbus 4. The arrows denote the spectral regions sampled by the VTPR instrument (see also Fig. 4.1).

about 690 cm^{-1}, the temperature shows a minimum, which is related to the colder tropopause. Decreasing the wavenumber beyond 690 cm^{-1}, however, increases the temperature. This is due to the increase in temperature in the stratosphere, since the observations near the band center see only the very top layers of the atmosphere. On the basis of the sounding principle discussed above, we can select a set of sounding wavenumbers such that temperature profiles in the troposphere and lower stratosphere can be largely covered. The arrows in Fig. 7.14 indicate an example of such a selection.

The selection of appropriate sounding wavenumbers for temperature retrieval requires an understanding of the behavior of weighting functions. In practice, an ideal weighting function would be a delta function. Weighting functions can be constructed

using known line-by-line absorption data and a program for transmittance calculations via Eq. (7.4.10). As discussed in Section 1.3, the absorption coefficient in the troposphere is slightly dependent on temperature, which can introduce complications in the retrieval of temperature profiles. However, the temperature dependence of transmittances can be accounted for in the retrieval process by building a set of values corresponding to a number of standard atmospheric profiles from which a search can be carried out to obtain the best estimated transmittances for a given temperature profile.

The computation of spectral transmittances through an inhomogeneous atmosphere for satellite remote sensing applications first requires a line-by-line integration effort, prior to the development of simplified or parameterized programs to speed up the computational requirement. In Section 4.2.3, we outlined the procedure for line-by-line calculations in which line shape, strength, and position, along with pressure and temperature dependence, are properly taken into account.

Sounding of the atmospheric temperature for meteorological purposes utilizes the 15 μm and 4.3 μm CO_2 bands. The latter is used at nighttime to enhance temperature retrieval. As discussed in Section 4.2, the 15 μm CO_2 band consists of a number of bands produced by vibrational transitions including the ν_2 fundamentals, combination bands, hot bands, and bands associated with isotopes, with strong Q-branch rotational lines located at the center of the band and P- and R-branch lines almost equally spaced on each side of the band center.

An example of the weighting function and transmittance profiles is shown in Fig. 7.15 for a set of Vertical Temperature Profile Radiometers (VTPR) on the
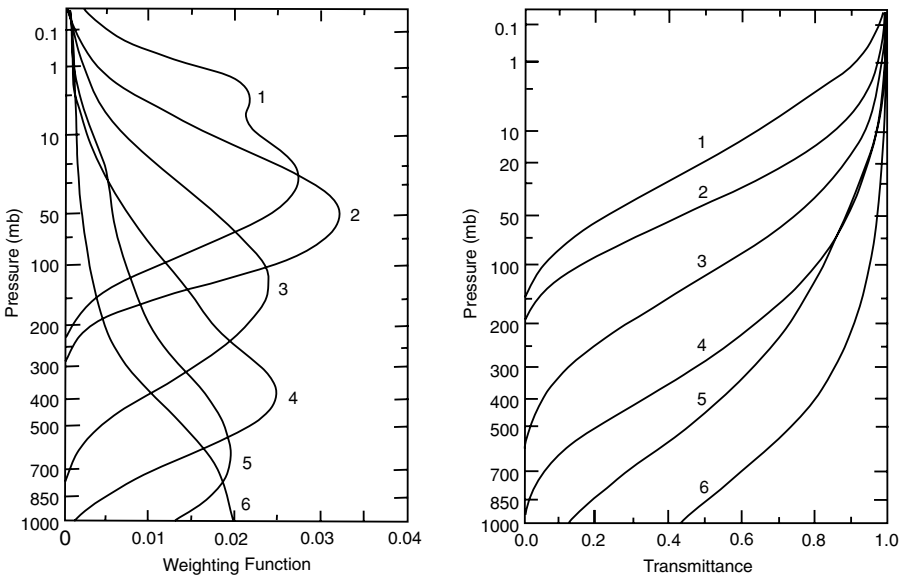
Figure 7.15   The weighting function and transmittance for the NOAA 2 Vertical Temperature Profile Radiometer (VTPR).

NOAA 2 satellite, the first satellite experiment to measure atmospheric temperatures for operational meteorological use. The VTPR consists of 6 channels in the $15 \mu m$ CO_2 band, with the nominal center wavenumbers of 668.5, 677.5, 695.0, 708.8, 725.0, and 745.0 cm^{-1} for channels 1–6, respectively. The weighting function curves represent the part of the atmosphere where the upwelling radiance arises. Each peak denotes the maximum contribution to the upwelling radiance. Clearly, radiance is contributed mostly from progressively lower levels as the wavenumber moves from the center to the wing of the band. It is also apparent that the weighting functions overlap somewhat, allowing finite radiance data to define the temperature profile adequately. Operational and practical temperature profile retrievals do not follow the direct inversion methods discussed in Section 7.2.1, but rather employ the nonlinear iteration and physical–statistical methods outlined in the following.

7.4.3.1   NONLINEAR ITERATION METHOD

The difficulty in reconstructing a temperature profile from radiance observations is due to the fact that the Fredholm equation with fixed limits may not always have a solution for an arbitrary function. Because the radiances are derived from measurements that are only approximate, the reduction of the inversion problem to a linear system is mathematically improper and a nonlinear approach to the forward radiative transfer equation appears necessary. Recalling Eq. (7.4.9), we write

$$I_i = B_i(T_s)T_i(p_s) + \int_{p_s}^{0} B_i[T(p)]\frac{\partial T_i(p)}{\partial \ln p} \, d\ln p, \qquad (7.4.25a)$$

where i denotes the number of spectral channels and the weighting function is expressed in logarithmic scale of the pressure. The Planck function in the wavenumber domain may be rewritten in the form

$$B_i(T) = a v_i^3 / (e^{bv_i/T} - 1), \qquad (7.4.25b)$$

with $a = 2hc^2$ and $b = hc/K$. In reference to the weighting function illustrated in Fig. 7.15, we note that for a given wavenumber, the integrand reaches a maximum at different pressure levels. From the mean value theorem, the observed radiance can be approximated by

$$\hat{I}_i - B_i(T_s)T_i(p_s) \cong B_i[T(p_i)]\left[\frac{\partial T_i(p)}{\partial \ln p}\right]_{p_i} \Delta_i \ln p, \qquad (7.4.26)$$

where p_i denotes the pressure level at which the maximum weighting function is located, and $\Delta_i \ln p$ is the pressure difference at the ith level and is defined as the effective width of the weighting function. Let the guessed temperature at the p_i level be $T^*(p_i)$. Then, the expected radiance is given by

$$I_i^* - B_i(T_s)T_i^*(p_s) = B_i[T^s(p_i)]\left[\frac{\partial T_i^*(p)}{\partial \ln p}\right]_{p_i} \Delta_i \ln p. \qquad (7.4.27)$$

On dividing Eq. (7.4.26) by Eq. (7.4.27) and noting that the dependence of the Planck function is much larger than that of the weighting function, we obtain

$$\frac{\hat{I}_i - B_i(T_s)T_i(p_s)}{I_i^* - B_i(T_s)T_i^*(p_s)} \cong \frac{B_i[T(p_i)]}{B_i[T^*(p_i)]}. \tag{7.4.28}$$

When the surface contribution to the upwelling radiance is small or dominant, Eq. (7.4.28) may be approximated by

$$\frac{\hat{I}_i}{I_i^*} \cong \frac{B_i[T(p_i)]}{B_i[T^*(p_i)]}. \tag{7.4.29}$$

This is referred to as the relaxation equation (Chahine, 1970; Smith, 1970). We may now devise iterative procedures to retrieve the temperature profile $T^{(n)}(p_i)$ at level p_i where n is the order of iterations as follows:

(a) Make an initial guess for $T^{(n)}(p_i)$, $n = 0$.
(b) Substitute $T^{(n)}(p_i)$ into Eq. (7.4.25a) and use an accurate quadrature formula to compute the expected upwelling radiance $I_i^{(n)}$ for each sounding channel.
(c) Compare the computed radiance values $I_i^{(n)}$ with the measured data $\hat{I}_i$. If the residuals $R_i^{(n)} = |\hat{I}_i - I_i^n|/\hat{I}_i$ are less than a preset small value (e.g., 10^{-4}) for each sounding channel, then $T^{(n)}(p_i)$ is a solution. If not, continue the iteration.
(d) Apply the relaxation equation (7.4.29) M times to generate a new guess for the temperature values $T^{(n+1)}(p_i)$ at the selected i pressure level, i.e., force the temperature profile to match the observed radiances. From Eqs. (7.4.29) and (7.4.25b), we have

$$T^{(n+1)}(p_i) = b\nu_i/\ln\{1 - [1 - \exp(b\nu_i/T^{(n)}(p_i))]I_i^{(n)}/\hat{I}_i\}, \quad i = 1, 2, \dots, M.$$

(e) Carry out the interpolation between the temperature value at each given level p_i to obtain the desired profile (it suffices to use linear interpolation in the retrieval).
(f) Finally, go back to step (b) and repeat until the residuals are less than a preset criterion.

The relaxation method could, in principle, fail to converge or converge to the wrong solution. The former will be obvious and the solution can be rejected. The latter requires numerical experiments to demonstrate whether the method has converged to a correct solution. Exercise 7.6 is an example of using the relaxation method to derive a temperature profile.

7.4.3.2   MINIMUM VARIANCE METHOD: HYBRID RETRIEVAL

In many remote sensing problems, the errors encountered are statistical. Hence, it is desirable to take into account the statistical nature of the measurement errors and available information when considering the inversion problem. In the statistical approach, it is customary to assume that the deviation of predicted (or estimated) parameters (in the present case the temperature) from the climatological mean can be expressed as a linear combination of the deviation of measured data (in this case the

radiance). Under this assumption, we have

$$\hat{T}_j - \bar{T}_j = \sum_{i=j}^{M} D_{ji}(\hat{I}_i - \bar{I}_i), \qquad j = 1, 2, \ldots, N, \tag{7.4.30}$$

where j denotes the atmospheric level, $\hat{T}_j$ is the predicted value of T_j, which represents the "true" temperature, $\hat{I}_i$ is the observed radiance, $\|D_{ji}\|$ represents a certain predictor matrix, and the climatological means of T_j and I_i (the "true" radiance) are defined by a statistical ensemble, denoted by x, of available temperature profiles and radiances in the forms

$$\bar{T}_j = \frac{1}{X} \sum_{x=1}^{X} T_{jx}, \quad \bar{I}_i = \frac{1}{X} \sum_{x=1}^{X} I_{ix}. \tag{7.4.31}$$

In matrix from, Eq. (7.4.30) can be written

$$\hat{\mathbf{T}} - \bar{\mathbf{T}} = \mathbf{D}(\hat{\mathbf{I}} - \bar{\mathbf{I}}). \tag{7.4.32}$$

Thus, once $\mathbf{D}$ is known the predicted temperature vector can be determined from the observed radiance vector.

We wish to find a linear predictor which will give the minimum square deviation of the predictor profile from the true profile defined by

$$\sum_{x=1}^{X}(T_{jx} - \hat{T}_{jx})^2 = \sum_{x=1}^{X}\left[(T_{jx} - \bar{T}_j) - (\hat{T}_{jx} - \bar{T}_j)\right]^2, \qquad j = 1, 2, \ldots, N. \tag{7.4.33}$$

Using Eq. (7.4.30) and searching the minimum with respect to the linear predictor yield

$$\frac{\partial}{\partial D_{jk}}\left\{ \sum_{x=1}^{X}\left[(T_{jx} - \bar{T}_j) - \sum_{i=j}^{M} D_{ji}(\hat{I}_{ix} - \bar{I}_i)\right]^2\right\} = 0,$$

$$j = 1, 2, \ldots, N, \qquad k = 1, 2, \ldots, M. \tag{7.4.34}$$

It follows that

$$\sum_{x=1}^{X}(T_{jx} - \bar{T}_j)(\hat{I}_{kx} - \bar{I}_k) - \sum_{i=1}^{M} D_{ji}\sum_{x=1}^{X}(\hat{I}_{ix} - \bar{I}_i)(\hat{I}_{kx} - \bar{I}_k) = 0. \tag{7.4.35a}$$

In matrix form, we write

$$\sum_{x=1}^{X}(\mathbf{T}_x - \bar{\mathbf{T}})(\hat{\mathbf{I}}_x - \bar{\mathbf{I}})^{\mathrm{T}} - \mathbf{D}\sum_{x=1}^{X}(\hat{\mathbf{I}}_x - \bar{\mathbf{I}})(\hat{\mathbf{I}}_x - \bar{\mathbf{I}})^{\mathrm{T}} = 0, \tag{7.4.35b}$$

where $\mathbf{T}$ is an $N \times 1$ matrix, $\mathbf{I}$ is an $M \times 1$ matrix, and $\mathbf{D}$ is an $N \times M$ matrix. Further, let the experimental random error vector due to system noise of a radiometer be ε. Thus, the observed radiance vector is then

$$\hat{\mathbf{I}} = \mathbf{I} + \varepsilon, \tag{7.4.36}$$

where $\mathbf{I}$ is the "true" radiance vector mentioned earlier. Substituting Eq. (7.4.36) into Eq. (7.4.35b) and noting that the measurement error vector is uncorrelated with the temperature and radiance vectors, we have

$$\sum_{x=1}^{X}(\mathbf{T}_x - \mathbf{T})(\mathbf{I}_x - \mathbf{I})^{\mathrm{T}} - \mathbf{D}\sum_{x=1}^{X}\left[(\mathbf{I}_x - \mathbf{I})(\mathbf{I}_x - \mathbf{I})^{\mathrm{T}} + \varepsilon\varepsilon^{\mathrm{T}}\right] = 0. \quad (7.4.37)$$

Define the covariance matrix for any two variables as follows:

$$\mathbf{C}(\mathbf{T}_x, \mathbf{I}_x) = \frac{1}{X}\sum_{x=1}^{X}\mathbf{T}_x\mathbf{I}_x^{\mathrm{T}}. \quad (7.4.38)$$

Thus, the predictor matrix can be expressed in terms of the covariance matrix in the form

$$\mathbf{D} = \mathbf{C}(\Delta\mathbf{T}_x, \Delta\mathbf{I}_x)[\mathbf{C}(\Delta\mathbf{I}_x, \Delta\mathbf{I}_x) + \mathbf{C}(\varepsilon_x, \varepsilon_x)]^{-1}, \quad (7.4.39)$$

where $\mathbf{C}(\varepsilon_x, \varepsilon_x)$ represents the noise covariance matrix, and

$$\Delta\mathbf{T}_x = \mathbf{T}_x - \bar{\mathbf{T}}, \quad \Delta\mathbf{I}_x = \mathbf{I}_x - \bar{\mathbf{I}}. \quad (7.4.40)$$

The covariance matrix can be constructed empirically by collecting coincidences of radiances derived from satellites and temperatures obtained from radiosondes or rocket soundings. Normally, the $\mathbf{D}$ matrix is built to give a dimension $N \times M$ such that $N > M$. In this manner, more temperature values can be inferred from a limited set of radiance observations. The $\mathbf{D}$ matrix is thus determined entirely from available data sets and information concerning the weighting function is not required. Although the statistical method does not require a good first-guess profile, acquiring a representative data set is crucial to its success. This data set must be large enough to ensure that the retrieval matrix is stable, must be collected for each satellite, and must be updated frequently to cover different seasons and latitudinal zones. The most advantageous aspect of the statistical method is the computational efficiency afforded by Eq. (7.4.32), which is important for operational purposes.

The minimum variance method utilizes the same approach as the statistical method but does not require a large data set and uses weighting functions similar to physical retrievals. Consider the basic forward radiative transfer equation denoted in Eq. (7.4.22) and write it in the form

$$I_i = \int_{p_s}^{0} B_i[T(p)]K_i(p)\,dp, \quad i = 1, 2, \ldots, M. \quad (7.4.41)$$

Let the climatological (or standard) temperature profile be $\bar{T}(p)$ and let the radiance corresponding to this profile be $\bar{I}$. Expanding the Planck function linearly with respect to this temperature, we have

$$I_i - \bar{I}_i = \int_{p_s}^{0}\frac{\partial B_i[T(p)]}{\partial p}K_i(p)\,dp[T(p) - \bar{T}(p)]. \quad (7.4.42)$$

Using the quadrature summation for the integration and employing the notation defined in Eq. (7.4.40), we have

$$\Delta I_i = A_{ij}\Delta T_j, \tag{7.4.43a}$$

where

$$A_{ij} = \sum_{j=1}^{N}\left\{\frac{\partial B_i[T(p)]}{\partial p}K_i(p)\right\}_{p_j}\Delta p_j. \tag{7.4.43b}$$

In matrix form, we have

$$\mathbf{\Delta I} = \mathbf{A}\,\mathbf{\Delta T}. \tag{7.4.44}$$

Substituting Eq. (7.4.44), which correlates the temperature deviation with the radiance deviation via a matrix **A,** into the predictor matrix derived in Eq. (7.4.39), we have

$$\mathbf{D} = \mathbf{C_T}\mathbf{A}^{\mathrm{T}}\big(\mathbf{AC_T}\,\mathbf{A}^{\mathrm{T}} + \mathbf{C_\varepsilon}\big)^{-1}, \tag{7.4.45a}$$

$$\mathbf{C_T} = \frac{1}{X}\sum_{x=1}^{X}\Delta\mathbf{T}_x\Delta\mathbf{T}_x^{\mathrm{T}}, \tag{7.4.45b}$$

$$\mathbf{C_\varepsilon} = \frac{1}{X}\sum_{x=1}^{X}\varepsilon_x\varepsilon_x^{\mathrm{T}}. \tag{7.4.45c}$$

Equations (7.4.32) and (7.4.45a)–(7.4.45c) constitute the so-called *minimum variance method*. The components in the predictor matrix can be determined separately; **A** can be calculated from the transmittances; $\mathbf{C_T}$ can be determined from a sample of radiosonde soundings; and $\mathbf{C_\varepsilon}$ can be obtained from instrument calibration in the laboratory. In this manner, a large data set of collocated radiosonde and satellite soundings is no longer required.

Although operational retrievals of temperature profiles have been routinely performed over the past 25-plus years, the techniques and procedures used have undergone constant refinement and improvement. This is a result of the complicated nature of the problem, in which consistent and accurate performance is difficult to achieve. It is particularly so when clouds are involved. Radiances from operational satellites, NOAA and GOES, are converted into temperature soundings through an involved algorithm in which retrieval, such as use of the minimum variance method, is only a small part of the process. The NOAA satellites have three sounders: the High Resolution Infrared Radiation Sounder 2 (HIRS/2), the Microwave Sounding Unit (MSU), and the Stratospheric Sounding Unit (SSU). Together they are referred to as the TIROS N Operational Vertical Sounder (TOVS). The TOVS data have been routinely processed to obtain temperature and humidity soundings in connection with numerical weather prediction.
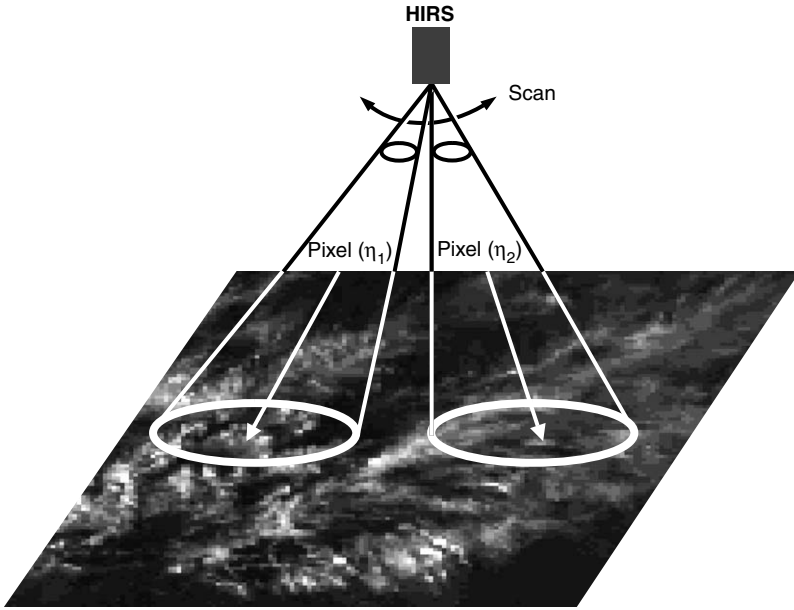
Figure 7.16   Configuration for the adjacent pixels ($\sim$20 km) in a partly cloudy condition for the removal of cloud contribution in the temperature retrieval algorithm using HIRS channels. η_1 and η_2 denote cloud covers in the two pixels. The picture displays AVHRR visible images ($\sim$1 km).

7.4.3.3   CLOUD REMOVAL

In many temperature retrieval cases, the major problem is due to the presence of clouds in the field of view (FOV) of the sounding radiometers. It is, therefore, critically important to remove the cloud effect prior to the temperature retrieval. Consider a pair of adjacent pixels (scan spots), as shown in Fig. 7.16, and let the radiances for the ith channel over clear and cloudy areas be I_i^{clr} and I_i^{cld}, respectively. Then, the radiances for pixels 1 and 2 are given by

$$\hat{I}_{i1} = (1 - \eta_1)I_{i1}^{\mathrm{clr}} + \eta_1\, I_{i1}^{\mathrm{cld}},$$
$$\hat{I}_{i2} = (1 - \eta_2)I_{i2}^{\mathrm{clr}} + \eta_2\, I_{i2}^{\mathrm{cld}}, \tag{7.4.46}$$

where η_1 and η_2 are the effective cloud covers, the product of cloud cover and emissivity, for the respective pixels. If the optical properties of the clouds in the adjacent pixels are the same for the sounding channels, and the temperature fields in the adjacent pixels are also the same, then $I_{i1}^{\mathrm{clr}} = I_{i2}^{\mathrm{clr}} = I_i^{\mathrm{clr}}$, and $I_{i1}^{\mathrm{cld}} = I_{i2}^{\mathrm{cld}} = I_i^{\mathrm{cld}}$. Thus, we obtain

$$N^* = \frac{\eta_1}{\eta_2} = \frac{I_i^{\mathrm{clr}} - \hat{I}_{i1}}{I_i^{\mathrm{clr}} - \hat{I}_{i2}}. \tag{7.4.47}$$

If we know in advance the N^*, which is independent of channels, we can then determine the clear column radiance from the following equation:

$$I_i^{\text{clr}} = \frac{\hat{I}_{i1} - \hat{I}_{i2} N^*}{1 - N^*}.$$ (7.4.48)

The preceding outlines the essence of the N^* method for the removal of cloud contribution in the FOV of the radiometer. It is subject to the following restrictions: (a) the assumption that adjacent clouds are of the same height, temperature, and optical properties; (b) the effective cloud covers η_1 and η_2 must be different; otherwise, $N^* = 1$, and the solution for I_i^{clr} is singularity; and (c) additional information is required to determine I_i^{clr} from the value of N^*.

The conventional approach is to use an independent channel, say M, in the microwave to determine N^* such that

$$N^* = \frac{I_M^{\text{clr}} - \hat{I}_{M1}}{I_M^{\text{clr}} - \hat{I}_{M2}}.$$ (7.4.49)

However, microwave measurements are usually subject to errors caused by poor resolution, contamination by precipitating clouds, and uncertainty in the estimated surface emissivity. As shown in McMillan and Dean (1982), many empirical procedures and threshold tests are required for the processing of TOVS soundings from NOAA satellites to detect and correct for clouds in connection with temperature retrieval.

Figure 7.17 shows an example of the temperature retrieval accuracy from TOVS data with respect to radiosonde soundings during the entire global weather experiment year in 1979 in terms of root-mean-square differences (Smith, 1991). The collocation and time differences between the two data sources are within about 200 km and ± 3 hours. The error differences were due primarily to the poor vertical resolution of the TOVS compared to that of the radiosondes. In clear and partly cloudy conditions, error differences are 2–3 K and they increase in overcast cloud conditions. The reduced accuracy of overcast retrievals is due to the limited number of tropospheric microwave sounding channels and their poor vertical resolution, which result in ineffective cloud removal (see Section 7.5.3 for microwave temperature retrievals).

Many analyses appear to indicate that poor vertical resolution can be the primary source of error that limits the use of current satellite soundings for weather analysis/forecast operations. More advanced sounding instruments that have been developed are the Advanced Infrared Radiation Sounder (AIRS) (Chahine et al., 1984), which will be on board the NASA EOS/Aqua satellite, and the High spectral resolution Interferometer Sounder (HIS), cited in Section 4.2, which is being planned for both the geostationary and polar platforms. The former is a grating spectrometer; the latter is a Michelson interferometer covering a spectral region from about 3.5 to about 19 μm with high spectral resolution ($\lambda/\Delta\lambda > 2000$) (Smith et al., 1995). The goal of future atmospheric infrared sounder observations is to provide an overall accuracy of about 1 K with a vertical resolution of 1 km in temperature retrieval. Clearly, the
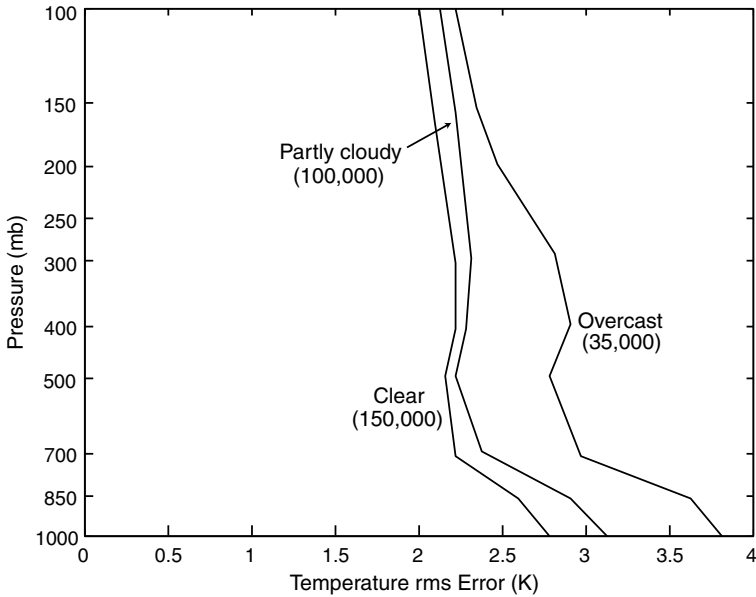
Figure 7.17   Root-mean-square differences between TOVS and radiosonde soundings during the entire global weather experiment year in 1979. The collocation and time differences for the two data sets are generally within 200 km and 3 hours. The number of cases employed in the retrieval for clear, partly cloudy, and overcast conditions are indicated in the diagram (data taken from Smith, 1991).

search for efficient and physically based retrieval programs employing a finite number of spectral channels to achieve this goal is still a subject of ongoing research and development.

7.4.4   Remote Sensing of Water Vapor and Trace Gas Profiles

7.4.4.1   WATER VAPOR FROM THE 6.3 μm VIBRATIONAL–ROTATIONAL BAND
Recalling Eq. (7.4.9) and performing integration by parts on the integral term, we find

$$I_{\bar{\nu}} = B_{\bar{\nu}}[T(0)] - \int_{p_s}^{0} T_{\bar{\nu}}(p) \frac{\partial B_{\bar{\nu}}(p)}{\partial p}\, dp, \qquad (7.4.50)$$

where $T(0)$ denotes the temperature at the top of the atmosphere and the spectral transmittance

$$T_{\bar{\nu}}(p) = \int_{\Delta\nu} \exp\left[-\int_{0}^{p} k_{\nu}(p)\, du(p) \right] \frac{d\nu}{\Delta\nu}. \qquad (7.4.51)$$

We have used the path length $u(g\ cm^{-2})$ to denote the integration in the transmittance.

If the temperature profile has been retrieved from the 15 and 4.3 μm CO_2 bands, then the remaining unknown is the transmittance. Further, if we select wavenumbers

in the 6.3 μm vibrational–rotational band of water vapor at which the absorption lines are given, then the only unknown is the path-length profile. The basic radiative transfer equation for the inference of gaseous profiles is more complicated than that for temperature retrieval. There is no clear-cut mathematical analysis that can be followed to invert the gaseous density profile. Nevertheless, one may devise numerical procedures for the determination of water vapor path-length profiles (Smith, 1991). The determination of water vapor parameters from satellite remote sensing, which would provide the required data for an understanding of the global hydrological cycle of the earth's atmosphere, is a critical task. Yet the accuracy of retrieval from passive sounders has been limited, partly because of the highly varying humidity field and the small amounts of water substance in the upper troposphere.

In principle, the concentration of greenhouse gases, O_3, N_2O, CH_4, and CFCs, can be inferred from nadir-looking spectrometers. Ozone exhibits vibrational–rotational lines in the 9.6 μm band. However, because of the structure of the O_3 concentration, having its maximum located in the stratosphere, the nadir-looking radiometers based on thermal emission have not been designed for retrieval purposes. Instead, reflected sunlight has been used for monitoring the total ozone concentration. N_2O and CH_4 have line structures in the 7.9 and 7.6 μm regions, as do other minor gases, and because of the relatively small mixing ratios, remote sensing of their concentrations utilizes the limb scanning technique.

7.4.4.2   LIMB SCANNING TECHNIQUE

A collimated radiometer or spectrometer, viewing horizontally, recovers radiation from the atmosphere over a layer that is relatively narrow in height, referred to as limb radiation. The general approach for obtaining profiles of trace constituents from limb measurements is known as limb sounding. A series of satellite experiments have used this method to retrieve the trace constituent profiles in the middle atmosphere, specifically the stratosphere.

The geometry of limb viewing is illustrated in Fig. 7.18. A radiometer that receives radiation emitted by the atmosphere along a ray path is identified by the tangent height closest to the surface. The atmosphere can be scanned by sweeping the view direction vertically or horizontally. There are a number of advantageous features of limb scanning for atmospheric probes: (a) Emission originates in the few kilometers immediately above the tangent point because of the rapid decrease of atmospheric density and pressure. Thus, a high inherent vertical resolution may be obtained. (b) All radiation received comes solely from the atmosphere. Variation of a changing underlying surface, which occurs when a nadir-viewing instrument is utilized, is not present. (c) A large degree of opacity is involved along a horizontal path. Hence, the limb technique is particularly useful for the determination of minor gases in the middle atmosphere. (d) The viewing direction from the satellite can be oriented in any azimuthal direction relative to the satellite motion and covers a large area. One disadvantage of this method is the interference of high clouds along the ray path, which act as bodies of infinite opacity and can produce considerable uncertainty in the emitted radiation. In addition, the sharp vertical weighting function, related to a
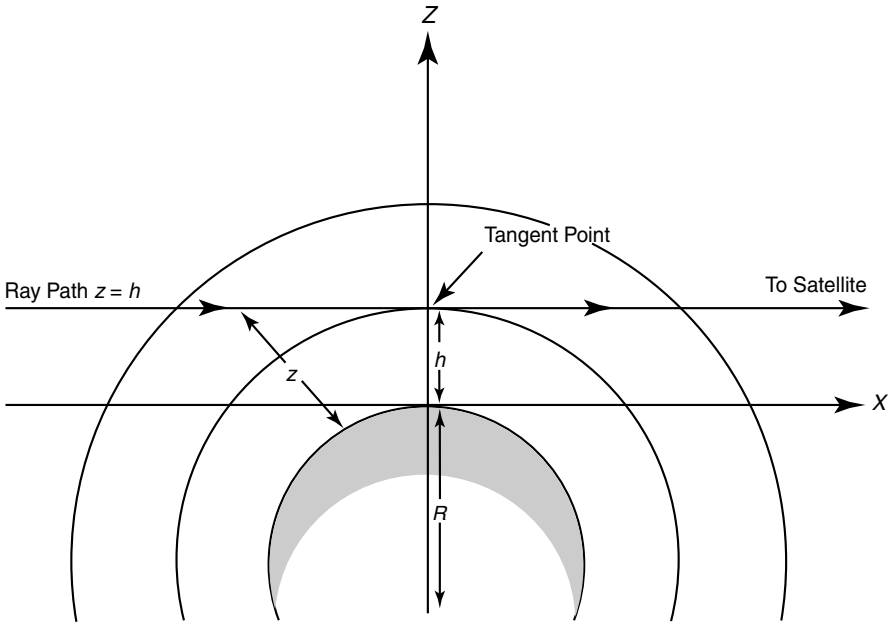
Figure 7.18   The geometry of limb viewing (see also Fig. 7.4). The long-path emission and absorption without the interference from the underlying surface are particularly useful for the inference of minor gases in the middle atmosphere.

horizontal region stretching 200 km or more along the ray path, leads to problems of interpretation of large changes in the atmospheric state over this distance. For these reasons, the limb scanning technique has been shown to be most useful for the inference of composition and structure of the stratosphere and mesosphere.

In reference to the limb viewing geometry shown in Fig. 7.18, the solution of the fundamental radiative transfer equation for a nonscattering atmosphere in local thermodynamic equilibrium can be expressed by

$$I_{\bar{\nu}}(h) = \int_{-\infty}^{\infty} B_{\bar{\nu}}[T(x)] \frac{\partial T_{\bar{\nu}}(h, x)}{\partial x}\, dx, \qquad (7.4.52)$$

where x is the distance coordinate along the ray path with the origin at the tangent point. The principle of the temperature and gaseous profile inversion problems is similar to that discussed in the previous sections. However, because of the spherical geometry involved, it is necessary to change from variable x to z in order to obtain the temperature and gaseous profiles as functions of height in the atmosphere. Thus, the limb viewing radiance may be written in the form

$$I_{\bar{\nu}}(h) = \int_{0}^{\infty} B_{\bar{\nu}}[T(z)] K_{\bar{\nu}}(h, z)\, dz, \qquad (7.4.53)$$
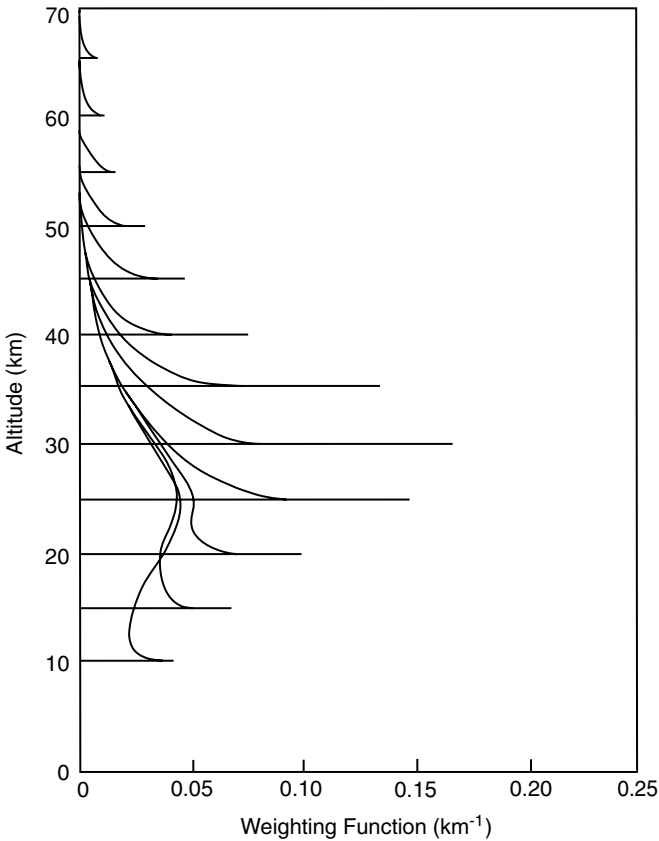
Figure 7.19 Limb viewing weighting function for the ideal case of an instrument with an infinitesimal vertical field of view for the spectral band 585–705 cm^{-1} covering most of the 15 μm band of CO_2 (after Gille and House, 1971).

where $K_{\bar{\nu}}$ represents the weighting function, which is a function of the geometrical factors and the spectral band model used with respect to height z. Figure 7.19 shows the limb viewing weighting function for a hypothetical instrument with an infinitesimal vertical field-of-view for a wide spectral band 585–705 cm^{-1} covering most of the 15 μm CO_2 band. For tangent heights above 25 km, the major part of the contribution comes from within about 3 km of the tangent height. Below 25 km, the weighting function takes on the broader shape of the nadir-viewing weighting functions, although a spike still remains at that tangent point. In principle, inversion of limb radiance measurements may be carried out utilizing the same techniques as in nadir-looking radiance observations.

The most significant application of the limb scanning radiometer has been in the determination of the temperature structure and minor gaseous concentrations of the

middle atmosphere, where nadir-looking radiometers have considerable difficulty deriving sufficient information for their recovery. Pioneering work on the retrieval of temperatures from limb measurements has been carried out by Gille and House (1971). Three infrared limb sounders have flown on the Nimbus satellites, beginning on Nimbus 6, for temperature, water vapor, and other trace gas studies. More recently, the NASA Upper Atmosphere Research Satellite (UARS) was launched in September, 1991, which included the Cryogenic Limb Array Etalon Spectrometer for atmospheric infrared emission measurements. In addition to measurements of the radiation emitted by the Q branch of CO_2 as a function of relative altitude for temperature retrievals, this spectrometer also measured the emission lines of O_3, CH_4, N_2O, and other trace gases associated with ozone destruction (Gille *et al.,* 1996).

In the stratosphere, the geostrophic approximation, in which the geostrophic wind is related to the horizontal gradient of the thickness, may be applicable to a good approximation. The horizontal equation of motion is given by

$$\left(\frac{d\mathbf{v}}{dt}\right)_H = g\nabla_H p - f\mathbf{k} \times \mathbf{v}, \tag{7.4.54a}$$

where $\nabla_H p = \nabla_p z$ is the pressure gradient force, $\mathbf{k}$ is the unit vector in the vertical direction, and f is the Coriolis parameter. The balance between the pressure gradient force and the Coriolis force gives the geostrophic wind velocity:

$$\mathbf{v}_g = -(g/f)\nabla_p z \times \mathbf{k}. \tag{7.4.54b}$$

Let $\mathbf{v}_{g1}$ and $\mathbf{v}_{g2}$ correspond to heights z_1 and z_2, respectively, such that $z_2 > z_1$. Then, the variation of the geostrophic wind velocity is

$$\Delta\mathbf{v}_g = \mathbf{v}_{g2} - \mathbf{v}_{g1} = -(g/f)\nabla_p(\Delta z) \times \mathbf{k}. \tag{7.4.54c}$$

From the hydrostatic equation $dp = -\rho g\,dz$, where ρ is the air density, and the equation of state $p = \rho RT$, we obtain

$$\Delta z = \int_{z_1}^{z_2} dz = z_2 - z_1 = -\frac{R}{g}\int_{p_1}^{p_2} T(p)\frac{dp}{p}, \tag{7.4.54d}$$

where R is the gas constant. Consequently, if temperatures are known between p_1 and p_2, then from Eqs. (7.4.54c) and (7.4.54d), the vertical change in the geostrophic wind between two pressure surfaces can be estimated. If the geostrophic wind at pressure level p_1 is known, an estimate of the geostrophic wind at pressure p_2 can then be derived.

The limb scanning radiometer is capable of determining the vertical temperature from about 10 to 1 mb. The temperature data from satellites can then provide the required wind data essential to the understanding of stratospheric dynamics.

7.4.5   Infrared Remote Sensing of Clouds

7.4.5.1   CARBON DIOXIDE SLICING TECHNIQUE FOR CLOUD TOP PRESSURE AND EMISSIVITY

Consider a FOV consisting of a cloud layer located at a pressure level p_c with a temperature T_c. Let the fraction of cloud cover be η and the cloud emissivity be ε_ν. For purposes of illustration, we shall consider monochromatic radiative transfer such that the transmittance multiplication follows exponential operations. The radiance observed at the satellite can be written in the form

$$\hat{I}_\nu = (1 - \eta\varepsilon_\nu)\left[B_\nu(T_s)T_\nu(p_s, p_c) + \int_{p_s}^{p_c} B_\nu(p)\frac{\partial T_\nu(p, p_c)}{\partial p}\, dp \right]T_\nu(p_c, 0)$$

$$+ \eta\varepsilon_\nu B_\nu(T_c)T_\nu(p_c, 0) + \int_{p_c}^{0} B_\nu(p)\frac{\partial T_\nu(p, 0)}{\partial p}\, dp, \tag{7.4.55}$$

where we have sliced the atmosphere above and below the cloud and defined the monochromatic transmittance in the form

$$T_\nu(p_1, p_2) = \exp\left(-\frac{1}{g}\int_{p_1}^{p_2} k_\nu(p)q(p)\, dp \right). \tag{7.4.56a}$$

The four terms in Eq. (7.4.55) are the radiation emitted from the surface, the contribution from the atmosphere below the cloud, the cloud contribution, and the contribution from the atmosphere above the cloud; all at TOA. Based on the definition of monochromatic transmittance, we have

$$T_\nu(p_s, p_c) \times T_\nu(p_c, 0) = T_\nu(p_s, 0), \quad T_\nu(p, p_c) \times T_\nu(p_c, 0) = T_\nu(p, 0), \tag{7.4.56b}$$

and so on.

For a clear FOV, $\eta\varepsilon_\nu = 0$ and the satellite measured radiance is

$$I_\nu^{clr} = B_\nu(T_s)T_\nu(p_s, 0) + \int_{p_s}^{0} B_\nu(p)\frac{\partial T_\nu(p, 0)}{\partial p}\, dp. \tag{7.4.57}$$

Subtracting the clear FOV radiance from the cloudy FOV radiance leads to the following result:

$$\hat{I}_\nu - I_\nu^{clr} = -\eta\varepsilon_\nu\, B_\nu(T_s)T_\nu(p_s, p_c)$$

$$-\eta\varepsilon_\nu \int_{p_s}^{p_c} B_\nu(p)\frac{\partial T_\nu(p, 0)}{\partial p}\, dp + \eta\varepsilon_\nu B_\nu(T_c)T_\nu(p_c, 0). \tag{7.4.58a}$$

This is the cloud signal in the satellite measured radiances for wavenumber ν; it is the radiance difference of the cloudy FOV from the neighboring clear FOVs. After performing integration by parts, we obtain

$$\hat{I}_\nu - I_\nu^{clr} = \eta\varepsilon_\nu \int_{p_s}^{p_c} T_\nu(p, 0)\frac{\partial B_\nu(p)}{\partial p}\, dp. \tag{7.4.58b}$$

Strictly speaking, this equation is valid only under the assumption of monochromatic radiation in which the transmittance multiplication follows the exponential rule. We may now apply this equation to two spectral bands in the 15 μm CO_2 band, denoted by 1 and 2. It follows that

$$\frac{\hat{I}_1 - I_1^{\text{clr}}}{\hat{I}_2 - I_2^{\text{clr}}} = \frac{\int_{p_s}^{p_c} T_1(p, 0)\frac{\partial B_1(p)}{\partial p}\, dp}{\int_{p_s}^{p_c} T_2(p, 0)\frac{\partial B_2(p)}{\partial p}\, dp}, \tag{7.4.59}$$

where we have set $\varepsilon_1 \cong \varepsilon_2$, a good approximation for wavenumbers close enough together in the 15 μm CO_2 band. The idea of employing the ratio of the cloud signal for two CO_2 channels viewing the same FOV to determine the cloud top pressure appears to have been developed by Smith and Platt (1978).

The left side of Eq. (7.4.59) is determined from the satellite-observed radiances in a given FOV, as well as from the clear-air radiances that can be inferred from analyses of satellite clear radiance observations. The temperature profile is supposed to be retrieved from the channels in the 4.3 and 15 μm CO_2 bands. It follows that the cloud-top pressure p_c, or height, can be inferred numerically. This is the essence of the CO_2 slicing method. Once the cloud-top pressure has been determined, an effective emissivity $\eta\varepsilon$ can be evaluated from the infrared window channel. In the window (let the wavenumber index be 3), we can write

$$\hat{I}_3 = \eta\varepsilon_3 B_3(T_c) + (1 - \eta\varepsilon_3)I_3^{\text{clr}}, \tag{7.4.60a}$$

where $B_3(T_c)$ is the blackbody cloud radiance in the window, which can be determined once the cloud height is given. Thus, the effective emissivity is

$$\eta\varepsilon_3 = \frac{\hat{I}_3 - I_3^{\text{clr}}}{B_3(T_c) - I_3^{\text{clr}}}. \tag{7.4.60b}$$

The CO_2 slicing method is most applicable to high-level cirrus clouds because of the nature of the weighting functions in the 15 μm CO_2 band. Wylie *et al.* (1994) used this method to determine cirrus cloud statistics from NOAA's polar-orbiting High-Resolution Infrared Radiation Sounder (HIRS) multispectral data in terms of cloud cover, height, and effective emissivity. Many processing procedures are required to produce a global map of cloud parameters. Displayed in Fig. 7.20 is the geographical distribution of cirrus clouds in the Northern Hemisphere summer (June, July, August) and winter (December, January, February) (referred to as boreal summer and winter) seasons during a 4-year period, June 1989 through May 1993. The intertropical convergence zone (ITCZ) is a region of more frequent cirrus, shown in the darker bands in the tropics, as are the midlatitude storm belts. The ITCZ moves north in the summer season. The subtropical high-pressure systems are evident in the region of less frequent cirrus cover. Over the Indonesian region, the ITCZ expands in coverage from winter to summer, whereas in the central Pacific Ocean, it shows extension in
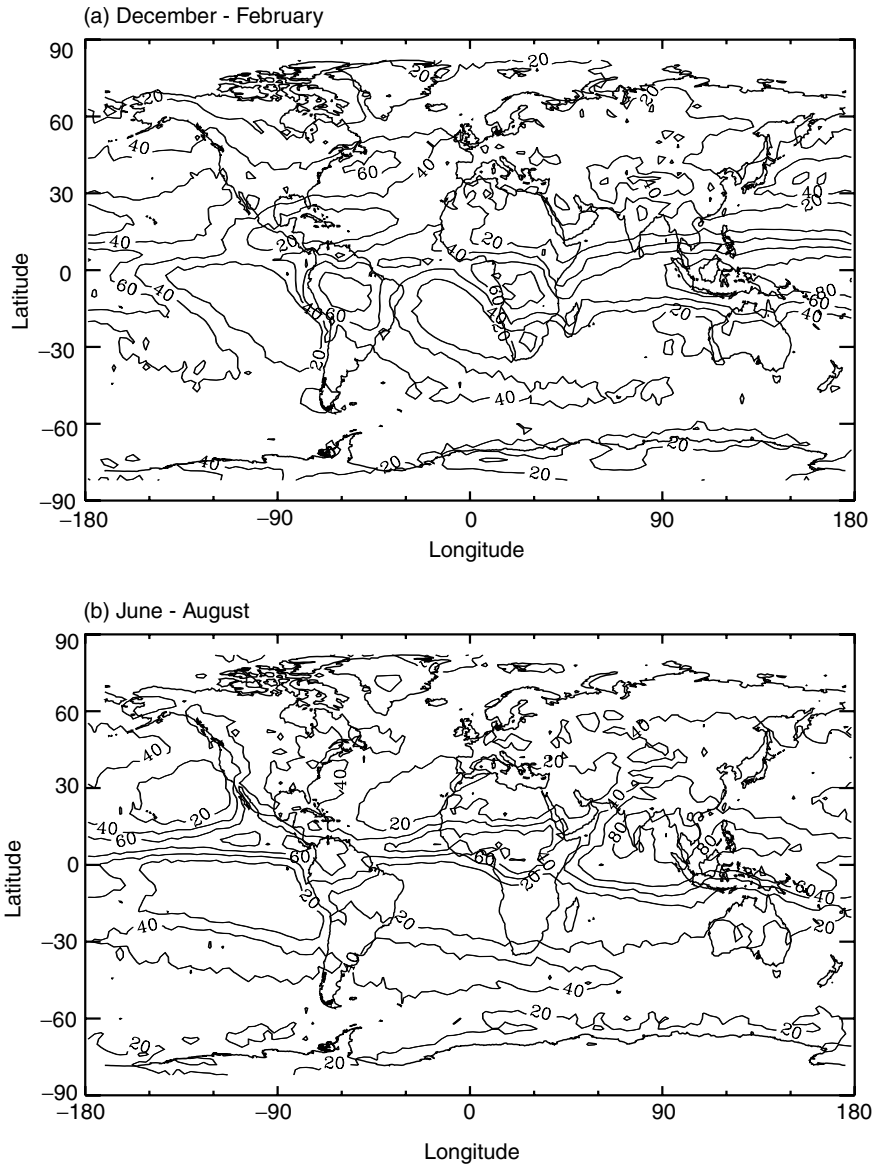
Figure 7.20 The frequency of clouds (%) above 6 km determined from the HIRS data using the CO_2 slicing technique. The 4-year average results are presented for (a) winter months and (b) summer months (data taken from Wylie *et al.*,1994).

both hemispheres during the winter months (of the Northern Hemisphere). In the Southern Hemisphere, the eastern Pacific Ocean off South America and the eastern Atlantic Ocean off Africa remain relatively free of cirrus clouds throughout the year. Finally, the North American cirrus cloud cover shows little seasonal change. Perhaps the most significant results of the cirrus cloud study based on the CO_2 slicing method are the ubiquitous cirrus occurrence in the tropics (more than 70%), and a noticeable increase in cirrus cover during the 4-year period.

7.4.5.2   EMITTED RADIANCE FOR CLOUD COVER

Many attempts have been made to classify the global distribution of clouds based on the emitted IR radiation in the 10 μm window (plus the reflected visible radiation during daytime). In general, the cloud tops are colder than the surfaces and so the emitted cloud radiances are smaller than the emitted clear radiances. Thus, it is practical to set various types of thresholds to distinguish between clear and cloudy conditions and to obtain cloud cover information. The classification of cloudy pixels by measured radiances is a complex decision-making process based on computer codes. The best example is the International Satellite Cloud Climatology Program (ISCCP), which is a project of the World Meteorological Organization developed in the 1980s (Schiffer and Rossow, 1983; Rossow and Schiffer, 1999). Its goal is to collect global visible and 11-μm data from polar and geostationary satellites and to process them into a cloud climatology.

An accurate global cloud climatology is essential to the development of physically based climate models; it can serve both as a source of input data for models that prescribe cloudiness and as a means of validating models that predict cloud formation. The ISCCP cloud products have also been used by researchers to understand the relationship of large-scale cloud patterns and dynamic processes in the atmosphere and the oceans. Zonally averaged cloud cover for January, July, and annual conditions derived from ISCCP are illustrated in Fig. 7.21. The equatorial maximum associated with the ITCZ lies between 0 and 10°N in July and 0 and 10°S in January. The minima, associated with subtropical anticyclones and desert areas, occur in both hemispheres at about 15° in the winter and at about 25° in the summer. The cloud amount tends to decrease poleward of 60°S in both January and July, but increases in January poleward of 60°N. The most cloudy latitude band is 60–70°S and in July, the area of cloudiness at 80–90°N is largely associated with Arctic stratus.

7.4.5.3   RETRIEVAL OF CIRRUS CLOUD OPTICAL DEPTH AND TEMPERATURE

As shown in Subsection 7.4.5.1 and presented in Section 5.1, cirrus clouds are globally distributed and have a high percentage of coverage. They are primarily present in the upper troposphere and lower stratosphere and are exclusively composed of nonspherical ice crystals. These clouds are normally transparent in the thermal IR wavelengths, particularly in the 3.7 and 10 μm windows, and consequently cannot be treated as blackbodies in radiative transfer in conjunction with remote sensing applications. In Section 7.3.5, we discussed the principles for the detection and determination of cirrus cloud optical depth and mean effective ice crystal size from
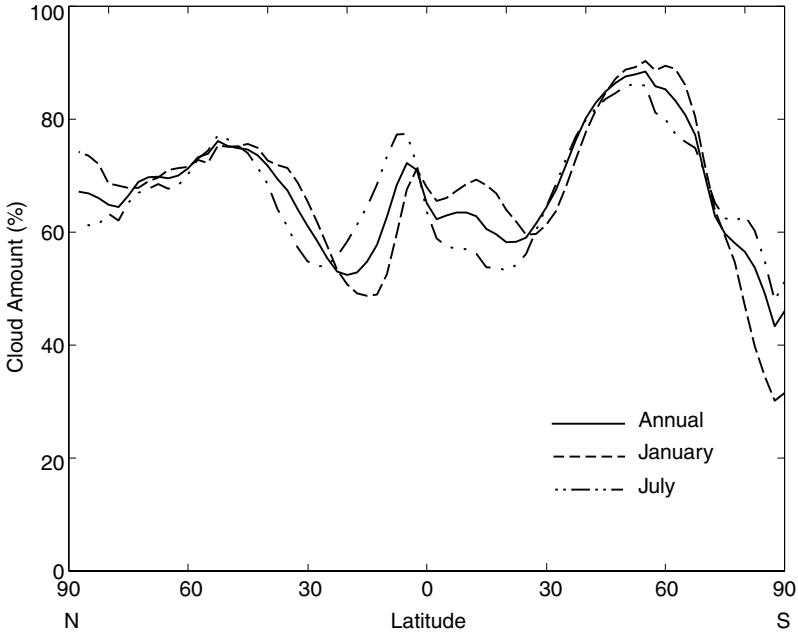
Figure 7.21   Meridional profiles of zonally averaged cloud amount for January, July, and annual conditions based on results derived from the International Satellite Cloud Climatology Program (ISCCP) data over a period from 1983 to 1994 (data available from William Rossow of the ISCCP office).

reflected sunlight. In the following, we introduce the principle of the retrieval of cirrus cloud parameters utilizing IR techniques.

Consider the upwelling radiance I at TOA in cirrus cloudy atmospheres. It is contributed by the radiance emitted from the cloud and the radiance transmitted through the cloud from the atmosphere and surface below. By selecting two appropriate window wavelengths, we have

$$I_i = (1 - \varepsilon_i)I_{ai} + \varepsilon_i B_i(T_c), \qquad i = 1, 2, \tag{7.4.61}$$

where I_{ai} are the upwelling radiances reaching the cloud base for the two wavelengths; ε_i are the cloud IR emissivities; T_c is the mean effective cloud temperature; and B_i are the Planck radiances. The objective of retrieval is to infer ε and T_c from two radiance measurements. However, there are a number of unknown variables involved in this basic equation that require parameterization.

From radiative transfer calculations of IR wavelengths, it is appropriate to approximate cloud emissivity in terms of visible optical depth τ in the form

$$\varepsilon_i \cong 1 - e^{-k_i\tau}, \tag{7.4.62}$$

where k_i represents the effective extinction coefficient. Thus, ε_1 and ε_2 can be correlated by introducing the ratio k_2/k_1, which is a function of the ice-crystal size distribution that needs to be prescribed prior to the retrieval exercise. The Planck

functions of two wavelengths can also be correlated by polynomials as follows:

$$B_1(T_c) = \sum_{n=0}^{3} a_n [B_2(T_c)]^n, \tag{7.4.63}$$

where a_n are fitting coefficients. Furthermore, values of the upwelling radiances reaching the cloud base must also be known. This can be accomplished by searching the clear pixels close to the cloud. With all the preceding considerations, radiance measurements in two wavelengths can be used to determine two variables: cloud visible optical depth and temperature, which may be related to ice crystal size.

The two channel IR technique requires that the optical properties of cirrus vary substantially between the two wavelengths. Huang and Liou (1984) developed a method to infer cirrus cloud optical depth and temperature based on AVHRR 3.7 and 10.9 μm channels. Inoue (1985) used the AVHRR radiance difference between 11 and 12 μm to determine the IR emissivity of cirrus clouds by assuming an implicit mean particle size. Liou *et al.* (1990b) used 6.5 and 11 μm radiometers to infer the temperature and optical depth of tropical cirrus. Cloud microphysics observations have shown that ice-crystal size distribution is correlated with cloud temperature. It is feasible, therefore, to estimate the mean effective ice crystal size once the cloud temperature is determined from satellite data (Ou *et al.,* 1995). Application of the 3.7 μm channel to daytime data requires the removal of the solar reflection component. A combination of the emitted 10 μm radiance and the reflected visible intensity has also been used to infer cirrus optical depth and temperature by means of a correlation between the IR emissivity and visible optical depth (Minnis *et al.,* 1993).

Independent validation of any remote-sensing algorithm for a variety of atmospheric conditions is essential to ensure its success. In the case of clouds, validation of retrieved ice-crystal size, which is an integrated value in the vertical, requires *in situ* aircraft measurements that must be collocated and coincident with the radiometric pixel observations. In recent years, airborne optical probes and replicator systems have been well developed for the collection of ice particles. Since aircraft observations are generally limited in time and space, especially in the vertical direction, validation of satellite-retrieved cloud particle sizing is a subject requiring considerable research and development. Cloud optical depth is a product of cloud thickness and the extinction coefficient. Cloud thickness can be inferred precisely from lidar backscattering for single cirrus and/or from radiosondes. The extinction coefficient requires light-scattering calculations involving ice-crystal size distributions. A consistent and reliable validation program for the remote sensing of clouds is evidently an involved and challenging task.

7.4.5.4 INFORMATION CONTENT IN INFRARED LINE SPECTRUM

In the analysis of the satellite Infrared Radiation Interferometer Spectrometer (IRIS) data, Prabhakara *et al.* (1993) have illustrated that thin cirrus in the tropics can be identified in the 8–12 μm window region. The IRIS instrument was on

board Nimbus 3 (Conrath *et al.,* 1970) and 4 (Kunde *et al.,* 1974) and for the first time provided the infrared emission of the atmosphere from about 400 to 1600 cm^{-1} with a 5 cm^{-1} spectral resolution (see also Section 4.1). Although the interferometer spectrometer experiment was not followed in the United States satellite program, three Fourier spectrometers in the spectral range 400–1600 cm^{-1} were included in the European METEOR satellites in the late 1970s. In the analysis of the emission spectra measured in clear and cirrus cloudy atmospheres, Spänkuch and Döhler (1985) have found that the presence of cirrus clouds significantly reduces the upwelling radiances in the entire spectral region, except in the center of the CO_2 15 μm band.

Smith *et al.* (1995) have developed a high spectral resolution infrared spectrometer, referred to as HIS, intended for satellite applications. This instrument was cited in Section 7.4.3 in the discussion of temperature retrievals. Sufficient information about cirrus clouds appears in the HIS spectrum. More recently, Smith *et al.* (1998) have displayed an interesting spectrum for a case involving a cold cirrus that was particularly evident in the 800–1000 cm^{-1} window region, which was presented in Fig. 4.12. As discussed in Section 4.6.1, clouds composed of water droplets such as low clouds or thin clouds containing large ice crystals behave as blackbodies or near blackbodies with little variation in the 10 μm window. However, thin cirrus, which are frequently subvisual, contain small ice crystals with maximum dimensions ranging from about 5 to 20 μm. These clouds have features in the 10 μm window associated with the absorption coefficients of ice crystals that can be used for their identification, as revealed in the extinction and single-scattering albedo spectra illustrated in Fig. 5.24. The development of verifiable schemes for the retrieval of the composition and structure of thin cirrus based on the IR line spectrum would be a subject of practical importance for climate studies.

7.4.6   Remote Sensing of Infrared Cooling Rate and Surface Flux

Infrared flux exchanges in the atmosphere and the atmospheric cooling produced by these exchanges are significant energy sources in the dynamic and thermodynamic processes of the atmosphere. In the troposphere under clear conditions, IR cooling is associated with absorption and emission of water vapor molecules. In particular, this cooling is primarily produced by the H_2O rotational band and H_2O continuum (near the surface). In the middle atmosphere, cooling is largely produced by the CO_2 15 μm band and, to a lesser degree, by the O_3 9.6 μm band, as explained in Section 4.5. The input data for IR calculations in clear atmospheres are basically temperature and water vapor profiles. For middle-atmosphere applications, other absorbing gaseous profiles are also needed. If these profiles can be reliably determined from the radiance data derived from satellite measurements, IR fluxes and cooling rates can be computed using radiative transfer methodology. Radiation algorithms for cooling rate calculations usually require very large amounts of computer time, and substantial discrepancies occur between different algorithms,

even with the same input data. Direct measurement of IR fluxes and cooling rates from space could have, therefore, important advantages. In the following, we present the general principles for the determination of cooling-rate profiles and surface IR fluxes.

Consider the IR cooling rate defined in Eq. (4.7.1) and rewrite this equation in the form

$$\frac{dF(z)}{dz} = -\rho C_p \dot{T}(z), \tag{7.4.64}$$

where $\dot{T} = (\partial T/\partial t)_{ir}$, denoting the cooling rate for a spectral band $\Delta \nu$, ρ is the air density, and C_p is the specific heat at constant pressure. There is no obvious relationship between atmospheric cooling rates and radiances at TOA. However, a direct link between the two must exist in view of the fact that both are related to temperature and specific gaseous profiles. As demonstrated in Section 7.4.3, the retrieval of the profile of any physical parameter requires a set of weighting functions that adequately cover the atmosphere. Thus, we may consider the following weighting function in the z-coordinate:

$$K_j(z) = \rho C_p T_j[\xi(z)/\mu], \tag{7.4.65}$$

where the transmittance for a spectral subband (or channel) $\Delta \nu_j$ corresponding to an absorber path length $\xi(z)$ is defined by

$$T_j(\xi/\mu) = \int_{\Delta \nu_j} e^{-k_\nu \xi/\mu} \frac{d\nu}{\Delta \nu_j} = \int_0^1 e^{-k(g)\xi/\mu} dg. \tag{7.4.66}$$

The correlated k-distribution method (Section 4.3) can be used to determine the absorption coefficients. It should be noted that the spectral band is related to the spectral subband by $\Delta \nu = \sum_j \Delta \nu_j$.

A convolution of the cooling-rate profile and weighting function defined in Eq. (7.4.65) leads to

$$\int_0^{z_\infty} \dot{T}(z) K_j(z) \, dz = -\int_0^{z_\infty} T_j[\xi(z)/\mu] \frac{dF(z)}{dz} \, dz. \tag{7.4.67}$$

In reference to Eqs. (4.2.10a,b), the spectral upward and downward fluxes in the optical-depth (or height) domain may be written in the path-length domain in the forms

$$F^\uparrow(\xi) = \pi B(\xi_*) T^f(\xi_* - \xi) - \int_\xi^{\xi_*} \pi B(\xi') \frac{d}{d\xi'} T^f(\xi' - \xi) \, d\xi', \tag{7.4.68a}$$

$$F^\downarrow(\xi) = \int_0^\xi \pi B(\xi') \frac{d}{d\xi'} T^f(\xi - \xi') \, d\xi', \tag{7.4.68b}$$

where ξ is the absorber path length in which $\xi = 0$ at TOA and $\xi = \xi_*$ at the surface, and the spectral interval subscript is omitted for simplicity of presentation.

The spectral diffuse transmittance is defined in Eq. (4.2.9) and can be written in the form

$$T^f(\xi) = 2 \int_0^1 T(\xi/\mu)\mu \, d\mu, \tag{7.4.68c}$$

where $T(\xi/\mu)$ is the spectral transmittance. The net flux divergence is then given by

$$
\begin{aligned}
\frac{dF(\xi)}{d\xi} &= \frac{d}{d\xi}[F^\uparrow(\xi) - F^\downarrow(\xi)] \\
&= 2\pi B(\xi_*) \int_0^1 \frac{\partial}{\partial \xi} T[(\xi_* - \xi)/\mu]\mu \, d\mu \\
&\quad -2\pi \int_0^1 \mu d\mu \left\{ \int_\xi^{\xi_*} B(\xi') \frac{\partial}{\partial \xi} \left(\frac{d}{d\xi'} T[(\xi' - \xi)/\mu] \right) d\xi' \right. \\
&\qquad \left. -B(\xi) \frac{d}{d\xi'} T[(\xi' - \xi)/\mu]\Big|_{\xi'=\xi} \right\} \\
&\quad -2\pi \int_0^1 \mu d\mu \left\{ \int_0^\xi B(\xi') \frac{\partial}{\partial \xi} \left(\frac{d}{d\xi'} T[(\xi - \xi')/\mu] \right) d\xi' \right. \\
&\qquad \left. +B(\xi) \frac{d}{d\xi'} T[(\xi - \xi')/\mu]\Big|_{\xi'=\xi} \right\}. \tag{7.4.69a}
\end{aligned}
$$

The spectral transmittance defined in the g-domain as shown in Eq. (7.4.66) is a linear operator involving the exponential function. For the purpose of this analysis, it suffices to let $T(\xi/\mu) = \exp(-k\xi/\mu)$. Thus, using this expression the net flux divergence after differential operations can then be expressed by

$$
\begin{aligned}
\frac{dF(\xi)}{d\xi} &= 2\pi k B(\xi_*) \int_0^1 e^{-k(\xi_* - \xi)/\mu} \, d\mu + 2\pi k^2 \int_\xi^{\xi_*} B(\xi') \int_0^1 e^{-k(\xi' - \xi)/\mu} \frac{d\mu}{\mu} \, d\xi' \\
&\quad + 2\pi k^2 \int_0^\xi B(\xi') \int_0^1 e^{-k(\xi - \xi')/\mu} \frac{d\mu}{\mu} \, d\xi' - 4\pi k B(\xi). \tag{7.4.69b}
\end{aligned}
$$

The spectral subband (channel) transmittance associated with the kernel function may also be written in a simple exponential function, $T_j(\xi/\mu) = \exp(-k_j\xi/\mu)$. It follows that

$$\int_0^{\xi_*} T_j(\xi/\mu) \frac{dF(\xi)}{d\xi} \, d\xi = \int_0^{\xi_*} e^{-k_j\xi/\mu} \frac{dF(\xi)}{d\xi} \, d\xi. \tag{7.4.70}$$

On substituting Eq. (7.4.69b) into Eq. (7.4.70), we first find that the first term on the right-hand side of Eq. (7.4.69b) approaches zero, when we set $\xi_* \to \infty$. Second,

the fourth term

$$\vartheta_4 = -4\pi k \int_0^{\xi_*} B(\xi) e^{-k_j \xi / \mu} \, d\xi = -4\pi \chi_j I_j(\mu), \qquad (7.4.71a)$$

based on the definition of the emergent channel radiance, where the coefficient $\chi_j = (k/k_j)\mu$. Here, we have omitted the surface term in this remote sounding application, although it could be included in the analysis. The solutions for the second and third terms require intricate mathematical analyses involving the interchange of the order of integrations and the use of the mean value theorem (Liou and Xue, 1988). The final results are, respectively,

$$\vartheta_2 = 2\pi \chi_j \ln(1 - 1/\chi_j)[\bar{\mu} I(\bar{\mu}) - \chi_j I_j(\mu)], \qquad (7.4.71b)$$

$$\vartheta_3 = 2\pi \chi_j^2 \ln(1 + 1/\chi_j) I_j(\mu), \qquad (7.4.71c)$$

where the beam-flux parameter determined from the mean value theorem using the isothermal atmosphere approximation is given by

$$\bar{\mu} = [1 + \chi_j \ln(1 - 1/\chi_j)]/ \ln(1 - 1/\chi_j). \qquad (7.4.71d)$$

In practice, the spectral and channel transmittances in the g-space may be written in summation form as follows:

$$T(\xi/\mu) = \sum_m e^{-k_m \xi / \mu} \Delta g_m, \qquad (7.4.72a)$$

$$T_j(\xi/\mu) = \sum_n e^{-k_n \xi / \mu} \Delta g_n. \qquad (7.4.72b)$$

Since the summation operators are linear, we may carry out the preceding analyses employing Eqs. (7.4.72a,b) with the final results involving two double summations for all the terms in Eq. (7.4.70). Consequently, we may write

$$\int_0^{\xi_*} T_j(\xi/\mu) \frac{dF(\xi)}{d\xi} \, d\xi = I(\bar{\mu})\alpha + I_j(\mu)\beta, \qquad (7.4.73a)$$

where $I(\bar{\mu})$ is the radiance for the spectral band at a mean angle $\cos^{-1} \bar{\mu}$, $I_j(\mu)$ is the channel radiance at an emergent angle, $\cos^{-1} \mu$, and α and β are coefficients associated with the absorption coefficients that can be determined exactly from numerical means. The value of $\bar{\mu}$ varies from about 0.5 to 0.6. In this manner, the net flux divergence is transformed into spectral and channel radiances via a convolution with the channel transmittance.

Combining Eqs. (7.4.67) and (7.4.73a), we obtain the basic Fredholm equation of the first kind for the remote sounding of cooling rate profiles from space:

$$I(\bar{\mu})\alpha + I_j(\mu)\beta = -\int_0^{z_\infty} \dot{T}(z) K_j(z) \, dz. \qquad (7.4.73b)$$

Two sets of measurements are needed to drive the spectral cooling rate profiles: spectral radiance $I(\bar{\mu})$ and channel radiance $I_j(\mu)$, which can both be measured at
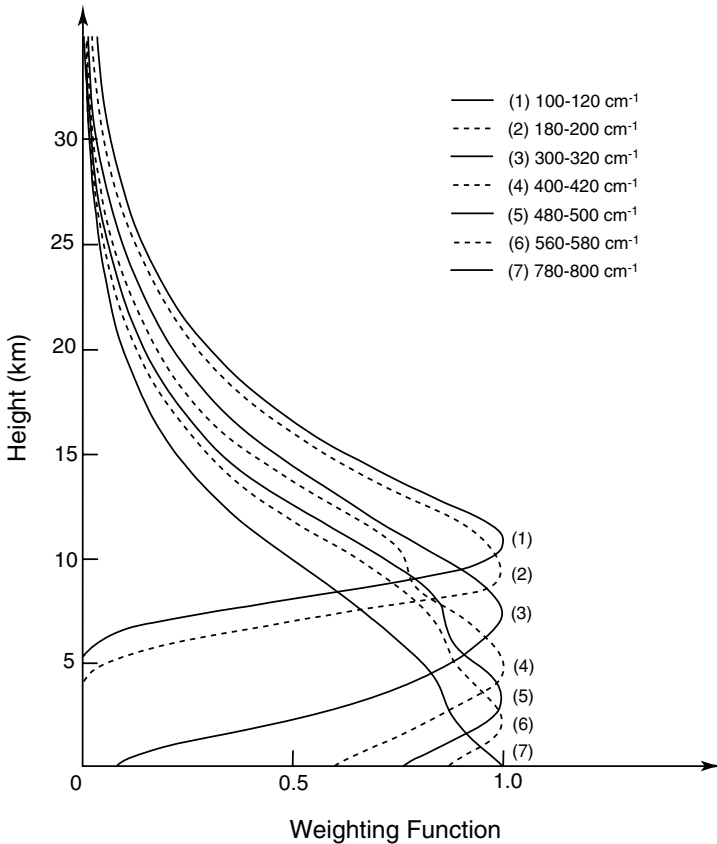
Figure 7.22   Normalized weighting functions for atmospheric cooling rate retrieval in the H_2O rotational band (20–800 cm^{-1}). The weighting functions presented here are products of air density and channel transmittances for a 20 cm^{-1} interval using standard atmospheric temperature and water vapor profiles (data taken from Liou, 1992).

an emergent angle, $\cos^{-1}\bar{\mu}$. We must now examine the required weighting functions. A number of weighting functions based on Eq. (7.4.65) have been calculated in the H_2O rotational band for a 20 cm^{-1} interval using a line-by-line program. Illustrated in Fig. 7.22 are seven weighting functions whose peaks are spaced within $\sim$2 km in height. The four uppermost and three lowermost channels could be used to retrieve cooling rates for spectral intervals of 20–500 cm^{-1} and 500–800 cm^{-1}, respectively. The preceding theory would be ideal for the retrieval of the cooling rate produced by CO_2 in the middle atmosphere, since clouds would have little effect on the upwelling radiances at TOA. However, appropriate weighing functions must be determined in the stratosphere and lower mesosphere for practical applications.

The determination of surface radiative fluxes from available satellite radiance data has been a subject of considerable interest in recent years in view of the important role that surface–atmosphere interactions play in climate (Section 8.2.5). A direct

observational study of surface IR fluxes could be conducted using the following approach. Based on the definition of cooling rate in Eq. (7.4.64), we may perform an integration of this equation from the surface to TOA to obtain

$$F(0) = F(z_\infty) + \int_0^{z_\infty} \rho C_p \dot{T}(z)\, dz. \tag{7.4.74}$$

The broadband IR flux at TOA, $F(z_\infty)$, has been routinely derived from satellite measurements (Section 8.2.2). Thus, if the cooling rate profile $\dot{T}(z)$ is retrievable from radiance observations, we may estimate the surface radiative flux $F(0)$ based solely on measurements from space. Furthermore, we note that if the contributions due to O_3, CO_2, and H_2O–CO_2 overlap are omitted, the total cooling would be accurate within $\sim$4 % depending on the temperature profile. Thus, to a good approximation, total atmospheric cooling may be obtained by measuring the cooling profile produced by water vapor.

The history of infrared sounding from space, as discussed in this section, has been characterized by the slow evolution of retrieval methodologies and of the selection of sounding channels. To be successful in cooling-rate retrieval, many additional aspects of practical issues such as cloud problems and infrared technology will need to be examined. The HIS data presented in Subsection 7.4.5.4 could be useful for the exploration of IR cooling rate retrieval, particularly if the spectrum coverage is extended to the rotational band of water vapor. The HIS instrument is capable of scanning from space and could provide the needed spectral radiances for conversion to fluxes.

7.5   Remote Sensing Using Emitted Microwave Radiation

7.5.1   Microwave Spectrum and Microwave Radiative Transfer

In the microwave region, only water vapor and molecular oxygen exhibit significant absorption lines. It is customary to use the frequency unit GHz in the discussion of microwave radiative transfer. Note that 1 GHz $= 10^9$ cycles/sec, and 1 cm $=$ 30 GHz. Below 40 GHz only the weakly absorbing pressure-broadened 22.235 GHz water vapor line is dominant. This resonance absorption line arises from transitions between nuclear spin rotational states. At about 31.4 GHz, air is relatively transparent, which is the window between the resonance water vapor line. The oxygen molecule has a magnetic dipole moment arising from the combined spins of two impaired electrons in its electronic ground state. Changes in the orientation of the electronic spin relative to the orientation of the molecular rotation produce a strong band of magnetic dipole transitions near 60 GHz and a single transition at 118.75 GHz. For frequencies greater than 120 GHz, water vapor absorption again becomes dominant because of the strongly absorbing line at 183 GHz. Figure 7.23 illustrates vertical atmospheric transmittance as a function of frequency and wavelength for a standard atmosphere, clearly identifying the aforementioned absorption lines.
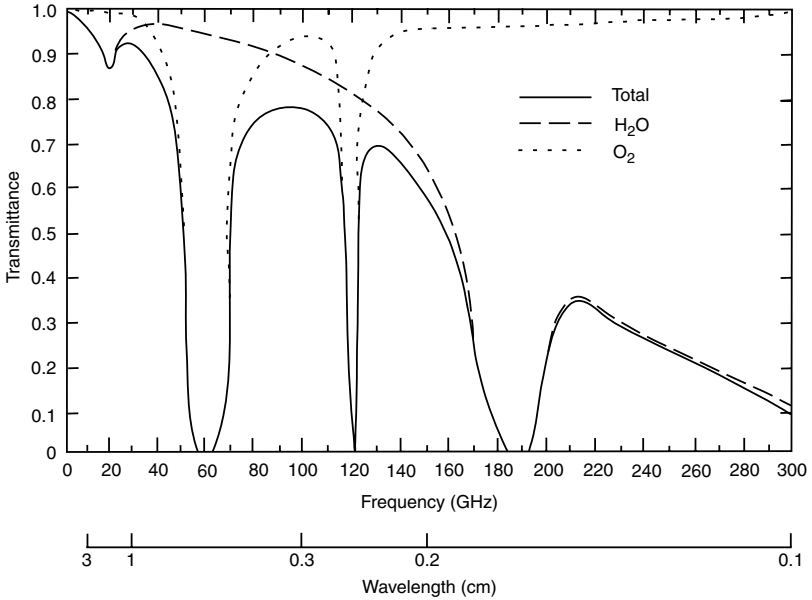
Figure 7.23   Demonstrative atmospheric transmittances (total, H_2O, and O_2) as a function of frequency and wavelength in the microwave region.

In view of the size of water droplets ($\sim 10~\mu m$) and ice crystals ($\sim 100~\mu m$) in the atmosphere, their effects on the scattering of microwaves (~ 1 cm) are generally small. Thus, we may, as a good approximation, neglect the scattering contribution in the discussion of microwave radiative transfer. In reference to Section 7.4.1, the solution of the radiative transfer equation for a nonscattering atmosphere in local thermodynamic equilibrium is given by

$$I_{\tilde{v}}(0) = I_{\tilde{v}}(p_s)T_{\tilde{v}}(p_s, 0) + \int_{p_s}^{0} B_{\tilde{v}}[T(p)]\frac{\partial T_{\tilde{v}}(p, 0)}{\partial p}\, dp, \qquad (7.5.1)$$

where $\tilde{v}$ denotes the frequency, $I_{\tilde{v}}(p_s)$ represents the radiance contribution from the surface, and the transmittance, $T_{\tilde{v}}(p, 0)$, is expressed with respect to TOA. Emissivity in the microwave region is normally less than unity, so there is a significant reflection contribution from the surface. The radiance emitted from the surface is given by

$$I_{\tilde{v}}(p_s) = \varepsilon_{\tilde{v}} B_{\tilde{v}}(T_s) + (1 - \varepsilon_{\tilde{v}})\int_{0}^{p_s} B_{\tilde{v}}[T(p)]\frac{\partial T_{\tilde{v}}(p_s, p)}{\partial p}\, dp, \qquad (7.5.2)$$

where the first term on the right-hand side denotes the surface emission contribution and the second term represents the emission contribution from the entire atmosphere to the surface that is reflected back to the atmosphere at the same frequency. The transmittance, $T_{\tilde{v}}(p_s, p)$, is now expressed with respect to the surface.

Inserting the lower boundary condition defined in Eq. (7.5.2), the upwelling radiance can now be expressed as

$$I_{\tilde{v}} = \varepsilon_{\tilde{v}} B_{\tilde{v}}(T_s) T_{\tilde{v}}(p_s, 0) + (1 - \varepsilon_{\tilde{v}}) T_{\tilde{v}}(p_s, 0) \int_0^{p_s} B_{\tilde{v}}[T(p)] \frac{\partial T_{\tilde{v}}(p_s, p)}{\partial p} \, dp$$

$$+ \int_{p_s}^0 B_{\tilde{v}}[T(p)] \frac{\partial T_{\tilde{v}}(p, 0)}{\partial p} \, dp. \tag{7.5.3}$$

The Planck function in the frequency domain is given by

$$B_{\tilde{v}}(T) = 2h\tilde{v}^3 / \left[c^2 \left(e^{h\tilde{v}/KT} - 1 \right) \right]. \tag{7.5.4a}$$

In the microwave region, $h\tilde{v}/KT \ll 1$, the Planck function can be approximated by

$$B_{\tilde{v}}(T) \cong (2K\tilde{v}^2/c^2)T. \tag{7.5.4b}$$

Thus, Planck radiance is linearly proportional to temperature, referred to as the *Rayleigh–Jeans law* discussed in Section 1.2. Analogous to the above approximation, we can define an equivalent brightness temperature T_B such that

$$I_{\tilde{v}} = (2K\tilde{v}^2/c^2)T_B(\tilde{v}). \tag{7.5.5}$$

Substituting Eqs. (7.5.4b) and (7.5.5) into Eq. (7.5.3), the solution of microwave radiative transfer may now be written in terms of temperature as follows:

$$T_B(\tilde{v}) = \varepsilon_{\tilde{v}} T_s T_{\tilde{v}}(p_s, 0) + (1 - \varepsilon_{\tilde{v}}) T_{\tilde{v}}(p_s, 0) \int_0^{p_s} T(p) \frac{\partial T_{\tilde{v}}(p_s, p)}{\partial p} \, dp$$

$$+ \int_{p_s}^0 T(p) \frac{\partial T_{\tilde{v}}(p, 0)}{\partial p} \, dp. \tag{7.5.6}$$

The contribution of each term to the brightness temperature at TOA is illustrated in Fig. 7.24. The first, second, and third terms represent the surface emission contribution, the reflection contribution from the surface into the atmosphere, and the atmospheric emission contribution, respectively.

The transmittance used for satellite remote sensing is generally expressed with respect to TOA; i.e., $T_{\tilde{v}}(p) = T_{\tilde{v}}(p, 0)$. Thus, for computational purposes, it is desirable to express $T_{\tilde{v}}(p_s, p)$ in terms of $T_{\tilde{v}}(p, 0)$. For monochromatic frequencies, transmittance is an exponential function of optical depth [see Eq. (7.4.4)]. Hence, we can write

$$T_{\tilde{v}}(p_s, p) = \exp\left[-\frac{1}{g} \int_p^{p_s} k_{\tilde{v}}(p')q(p') \, dp' \right]$$

$$= \exp\left[-\frac{1}{g} \int_0^{p_s} k_{\tilde{v}}(p')q(p') \, dp' + \frac{1}{g} \int_0^p k_{\tilde{v}}(p')q(p') \, dp' \right]$$

$$= T_{\tilde{v}}(p_s, 0)/T_{\tilde{v}}(p, 0), \tag{7.5.7}$$
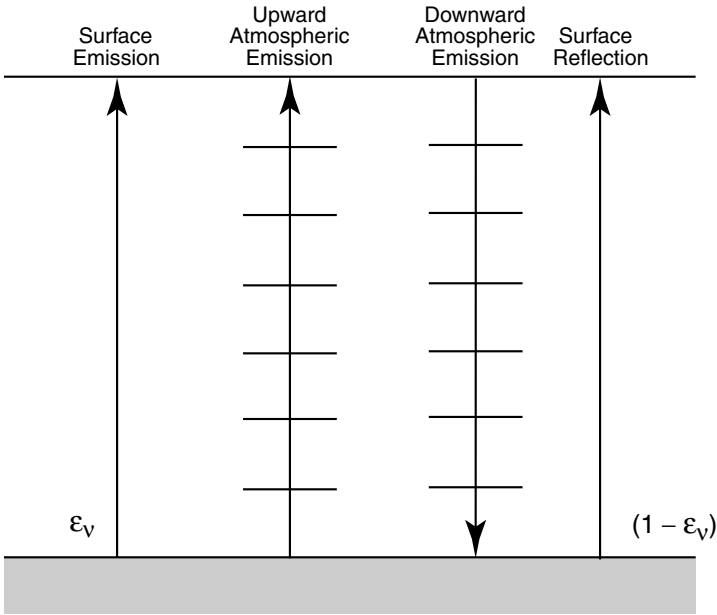
Figure 7.24 Contributions of brightness temperature at the top of a clear atmosphere, represented by the terms in Eq. (7.5.6), including surface emission and reflection, and atmospheric contributions.

where $T_{\tilde{\nu}}(p_s, 0)$, the transmittance of the entire atmosphere, is a constant value. Thus,

$$\frac{\partial T_{\tilde{\nu}}(p_s, p)}{\partial p} = -\frac{T_{\tilde{\nu}}(p_s, 0)}{[T_{\tilde{\nu}}(p, 0)]^2} \frac{\partial T_{\tilde{\nu}}(p, 0)}{\partial p}. \tag{7.5.8}$$

Inserting Eq. (7.5.8) into Eq. (7.5.6), rearranging terms, and letting $T_{\tilde{\nu}}(p, 0) = T_{\tilde{\nu}}(p)$, we have

$$T_B(\tilde{\nu}) = \varepsilon_{\tilde{\nu}} T_s T_{\tilde{\nu}}(p_s) + \int_{p_s}^{0} J_{\tilde{\nu}}(p) \frac{\partial T_{\tilde{\nu}}(p)}{\partial p} \, dp, \tag{7.5.9}$$

where the atmospheric source term is given by

$$J_{\tilde{\nu}}(p) = \left\{ 1 + (1 - \varepsilon_{\tilde{\nu}})[T_{\tilde{\nu}}(p_s)/T_{\tilde{\nu}}(p)]^2 \right\} T(p). \tag{7.5.10}$$

In microwave sounding, transmittances are computed by including the antenna gain characteristics.

The use of microwaves for atmospheric sounding from a satellite platform differs from that of IR techniques in the consideration of surface emissivity. The surface emissivity appearing in the first term of Eq. (7.5.6) has a substantial effect on the brightness temperature value. In the microwave spectrum, emissivity values of the earth's surface vary over a considerable range, from about 0.4 to 1.0. Over land, emissivity depends on the moisture content of the soil. The wetting of a soil surface results in a rapid decrease in emissivity. The emissivity of dry soil is on the order
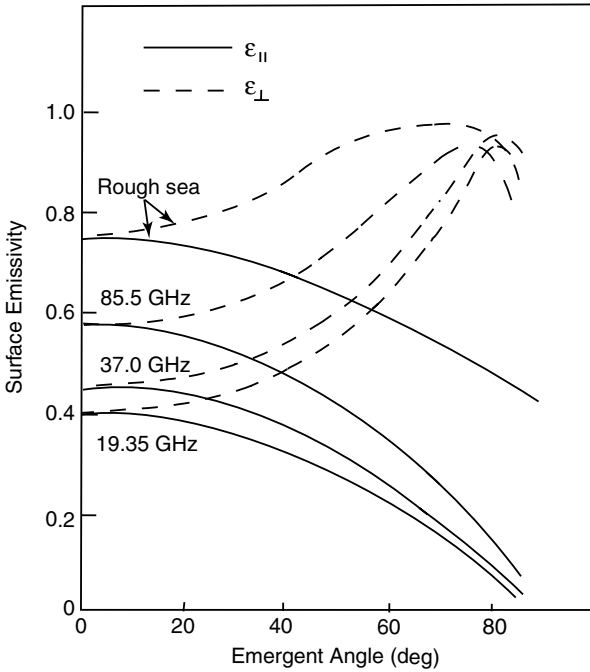
Figure 7.25   Emissivities of calm and rough ocean surfaces for a number of microwave frequencies. The solid and dashed curves represent the horizontal $\varepsilon_\parallel$ and vertical $\varepsilon_\perp$ polarization components, respectively (data taken from Huang and Liou, 1983).

of 0.95 to 0.97, whereas for wet bare soil it is about 0.80 to 0.90, depending on the frequency. The emissivity of the sea surface typically ranges between 0.4 and 0.6, depending upon such variables as salinity, sea ice, surface roughness, and sea foam. In addition, there is a frequency dependence with higher frequencies displaying higher emissivity values.

Emission from the oceans is polarized. For a calm ocean surface, we can apply the Fresnel reflection formula (Section 5.3.2) to compute the vertical and horizontal emissivity components using the real and imaginary refractive indices for pure water. Shown in Fig. 7.25 are the vertically ($\varepsilon_\perp$) and horizontally ($\varepsilon_\parallel$) polarized emissivities of calm ocean surfaces for 19.35, 37, and 85.5 GHz. These emissivities increase with increasing frequency. The results for rough sea were derived from empirical fitting to a limited set of observations and are applicable to 37 and 85.5 GHz. Substantial differences between the two components are shown for large emergent angles.

Equation (7.5.1) does not include the scattering and polarization properties of precipitating particles. We may approach the fundamental microwave radiative transfer by incorporating the multiple scattering and emission terms, as presented in Chapter 6. Emission from a homogeneous earth surface is either isotropic (unpolarized land), or zenith-angle dependent (polarized ocean). Further, under the plane-parallel and horizontally homogeneous assumption, the transfer of microwave radiation can be treated as azimuthally independent. In this case, the analysis of the vertically and

horizontally polarized brightness temperatures is independent of the circular polarization and requires only the P_{11}, P_{12}, P_{21}, and P_{22} elements in the scattering phase matrix (see Section 6.6 for definition). Microwave radiative transfer involving finite and inhomogeneous clouds and precipitation can be solved following the methodology presented in Section 6.7.

7.5.2   Rainfall Rate and Water Vapor Determination from Microwave Emission

One of the important applications of microwave radiative transfer has been the determination of atmospheric liquid water and water vapor amounts, since microwaves see through heavy clouds and precipitation, which are largely opaque in the infrared wavelengths.

Based on Lorenz–Mie scattering calculations for spheres, the radiative properties of clouds and precipitation in the microwave spectrum have the following features: (a) ice particles essentially do not absorb microwave radiation but only scatter it; (b) water and rain droplets absorb and scatter with absorption dominating; (c) scattering and absorption of ice and water both increase with frequency. With respect to the scattering and absorption properties of ice and water, the microwave spectrum can be divided into three regions. Below about 22 GHz, absorption is the main process for microwave radiative transfer. Although scattering also occurs, it is of secondary significance. Around 60 GHz, scattering is more important than absorption. Between 20 and 60 GHz, both scattering and absorption processes are important. The scattering and absorption of ice and water particles depend on their sizes relative to the microwave frequency. For frequencies in the 22 and 31 GHz regions (longer wavelengths), the scattering of ice crystals is negligible so that ice clouds are transparent to these microwave frequencies, making them ideal for the detection of rainfall rates.

In order to derive the liquid water and water vapor amounts, it is necessary to develop parameterization equations in which the liquid water content and water vapor amount are explicitly defined. In reference to Eq. (7.5.6), we perform integration by parts on the two integral terms to obtain

$$T_B(\tilde{v}) = \left[T(0) - \int_{p_s}^0 T_{\tilde{v}}(p) \frac{\partial T(p)}{\partial p} \, dp \right] - (1 - \varepsilon_{\tilde{v}}) T_{\tilde{v}}^2(p_s)$$
$$\times \left[T(0) + \frac{1}{T_{\tilde{v}}(p_s)} \int_0^{p_s} T_{\tilde{v}}(p_s) \frac{\partial T(p)}{\partial p} \, dp \right], \qquad (7.5.11a)$$

where, for simplicity, we use one pressure variable in the argument of the transmittance. Moreover, we define

$$x_{\tilde{v}} = \frac{T(0)}{T_s} - \frac{1}{T_s} \int_{p_s}^0 T_{\tilde{v}}(p) \frac{\partial T(p)}{\partial p} \, dp$$
$$= 1 + \frac{1}{T_s} \int_{p_s}^0 [1 - T_{\tilde{v}}(p)] \frac{\partial T(p)}{\partial p} \, dp, \qquad (7.5.11b)$$

$$y_{\tilde{v}} = \frac{T(0)}{T_s} + 1 - \frac{1}{T_{\tilde{v}}(p_s)T_s} \int_0^{p_s} T_{\tilde{v}}(p)\frac{\partial T(p)}{\partial p}\,dp$$

$$= \frac{1}{T_{\tilde{v}}(p_s)} + \frac{T(0)}{T_s}\left[1 - \frac{1}{T_{\tilde{v}}(p_s)}\right] - \frac{1}{T_{\tilde{v}}(p_s)T_s}\int_0^{p_s}[1 - T_{\tilde{v}}(p)]\frac{\partial T(p)}{\partial p}\,dp.$$

$$(7.5.11c)$$

Hence, Eq. (7.5.11a) can be rewritten as follows:

$$T_B(\tilde{v}) = T_s\big[x_{\tilde{v}} - y_{\tilde{v}}T_{\tilde{v}}^2(p_s)(1 - \varepsilon_{\tilde{v}})\big]. \qquad (7.5.12)$$

Upon inspection of the microwave spectrum, we find that below about 40 GHz, the transmittance $T_{\tilde{v}}(p_s) \approx 1$, so that $x_{\tilde{v}} \approx y_{\tilde{v}} \approx 1$. As a result of this simplification, the brightness temperature can be approximated by

$$T_B(\tilde{v}) \cong T_s\big[1 - T_{\tilde{v}}^2(p_s)(1 - \varepsilon_{\tilde{v}})\big]. \qquad (7.5.13a)$$

The transmittance for frequencies lower than 40 GHz is mainly due to the absorption of water vapor and liquid water. It may be expressed by

$$T_{\tilde{v}}(p_s) = T_{\tilde{v}}(\text{vapor})\,T_{\tilde{v}}(\text{liquid}). \qquad (7.5.13b)$$

For frequencies lower than about 40 GHz, the transmittance of liquid water can be approximated by

$$T_{\tilde{v}}(\text{liquid}) \cong \exp(-Q/Q_0) \approx 1 - Q/Q_0(\tilde{v}), \qquad (7.5.13c)$$

where Q is the liquid water path (g cm^{-2}) and $Q_0(\tilde{v})$ is a constant that depends on the frequency and cloud temperature. In a similar manner, if we select a frequency at about 22 GHz, the water vapor transmittance can be approximated by

$$T_{\tilde{v}}(\text{vapor}) \approx 1 - W/W_0(\tilde{v}), \qquad (7.5.13d)$$

where W is the total water vapor path length (g cm^{-2}) and, again, $W_0(\tilde{v})$ is a constant. Inserting these two expressions in Eq. (7.5.13a), and neglecting second-order terms involving Q and W, we obtain

$$T_B(\tilde{v}) \cong \varepsilon_{\tilde{v}}T_s + 2(1 - \varepsilon_{\tilde{v}})T_s(Q/Q_0 + W/W_0). \qquad (7.5.14)$$

Assuming that the surface temperature T_s and the surface emissivity $\varepsilon_{\tilde{v}}$ are known parameters, two brightness temperature observations at about 40 and 22 GHz can be used to determine Q and W as follows:

$$Q = q_0 + q_1 T_B(\tilde{v}_1) + q_2 T_B(\tilde{v}_2),$$

$$W = w_0 + w_1 T_B(\tilde{v}_1) + w_2 T_B(\tilde{v}_2), \qquad (7.5.15)$$

where w_i and q_i are coefficients related to the frequencies chosen, the surface temperature, the emissivity, and the empirical parameters Q_0 and W_0. They are normally determined statistically from a sample of known brightness temperatures and the liquid water and water vapor amounts in known atmospheric profiles.

The rainfall rate is related to raindrop size distribution, which can be measured at the surface. Based on measurements, Marshall and Palmer (1948) have suggested that droplet size distribution can be fitted by an appropriate negative exponential form given by

$$n(D) = n_0 e^{-\Lambda D}, \tag{7.5.16a}$$

where the raindrop diameter $D = 2a$, a is the radius, and $n(D)dD$ represents the number of droplets per unit volume with the diameter between D and $D + dD$ in units of cm. The slope factor Λ was found to depend only on the rainfall rate R, which is measured in terms of mm hr^{-1} and is given by

$$\Lambda = 41 R^{-0.21}. \tag{7.5.16b}$$

In Eq. (7.5.16a), the intercept parameter $n_0 = 0.08$ cm^{-4}. Thus, once the rainfall rate is given, a Marshall and Palmer size distribution is determined. The liquid water path (LWP, g cm^{-2}) is the product of LWC and cloud thickness, Δz, defined in Eq. (7.3.13b). Using the Marshall and Palmer size distribution, we can show that

$$\text{LWP} = k R^{-0.84}, \tag{7.5.16c}$$

with k a certain coefficient. Thus, LWP is directly related to the rainfall rate.

In the case of pure absorption and emission, the LWP (or Q) is directly connected to the brightness temperature via Eq. (7.5.15). Using the relationship between LWP and the rainfall rate, we have

$$T_B(\tilde{\nu}) \cong b + c R^{0.84}, \tag{7.5.17}$$

where b and c are empirical coefficients to be determined from known atmospheric and surface conditions. The preceding analysis did not account for the scattering contribution of cloud particles, which can be important if their sizes are comparable to or larger than the emission wavelength. In this case, a full radiative transfer solution including scattering contributions should be developed in conjunction with the retrieval of rainfall rates.

The inference of water content and total water vapor amount from Nimbus E Microwave Spectrometer (NEMS) data at 22.235 and 31.4 GHz channels on Nimbus 5 (1972) over the oceans was first reported by Staelin *et al.* (1976) and Grody (1976). The coefficients were obtained by a multidimensional regression analysis based upon computed brightness temperatures with known atmospheric temperature and water vapor profiles derived from radiosondes. Computations including the contribution of clouds and precipitation to the brightness temperature did not include the scattering contribution due to clouds and rain drops. These authors have demonstrated the feasibility of mapping the large-scale features of liquid water and total water vapor patterns from the water vapor (22.235 GHz) and window (31.4 GHz) channel data. Liou and Duff (1979) attempted to derive the liquid water content from Nimbus 6 (1975) Scanning Microwave Spectrometer (SCAMS) data
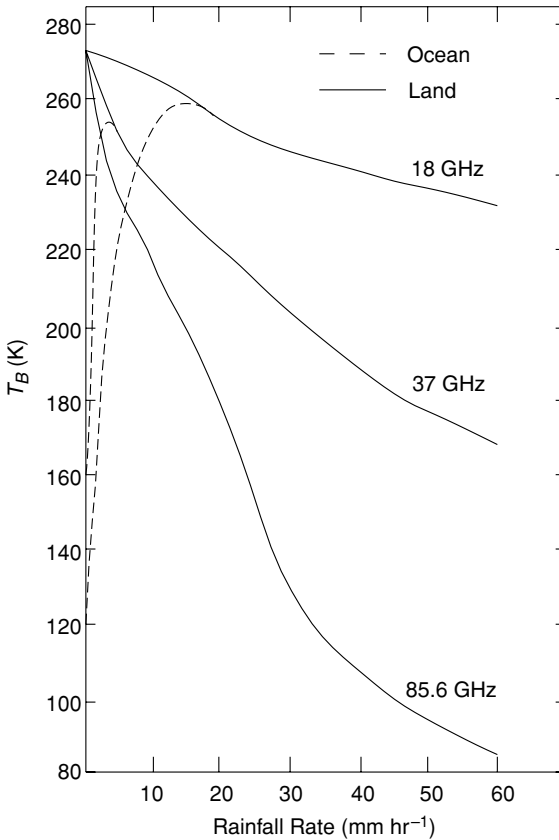
Figure 7.26   Brightness temperature as a function of rainfall rate over the ocean and land for three frequencies (data taken from Spencer *et al.,* 1989).

over land utilizing Eq. (7.5.15). Effects of the scattering and absorption properties of hydrometeors in an inhomogeneous absorbing gaseous atmosphere were included in the calculations along with an empirical means to determine the surface emissivity characteristics.

The mapping of rainfall rates over the oceans using microwave technology from satellites was first proposed by Wilheit and Chang (1977). Effects of the scattering and polarization of precipitation and clouds on the brightness temperatures of microwave window frequencies (19, 37, and 85 GHz) have been reported in Huang and Liou (1983), Mugnai and Smith (1988), and Gasiewski (1993). Shown in Fig. 7.26 is an example of the brightness temperature, T_B, as a function of rainfall rate for three frequencies, 18, 37, and 85.6 GHz, over land and the ocean (Spencer *et al.,* 1989). Over land, T_B decreases with the rainfall rate, an effect that is exacerbated at higher frequencies. Over the ocean, because of lower surface emissivity, T_B initially increases with the rainfall rate.

7.5.3 Temperature Retrieval from Microwave Sounders

The basic concept of inferring atmospheric temperatures from satellite observations of thermal microwave emission in the oxygen spectrum was first developed by Meeks and Lilley (1963), who presented microwave weighting functions for retrieval purposes. The prime advantage of microwave over infrared sounders is that the longer wavelength microwaves are much less influenced by clouds and precipitation. Consequently, microwave sounders can be more effectively utilized when clouds are present.

The first application of microwave techniques for temperature profile determination from an orbiting satellite was the Nimbus 5 Microwave Spectrometer experiment which was designed to evaluate passive microwave techniques for use on operational meteorological satellites. It consisted of three channels centered at frequencies 53.65, 54.90, and 58.80 GHz in the oxygen band (Water *et al.,* 1975). An operational version of the SCAMS, the Microwave Sounding Unit (MSU), was placed on the NOAA series beginning in 1978. It consists of four channels: 50.30, 53.74, 54.96, and 57.95 GHz. The high stability of the MSU instruments has been used to monitor global temperature trends. The approach to recovering temperature profiles using the microwave sounder has generally followed the statistical inversion principle discussed in Subsection 7.4.3.2. Basically, the predictor **D** matrix is derived from *a priori* atmospheric data provided by radiosonde observations.

A more advanced instrument, called the Special Sensor Microwave/Temperature (SSM/T) sounder, was developed as part of the Defense Meteorological Satellite Program (DMSP) aboard the Air Force Block 5D satellite system, first launched in June 1979, and subsequently in 1983, 1987, and 1991. This microwave sensor contains seven channels at 50.5, 53.2, 54.35, 54.9, 58.4, 58.825, and 59.4 GHz. Because of the surface reflectivity effect, the weighting function $K_{\tilde{v}}(p)$ is defined by

$$T_B(\tilde{v}) = \varepsilon_{\tilde{v}} T_s T_{\tilde{v}}(p_s) + \int_{p_s}^{0} T(p) K_{\tilde{v}}(p) \, dp. \qquad (7.5.18a)$$

From Eq. (7.5.6), we have

$$K_{\tilde{v}}(p) = 1 + (1 - \varepsilon_{\tilde{v}}) \left[\frac{T_{\tilde{v}}(p_s)}{T_{\tilde{v}}(p)} \right]^2 \frac{\partial T_{\tilde{v}}(p)}{\partial p}. \qquad (7.5.18b)$$

Displayed in Fig. 7.27a are the weighting functions of seven SSM/T channels for an incident (emergent) angle of $0°$. Absorption due to molecular oxygen and water vapor, along with antenna gain characteristics, were included in the transmittance calculations. Channel 1 is a window channel that responds strongly to the earth's surface characteristics, dense clouds, and rain. It is used to correct the other channels for these background effects. The weighting function peaks of channels 1 to 4 are below about 10 km, and therefore, dense clouds and precipitation would have some effect on the temperature retrieval.

As a result of the surface emissivity's effect on the brightness temperature, it is desirable to remove it in the statistical method of temperature retrieval so that
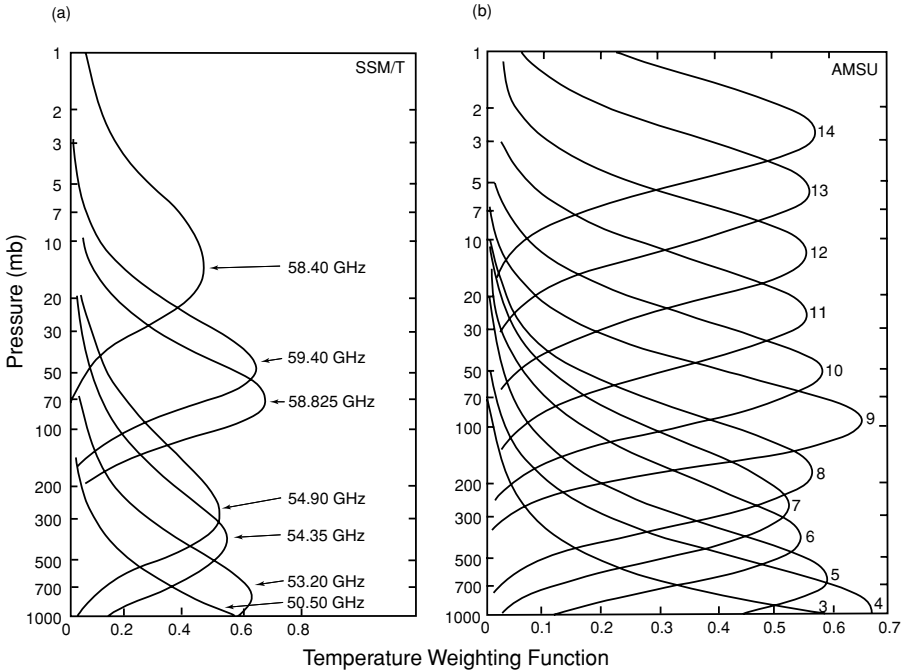
Figure 7.27   (a) Weighting functions at nadir position over land for seven SSM/T channels; (b) weighting functions at nadir viewing over land for 12 AMSU channels (data taken from Grody, 1993).

the predictor matrix $\mathbf{D}$ can be constructed over all surface conditions (Rigone and Stogryn, 1977). For simplicity of analysis, we define

$$T_d(\tilde{\nu}) = \int_0^{p_s} T(p) \frac{\partial T_{\tilde{\nu}}(p)}{\partial p} \, dp,$$

$$T_u(\tilde{\nu}) = \int_{p_s}^0 T(p) \frac{\partial T_{\tilde{\nu}}(p)}{\partial p} \, dp, \qquad (7.5.19a)$$

so that Eq. (7.5.6) can be written in the form

$$T_B(\tilde{\nu}) = \varepsilon_{\tilde{\nu}} T_s T_{\tilde{\nu}}(p_s)[1 - T_d(\tilde{\nu})/T_s] + T_a(\tilde{\nu}), \qquad (7.5.19b)$$

where

$$T_a(\tilde{\nu}) = T_u(\tilde{\nu}) + T_d(\tilde{\nu}) T_{\tilde{\nu}}(p_s). \qquad (7.5.19c)$$

The second term on the right-hand side of Eq. (7.5.19b) denotes the contribution to the brightness temperature caused solely by the atmosphere while the surface effects are confined in the first term. The channel in the wing whose weighting function peaks at the surface is used to remove the surface contribution for other channels. Based on Eq. (7.5.19b), we may define the contribution to the brightness temperature caused primarily by the atmosphere for channels 2 to 7 in the

form

$$T_a(\tilde{v}_j) = T_B(\tilde{v}_j) - [T_B(\tilde{v}_1) - T_a(\tilde{v}_1)]a(\tilde{v}_j), \qquad j = 2, 3, \ldots, 7, \quad (7.5.20)$$

where

$$a(v_j) = \frac{\varepsilon_{\tilde{v}_j} T_s T_{\tilde{v}_j}(p_s)[1 - T_d(\tilde{v}_j)/T_s]}{\varepsilon_{\tilde{v}_1} T_s T_{\tilde{v}_1}(p_s)[1 - T_d(\tilde{v}_1)/T_s]}, \quad \text{and} \quad a(\tilde{v}_1) = 1.$$

The statistical method described in Subsection 7.4.3.2 assumes a correlation between the atmospheric temperature and measured data, T_a in the present case, defined by Eq. (7.5.20). It follows that

$$\begin{aligned}
(\hat{T}_i - \bar{T}_i) &= \sum_j D_{ij}(\hat{T}_{aj} - \bar{T}_{aj}) \\
&= \sum_j D_{ij}[\hat{T}_{Bj} - (\hat{T}_{B1} - \hat{T}_{a1})a_j - \bar{T}_{aj}] \\
&= \sum_{j \neq 1} D_{ij}\hat{T}_{Bj} - \bar{T}_{B1} \sum_j D_{ij}a_j + \sum_j D_{ij}(\hat{T}_{a1}a_j - \bar{T}_{aj}). \quad (7.5.21)
\end{aligned}$$

Note that T_{B1} is not defined in Eq. (7.5.20), and so the first term contains $j = 2, \ldots, 7$. In matrix notation, we have

$$\hat{\mathbf{T}} = \mathbf{D}'\hat{\mathbf{T}}_B + \mathbf{R}, \qquad (7.5.22)$$

where

$$\mathbf{R} = \bar{\mathbf{T}} + \hat{T}_{a1}\mathbf{Da} - \mathbf{D}\bar{\mathbf{T}}_a, \qquad (7.5.23)$$

and $\mathbf{D}'$ is a matrix whose first column is $-\mathbf{Da}$ and whose remaining columns are the columns of $\mathbf{D}$. The retrieval technique contains elements that are dependent mainly on the atmosphere. Thus, it should be valid over land, water, or mixed surface conditions. $\mathbf{D}$ and $\mathbf{R}$ can be determined from a large number of upper air soundings for a wide range of meteorological conditions that have been achieved over the years, while the brightness temperatures can be calculated for a given atmospheric condition.

Microwave temperature retrieval can be affected by the surface emissivity, as pointed out previously. At the frequencies around 50 GHz, the emissivity over land can vary between about 1 for dry or vegetated land to less than 0.8 for snow cover, bare wet land, rivers, and lakes. The emissivity for oceans varies because of wind-induced foam and surface roughness (see Fig. 7.25). Within the field of view of the microwave instrument, about 25–100 km resolution, the effect of emissivity is relatively small and can be taken into account by empirical approaches as shown above. The effects of nonprecipitating clouds on temperature retrieval are also small. Over the ocean, clouds can either increase or decrease the brightness temperature, depending on their position and LWC. Over high-emissivity land surfaces, clouds normally reduce the brightness temperature, with a maximum effect produced by high clouds having substantial liquid water. Effects due to precipitation are much more pronounced, as illustrated in Liou *et al.* (1981).
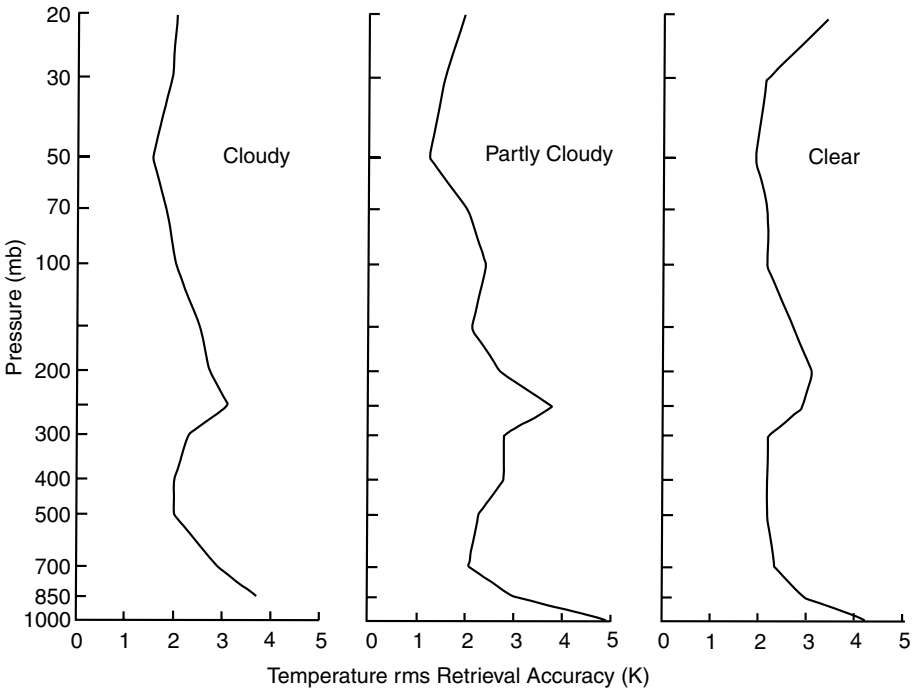
Figure 7.28   Differences between SSM/T retrieval results and the radiosondes stratified according to cloud conditions for November 14–17, 1983, in terms of temperature rms (data taken from Grody, 1993).

Figure 7.28 illustrates an example of the temperature retrieval accuracy based on the SSM/T sounder. The retrieved profiles in terms of the root-mean-square (rms) error are shown relative to radiosonde observations that were collected within a maximum time difference of 3 hours and a 1° latitude–longitude window between November 14 and 17, 1983, for midlatitudes (30–60°N) at pressure levels down to 20 mb (Grody, 1993). Results are shown separately for clear, partly cloudy, and completely cloudy conditions. The rms errors are generally within 2–3 K for pressure levels between 700 and 20 mb, independent of cloud cover. The largest temperature errors occur at pressure levels at which the weighting functions are inadequate and there is no strong correlation between level temperatures and the weighted temperatures sensed by the microwave instrument.

The latest generation of microwave sounders for research and development are the Advanced Microwave Sounding Unit (AMSU) with 12 oxygen channels, flown aboard the NOAA polar-orbiting satellites (1998), and the Air Force SSM/IS instrument. Figure 7.27b displays the weighting functions for the 12 oxygen channels at nadir over land. AMSU also contains channels near the 183 and 22 GHz H_2O lines for water vapor soundings and window channels at 31, 90, and 157 GHz for the inference of precipitation and surface parameters.

7.6   Remote Sensing Using Laser and Microwave Energy

During the 1960s, the advent of the laser as a source of energy opened up a number of possibilities for new remote sensing techniques of probing the atmosphere. Laser energy at optical frequencies is highly monochromatic and coherent. With the development of Q-switching techniques, very short pulses of high power could be generated. The recognition of the applicability of high power pulsed laser energy by a number of atmospheric scientists (Schotland, 1969; Collis, 1969) has prompted the development of backscattering lidar techniques for the detection of the composition and structure of clouds, aerosols, and minor gases in the atmosphere. The word *lidar,* which originally stood for *L*ight *D*etection *A*nd *R*anging, is an acronym analogous to radar that utilizes an energy source in the microwave region. In a sense, lidar can best be described as a laser radar. Since the development of laser radar, which employs the same backscattering principle as microwave radar, advanced techniques for atmospheric probes have proved fruitful. These include the use of multiple wavelength lidars for determining the composition of minor gases by means of differential absorption techniques, the use of Doppler techniques for determining the motion of particulates and molecules, the use of depolarization techniques for inferring the water and ice content in clouds, and the use of the Raman scattering technique, in which a weak scattering occurs at a shifted wavelength for water vapor measurements. In line with the discussions of the theory of light scattering and radiative transfer introduced in Chapters 5 and 6, we first present the basic lidar (or radar) equation that is fundamental to all backscattering techniques and then discuss two useful techniques involving the use of absorption and depolarization properties for atmospheric studies. That is followed by a discussion of the use of millimeter (mm)-wave radars for cloud study.

7.6.1   Backscattering Equation: Theoretical Foundation

From the light-scattering theory developed in Sections 3.3 and 5.2, the scattered intensity for a single particle or molecule can be expressed as [see Eqs. (5.2.84) and (5.2.111a,b), also Eq. (3.3.15)]

$$I^s_{\parallel,\perp} = I^i_{\parallel,\perp} \frac{\sigma_s}{r^2} \frac{P_{2,1}(\Theta)}{4\pi},\tag{7.6.1}$$

where the subscripts $\parallel$ (2) and $\perp$ (1) denote the light beam parallel and perpendicular to the scattering plane, respectively, r is the distance at which scattering takes place, σ_s is the scattering cross section, and $P(\Theta)$ is the phase function. For the convenience of the following discussion, we shall neglect the subscripts on the flux density and phase function. The backscattered ($\Theta = \pi$) flux density due to a single particle from Eq. (7.6.1) can be expressed by

$$F^s(\pi) = F^i \frac{\sigma_s P(\pi)}{r^2 \, 4\pi}.\tag{7.6.2}$$

For backscattering applications, it is necessary to define the *backscattering cross section* σ_π as the area that, when multiplied by the incident flux density, gives the

total power radiated by an isotropic source such that it radiates the same power in the backward direction as the scatterer. Thus,

$$F^i \sigma_\pi = F^s(\pi) 4\pi r^2, \tag{7.6.3}$$

where $4\pi r^2$ represents the surface area of a sphere. It follows that the backscattering cross section is given by

$$\sigma_\pi = \sigma_s P(\pi). \tag{7.6.4}$$

Let P_t denote the transmitted power so that the incident flux density can be expressed by

$$F^i = P_t/A_t, \tag{7.6.5}$$

where A_t is the cross-sectional area at distance r. If A_r is the collecting aperture, then the backscattered power received is given by

$$P_{r0} = F^s(\pi)A_r = \frac{P_t}{A_t} \frac{\sigma_s P(\pi)}{4\pi r^2} A_r. \tag{7.6.6a}$$

In reference to the (volume) scattering coefficient given by Eq. (5.2.116), we may define an average scattering cross section as $\bar{\sigma}_s = \beta_s/N$, where N denotes the particle number density. Hence, after the particle-size distribution integration, the average backscattered power is given by

$$\bar{P}_{r0} = P_t \frac{A_r}{r^2} \frac{P(\pi)}{4\pi} \frac{\bar{\sigma}_s}{A_t}. \tag{7.6.6b}$$

Let the pulse length transmitted by a lidar (or radar) system be Δh, as shown in Fig. 7.29. For a given instant of time, the lidar receiver collects the scattered energy from half of the pulse length, so that the bottom pulse will undergo roundtrip backscattering and return simultaneously as the top pulse. The total number of particles within the effective scattering volume is, therefore, $N A_t \Delta h/2$. Upon utilizing the definition of the volume scattering coefficient, the backscattered power is now given by

$$\bar{P}_{r0} = P_t \frac{A_r P(\pi)}{r^2 \, 4\pi} \beta_s \frac{\Delta h}{2}. \tag{7.6.6c}$$

During the backscattering event, the energy pulse also undergoes attenuation. On the basis of the Beer–Bouguer–Lambert law introduced in Section 1.4.2, the actual backscattered power at the point of the receiver is

$$\bar{P}_r = \bar{P}_{r0} \exp\left\{ -2 \int_0^r \beta_e(r')\,dr' \right\}, \tag{7.6.7a}$$

where 2 represents the roundtrip attenuation effect, $r = 0$ is the position corresponding to $\bar{P}_{r0}$, and β_e is the (volume) extinction coefficient including both scattering and absorption effects. If we insert Eq. (7.6.6c) into Eq. (7.6.7a), the backscattered power observed by the receiver can now be written in the form

$$\bar{P}_r(r) = P_t \frac{A_r P(\pi)}{r^2 \, 4\pi} \beta_s \frac{\Delta h}{2} \exp\left\{ -2 \int_0^r \beta_e(r')\,dr' \right\}. \tag{7.6.7b}$$
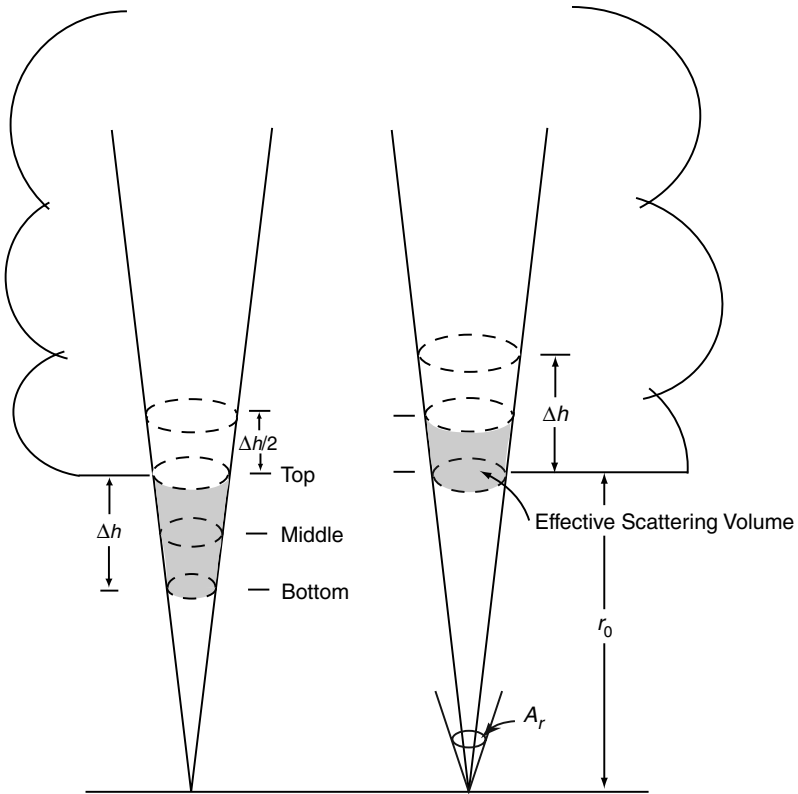
Figure 7.29   The backscattering geometry of a pulsed lidar system where r_0 denotes a reference range, Δh is the pulse length, and A_r is the receiver aperture. The effective length under which the pulsed laser light returns to the receiver is $\Delta h/2$.

In this development, we have neglected the energy gains corresponding to the transmitter and receiver. For lidar or radar applications, it is customary to use the volume backscattering coefficient similar to that given in Eq. (7.6.4), i.e., $\beta_\pi = P(\pi)\beta_s$. Thus, we rewrite Eq. (7.6.7b) to obtain

$$\bar{P}_r(r) = \frac{P_t C A_r \beta_\pi(r) \Delta h}{8\pi r^2} \exp\left\{ -2 \int_0^r \beta_e(r')\, dr' \right\}, \qquad (7.6.7c)$$

where we have added a known instrument factor C to the equation. This is the basic lidar (or radar) equation, where the collecting aperture A_r, the transmitted power P_t, and the pulse Δh are known parameters and the range r is a function of time t. However, there are two unknown parameters, β_π and β_e, which relate to the optical properties and concentration of particles and/or molecules. It is not possible to investigate the information content of the return power in absolute terms unless the volume backscattering coefficient, β_π, and the volume extinction coefficient, β_e, are uniquely correlated.

For Rayleigh scattering, the phase function for unpolarized light in the backscattering direction from Eq. (3.3.11) is $P^R(\pi) = 3(1 + \cos^2 \pi)/4 = 1.5$. Thus, $\beta_\pi^R = 1.5\beta_e^R$. It is clear that the ratio of backscattering to the extinction coefficient for Rayleigh molecules is a constant of 1.5 and is not subject to fluctuations. For a single Lorenz–Mie spherical particle, however, the phase function depends on the size parameter $2\pi a/\lambda$ and is characterized by strong forward scattering (see, e.g., Fig. 5.8). The backscattering phase function $P(\pi)$ as a function of the size parameter fluctuates greatly and is normally less than unity. For a sample of polydispersed spheres, the fluctuations tend to average out (see Figs. 5.13 and 5.15), and useful approximate values can be determined for the term. For water clouds, it has been found that for many cases $\beta_\pi \approx 0.625\beta_e$, a reasonable approximation also applicable to spherical aerosols. In general, it has been assumed that $\beta_\pi \cong a\beta_e^b$, where a and b are certain coefficients to be determined from light-scattering calculations.

The inversion of lidar backscattering data to obtain the volume extinction coefficient profile is the subject of continued research in the lidar field. It begins with the definition of the signal variable defined by

$$S(r) = \ln[r^2 \bar{P}_r(r)]. \tag{7.6.8a}$$

In differential form, we have

$$\frac{dS(r)}{dr} = \frac{1}{\beta_\pi}\frac{d\beta_\pi}{dr} - 2\beta_e(r). \tag{7.6.8b}$$

Having the relation between β_π and β_e established, techniques can be developed to infer $\beta_e(r)$ from $S(r)$, subject to the definition of a reference range *a priori*. The volume extinction coefficients can in turn be related to the sizes and concentrations of aerosols and cloud particles. In the following, we present two specific techniques for the detection of gaseous profiles and for cloud and aerosol studies.

7.6.2 Lidar Differential Absorption and Depolarization Techniques

7.6.2.1 DIFFERENTIAL ABSORPTION TECHNIQUE

The differential absorption lidar (DIAL) uses the gas absorption properties in two wavelengths and requires a tunable laser to produce the peak of an absorption line of the gas of interest and a second wavelength in a low-absorption region. In clear atmospheres, the extinction coefficient for a given wavelength is contributed by the extinction due to aerosols and the absorption caused by the pertinent gas. Thus, we can write

$$\beta_e(\lambda) = \beta_{e,A}(\lambda) + \rho_a k(\lambda), \tag{7.6.9}$$

where ρ_a is the gas density and $k(\lambda)$ is the absorption coefficient. We may select two wavelengths at which the optical properties of aerosols are about the same. Further, we can perform a logarithmic operation on Eq. (7.6.7c) to obtain the normalized backscattered power as follows:

$$\ln[\bar{P}_r(r)/P_t]_i = \ln\left[\frac{C A_r \beta_\pi(r)\Delta h}{8\pi r^2}\right]_i - 2\int_0^r \beta_e(\lambda_i)\,dr, \qquad i = 1, 2. \tag{7.6.10}$$

Since λ_1 is close to λ_2, the scattering and extinction properties of aerosols are about the same so that $\beta_\pi(\lambda_1) \cong \beta_\pi(\lambda_2)$ and $\beta_{e,A}(\lambda_1) \cong \beta_{e,A}(\lambda_2)$. We can then carry out the subtraction operation to obtain

$$\ln \hat{P}_{1,2} \cong -2 \int_0^r \rho_a(r)[k(\lambda_1) - k(\lambda_2)] \, dr, \qquad (7.6.11a)$$

where

$$\hat{P}_{1,2} = [\bar{P}_r(r)/P_t]_1/[\bar{P}_r(r)/P_t]_2. \qquad (7.6.11b)$$

Thus, once the absorption coefficients are known, the backscattered measurements at λ_1 and λ_2 can provide the gas density profile. Recent technology has produced lasers that can be accurately tuned to the absorption lines of various gases such as H_2O, NO_2, SO_2, and O_3. This has led to the development of DIAL systems for the measurement of their profiles in the atmosphere. The capability of the DIAL system has been presented in a summary paper by Browell (1994).

7.6.2.2   PRINCIPLE OF DEPOLARIZATION

In the preceding discussion, we neglected polarization information in the development of the lidar equation. In reference to Eq. (7.6.6a), the transmitted power P_t can be generated to be vertically ($P_{t,\perp}$) or horizontally ($P_{t,\parallel}$) polarized, while the detector can be devised so that both polarization components ($\bar{P}_{r,\parallel}, \bar{P}_{r,\perp}$) are measured. This provides additional information about particle characteristics and is referred to as the *depolarization technique*. The principle of depolarization associated with a laser beam that is backscattered by cloud particles can be described by the laws of geometric optics. These laws follow the physical principle that the size of a particle is much larger than the incident wavelength so that light beams can be localized as geometric rays. Each ray that hits the particle will undergo reflection and refraction and will pursue its own specific path along a straight line according to Snell's law at the particle surface (Sections 5.3 and 5.4).

We shall use the geometric optics principle and apply it to the scattering of spherical droplets and hexagonal ice crystals, as shown in Fig. 7.30. Backscattering from a symmetrical sphere can arise from a number of possibilities: the edge rays; a central ray that undergoes external reflection; and two refractions, with the first making the maximum contribution (see Section 5.3). Because no additional reference plane is formed for the 180° backscattered rays, the light beam would retain the polarization state of the incident energy. However, the backscattered energy from a nonspherical particle can be produced by internal reflections that rotate the initial vibration plane of the electric vector and produce depolarization. Shown in Fig. 7.30 are two types of ice crystals: solid column and plate. For these ice crystals to produce backscattering, light rays must undergo internal reflections more than twice, except for normal incidence. In the case of normal incidence, a laser beam transmitting through a plate will retain its polarization state because of symmetry. The incident skew rays that produce backscattering undergo internal reflections that transform the electric vectors from the original reference plane to the planes that contain the incident and refracted rays. Consider a vertically polarized incident laser beam. A cross-polarization component
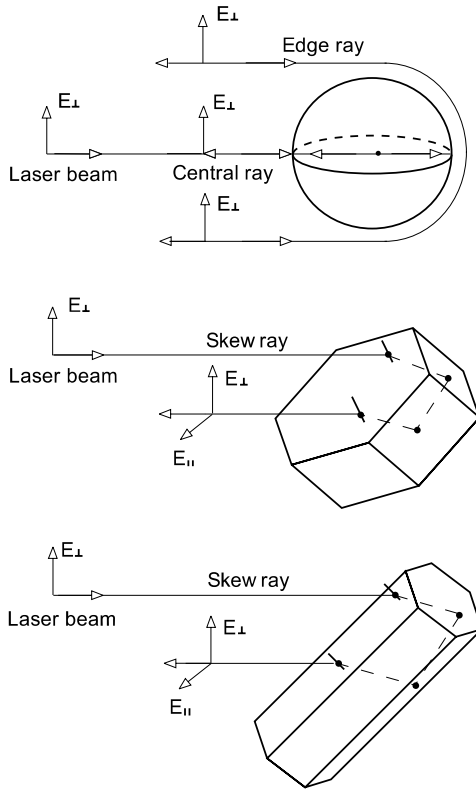
Figure 7.30   Principle of backscattering depolarization from a spherical water droplet, a hexagonal plate, and a solid column. The incident laser beam is vertically polarized. The backscattered beam from a sphere retains the polarization state because of its geometrical symmetry. The backscattered beams from nonspherical particles produce a cross-polarization component for light rays that are not perpendicular to the major axis of the ice particle.

can be produced after the internal reflections. The ratio of this component to the component that retains the same polarization as the incident light beam is referred to as the *backscattering depolarization ratio* defined by

$$\delta = \bar{P}_{r,\parallel}/\bar{P}_{r,\perp}. \tag{7.6.12}$$

The theoretical foundation of depolarization from nonspherical ice crystals was first discussed by Liou and Lahore (1974). Sassen (1976, 1991) presented the depolarization ratio for various types of ice crystals and water droplets derived from laboratory and field studies. For water droplets, this ratio is close to zero, as is predicted from the theory. For ice crystals, the depolarization ratio varies from 50 to 70%, depending on their shape and size. This provides a powerful means of distinguishing between water and ice clouds in the atmosphere. The backscattering depolarization technique can also be used to probe the orientation properties of ice particles. In the case of a vertically pointing lidar involving horizontally oriented plate crystals, the depolarization ratio would be zero because of the symmetry of the two parallel

crystal surfaces with respect to the laser beam. Lidar experiments have shown that the backscattering depolarization ratio in this case is close to zero, whereas it increases significantly as the lidar scans a few degrees off the vertical (Platt *et al.,* 1978).

An example of backscattering depolarization studies of cirrus clouds is illustrated in Fig. 7.31. Sassen and Hsueh (1998) and Freudenthaler *et al.* (1996) presented the depolarization ratios determined from high-resolution 0.532/1.06 μm polarization lidar for contrail cirrus. The former authors showed that the lidar depolarization ratio in persisting contrails ranged from about 0.3 to 0.7; the latter authors observed
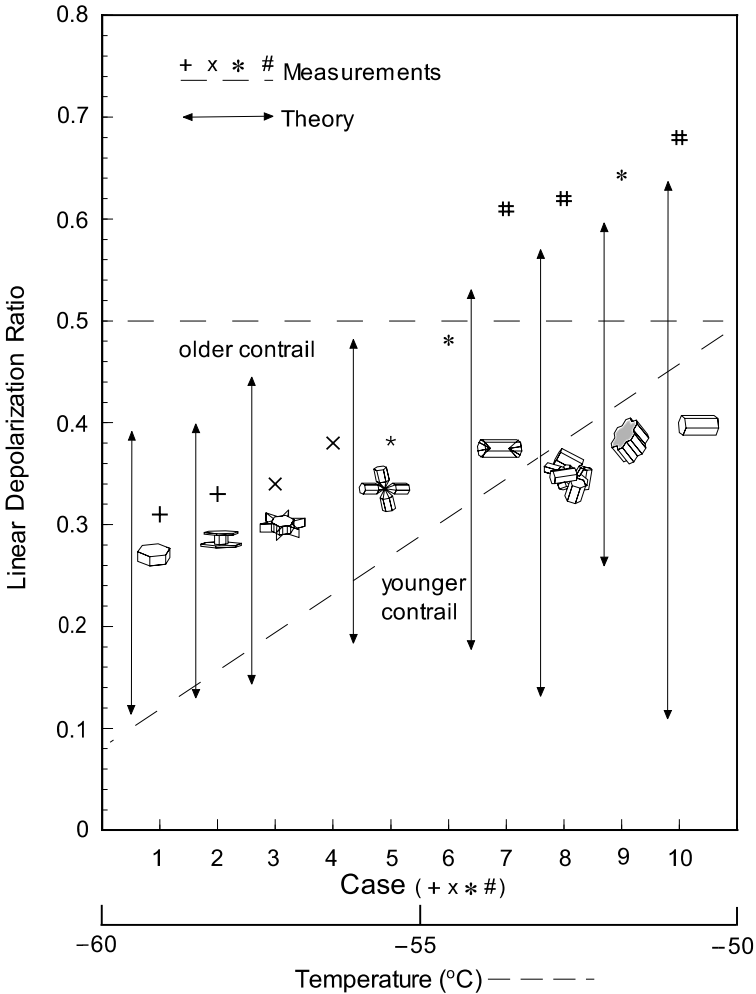
Figure 7.31   Depolarization ratios determined from high resolution polarization lidar for contrail cirrus (Sassen and Hsueh, 1998, case; Freudenthaler *et al.,*1996, temperature) and computed from a unified theory for light scattering by ice crystals (Section 5.4) with shapes ranging from single and double plates, solid and hollow columns, dendrites, bullet rosettes, aggregates, and irregular surface, the sizes of which span from a few micrometers to the geometric optics limit (after Liou *et al.,* 2000).

this ratio from 0.1 to 0.5 for contrails with temperatures ranging from -60 to $-50°C$, depending on the stage of their growth. For purposes of interpretation, displayed in this figure are backscattering depolarization ratios for various ice crystal sizes and shapes computed from the unified theory for light scattering by ice crystals described in Section 5.4. The vertical bars indicate the results for ice crystals of a few micrometers to the geometric optics region. The depolarization ratio generally becomes larger for larger ice particles and reaches a maximum of about 0.6 for size parameters in the geometric optics limit.

One complication in the backscattering and depolarization experiments has been the effect of multiple scattering. Through multiple scattering events, the incident electric vector is transferred from the initial reference plane to the plane of scattering, and therefore, partial depolarization is produced, even for clouds consisting of spherical water droplets. The relative significance of multiple scattering in backscattering experiments involving clouds is generally associated with the field of view of the detector. Multiple backscattering from cloud particles has been a subject of extensive research since the 1970s. The most recent reference on this subject is Wandinger (1998).

7.6.3   Millimeter-Wave Radar for Cloud Study

The conventional meteorological radars typically operate at centimeter (cm) wavelengths (e.g., 10 cm and 3.21 cm radars) to circumvent problems in the interpretation of backscattering signals from precipitating clouds that contain large raindrops and snowflakes on the order of millimeter (mm) to cm sizes. Nonprecipitating clouds generally consist of particles that are much smaller than mm sizes and do not produce significant radar backscattering signals for their detection. Radar meteorology is a well-developed discipline and will not be covered in this text. However, the technical development of mm-wave radar and its application to cloud studies is a subject of contemporary research. In particular, when mm-wave radar data are combined with lidar measurements, comprehensive cloud structure and composition can be inferred.

The selection of wavelength for a radar operated in the earth's atmosphere is restricted to those spectral regions where absorption by atmospheric gases is smaller. In reference to Fig. 7.23, the windows of H_2O and O_2 absorption are located at 35 GHz (8.6 mm, K_a-band), 94 GHz (3.2 mm, W-band), 140 GHz (2.14 mm, F-band), and 220 GHz (1.36 mm, G-band). The first mm-wave radar designed for meteorological use was a 35 GHz system for cloud deck monitoring. It was subsequently equipped with Doppler and polarization capabilities and employed for the observation of clouds. A 94 GHz system with a shorter wavelength was further shown to provide effective observations of the liquid/ice water content of clouds (Lhermittee, 1990). Millimeter-wave radars have two distinct advantages over cm-wave radars in that very narrow beams can be produced with small antenna size and the gain from Rayleigh scattering can reduce the need for high-power transmitters. In recent years, both 35 and 94 GHz radars have been well developed for the detection of nonprecipitating clouds.

Since cloud particles are normally much smaller than mm wavelengths, the Rayleigh scattering theory presented in Section 3.3.1 may be followed. The scattering

cross section for a cloud particle may be expressed by

$$\sigma_s = \frac{128\pi^2}{3\lambda^4}\left(\frac{3}{4\pi N_c}\right)^2\left|\frac{m^2-1}{m^2+2}\right|^2, \tag{7.6.13a}$$

where the number density of spherical particles $N_c = 1/V$ and the volume $V = 4\pi a^3/3$. Letting $K = (m^2 - 1)/(m^2 + 2)$, we have

$$\sigma_s = \frac{128\pi^2}{3\lambda^4}a^6|K|^2. \tag{7.6.13b}$$

Referring to Eq. (7.6.4), noting the Rayleigh backscattering phase function $P^R(\pi) = 1.5$, and using the diameter $D = 2a$ for a cloud particle, the backscattering cross section (in units of cm^2) can then be written in the form

$$\sigma_\pi = \sigma_s \cdot P^R(\pi) = \frac{\pi^5}{\lambda^4}|K|^2 D^6. \tag{7.6.14}$$

Let the cloud particle size distribution be $n(D)$. The backscattering coefficient (in units of cm^{-1}) is then

$$\beta_\pi = \frac{\pi^5}{\lambda^4}|K|^2 Z, \tag{7.6.15}$$

where the radar reflectivity factor

$$Z = \int D^6 n(D)\,dD. \tag{7.6.16}$$

The average return power without accounting for attenuation can now be written in the form

$$\bar{P}_r(r) = \frac{C}{r^2}|K|^2 Z, \tag{7.6.17}$$

where C is a constant associated with the radar characteristics. Measurements of $\bar{P}_r(r)$ versus r allow the calculation of Z, which in turn can be related to the cloud properties.

The radar reflectivity factor defined in Eq. (7.6.16) is related to the particle diameter to the sixth power. As discussed in Section 7.3.5, the liquid (or ice) water content (LWC/IWC) is associated with the particle radius to the third power. Thus, one may correlate the radar reflectivity factor with the cloud LWC/IWC. The 35 GHz cloud radar has been developed specifically for the determination of the vertical profile of IWC in high-level cirrus clouds (Mace *et al.*, 1998). Moreover, with the Doppler technique, some information about the ice-crystal size based on the fall velocity may also be inferred. The development of a combination of the reflectivity and Doppler spectrum technique for the simultaneous retrieval of the vertical profile of IWC and ice-crystal size in cirrus is still in its embryonic stage. In Fig. 7.32, we demonstrate a time series of the retrieved IWC and ice crystal size in terms of D_e, defined in Eq. (5.1.1), as functions of height for a cirrus cloud case observed in Oklahoma on April 18, 1997, based on the backscattering return of a 35 GHz Doppler radar. The cloud top heights were nearly constant between about 10 to 11 km, but the cloud base heights varied significantly. The retrieved IWC and D_e ranged from 0 to 0.6 g m^{-3} and 0 to 170 μm, respectively. Substantial variabilities in both the
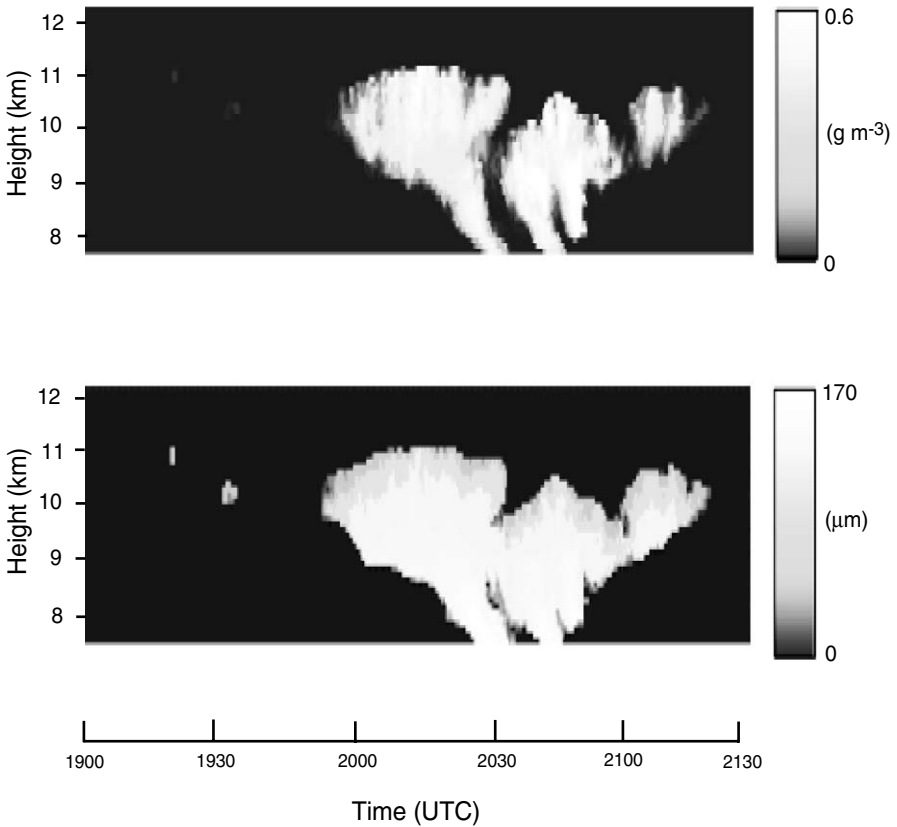
Figure 7.32 Time series of the ice water content (IWC, top panel) and mean effective ice crystal size (D_e, lower panel), determined from the backscattering return of a 35 GHz Doppler radar, as functions of height for a cirrus cloud case observed in Oklahoma on April 18, 1997 (data provided by Gerald Mace of the University of Utah).

vertical and horizontal directions are clearly displayed. It is evident that the successful development of mm-wave cloud radar will revolutionize our understanding of the structure of inhomogeneous clouds, particularly for high-level cirrus.

Exercises

7.1 (a) Derive Eq. (7.2.7) from Eqs. (7.2.5) and (7.2.6) and show that

$$k = \pi (2\pi)^{\nu^* - 2} \int_0^{z_\infty} C(z)\, dz \int_{x_1}^{x_2} \frac{Q_e(x)\, dx}{x^{\nu^* - 1}},$$

where the extinction efficiency Q_e is expressed in terms of the size parameter x. (b) Direct solar radiation measurements are made with a multiple wavelength radiometer. The aerosol optical depths derived from the observations are 0.17

and 0.1 at 0.63 and 0.86 μm, respectively. Assuming a Junge size distribution and a constant k, what would be the shaping factor ν^*?

7.2 Consider the following Fredholm equation of the first kind:

$$g(k) = \int_0^1 e^{-kx} f(x) \, dx,$$

where the kernel is given by a simple exponential function. Let the unknown function be given in the form

$$f(x) = x + 4x\left(x - \frac{1}{2}\right)^2.$$

(a) Derive an analytical expression for $g(k)$ and compute $g(k_i)$ for k_i in the interval $(0, 10)$ using $\Delta k_i = 0.5$ ($i = 1, 2, \ldots, 20$).

(b) Write the integral equation in summation form as

$$g(k_i) = \sum_{j=1}^{20} f(x_j)e^{-k_i x_j} \Delta x_j, \qquad i = 1, 2, \ldots, 20.$$

Letting $\Delta x_i = 0.05$, compute $g(k_i)$ again and compare with those computed from the exact integration.

(c) Let

$$A_{ij} = e^{-k_i x_j} \Delta x_j.$$

Compute $\|A_{ij}\|$, which is a 20×20 matrix. Use the direct linear inversion method to recover $f(x_j)$. Compare the retrieval results with the exact values.

7.3 Prove Eq. (7.2.25) from Eq. (7.2.23) by assuming $M = N = 2$. Using the first, second, and third differences, derive the **H** matrices similar to that shown in Eq. (7.2.26).

7.4 A seven-channel sunphotometer is used to measure aerosol optical depth following the procedure discussed in Section 7.2.1. The measured optical depths corresponding to the wavelengths depicted in Fig. 7.3 are $\tau_1(0.382\mu\text{m}) = 0.645$, $\tau_2(0.451\mu\text{m}) = 0.562$, $\tau_3(0.501\mu\text{m}) = 0.513$, $\tau_4(0.526\mu\text{m}) = 0.491$, $\tau_5(0.778\mu\text{m}) = 0.338$, $\tau_6(0.861\mu\text{m}) = 0.306$, and $\tau_7(1.060\mu\text{m}) = 0.249$. Using the weighting functions given in this figure, determine the aerosol size distribution from the following procedures:

(a) Use Eq. (7.2.12) and expand $f(a)$ in a seven-term Legendre polynomial defined by

$$P_j(a) = \frac{1}{2^j j!} \frac{d^j}{da^j}(a^2 - 1)^j.$$

(b) Then, from Eq. (7.2.13b), compute

$$A_{ij} = \int_{a_1}^{a_2} P_j(a) K_i(a) \, da, \qquad i = 1, 2, \ldots, 7,$$

where $a_1 = 0.01$ μm and $a_2 = 1$ μm. Use the data presented in Table 7.2 and a quadrature integration method in the computation.

Table 7.2

Weighting Functions as Functions of Aerosol Radius for Seven Sunphotometer Wavelengths

Radius (μm)	Weighting functions						
	1	2	3	4	5	6	7
0.0100	4.4328	2.2811	1.4978	1.2327	0.2575	0.1717	0.0747
0.0176	13.7007	7.0492	4.6281	3.8086	0.7953	0.5302	0.2307
0.0260	29.9713	15.4296	10.1313	8.3375	1.7407	1.1602	0.5049
0.0308	42.0547	21.6767	14.2381	11.7184	2.4469	1.6309	0.7097
0.0466	93.8568	49.0855	32.3937	26.7002	5.5989	3.7324	1.6241
0.0509	110.4079	58.2449	38.5510	31.8052	6.6892	4.4599	1.9409
0.0690	177.1635	100.1537	68.0051	56.5812	12.2394	8.1761	3.5645
0.0766	198.6910	117.1921	81.0726	67.8926	15.0261	10.0546	4.3907
0.0878	222.1855	138.4621	99.1376	84.1119	19.5691	13.1463	5.7637
0.0989	244.8080	153.7768	113.6846	97.9175	24.3999	16.4917	7.2761
0.1013	251.0553	156.6501	116.4091	100.5890	25.5108	17.2721	7.6341
0.1107	278.9088	167.4753	125.3450	109.3831	29.8075	20.3415	9.0670
0.1273	309.0414	193.7464	139.8560	121.7655	37.3108	25.9747	11.8390
0.1306	307.8837	200.0057	143.3598	124.3373	38.7264	27.0920	12.4191
0.1496	287.9718	221.6949	168.0708	143.8293	45.9078	33.1677	15.8485
0.1586	288.5376	218.4273	176.9670	154.0849	48.6430	35.7095	17.4849
0.1668	293.5934	212.1231	179.6682	160.7122	50.7921	37.7829	18.9532
0.1759	292.5142	206.8395	177.1622	162.9790	52.8937	39.7875	20.5186
0.2157	246.2814	203.6015	169.9123	151.7108	63.7010	46.7642	26.1580
0.2357	221.1176	182.9026	167.6034	155.3636	70.7781	51.2328	28.1459
0.2469	201.5677	177.9200	159.1576	152.3555	73.4846	54.2184	29.1279
0.2549	192.6196	176.6381	152.7078	147.4533	74.3904	56.3492	29.8385
0.2600	189.4106	175.0724	149.2663	143.8085	74.4998	57.5937	30.3113
0.2777	172.3674	159.3578	143.7655	133.2944	72.8265	60.4652	32.1875
0.2871	156.2557	148.8443	142.3785	130.9186	71.3910	60.8324	33.3749
0.2917	148.8346	144.4151	140.8010	130.4095	70.7245	60.7327	34.0098
0.3075	133.4110	135.7173	129.9356	127.2397	69.2935	59.4388	36.2900
0.3245	119.9005	126.8022	116.7393	116.6808	69.6569	57.5826	38.5157
0.3372	101.8229	113.8690	111.0767	107.9117	70.6203	56.6867	39.6402
0.3521	87.9238	100.7208	107.1272	101.2790	70.9368	56.5918	40.1344
0.3882	60.6246	81.6000	83.1148	87.1008	64.9331	57.9546	38.6334
0.4093	52.7835	65.6209	76.3747	74.6244	61.1072	56.2177	37.5694
0.4263	43.3359	61.1699	69.2994	69.8766	59.7750	53.5649	37.2795
0.4441	32.8936	52.4110	58.1053	64.5868	59.1438	50.8691	37.5811
0.4628	32.5212	41.2741	50.9002	54.4527	56.9318	49.1611	38.1275
0.4953	21.5198	34.6967	41.5344	45.1617	49.9987	48.1017	37.6474
0.5085	23.7100	28.7729	35.4682	42.3528	47.6895	46.9398	36.7315
0.5286	19.8040	23.0721	30.7868	34.1439	45.7240	43.9804	35.0077
0.5419	17.9112	23.8040	30.2158	29.7947	44.8635	41.7830	33.9158
0.5681	19.3692	17.9502	22.3759	27.5200	41.0542	38.4228	32.4880
0.6218	19.5171	14.4908	16.2054	16.9793	32.4441	34.4953	31.2234
0.6679	19.6098	13.9251	13.7338	12.4354	27.7156	28.0536	27.4974
0.7033	17.9602	13.5937	10.6527	12.3184	22.3268	25.8595	25.1727
0.7363	16.7441	13.8903	11.4802	9.7987	20.5210	23.0045	24.2301
0.7768	14.0616	13.6846	10.5095	10.2743	16.2295	18.3615	21.8955

Continues

Continued

Radius	Weighting functions						
(μm)	1	2	3	4	5	6	7
0.8101	13.0399	13.0259	11.3784	9.3696	13.0756	16.8683	19.2823
0.8485	9.7714	12.8928	11.9151	10.2819	12.1259	14.3272	17.4424
0.8695	8.8943	12.0394	11.2795	9.7739	10.3275	12.3727	16.9763
0.8917	7.7516	11.1963	10.6176	10.8465	8.5943	10.8745	16.0946
0.9151	7.2615	10.1764	11.0882	10.3081	7.7314	10.4032	14.5823
0.9396	6.5278	9.2237	10.6382	9.5603	7.9324	9.8656	12.9847
0.9523	6.7532	9.3616	9.8044	10.1288	7.3905	9.0824	12.3006
0.9784	6.3343	7.9969	9.5029	9.6135	6.1033	7.4553	11.3872
0.9918	6.2839	7.4044	9.1160	9.3268	5.5344	6.7934	11.1889
1.0055	6.2318	7.1540	8.9166	8.7280	5.2022	6.3777	11.0159

(c) Based on Eq. (7.2.27), compute $\mathbf{f}$ ($j = 1, 2, \ldots 7$) by using the $\mathbf{H}$ derived for the second difference obtained from Exercise 7.3. The aerosol radii corresponding to the peaks of the weighting function are, respectively, $a_1 = 0.127$, $a_2 = 0.150$, $a_3 = 0.167$, $a_4 = 0.176$, $a_5 = 0.260$, $a_6 = 0.287$, and $a_7 = 0.352$ for wavelengths ranging from 0.382 to 1.060 μm. Use $\gamma = 10^{-3}$ and 10^{-6} in the inversion exercise.

(d) Compute the column aerosol size distribution from Eq. (7.2.9a) by assuming $v^* = 3$ and compare your results with those displayed in Fig. 5.1.

7.5 Observations from satellite visible imagery have shown that ships can influence the composition and radiative properties of shallow clouds (stratus and stratocumulus) over the oceans via their stack effluents. It is hypothesized that the additional aerosols serve as effective condensation nuclei upon which small water droplets form. This is referred to as the indirect aerosol effect in climate studies. Consider Eqs. (7.3.13b) and (7.3.13c) and use a mean effective radius $\bar{a}$ in the integrand. Define the total number density of water droplets as $N = \int n(a)\,da(\# \text{ cm}^{-3})$. What would be the increase in optical depth if N increases by a factor of 2? Remote sensing of the optical depths and droplet sizes of disturbed and pristine stratiform clouds over the oceans is a subject of considerable importance for the study of aerosol–cloud interactions.

7.6 Given the temperature profile and transmittances for the six VTPR channels (669.0, 676.7, 694.7, 708.7, 723.6, and 746.7 cm^{-1}) in Table 7.3:

(a) Compute and plot the weighting function $\Delta T_i(p)/\Delta \ln p$ as a function of the pressure on a logarithmic scale. What is the physical meaning of the weighting function?

(b) Assuming that the calculated radiances are the values observed from the NOAA 4 VTPR instrument, recover the temperature profile from these radiances utilizing the relaxation method outlined in Subsection 7.4.3.1. Use a linear

Table 7.3

Atmospheric Profiles and Transmission Functions for VTPR Channels

Pressure mb	Temperature (K)	Transmittances					
		1	2	3	4	5	6
0.8	270.7	.9198	.9817	.9890	.9922	.9931	.9968
1.4	265.0	.8846	.9733	.9837	.9891	.9906	.9953
2.1	256.4	.8429	.9637	.9777	.9861	.9879	.9940
3.1	248.3	.7979	.9508	.9704	.9817	.9848	.9925
4.4	241.9	.7520	.9344	.9612	.9783	.9810	.9906
5.9	238.2	.7061	.9139	.9497	.9732	.9763	.9885
10.3	232.7	.6094	.8591	.9188	.9597	.9645	.9828
13.1	228.8	.5562	.8239	.8980	.9506	.9570	.9790
16.5	226.3	.5001	.7831	.8740	.9403	.9485	.9747
20.4	222.5	.4423	.7369	.8470	.9290	.9392	.9701
24.9	221.1	.3840	.6853	.8168	.9167	.9290	.9652
30.2	219.5	.3266	.6291	.7831	.9033	.9180	.9600
36.1	219.0	.2716	.5691	.7458	.8887	.9065	.9546
42.9	218.5	.2203	.5064	.7051	.8728	.8945	.9489
50.5	217.9	.1738	.4424	.6609	.8555	.8821	.9431
59.1	217.4	.1329	.3785	.6136	.8366	.8695	.9370
68.6	216.8	.0980	.3160	.5638	.8162	.8567	.9307
79.2	217.3	.0693	.2563	.5119	.7941	.8437	.9241
90.9	218.0	.0468	.2008	.4584	.7699	.8304	.9173
103.8	218.8	.0299	.1510	.4043	.7436	.8163	.9101
117.9	219.7	.0179	.1080	.3508	.7152	.8029	.9026
133.3	220.7	.0100	.0727	.2988	.6847	.7884	.8946
150.2	221.7	.0052	.0456	.2496	.6520	.7731	.8861
168.5	222.6	.0024	.0264	.2042	.6175	.7570	.8771
188.4	223.6	.0010	.0139	.1634	.5812	.7397	.8674
209.9	225.2	.0004	.0066	.1275	.5431	.7212	.8569
233.1	227.5	.0001	.0028	.0968	.5033	.7011	.8454
258.0	229.7	.0000	.0011	.0711	.4615	.6792	.8325
284.8	230.2	.0000	.0004	.0508	.4195	.6561	.8187
313.6	231.8	.0000	.0001	.0354	.3782	.6321	.8043
344.3	232.8	.0000	.0000	.0237	.3365	.6064	.7883
377.2	234.2	.0000	.0000	.0151	.2940	.5782	.7701
412.2	235.5	.0000	.0000	.0090	.2514	.5475	.7493
449.5	236.9	.0000	.0000	.0050	.2099	.5142	.7253
489.2	241.6	.0000	.0000	.0026	.1706	.4785	.6992
531.2	245.4	.0000	.0000	.0012	.1343	.4402	.6687
575.8	249.0	.0000	.0000	.0005	.1017	.3993	.6326
623.1	252.8	.0000	.0000	.0002	.0740	.3565	.5917
673.0	256.8	.0000	.0000	.0000	.0516	.3127	.5467
725.7	260.5	.0000	.0000	.0000	.0346	.2689	.4983
781.3	263.9	.0000	.0000	.0000	.0221	.2261	.4476
839.9	267.5	.0000	.0000	.0000	.0134	.1852	.3952
901.5	272.1	.0000	.0000	.0000	.0076	.1456	.3371
966.3	277.0	.0000	.0000	.0000	.0039	.1064	.2682
1019.8	279.5	.0000	.0000	.0000	.0019	.0770	.2099

interpolation between the recovered temperatures and use the true temperatures for the surface and the top layer. Plot the retrieved temperature profile on a logarithmic scale and compare it with the true temperature profile.

Suggested Reading

Baker, J. (1990). *Planet Earth: The View from Space*. Harvard University Press, Cambridge, MA. This book, written for the lay public and beginning students interested in remote sensing, contains illustrative presentations of the objectives of remote sensing and the key atmospheric and surface parameters that can be determined from various types of satellite systems.

Houghton, J. T., Taylor, F. W., and Rodgers, C. D. (1984). *Remote Sounding of Atmospheres*. Cambridge University Press, Cambridge, U.K. Chapter 7 provides a systematic discussion of retrieval theory, particularly for temperature and minor gas profiles.

Janssen, M. A., ed. (1993). *Atmospheric Remote Sensing by Microwave Radiometry*. Wiley, New York. This monograph collects important information on the remote sensing of the atmosphere and the surface based on microwave emission. Chapters 3 and 6 are specifically relevant to the discussion presented in the text.

Kidder, S. Q., and Vonder Haar, T. H. (1995). *Satellite Meteorology*. Academic Press, San Diego. Chapter 2 contains a useful discussion of satellite orbits and viewing geometries and also gives a comprehensive list of meteorological satellite programs up to 1994.

Stephens, G. L. (1994). *Remote Sensing of the Lower Atmosphere*. Oxford University Press, New York. This is a useful text that covers remote sensing subjects of interest in atmospheric sciences. Specifically, Chapter 8 provides a concise presentation of active remote sensing of atmospheric particulates using radar and lidar.

Twomey, S. (1977). *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. Elsevier, New York. Twomey's book presents the fundamental theory of, and background materials on, the inversion of radiometric measurements.

Chapter 8 | Radiation and Climate

8.1 Introduction

Climate is usually defined as the average state of the atmosphere observed as weather in terms of the mean and its statistical deviations that measure the variability over a period of time. Thus, we may speak of the climate of a season, a year, or longer periods. We frequently use the terms climatological temperature, precipitation, and other atmospheric parameters (e.g., data averaged over a 30-year period) to represent the mean atmospheric state. Changes in climate involve interactions of the atmosphere with other parts of the climate system, including the oceans, ice, snow, and land, associated with natural variability and/or anthropogenic perturbations.

Natural variability includes changes in the solar constant associated with sunspot activities (Section 2.3), changes in the solar insolation distribution caused by the earth's orbit around the sun (Section 2.2), and changes in atmospheric composition due to volcanic eruptions (Section 8.4.4). These are referred to as *external forcings,* which may also occur as a result of human activities, such as increases in greenhouse gases (Sections 4.1 and 8.4) and anthropogenic aerosols (Section 8.6.3).

An understanding of the earth's climate and climate change must begin with a comprehensive understanding of the radiation emitted from the sun and the absorption of solar energy by the atmosphere and the surface, as well as of the emission of thermal infrared energy from the earth–atmosphere system, as presented in Chapters 2, 3, and 4, respectively. In Chapters 5 and 6, we detailed the scattering and absorption processes involving molecules, aerosols, and cloud particles. The atmospheric and surface data required for global weather and climate studies must be primarily gathered from satellites by means of remote sensing, the principles of which were discussed in Chapter 7.

Radiation equilibrium at the top of the atmosphere represents the fundamental mode of the climate system. The incoming solar energy that is absorbed by the earth–atmosphere system must be balanced by an equal amount of emitted thermal infrared energy so as to achieve climate equilibrium. Otherwise, the equilibrium temperature of the earth planet, the brightness as viewed from space, would undergo a continuous change until the energy balance was restored. In Section 4.1, we showed that the equilibrium temperature of the earth–atmosphere system, based on the Stefan–Boltzmann law, is given by $T_e = [S(1 - \bar{r})/4\sigma]^{1/4}$, where S is the solar

constant, representing the energy available from the sun, and $\bar{r}$ is the global albedo, representing the internal reflecting power of the earth–atmosphere system. For the present climate condition, T_e is about 255 K. As a result of the greenhouse effect of the earth's atmosphere and its convective nature, the present surface temperature is about 288 K. Thus, the surface temperature is 33 K warmer than the equilibrium temperature.

Global radiation budgets have been observed from space since the beginning of the meteorological satellite era. In Section 8.2, we present a description of the manner in which the radiation budget is derived. The climate system is also affected by the internal heating and cooling of the atmosphere, as well as by the surface radiative budget, which are subsequently discussed.

The physical understanding of climate processes and the prediction of future climate and climate change require numerical models that encompass all relevant climate system components, including the atmosphere, the oceans, land, ice, and snow, as well as their interactions and feedbacks. It is essential that the numerical models for climate studies be based upon well-established physical principles since we do not have future data to check and verify these models, which must be built solely on our knowledge of the present historical climate conditions. In what follows, we first present the role of radiation in two simplified climate models. One-dimensional radiative–convective climate models have been widely employed in the past to explain vertical temperature perturbations due to external radiative forcings such as the increase in greenhouse gases and the indirect role of clouds. The foundation of these models and some significant perturbation studies are presented in Sections 8.3 and 8.4. In Section 8.5, we present one-dimensional energy-balance models, which include surface temperature variation in the latitude and which have been used primarily to evaluate the effect of solar constant and solar insolation variations on climate, particularly in regard to the question of the advance and retreat of polar ice in the geological time scale.

To quantify the response of the climate system to changes in external forcings and to perfect the prediction of future climate and climate change, we must rely on global climate models that account for all of the essential interactions and feedbacks among the climate system components. Models and their components are based on the physical laws represented by mathematical equations that govern atmospheric and oceanic dynamics and physics. In Section 8.6, we first provide some introductory notes concerning the general circulation model, particularly in reference to the transfer of radiation in the earth–atmosphere system. We then discuss cloud radiative forcing induced by the representation of cloud cover and cloud microphysics in typical global climate models within the context of greenhouse warming. Subsequently, we present examples of the direct climate radiative forcings induced by anthropogenic aerosols and contrails. Finally, we point out that radiation and cloud interactions and feedbacks are relevant and important processes in the numerical simulation of interannual and decadal time-scale variability, and present an example associated with the El Niño–Southern Oscillation simulated from a global climate model that couples the atmosphere and the oceans.

8.2   Radiation Budget of the Earth–Atmosphere System

8.2.1   Observational Considerations

The earth radiation budget (ERB) at the top of the atmosphere (TOA) has been derived from satellite observations since the beginning of the meteorological satellite era. The first-generation ERB instruments included black and white hemispheric sensors using thermistor detectors to measure the sensor temperature on board the first U.S. meteorological satellite, Explorer 6, launched February 17, 1959 (Suomi, 1958). This first generation also included a five-channel scanning radiometer on board the TIROS satellite series. Each of the channels had a narrow instantaneous field of view, and the measurements were radiances (per solid angle) rather than flux densities.

In the second generation of satellite missions during the 1960s and 1970s, polar, sun-synchronous orbits provided the opportunity for daily global coverage of the earth. In addition, the duration of spacecraft measurements was extended to several years. Flat, nonscanning radiometers were installed on several research and Environmental Science Service Administration operational satellites and, later, medium and wide field-of-view radiometers were also deployed on the NOAA satellites. Medium- and high-resolution infrared radiometers were used on Nimbus satellites for the detection of shortwave (solar) and longwave (IR) radiation. Nimbus 2 and 3 contained five-channel medium-resolution scanning radiometers and provided the first observations of the ERB for the entire globe. The aperture of the radiometers was about $2.5°$, resulting in a spatial resolution of about 50 km near nadir and about 110 km at an angle of $40°$. Spectral band passes for these radiometers were 0.3–4 μm for shortwave and 5–50 μm for longwave. The computation of outgoing flux densities from measured radiances required conversion of measured filtered radiances to radiances covering the entire solar and thermal IR spectra, integration over all angles of measurements using bidirectional models, and estimation of the average flux density over a 24-hour period. NOAA polar-orbiting satellites performed scanning measurements in the visible and IR window regions. The transformation of these narrow spectral interval data to broadband estimates of flux densities required several assumptions and models. The narrowband scanning radiometers had a spatial resolution of about 4 km at nadir. Valuable data sets have been constructed from the narrowband measurements of ERB components.

The third generation of satellite observational systems led to the development of ERB instruments that measured direct solar irradiance, reflected shortwave radiation, and emitted longwave radiation. The Nimbus 6 and 7 satellites contained wide field-of-view and scanning radiometers that provided valuable observations of the ERB. The scanning measurements observed the directional and bidirectional reflecting and emitting properties of the earth–atmosphere system over both time and space and were important in developing directional models for the conversion of radiances to flux densities. The longest record of solar constant measurements was made available by Nimbus 7 observations.

The sun-synchronous, polar-orbiting satellites observed each location at about the same local time. As a consequence, the observations were insufficient to provide a more detailed quantitative estimate of the temporal and spatial sampling errors. Studies of the ERB using data from the geostationary satellites were especially useful because they provided a regular sample of the atmospheric diurnal cycle. This enabled a wide range of spatial and temporal radiation variations to be investigated. Observations from GOES and METEOSAT satellites have been used in numerous studies of the ERB. Radiometers on board geostationary satellites were confined to narrow bands and had spatial resolutions that varied from about 0.5 to 10 km at nadir. The processing of observed data from geosynchronous altitudes required assumptions similar to those used to interpret the measurements from NOAA polar-orbiting satellites.

In order to provide comprehensive data sets for studying the diurnal and annual cycles of the ERB, as well as the role of clouds in the ERB, the fourth generation of satellite observation systems, referred to as the Earth Radiation Budget Experiment (ERBE), was launched in the 1980s. The experiment consisted of scanning and non-scanning radiometers on three satellites. The NASA Earth Radiation Budget Satellite (ERBS) performed a $57°$ inclination orbital precession around the earth once every 2 months. The other two satellites were NOAA 9 and NOAA 10 operational meteorological satellites.

8.2.1.1   BLACK AND WHITE SENSORS BASED ON RADIATIVE EQUILIBRIUM

For historical reasons, we shall first discuss the use of a matched pair of spherical black and white sensors, the so-called Wisconsin sensors originally designed by Suomi (1958), which used thermistor detectors to measure the sensor temperature. In a short time after exposure to various radiative components involving the direct solar flux, shortwave flux reflected by the earth and atmosphere, and longwave flux emitted by the earth and atmosphere, each sensor achieves a radiative equilibrium. The absorptivity of the black sensor, A_b, was assumed to be the same for both shortwave and longwave radiation. However, the absorptivity of the white sensor for shortwave and longwave radiation was assumed to be different, denoted by A_w^s and A_w^l, respectively.

Let the temperatures measured by the black and white sensors be T_b and T_w, respectively. On the basis of the Stefan–Boltzmann and Kirchhoff laws (Section 1.2), radiative equilibrium equations for both sensors may be expressed by

$$4\pi a^2 A_b \, \sigma T_b^4 = \pi a^2 A_b (F_\odot + F_s' + F_{ir}'), \qquad (8.2.1)$$

and

$$4\pi a^2 A_w^l \, \sigma T_w^4 = \pi a^2 \left[A_w^s (F_\odot + F_s') + A_w^l F_{ir}' \right]. \qquad (8.2.2)$$

These two equations show that the emitted energy per unit time is equal to the absorbed energy per unit time, where $4\pi a^2$ and πa^2 represent the emission and absorption areas, respectively, of the two spherical sensors, each with radius a. The flux densities of

the reflected shortwave and longwave radiation for spherical sensors are defined by

$$F'_s = \int_0^\Omega I_s d\Omega, \quad F'_{ir} = \int_0^\Omega I_{ir} d\Omega, \tag{8.2.3}$$

where Ω is the solid angle by which the sensor sees the earth, I_s and I_{ir} are the radiant intensities reflected and emitted from the earth, respectively, and $F_\odot$ denotes the direct solar irradiance.

Upon solving the sum of the shortwave flux densities and the longwave flux density, we obtain

$$F_\odot + F'_s = \left[4\sigma A^l_w / \left(A^l_w - A^s_w\right)\right] \left(T_b^4 - T_w^4\right), \tag{8.2.4}$$

and

$$F'_{ir} = \left[4\sigma / \left(A^l_w - A^s_w\right)\right] \left(A^l_w T_b^4 - A^s_w T_w^4\right). \tag{8.2.5}$$

The direct solar irradiance, $F_\odot$, can be evaluated from the solar constant, which is specified prior to the experiment.

In order to derive the reflected solar flux density and the emitted thermal infrared flux density in terms of the measured values expressed in Eq. (8.2.3), the following evaluation procedures are required. According to the definition of the flux density and the isotropic radiation assumption discussed in Section 1.1, the reflected solar flux density is given by

$$F_s = \int_0^{2\pi} \int_0^{\pi/2} I_s(\theta, \phi) \cos \theta \sin \theta \, d\theta \, d\phi = \pi I_s, \tag{8.2.6a}$$

where θ and ϕ are the zenith and azimuthal angles, respectively. Thus, the planetary albedo, r, can be expressed by

$$r = \pi I_s / (F_\odot \cos \theta_0), \tag{8.2.6b}$$

where the denominator represents the solar flux density available at TOA normal to the plane-parallel stratification. On the basis of Eqs. (8.2.3) and (8.2.6b), we have

$$F'_s = \frac{F_\odot}{\pi} \int_0^\Omega r \cos \theta_0 \, d\Omega. \tag{8.2.7}$$

Assuming $\cos \theta_0$ does not vary significantly over the viewing area of the satellite, it can be removed from the integral. Moreover, we define the average planetary albedo of the viewing area as follows:

$$\bar{r} = \frac{1}{\Omega} \int_0^\Omega r \, d\Omega. \tag{8.2.8}$$

It follows that

$$\bar{r} = \pi F'_s / (F_\odot \cos \theta_0 \Omega). \tag{8.2.9}$$

In a similar manner, under the assumption of isotropic radiation, the emitted thermal infrared flux density is given by

$$F_{ir} = \pi I_{ir} = \sigma T_e^4, \tag{8.2.10}$$

where T_e denotes the equivalent blackbody temperature of the earth–atmosphere system. Upon defining the average equivalent blackbody temperature of the viewing area in the form

$$\overline{T_e^4} = \frac{1}{\Omega} \int_0^\Omega T_e^4 \, d\Omega, \tag{8.2.11}$$

we obtain

$$\overline{T_e^4} = \pi F_{ir}'/(\sigma \Omega). \tag{8.2.12}$$

The solid angle through which the sensor sees the earth is given by (see Exercise 1.2)

$$\Omega = 2\pi \left[1 - \frac{(2a_e h + h^2)^{1/2}}{a_e + h} \right]. \tag{8.2.13}$$

Hence, the average planetary albedo $\bar{r}$ and equivalent blackbody temperature T_e can be evaluated from the black and white sensors through F_s' and F_{ir}'. The average reflected solar flux density and emitted thermal infrared flux density, respectively, as functions of location and time may now be expressed by

$$\bar{F}_s = \bar{r} F_\odot \cos\theta_0 \tag{8.2.14}$$

and

$$\bar{F}_{ir} = \sigma \overline{T_e^4}. \tag{8.2.15}$$

In the following, we discuss the averaging process with respect to time and space.

8.2.1.2 SCANNING RADIOMETER AND ANGULAR MODELS

As discussed previously, the radiometric measurements determined from satellite platforms are radiances. The conversion of radiances (per solid angle) to flux densities is a major undertaking in satellite radiation budget analyses.

The daily average reflected flux density of solar radiation is defined by

$$
\begin{aligned}
F_s(\varphi, \lambda) &= \int_{\text{day}} F_s[\varphi, \lambda; \theta_0(t)] \, dt / \Delta t \\
&= \int_{\text{day}} dt \int_0^{2\pi} \int_0^{\pi/2} I_s[\varphi, \lambda; \theta, \phi; \theta_0(t)] \cos\theta \sin\theta \, d\theta \, d\phi / \Delta t, \quad (8.2.16)
\end{aligned}
$$

where φ and λ denote the latitude and longitude, respectively, and Δt is the time interval. For a given location (φ, λ), the broadband scanning radiometer measures the scattered radiance or intensity (energy/area/time/solid angle), which depends on the zenithal and azimuthal angles of the outgoing radiation as well as the position of the sun in terms of the solar zenith angle θ_0.

Since the scanning radiometer detects the reflected solar radiance only at a given scan angle, certain empirical adjustments are required in order to evaluate the daily reflected flux. The empirical anisotropic scattering function is defined as

$$X(\theta, \phi; \theta_0) = F_s(\theta_0)/[\pi I_s(\theta, \phi; \theta_0)]. \tag{8.2.17}$$

Prior to satellite experiments, the X function may be determined based on the radiative transfer theory as well as on aircraft and balloon observations for selected localities. With the assumption that values are independent of the locality, the daily average reflected solar flux can be evaluated by

$$F_s(\varphi, \lambda) = \int_{\text{day}} X[\theta, \phi; \theta_0(t)]\pi I_s[\varphi, \lambda; \theta, \phi; \theta_0(t)] \, dt/\Delta t. \tag{8.2.18}$$

It follows that the daily planetary albedo can now be defined by

$$r(\varphi, \lambda) = F_s(\varphi, \lambda)/Q(\varphi), \tag{8.2.19}$$

where the daily average solar insolation discussed in Section 2.2.3 is given by

$$Q(\varphi) = \int_{\text{day}} F_\odot \cos \theta_0(t) \, dt/\Delta t. \tag{8.2.20}$$

Similarly, an empirical function for the emitted infrared radiation may be defined. Because the outgoing infrared radiation is independent of both the azimuthal angle and the sun's position, the empirical function may be written in the form

$$X(\theta) = F_{ir}/[\pi I_{ir}(\theta)]. \tag{8.2.21}$$

Again, this function is to be obtained prior to satellite experiments. Thus, the daily average thermal infrared flux density can be evaluated by

$$\begin{aligned} F_{ir}(\varphi, \lambda) &= \int_{\text{day}} F_{ir}(\varphi, \lambda; t) \, dt/\Delta t \\ &= 2\pi \int_{\text{day}} dt \int_0^{\pi/2} I_{ir}(\varphi, \lambda; \theta; t) \cos \theta \sin \theta \, d\theta/\Delta t \\ &= \int_{\text{day}} X(\theta)\pi I_{ir}(\varphi, \lambda; \theta; t) \, dt/\Delta t. \end{aligned} \tag{8.2.22}$$

Note that radiances measured by the scanning radiometer normally are corrected to the nadir angle ($\theta = 0°$).

We may now define the radiation balance equation using Eqs. (8.2.19), (8.2.20), and (8.2.22). For a given locality with latitude φ and longitude λ, the net daily flux density may be expressed by

$$R(\varphi, \lambda) = Q(\varphi)[1 - r(\varphi, \lambda)] - F_{ir}(\varphi, \lambda). \tag{8.2.23}$$

To derive the zonally averaged quantities, we perform the integration over the longitudinal direction to give

$$R(\varphi) = Q(\varphi)[1 - r(\varphi)] - F_{ir}(\varphi), \tag{8.2.24}$$

where the operator

$$R(\varphi) = \int_{\lambda} R(\varphi, \lambda) \, d\lambda / \Delta\lambda.$$

Moreover, the global value may be evaluated by carrying out the integration over the latitudinal direction as follows:

$$\bar{R} = \bar{Q}(1 - \bar{r}) - \bar{F}_{ir}, \qquad (8.2.25)$$

where the operator

$$\bar{X} = \int_{\varphi} \int_{\lambda} X \, d\lambda \, d\varphi / (\Delta\lambda \Delta\varphi),$$

and X can be R, Q, or F_{ir}. Finally, time averaging can also be carried out to obtain the monthly and annual radiation budget values.

8.2.2   Radiation Budget Viewed from Space

The ERB is usually presented in terms of the emitted longwave (or IR) flux, F_{ir}, referred to as outgoing longwave radiation (OLR); the planetary albedo (or, simply, albedo), r, defined as the ratio of the reflected solar flux to the incident flux at TOA; and the net radiative flux, defined by

$$F = (1 - r)Q - F_{ir}, \qquad (8.2.26)$$

where we have used flux instead of flux density and replaced the notation R by F. The first term on the right-hand side of Eq. (8.2.26) represents the absorbed solar flux within the earth–atmosphere system. The globally and annually averaged albedo and OLR have been derived by a number of researchers from various data sources. In the following, we present a number of radiation budget results derived from satellite broadband radiation measurements.

Global distributions of OLR, albedo, and net flux averaged for the mean annual and seasonal conditions have been presented in numerous publications. Shown in Fig. 8.1 are maps of the mean annual condition. The albedos reveal a distinct land–ocean contrast equatorward of 30°N and 30°S, as well as high convective clouds, in particular Asian monsoons, where high albedo and low emission are evident. Poleward of 30°N and 30°S, the radiation budgets are relatively uniform zonally, especially in the Southern Hemisphere. At low latitudes, regions of relative energy gains and losses are evident for a given zone. Distinct variations are shown in the net fluxes in the tropical and subtropical zones where the deserts of Africa and Arabia appear as negative or small positive anomalies. The convective regions near Asia show large positive anomalies. In general, the albedos are negatively correlated with OLR, principally because of the presence of clouds. The exception is over desert regions, where cloud cover is at a minimum and the surface is relatively bright and warm. A net radiative gain is evident throughout almost the entire zone between ∼40°N and ∼40°S and is flanked by radiation sinks that generally deepen toward the poles. The radiation budgets of the earth and the atmosphere are largely regulated by clouds and temperature fields.
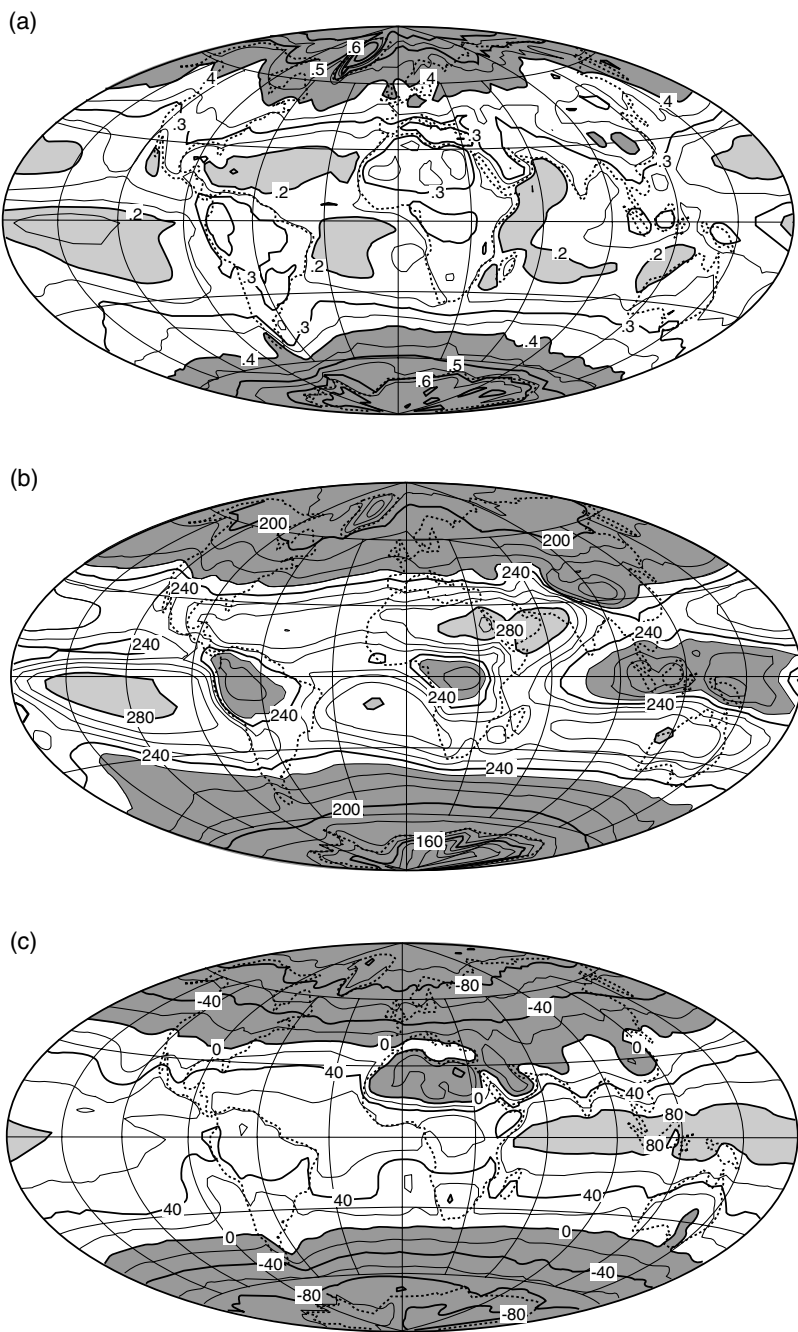
Figure 8.1 Global maps of (a) mean annual planetary albedo, (b) outgoing longwave radiation (W m^{-2}), and (c) net radiative flux (W m^{-2}), in a Hammer equal-area projection (data taken from Hartmann, 1994). In these diagrams, heavier shaded areas denote albedos larger than 0.4, longwave fluxes smaller than 230 W m^{-2}, and negative net radiative fluxes.
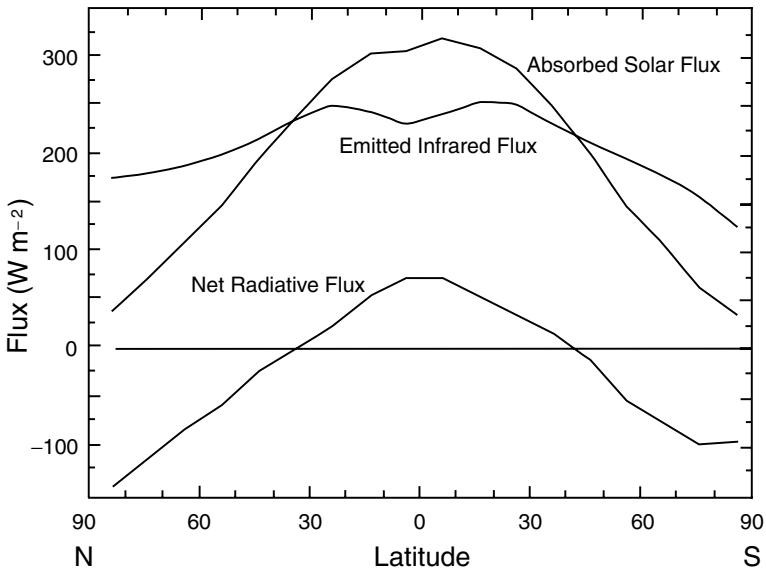
Figure 8.2 Zonally averaged components of the annual mean absorbed solar flux, emitted thermal infrared flux (or OLR), and net radiative flux at the top of the atmosphere, derived from satellite broadband radiation measurements. These patterns were originally presented by Vonder Haar and Suomi (1971), Stephens *et al.* (1981), and more recently by Hartmann (1994).

In Fig. 8.2 we present the latitudinal distribution of annual solar, thermal infrared, and net radiative fluxes at TOA. The annual net radiative flux pattern is approximately symmetric between the Northern and Southern Hemispheres, with a maximum occurring at the equator. This maximum is a result of a minimum of OLR due to the towering cumulus clouds associated with the intertropical convergence zone and the large amount of solar flux absorbed in the equatorial regions, as shown in the figure. In the polar regions, large negative net fluxes are due to the high albedo of snow and ice. The net radiative flux pattern is also associated in part with the variation of the mean solar zenith angle with latitude. The results clearly show that there are gains of radiative energy in the tropics and subtropical regions but losses in the polar regions, a pattern that is essential to the production of general circulation in the atmosphere.

8.2.3 Cloud Radiative Forcing Derived from ERB Data

Clouds regularly occupy at least 50% of the sky on a global scale and are the most important regulators of the ERB. The presence of clouds greatly increases the percentage of solar flux that is reflected back to space. This effect, known as the *solar albedo effect,* reduces the solar flux available to the earth–atmosphere system and results in a cooling of the system. On the other hand, clouds reduce the thermal radiation emitted to space by absorbing the IR flux from the earth's surface and the atmosphere below the cloud layer, and by emitting thermal radiation at normally colder cloud-top

temperatures. This effect, known as the *IR greenhouse effect,* increases the radiation budget, which, in general, results in a warming of the earth–atmosphere system. Thus, the net radiation available to the earth–atmosphere system, as well as the differential heating of the system, is strongly regulated by clouds in terms of their horizontal extent, vertical position, thermodynamic phase, liquid/ice content, and particle size distribution.

Many attempts have been made to diagnose cloud effects from the ERB data inferred from satellites. In reference to Eq. (8.2.26) and considering the specific effect of cloud cover η, partial differentiation leads to

$$\frac{\partial F}{\partial \eta} = -Q\frac{\partial r}{\partial \eta} - \frac{\partial F_{ir}}{\partial \eta} = \frac{\partial F_s}{\partial \eta} - \frac{\partial F_{ir}}{\partial \eta}. \tag{8.2.27a}$$

Or, we may write

$$\frac{\partial F}{\partial \eta} = \frac{\partial F}{\partial F_{ir}}\frac{\partial F_{ir}}{\partial \eta} = -\frac{\partial F_{ir}}{\partial \eta}\left(Q\frac{\partial r}{\partial F_{ir}} + 1\right). \tag{8.2.27b}$$

The first and second terms on the right-hand side of these equations represent the effect of cloud cover on absorbed solar and emitted IR fluxes. To estimate these terms the method of regression on the observed data at a given grid point may be used. Since albedo increases with increasing cloud cover, $\partial r/\partial \eta$ is always positive. In contrast, IR flux decreases with increasing cloud cover so that $\partial F_{ir}/\partial \eta$ is negative. It follows that if $\partial F/\partial \eta$ is positive, the IR greenhouse effect is greater than the solar albedo effect, whereas the reverse is true if $\partial F/\partial \eta$ is negative.

Extracting the effects of clouds from ERB data requires identifying the ERB in clear-sky conditions. Consider a region that is partially covered by clouds. This region consists of an overcast region (cloudy) with a fractional area of coverage of η and a clear-sky region with a fractional area of coverage of $1 - \eta$. Let F be the observed OLR, absorbed solar flux, or net flux over the region. Then, we may write

$$F = (1 - \eta)F^{cl} + \eta F^{ov}, \tag{8.2.28a}$$

where F^{cl} and F^{ov} are the clear-sky and cloud fluxes, respectively. The effects of clouds on F may be evaluated from

$$C = F^{cl} - F = \eta(F^{cl} - F^{ov}). \tag{8.2.28b}$$

The term C is referred to as *cloud radiative forcing.* A separation of solar and IR effects from Eq. (8.2.28a) leads to

$$C_{ir} = F^{cl}_{ir} - F_{ir}, \tag{8.2.29a}$$

$$C_s = Q(r^{cl} - r), \tag{8.2.29b}$$

where F^{cl}_{ir} and r^{cl} are clear-sky OLR and albedo, respectively.

Clouds are almost always more reflective than the oceans and land, except in ice and snow conditions. Thus, when clouds are present, more solar flux is reflected back to space than with clear sky. Cloud solar forcing, C_s, which is the difference between
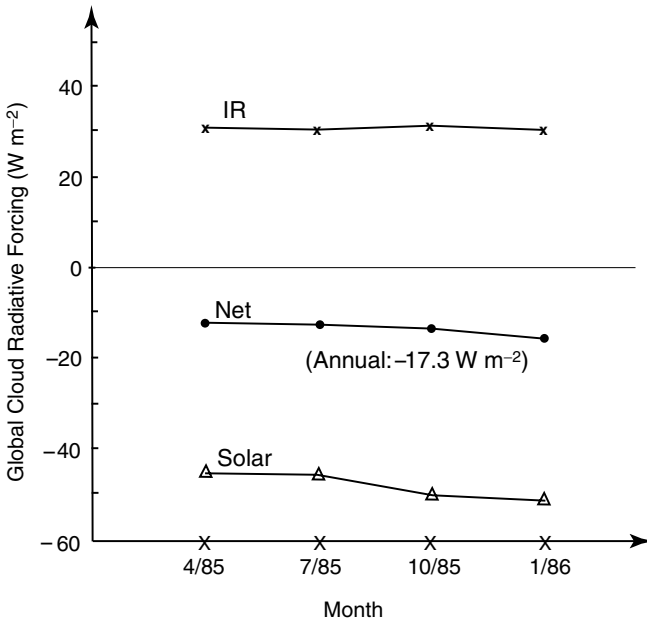
Figure 8.3   Global cloud radiative forcing in units of W m^{-2} as a function of months estimated from Earth Radiation Budget Experiment (ERBE) data (data taken from Harrison *et al.*, 1990).

the clear-sky and cloudy-sky reflected solar fluxes, gives a quantitative estimate of cloud effects on solar radiation. On the other hand, less thermal IR flux is emitted to space from a cloudy region than from clear sky. Cloud IR forcing, C_{ir}, which is the difference between the clear-sky and cloudy-sky emitted IR fluxes, is a measure of cloud effects on thermal IR radiation.

Estimates of the global distribution of cloud radiative forcing have been obtained from ERBE, which includes three satellites in different orbits: ERBS, NOAA 9, and NOAA 10. Figure 8.3 shows seasonal global cloud radiative forcing values based on data gathered from the scanning radiometers on board ERBS and NOAA 9. The global annual C_s and C_{ir} are -48 and 31 W m^{-2}, respectively, resulting in net cloud radiative forcing of -17 W m^{-2}. April has the smallest net cloud radiative forcing value, whereas January has the largest. Although the variation of this forcing between hemispheres not shown in this figure is significant during winter and summer months, the seasonal effect of cloud radiative forcing is generally small. On a global scale for all seasons, the albedo effect from clouds is more dominant than the greenhouse effect.

Cloud radiative forcing has significant regional characteristics. For a given region, cancellation of cloud solar and IR forcing, derived from ERB components at TOA, does not imply a negligible role of clouds in the regional climate. The vertical gradients of radiative heating and cooling produced by clouds presented in the next section, which are critical in weather and climate processes, are not accounted for in the preceding discussion of cloud radiative forcing.

8.2.4   Radiative Heating/Cooling Rates of the Atmosphere

Vertical profiles of the radiative heating and cooling rates of the atmosphere can, in principle, be determined from space based on spectral radiometric observations, as demonstrated in Section 7.4.6, a subject requiring considerable research and development. At this point, however, climatological radiative heating rates in the atmosphere must be computed from a radiative transfer program based on information about the composition and structure of the earth's atmosphere (Section 3.1). The solar input and surface albedo field are also needed in computing solar heating rates in the atmosphere.

In Sections 3.2 and 4.2, we discussed the absorption of solar and thermal infrared radiation in detail. Absorption of sunlight is chiefly produced by H_2O, covering the entire near-IR region, and O_3, covering the UV and visible regions. Absorption contributions also come from O_2 and other minor gases, including CH_4, NO_2, N_2O, $O_2 \cdot O_2$, $O_2 \cdot N_2$, and CO_2. In the thermal infrared region, absorption by H_2O essentially covers the entire spectrum. CO_2 and O_3 exhibit significant absorption in the 15 μm and 9.6 μm regions. H_2O, CO_2, and O_3 are radiatively active gases of primary importance to the earth's atmosphere. CH_4 and N_2O also show strong absorption bands in the 7–8 μm region.

Aerosols both absorb and scatter sunlight. The significance of absorption relative to scattering is determined by the particular aerosol's chemical composition and particle size distribution. Aerosols are usually considered to be important for their influence on solar radiation. Water droplets and ice crystals are relatively transparent in visible light, but absorb near-infrared radiation in the solar spectrum. Clouds have a profound influence on both solar and terrestrial radiation because of their large global coverage. Moreover, there are numerous types of surfaces, which vary significantly in their reflecting properties with respect to incoming solar flux.

In the first edition of this text, comprehensive radiative budget studies reported by Freeman and Liou (1979) were presented. In that study, the discrete-ordinates method for radiative transfer was used in conjunction with the exponential fitting of the band transmittance in carrying out spectral radiative budget calculations. It was pointed out that the radiative transfer method used for global calculations must be an efficient but accurate one in which the spectral dependence of radiation must be covered; the absorption, scattering, and emission of molecules and particulates should be treated adequately; and at the same time, the inhomogeneity of the atmosphere should be properly taken into account. Many simplified radiative transfer methodologies have been developed since the publication of the first edition. In particular, the innovation of the delta four-stream and delta two/four-stream approximations for radiative transfer in inhomogeneous atmospheres (Sections 6.5.3, 6.5.4, and 4.6.3) offers efficient and accurate calculations of atmospheric flux distributions. The development of the correlated k-distribution method (Section 4.3.2) for sorting absorption lines in inhomogeneous atmospheres allows the effective incorporation of gaseous absorption in multiple scattering calculations. Moreover, novel theories for the calculation of scattering and absorption properties of nonspherical ice crystals have been developed specifically for application to radiative transfer in cirrus clouds.

In what follows, we present meridional cross sections of atmospheric solar, IR, and net radiative heating rates based on the radiative transfer model developed by Fu and Liou (1993), which has been significantly improved by the NASA/Langley group working on the retrieval of radiative fluxes using satellite data (Charlock and Alberta, 1996). In brief, the model employs the delta-four-stream method for radiative transfer in inhomogeneous atmospheres containing Rayleigh molecules, absorbing gases, aerosols, and cloud particles. The solar and IR spectra are divided into a number of intervals according to the location of absorption bands based on which the incorporation of nongray gaseous absorption in multiple scattering atmospheres is accomplished by following the correlated k-distribution method. The single-scattering properties of spherical aerosols and water droplets and nonspherical ice crystals are determined from the light-scattering theories presented in various sections in Chapter 5.

The bulk of the input data required for global radiation computations are the atmospheric profiles, the geometric and physical properties of clouds, the global fractional cloudiness for each cloud type, the earth's surface albedo, the duration of sunlight, and the zenith angle of the sun. The atmospheric profile includes the vertical profiles of pressure, molecular and aerosol densities, water vapor, ozone, and other trace gases. The atmospheric profiles used in the present radiative budget calculations are based on mean annual atmospheric conditions involving water vapor, ozone, pressure, and temperature profiles.

Clouds are divided into high, middle, and low cloud types. The fractional cloud cover for each cloud type as a function of the latitude follows those provided by the ISCCP project (Section 7.4.5). The atmospheric aerosol model used is a light background concentration providing about 25 km surface visibility (d'Almeida *et al.,* 1991). For scattering calculations, all cloud (except cirrus) and aerosol particles are assumed to be spherical. The high cirrus are considered to be composed exclusively of nonspherical ice crystals randomly oriented in space. The radiative transfer program for clouds is driven by liquid/ice water path and mean effective particle radius/size, which are prescribed for each cloud type based on available observations (Liou, 1992). Within the IR spectrum, all clouds except cirrus are considered to be blackbodies.

The surface albedo of the earth is also an important parameter that determines the amount of the transmitted solar flux reaching the surface and reflecting back into the atmosphere to be absorbed or scattered, or to escape back into space as a component of the earth's global albedo. Values of the zonally averaged surface albedo are taken from data presented in Henderson-Sellers and Wilson (1983). The duration of sunlight and the solar zenith angle are important parameters in determining the radiation balance of the earth–atmosphere system. The solar zenith angle can be computed from the angles associated with the latitude, the declination of the sun, and the hour angle of the sun, as discussed in Section 2.3.1. The annual meridional cross sections of reflected solar and emitted IR fluxes at TOA were first compared to and matched those observed from satellites, as displayed in Fig. 8.2.

The differential solar heating of equatorial and polar regions provides the ultimate energy source for the general circulation of the atmosphere and the oceans and is also responsible for causing the climatic extremes between tropical and polar latitudes.
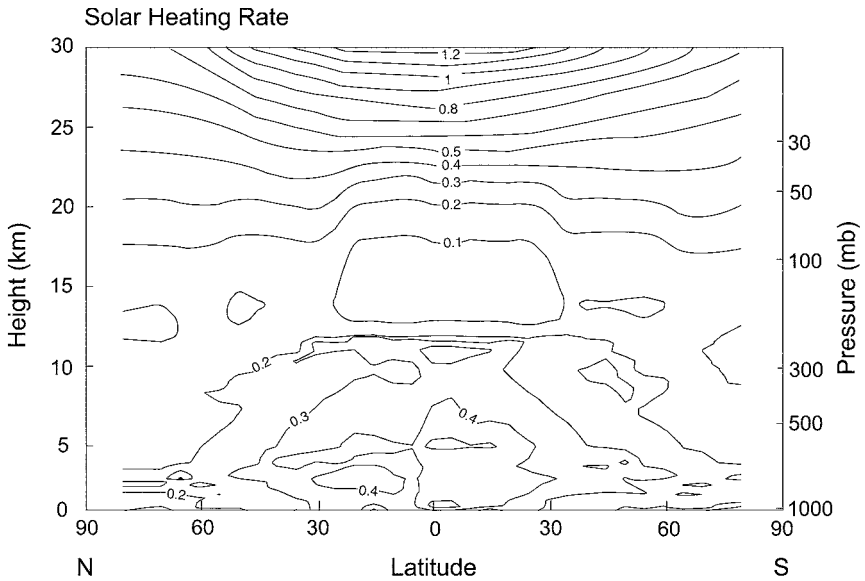
Figure 8.4   Annual meridional cross sections of the solar heating rate (K day^{-1}) of the atmosphere computed from a radiative transfer program using climatological temperature, cloud, gaseous, and surface albedo data. The input solar flux is 342 W m^{-2}. The contour line is 0.1 K day^{-1}.

Every portion of the earth in sunlit sky receives energy from the sun and is warmed to a greater or lesser degree. The primary factors that determine the degree of solar warming received by a particular region on average are the cloud cover, the water vapor and ozone contents of the atmosphere, the surface albedo, and the presence of aerosols in greater or lesser concentrations, as well as the latitude, which is related to the range of solar zenith angles experienced by the area.

The annual meridional cross sections of the solar heating rate are illustrated in Fig. 8.4. Maximum solar heating of about 0.5 K day^{-1} is observed at an altitude of about 4 to 5 km in the tropical and subtropical regions, using the solar flux of 342 W m^{-2} as input (solar constant of 1366 W m^{-2}). Broad, flat minima occur in the upper troposphere and lower stratosphere, followed by increasing solar heating in the stratosphere due exclusively to the presence of ozone. Atmospheric solar heating is primarily produced by the absorption of water vapor in the troposphere and of ozone in the stratosphere.

While the radiation from the sun warms the earth's atmosphere everywhere, the role of terrestrial IR radiation is more complex. In the main, thermal IR radiation serves to cool the atmosphere, radiating away to space an amount of energy equivalent to the solar input, maintaining the radiative balance. Under certain conditions, however, thermal IR radiation adds to the warming of the atmosphere at particular levels and locations.

The annual meridional cross sections of thermal infrared cooling rates are shown in Fig. 8.5. The maximum cooling takes place in the stratosphere, due exclusively
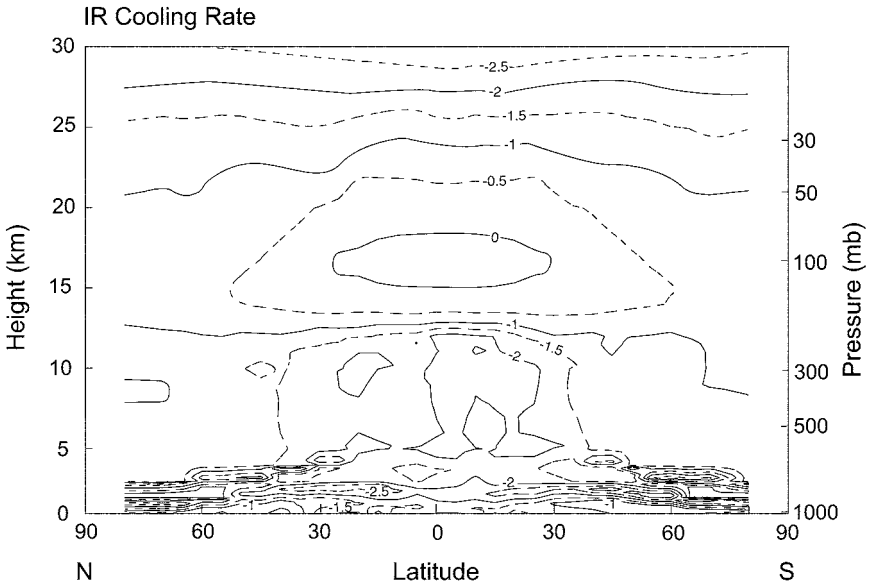
Figure 8.5 Annual meridional cross sections of the thermal infrared (IR) cooling rate (K day^{-1}) of the atmosphere computed from a radiative transfer program using climatological temperature, cloud, and gaseous data. The contour line is 0.5 K day^{-1}.

to carbon dioxide and ozone. Indeed, almost all cooling above the tropopause is produced by these two gases since above about 10 km the water vapor concentration decreases drastically to a negligible amount with respect to radiative transfer. Ozone is also responsible for the region of thermal heating found above the tropopause in tropical and subtropical latitudes. This heating is associated with the increase in ozone concentration with height to about 23 km, resulting in a convergence of flux into the region, where the heating is augmented by a similar region of heating due to carbon dioxide at the tropical tropopause, as well as by the higher temperatures found both above and below the tropopause. The effect of water vapor is to cool the clear atmosphere everywhere since there is an increase of flux with height as the water vapor concentration decreases. A secondary maximum of cooling occurs in tropical latitudes within the troposphere, associated with the large vertical gradients of water vapor and temperature. The effects of clouds, which tend to increase the cooling above their tops and decrease the cooling below their bases, also are evident in this region. Near the surface, large water vapor density and temperature gradients lead to another maximum of cooling in the tropics. This cooling is offset somewhat by the increase in warming below the cloud bases.

The net radiation budget was computed by summing the heating and cooling rates at each latitude and atmospheric layer. The net heating cross sections are presented in Fig. 8.6. Radiative cooling dominates solar heating almost everywhere. In the upper stratosphere, above 25 km, intense cooling due to ozone and carbon dioxide completely overshadows the solar heating by ozone. The high degree of cooling is

Net Radiative Heating/Cooling Rate
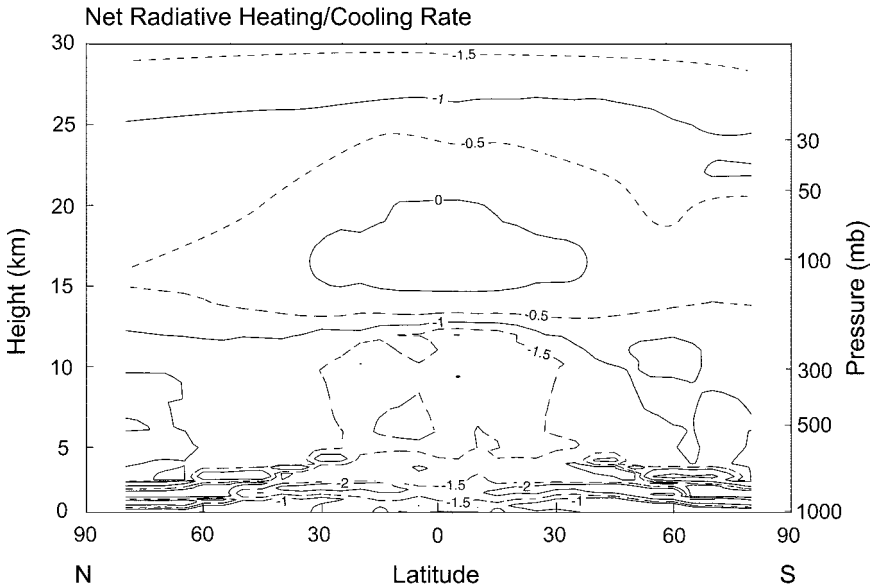
Figure 8.6   Annual meridional cross sections of the net radiative heating/cooling rate (K day^{-1}) of the atmosphere. The contour line is 0.5 K day^{-1}.

due, in part, to the effect of colder cloud tops. At the tropical tropopause, near 18 km, there is a small longwave heating that occurs in the region of minimum solar heating to produce a net heating. Below this region of heating is a region of maximum cooling associated with large vertical gradients of water vapor and temperature. Cooling produced by longwave radiation outweighs solar heating at every latitude, due primarily to water vapor, and thus has a maximum in the tropics. The presence of clouds tends to moderate the cooling in the lower atmosphere by reducing the cooling below their bases and by producing strong solar heating at their tops. The ubiquitous net radiative cooling in the earth's atmosphere is compensated for by other forms of energy in the atmosphere and by the transport of heat from the surface, as discussed in Section 8.5.1.

8.2.5   Radiation Budget at the Surface

Based on radiative equilibrium, the net radiative flux at the surface can be expressed by

$$F(0) = F_s^{\downarrow}(0)(1 - r_s) - \left[\varepsilon_s \sigma T_s^4 - \varepsilon_s F_{ir}^{\downarrow}(0)\right], \qquad (8.2.30)$$

where $F_s^{\downarrow}(0)$ denotes the solar flux reaching the surface, r_s is the surface albedo, ε_s is the surface emissivity, which in thermodynamic equilibrium is equal to absorptivity, T_s is the surface temperature, and $F_{ir}^{\downarrow}(0)$ denotes the downward thermal infrared flux emitted from the atmosphere to the surface. The first and second terms on the

right-hand side of Eq. (8.2.30) represent the net solar (shortwave) and infrared (longwave) fluxes at the surface, respectively. These fluxes cannot be determined directly from satellites.

The study of the surface radiation budget (SRB) is an important subject related to climate modeling and parameterization. It is an integral part of the energy budget at the surface, which will be discussed further in Section 8.5.1. Reliable SRB values in various temporal and spatial scales are required to evaluate the sensible and latent heat fluxes from the surface to the atmosphere, a critical part of the global hydrological and energy cycles. The SRB distribution over the globe is usually determined from a radiative transfer program using appropriate input cloud data that has been derived from satellites as well as other relevant profile data, including water vapor, ozone, aerosols, etc. Thus, the accuracy of SRB fields is dependent on the input data, particularly that regarding clouds, and on the radiation code employed in the calculations. The importance of the SRB in climate studies has been noted by Pinker *et al.* (1995).

As an example, we present in the following a mean annual net radiative flux distribution at the earth's surface (Fig. 8.7) estimated by Budyko (1986) based on some direct surface observations over land and the ocean. The radiation flux decreases with latitude from values of 160–180 W m^{-2} near the equator to values of about 20–40 W m^{-2} poleward of 60° latitude. Most of the globe shows that the net surface radiation is downward. That is, the surface gains radiative energy because of the domination of the net downward solar flux. However, over the polar regions in winter, a net radiation loss can exist at the surface when the solar flux tends to be very small or zero. In general, the net radiative fluxes are higher over the oceans than over the continents at the same latitude. The highest values shown are on the order of 180 W m^{-2} and occur in the tropical regions over the oceans in agreement with the distribution of the total solar radiation absorbed by the atmosphere and the oceans. Secondary equatorial maxima are found over the continents. The lowest values in the tropics occur over the deserts, because of high values of the surface albedo, low values of cloudiness and humidity, and high surface temperatures.

8.3 Radiative and Convective Atmospheres

8.3.1 Radiative Equilibrium

8.3.1.1 A GLOBAL MODEL

The simplest climate model for the earth–atmosphere system is to consider the earth and the atmosphere as a whole and to evaluate the global radiative equilibrium temperature from the balance of incoming solar and outgoing thermal infrared fluxes. Let the global albedo be $\bar{r}$, the solar constant be S, and the radius of the earth be a_e. Over a climatological time period, there should be a balance between the energy absorbed and the energy emitted so that a radiative equilibrium temperature is maintained. Thus we should have

$$\pi a_e^2(1 - \bar{r})S = 4\pi a_e^2 \sigma T_e^4, \tag{8.3.1}$$
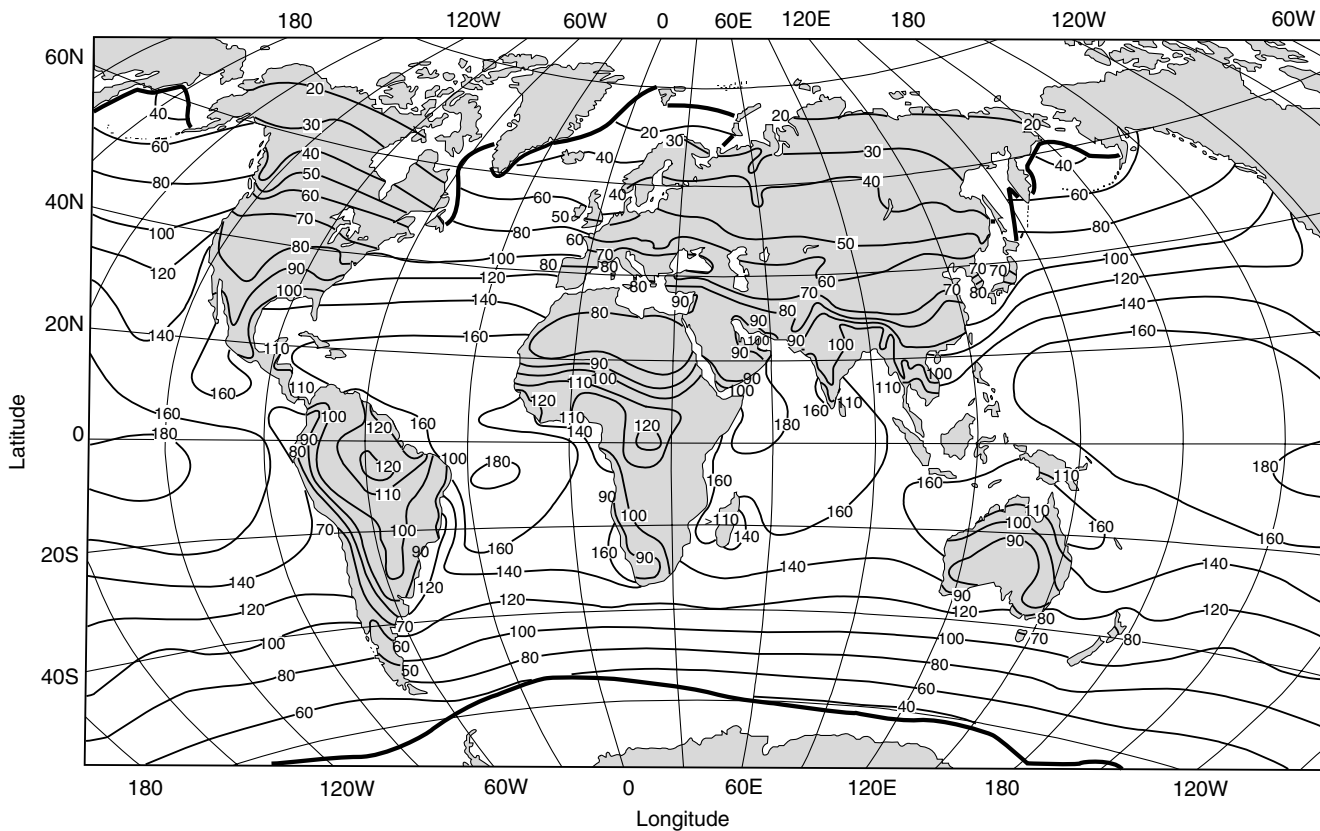
Figure 8.7 Global distribution of the mean radiation flux (W m^{-2}) at the earth's surface based on a number of direct surface observations over land and the ocean. The mean annual ice boundary is also shown by heavy lines (data taken from Budyko, 1986).

Q $Q\bar{r}$ $\bar{\epsilon}\sigma T_a^4$ $(1-\bar{\epsilon})\,\sigma T^4$

$A\,(SOL)$ $\bar{\epsilon}\,(IR)$ T_a

$Q\,(1-\bar{r}-\bar{A})$ $\bar{\epsilon}\sigma T_a^4$ σT^4 T
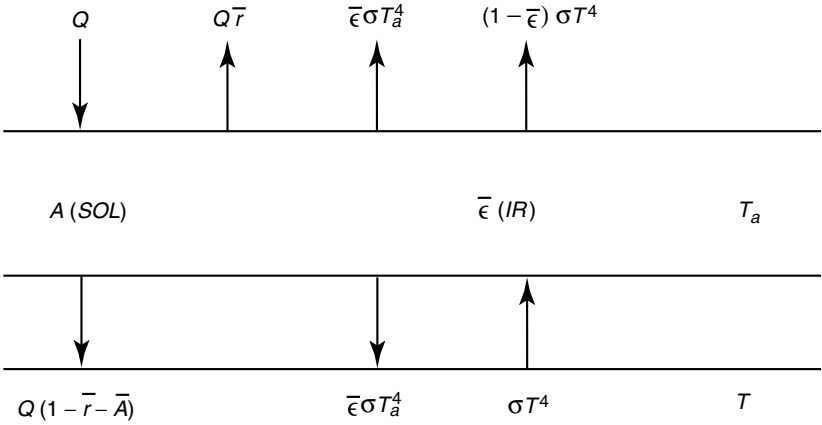
Figure 8.8   A two-layer radiative equilibrium model involving the surface and the atmosphere with temperatures denoted by T and T_a, respectively. The notations $\bar{A}$ and $\bar{\epsilon}$ denote the absorptivity and thermal infrared emissivity, respectively, and $\bar{r}$ is the reflectivity. The solar input is defined by Q.

where πa_e^2 represents the cross sectional area of the earth–atmosphere system that intercepts the incoming solar flux, and the spherical area $4\pi a_e^2$ denotes emission in all directions. It follows that the equilibrium temperature of the system is

$$T_e = [(1-\bar{r})S/4\sigma]^{1/4}. \tag{8.3.2}$$

With this simple equation, we may study the effect of changes in the global albedo and/or the solar constant on the equilibrium temperature of the entire system. However, the surface temperature, which is a fundamental parameter in climate studies, cannot be directly related to either the solar constant or the global albedo change. Information about the surface temperature must be related to the transparency and opacity of the atmosphere with respect to solar and thermal infrared radiation, respectively.

To include surface temperature and the radiative properties of the atmosphere in the simplest radiative equilibrium model, we may construct a two-layer model. Let the mean solar absorptivity and the thermal infrared emissivity of the earth's atmosphere be $\bar{A}$ and $\bar{\epsilon}$, respectively, and assume that the earth's surface is a blackbody with a temperature of T. In reference to Fig. 8.8, the energy balance equations at TOA and the surface may be written in the forms

$$Q(1-\bar{r}) - \bar{\epsilon}\sigma T_a^4 - (1-\bar{\epsilon})\sigma T^4 = 0, \tag{8.3.3}$$

$$Q(1-\bar{r}-\bar{A}) + \bar{\epsilon}\sigma T_a^4 - \sigma T^4 = 0. \tag{8.3.4}$$

Solutions for the surface and atmospheric temperatures are

$$T^4 = Q[2(1-\bar{r}) - \bar{A}]/[\sigma(2-\bar{\epsilon})], \tag{8.3.5}$$

$$T_a^4 = Q[\bar{A} + \bar{\epsilon}(1-\bar{r}-\bar{A}]/[\sigma\bar{\epsilon}(2-\bar{\epsilon})]. \tag{8.3.6}$$

These equations are highly nonlinear with many coupling terms. Thus, it is difficult to carry out sensitivity analyses concerning the effect of the radiative parameters on temperature values. However, if the absorptivity and emissivity of the atmosphere are assumed to be constants, the effect of solar constant changes on equilibrium surface and atmospheric temperatures may be studied.

8.3.1.2   A VERTICAL MODEL

Assuming that the atmosphere is motionless, the local time rate of change of temperature for a given atmospheric layer is associated with the net flux divergence in the form

$$\rho C_p \left(\frac{\partial T}{\partial t} \right)_{\text{RAD}} = -\frac{\partial}{\partial z}(F_s - F_{ir}), \tag{8.3.7}$$

where ρ is the air density, C_p is the specific heat at constant pressure, and F_s and F_{ir} denote the net solar and thermal infrared fluxes. Under steady-state conditions $\partial T/\partial t = 0$. Also, at TOA we require that $F_s(z_\infty) = F_{ir}(z_\infty)$; that is, the absorbed solar flux must be balanced by the thermal IR flux emitted by the surface and the atmosphere. With this boundary condition, an integration from z to z_∞ over Eq. (8.3.7) yields

$$F_s(z) = F_{ir}(z). \tag{8.3.8}$$

This equation represents the steady-state global radiative equilibrium condition. The broadband thermal IR flux may be expressed in terms of an integral form as follows (see Section 4.5):

$$
\begin{aligned}
F_{ir}(z) &= \int_0^{z_\infty} \sigma T^4(z') K(|z - z'|)\, dz' \\
&= \sigma T_s^4[1 - \varepsilon^f(z, T)] - \int_0^{z_\infty} \sigma T^4(z') \frac{d\varepsilon^f(|z - z'|, T(z'))}{dz'}\, dz', \tag{8.3.9}
\end{aligned}
$$

where K is a general kernel function associated with the weighting function, and ε^f is the broadband flux emissivity.

In a nongray atmosphere the radiative equilibrium temperature is determined by the balance between solar and thermal IR heating rates. The solar and IR heating rate profiles were illustrated in Sections 3.5 and 4.7. Radiative equilibrium calculations require information about the solar zenith angle, the length of the solar day, and the solar constant, as well as atmospheric composition and the earth's surface albedo.

To obtain the vertical temperature profile under radiative equilibrium, we may carry out an iterative and time-marching procedure as follows:

$$T^{(n+1)}(z) = T^{(n)}(z) + \left(\frac{\partial T}{\partial t} \right)_{\text{RAD}}^{(n)} \Delta t, \tag{8.3.10}$$

where n is the time step of the integration and Δt is the time interval. A numerical differencing scheme and an initial guess of the temperature are needed. Radiative

equilibrium is reached when the temperatures at the $(n + 1)$ and n time steps differ by a small preset value. Specifically, radiative equilibrium at the surface and at TOA must be satisfied.

The time-marching method, which involves considerable computational effort, will give the evolution of the temperature as a function of time under the radiative equilibrium condition. However, if one is primarily interested in the temperature profile in the equilibrium state, the simplification given in Eq. (8.3.8) may be used. The solar flux must be balanced by the thermal IR flux. Thus, we write

$$F_s(z) = F_{ir}(z) = \int_0^{z_\infty} \sigma T^4(z') K(|z - z'|) \, dz'. \tag{8.3.11}$$

The kernel function K in Eq. (8.3.9) depends largely on atmospheric gaseous profiles, principally those for H_2O, CO_2, and O_3, as well as on the distribution of cloud fields. As a good approximation, the kernel function may be considered to be independent of temperature. Thus, once the solar net flux profile has been given, the temperature profile may be determined from the following procedure.

The atmosphere can be divided into N finite vertical layers, so that Eq. (8.3.11) may be expressed in a finite difference form:

$$F_s(z_i) = \sum_{j=1}^N \sigma T^4(z_j) K(|z_i - z_j|) \Delta z_j, \qquad i = 1, 2, \ldots, N. \tag{8.3.12}$$

In compact matrix form, we have

$$\mathbf{F}_s = \mathbf{K} \cdot \sigma \mathbf{T}^4, \tag{8.3.13}$$

where $\mathbf{F}_s$ and $\sigma \mathbf{T}^4$ are N column vectors and $\mathbf{K}$ is an $N \times N$ matrix. Inverting the matrix leads to

$$\sigma \mathbf{T}^4 = \mathbf{K}^{-1} \mathbf{F}_s, \tag{8.3.14}$$

where $\mathbf{K}^{-1}$ denotes the inverse of the kernel function matrix.

Figure 8.9 shows the atmospheric temperature profiles under radiative equilibrium with and without cloud contributions. Without the contribution of clouds, the surface temperature under radiative equilibrium is $\sim$340 K and the temperature in the tropopause is $\sim$215 K. In cloudy conditions, the temperature in the troposphere decreases significantly because of the reflection of solar flux by clouds. In particular, a $\sim$30 K reduction in surface temperature is seen. The temperature profile for average cloudiness, shown in Fig. 8.9, is obtained by accounting for clear and cloudy areas. It is evident that the radiative equilibrium temperature is much too warm near the surface and too cold in the tropopause. On a mean annual basis, the earth–atmosphere system is in radiative equilibrium at TOA. However, it is clearly not in radiative equilibrium within the atmosphere or at the surface. In terms of the one-dimensional globally averaged condition, the only mechanism that can bring the system into thermodynamic equilibrium is the vertical transport of heat by means of eddies. The convective nature of the earth–atmosphere system is, thus, fundamental to weather and climate processes, as well as to their numerical modeling.
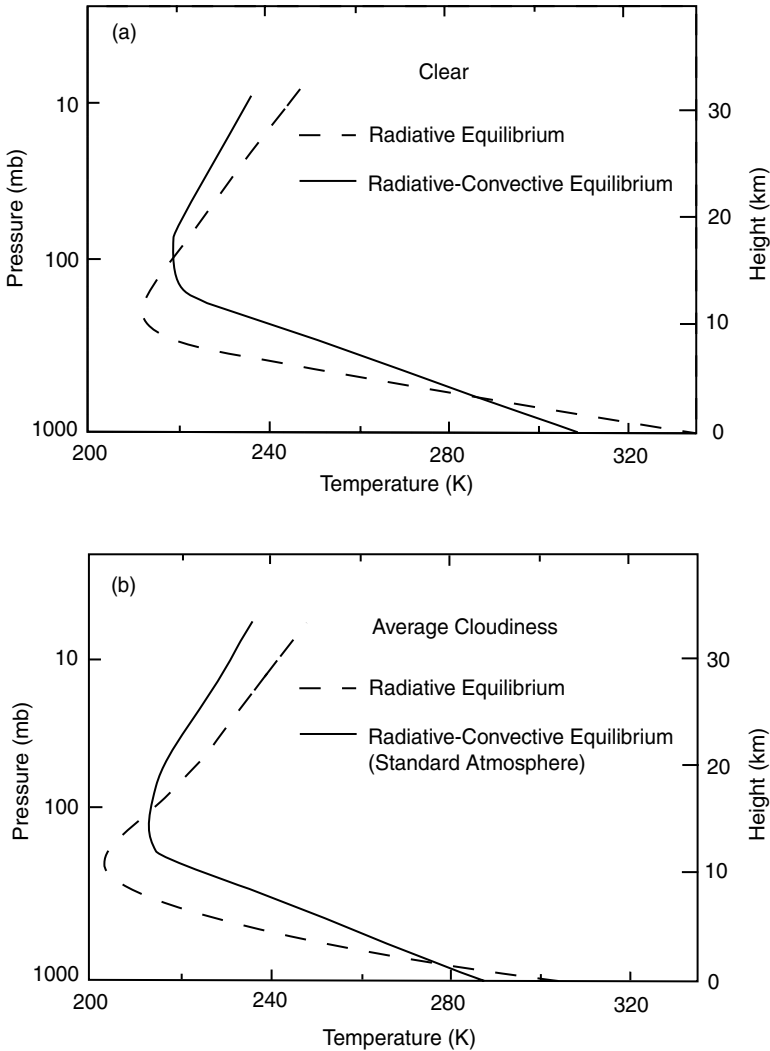
Figure 8.9   Vertical distributions of radiative and radiative–convective equilibrium temperatures in clear (a) and average cloud (b) conditions, simulated from a one-dimensional radiative–convective climate model.

8.3.2   Radiative and Convective Equilibrium

8.3.2.1   HEAT BUDGET OF THE EARTH–ATMOSPHERE SYSTEM

The source of energy that drives the earth's climate comes from the electromagnetic radiation emitted from the sun. The solar constant, that is, the solar irradiance (or flux) corresponding to the mean distance between the earth and the sun available on a 1 square meter area facing the sun at TOA, is about 1366 W m^{-2} (Lean and Rind,
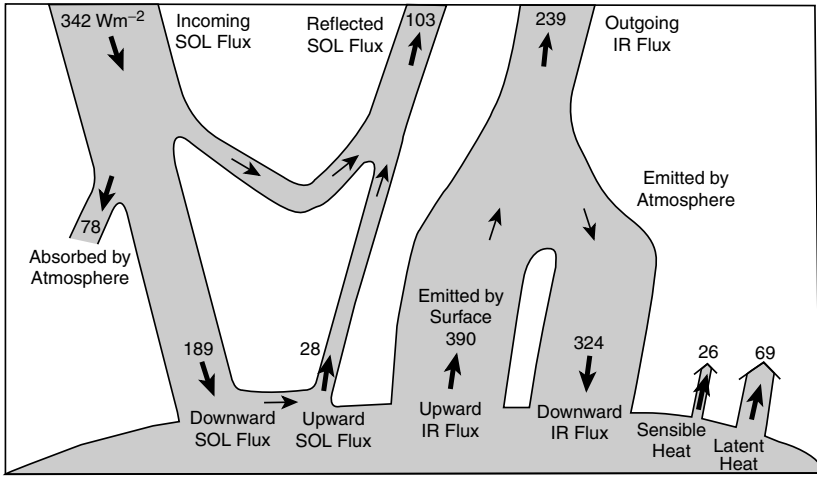
Figure 8.10   The heat balance of the earth and the atmosphere system. The solar (SOL) constant used is 1366 W m^{-2} so that the incoming solar flux for climatological energy balance is 342 W m^{-2} (round off the decimal point), while the global albedo is taken to be 30%. The "atmosphere" referred to in the graph contains molecules, aerosols, and clouds. The atmospheric thermal infrared (IR) flux is emitted both upward and downward. The upward IR flux from the surface is computed by using a climatological surface temperature of 288 K. At the top of the atmosphere, the energy is balanced by radiative flux exchange. At the surface, however, upward sensible and latent heat fluxes must be introduced to maintain energy balance. Absorption of the solar flux is obtained from the divergence of net solar fluxes at the top and the surface. The width of the shaded area with an arrow is approximately proportional to the flux value.

1998). The cross-section area at which the earth intercepts the solar flux is πa_e^2, but the area of the spherical earth is $4\pi a_e^2$, where a_e denotes the earth's radius. Thus, the effective average solar flux incident to a level surface at TOA is one-quarter of the solar constant, i.e., 342 W m^{-2} (round off the decimal point). Moreover, based on satellite radiometric observations and analyses, the best estimate of the global albedo is about 30% (Jacobowitz *et al.*, 1984). This albedo is a result of multiple scattering processes involving molecules, aerosols, clouds, and the underlying surface, as illustrated in Fig. 8.10. Interactions between multiple scattering and absorption processes in the atmosphere and surface reflection determine the amount of solar flux reflected back to space. As a result of this global albedo, about 239 W m^{-2} is available on average to warm the atmosphere and the surface.

To maintain global radiative equilibrium at TOA so that an equilibrium temperature can be defined, the earth–atmosphere system itself must radiate the same amount of energy, i.e., 239 W m^{-2}, back to space. Because the equilibrium temperature of the earth–atmosphere system is about 255 K, the emitted thermal radiation is in the infrared portion of the electromagnetic spectrum. The thermal infrared (or longwave) radiation is emitted from the surface as well as the atmosphere, as shown in Fig. 8.10. The 239 W m^{-2} outgoing infrared flux is the sum of the emitted surface flux attenuated to TOA plus the emission and absorption contributions within the atmosphere including clouds. The equilibrium temperature of the earth–atmosphere

system of 255 K is much colder than the climatological surface temperature of 288 K, which emits 390 W m^{-2} of infrared flux. As pointed out in Section 4.1, based on the standard lapse rate of 6.5 K km^{-1}, the effective scale height of the greenhouse effect of the earth's atmosphere is about 5 km.

To determine the remaining surface radiative components involving solar flux, downward infrared flux, and absorbed solar flux within the atmosphere, we use the standard atmospheric temperature and composition profiles, including all the trace gases listed in Table 3.3, and globally averaged cloud and background aerosol information noted above as inputs to a line-by-line radiative transfer model. The solar flux reaching the surface is about 189 W m^{-2}, while the reflected solar flux is 28 W m^{-2}, based on a globally averaged surface albedo of 15%. It follows that the absorbed solar flux at the surface is 161 W m^{-2}. On the other hand, the emitted downward infrared flux from the atmosphere reaching the surface is 324 W m^{-2}, leading to a net loss of thermal infrared flux by the earth's surface of 66 W m^{-2}. The net radiative energy gain at the surface due to the absorption of solar flux and the emission of thermal infrared flux is about 95 W m^{-2}. This gain is countered by the transport of sensible and latent heat fluxes out of the surface in order to maintain an overall heat balance. The absorbed solar flux within the atmosphere is about 78 W m^{-2}, which is obtained from the divergence of net solar fluxes at TOA and the surface.

The average annual ratio of sensible to latent heat loss at the surface, the so-called *Bowen ratio,* has a global value of about 0.27. Thus, the latent and sensible heat fluxes are 69 W m^{-2} and 26 W m^{-2}, respectively. As illustrated in Fig. 8.6, the atmosphere experiences a net radiative cooling that must be balanced by the latent heat of condensation released in precipitation processes and by the convection and conduction of sensible heat from the underlying surface. If there were no latent and sensible heat transfer, the earth's surface would have a temperature much higher than the observed value of 288 K.

8.3.2.2   CONVECTIVE ADJUSTMENT

The incoming solar flux at the top of the atmosphere must be balanced by the reflected solar and emitted infrared fluxes over a climatological time scale, since the only energy exchange with space is by means of radiative processes. At the surface, however, equilibrium must be achieved by the balance between net radiative fluxes and convective fluxes of sensible and latent heat, as explained in the previous subsection. Vertical fluxes of sensible and latent heat are governed by the motions involving various scales. In the molecular boundary layer, about 1 mm adjacent to the surface, the principal mechanisms for the transport of sensible and latent heat fluxes are conduction and diffusion. In the layer immediately above that, and within about a few tens of meters of the surface, defined as the surface layer, sensible and latent heat fluxes are transferred upward by means of eddies. Based on the theory of turbulence, the surface layer is characterized by strong vertical wind shear, with wind speed proportional to the logarithm of height. Above this layer, up to about 1 km, is the mixed layer, where convectively driven thermals assume the primary role of transporting the

vertical sensible and latent heat fluxes. From the mixed layer to the tropopause, the upward transport of sensible and latent heat is governed by deep cumulus convection, as well as by synoptic and planetary-scale circulations. In this region, essentially all conversion of latent to sensible heat takes place via irreversible condensation processes. The preceding scales of vertical motion associated with the transport of sensible and latent heat are generally classified as *convection*.

The simplest way to account for convection in a global model for climate studies is the *convective adjustment scheme* based on the concept of static stability, first introduced by Manabe and Wetherald (1967). To present this scheme, we shall begin with the first law of thermodynamics in the form

$$C_p \frac{dT}{dt} + g \frac{dz}{dt} + L \frac{dq}{dt} = Q_R, \tag{8.3.15a}$$

where L is the latent heat per unit mass, q is the specific humidity, and Q_R represents radiative heat exchange. Consider this law in the context of one-dimensional space. The local rate of temperature change may be expressed in terms of static stability and radiative flux divergence in the form

$$\rho C_p \frac{\partial T}{\partial t} = \rho C_p w (\gamma - \gamma_c) + \rho Q_R, \tag{8.3.15b}$$

where the vertical velocity $w = dz/dt$, the atmospheric lapse rate $\gamma = -\partial T/\partial z$, and

$$\gamma_c = \gamma_d + \frac{L}{C_p} \frac{dq}{dz}. \tag{8.3.16}$$

The dry adiabatic lapse rate, $\gamma_d = g/C_p$, while the last term in Eq. (8.3.16) represents a modification of the dry adiabatic lapse rate due to the saturation specific humidity gradient. Since specific humidity generally decreases with height, it has a negative value. The value of γ_c depends on the atmospheric humidity profile and varies from the tropics to the polar regions.

The local rate of temperature change, as shown in Eq. (8.3.15b), depends on the convective nature of the atmosphere, which is governed by atmospheric stability and the radiative flux exchange in the atmosphere. For the latter, let the radiative flux divergence in the vertical, ρQ_R, be denoted by $-\partial F/\partial z$. For atmospheric stability, the divergence of the convective flux may be written as

$$-\frac{\partial F_v}{\partial z} = \rho C_p w (\gamma - \gamma_c). \tag{8.3.17}$$

If $\gamma \le \gamma_c$, there will be no transport of convective flux. We may separate convective and radiative contributions to temperature perturbations as follows:

$$\rho C_p \frac{\partial T}{\partial t} = \rho C_p \left[\left(\frac{\partial T}{\partial t} \right)_{\text{CON}} + \left(\frac{\partial T}{\partial t} \right)_{\text{RAD}} \right] = -\frac{\partial F_v}{\partial z} - \frac{\partial F}{\partial z}. \tag{8.3.18}$$

Since a large portion of solar flux is absorbed at the surface, the air immediately above it is mostly unstable. In a one-dimensional context, the vertical transport of sensible and latent heat fluxes from the surface to the atmosphere is principally due to eddy motion.

In the convective adjustment scheme, the atmosphere is divided into layers that include nonconvection, convection without contact with the surface, and convection in contact with the surface. For a nonconvective layer, there would be no variation in the vertical eddy flux, so that

$$-\frac{\partial F_v}{\partial z} = \rho C_p \left(\frac{\partial T}{\partial t}\right)_{\text{CON}} = 0, \tag{8.3.19}$$

resulting in no adjustment. In a convective layer that is not in contact with the surface, the vertical temperature profile is numerically adjusted by using a critical lapse rate, γ_c, under the condition that the total potential energy is conserved within the layer. That is,

$$\int_{z_b}^{z_t} \rho C_p \left(\frac{\partial T}{\partial t}\right)_{\text{CON}} dz = -\int_{z_b}^{z_t} \frac{\partial F_v}{\partial z} dz = F_v(z_b) - F_v(z_t) = 0, \tag{8.3.20}$$

where z_t and z_b denote the top and bottom heights, respectively, of the unstable layer. This implies that convection develops when the atmospheric lapse rate exceeds the critical lapse rate. Convection then transports heat upward until the critical lapse rate is established, resulting in a redistribution of temperature with the total energy conserved. When a convective layer is in contact with the surface, the heat flux from the surface must be considered. Thus,

$$\int_0^{z_t} \rho C_p \left(\frac{\partial T}{\partial t}\right)_{\text{CON}} dz = -\int_0^{z_t} \frac{\partial F_v}{\partial z} dz = F_v(0) - F_v(z_t) = F(0), \tag{8.3.21}$$

where $F(0)$ is the net radiative flux at the surface.

On the basis of Eqs. (8.3.19)–(8.3.21), an iterative procedure may be constructed for the computation of $(\partial T/\partial t)$. This procedure begins at the surface and progressively scans the layers above it until all layers of the supercritical lapse rate have been eliminated. This process is repeated at each time step. Atmospheric temperatures are first constructed based on the balance between the radiative heating and cooling rates. Then the surface temperature is calculated from the balance of the solar and IR net fluxes at the surface. Since solar fluxes heat the surface, there will be net upward fluxes, which are distributed in the layer above the surface according to Eq. (8.3.21). Adjusting the temperature in steps upward and using Eq. (8.3.20) whenever the layer is convectively unstable eliminates all the supercritical lapse rates. For applications to one-dimensional climate models, the critical lapse rate, γ_c, is usually assumed to be 6.5 K km^{-1} for the globally averaged condition. This number is based on the fact that the climatological atmospheric temperature profile in the troposphere has a lapse rate close to this value.

Manabe and Wetherald (1967) performed the first sensitivity experiment of the radiative–convective model described above to examine radiative forcings of carbon dioxide and solar input. Figure 8.9b shows the effects of clouds on the vertical

distribution of equilibrium temperature. With average cloudiness, this profile is close to the standard atmospheric profile. Temperatures without the inclusion of clouds are much higher than those with clouds; a difference of as much as 20 K is shown for the surface temperature. The temperature difference decreases with height.

8.4   Radiation in One-Dimensional Climate Models

8.4.1   Carbon Dioxide Greenhouse Effects

One of the major concerns in climate studies has been the impact on the earth's surface temperature of the steady increase in atmospheric carbon dioxide content produced by the rapid burning of fossil fuels. Since the beginning of the Industrial Revolution more than a century ago, carbon has been removed from the earth in the form of coal, petroleum, and natural gas. In burning processes, carbon dioxide is formed through the oxidation reaction, $C + O_2 \rightarrow CO_2$. Of all the CO_2 that has been produced, about half is believed to remain in the atmosphere, while the other half has dissolved in the oceans or been absorbed by the earth's biomass, primarily the forests.

As discussed in Section 3.2.3, CO_2 is virtually transparent to solar radiation. However, it is a strong absorber in the 15 μm band ($\sim$12–18 μm) of the thermal IR spectrum, as described in Section 4.2.1. The 15 μm band consists of the ν_2 fundamental, the combination bands, the hot bands, and the P, Q, and R branches of the rotational transitions (Fig. 4.2). An increase in atmospheric CO_2 content can cause the additional trapping of the outgoing thermal IR radiation emitted from the surface and lower atmosphere, thereby enhancing the greenhouse effect. Arrhenius (1896) was the first to recognize the potential effect of increased CO_2 concentration on the earth's heat budget and hence, atmospheric temperature. He estimated that an increase in the CO_2 concentration by a factor of 2.5 to 3 would result in a globally averaged temperature increase of 8–9°C, an estimate that was not too far from that produced by sophisticated computer models. In the following, we shall first discuss the carbon cycle.

Variations in CO_2 over different geographical areas are relatively small because the variability in the sources and sinks of CO_2 is small at the earth's surface. Atmospheric CO_2 is removed from and released into the atmosphere through a number of natural processes. A considerable amount of CO_2 dissolves annually into the oceans and returns to the atmosphere by a reverse process. A significant component of atmospheric CO_2 consumption appears to be photosynthesis, via the reaction $CO_2 + H_2O + h\tilde{\nu} \leftrightharpoons CH_2O + O_2$. The sedimentary layer of the earth's crust, the lithosphere, contains a considerable amount of carbon from which CO_2 is formed through oxidation processes. These CO_2 cycles appear to introduce little change in the total amount of CO_2 in the atmosphere.

Measurements of atmospheric CO_2 have been made at the South Pole since 1957 and at Mauna Loa, Hawaii, since 1958. The mean monthly concentration record from Mauna Loa, displayed in Fig. 8.11a, has been frequently employed to illustrate a steady increase in the atmospheric CO_2 concentration due largely to human
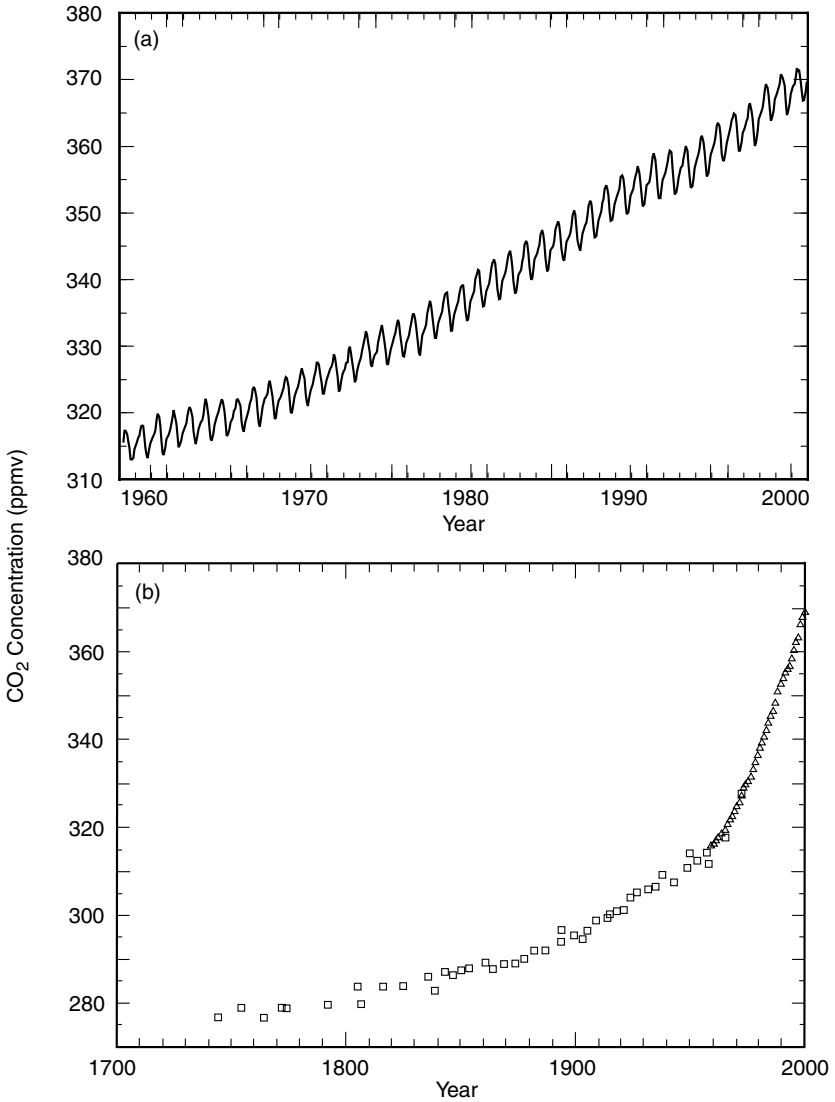
Figure 8.11   (a) Concentration of atmospheric CO_2 at Mauna Loa Observatory, Hawaii, expressed as a mole fraction in parts per million of dry air for the period 1958–2000 (courtesy of Pieter Tans, Environmental Research Laboratory, National Oceanic and Atmospheric Administration). (b) Atmospheric CO_2 concentration for the past 250 years as indicated by measurements in air trapped in ice core from Antarctica determined by Neftel *et al.* (1985) and extended to the present using the Mauna Loa record displayed in (a).

activities, in particular fossil fuel combustion, as well as land-use conversion. The rate of increase varies from 1958 to the present time, but an average of 0.4% per year is a good estimate. The atmospheric CO_2 concentration records prior to 1957 primarily come from ice-core measurements taken in Antarctica (Fig. 8.11b). Over the past 1000 years, CO_2 concentrations were relatively constant with a value of about 280 parts per million by volume (ppmv) and varied within about ±10 ppmv. The significant increase in CO_2 concentration since the onset of industrialization in about 1860 closely follows the increase in CO_2 emissions from fossil fuels.

Evaluation of the carbon budget of the atmosphere–ocean–land system is an ongoing, complex research program because of the many uncertainties in the various carbon processes and their lifetimes, particularly those processes involving the terrestrial biosphere. CO_2 is removed from the atmosphere by numerous processes that operate on different time scales, and is then transferred to a number of reservoirs. The fastest process is its uptake into vegetation and the surface layer of the oceans, which takes place over a few years. Transfer of CO_2 to soils and to the deep ocean operates on the century time scale. It has been estimated that within about 30 years, about 40–60% of the CO_2 currently released to the atmosphere will be removed. The average annual anthropogenic carbon budget has been estimated including emission of CO_2 from fossil fuel combustion and cement production of about 5.5 GtC/yr ($1 Gt = 10^{12}$ kg), and the net emissions from changes in land use of about 1.6 GtC/yr. Storage in the atmosphere is about 3.3 GtC/yr, while ocean uptake of CO_2 is about 2.0 GtC/yr. Other reservoirs include uptake by forest regrowth, nitrogen fertilization, CO_2 fertilization and climate effects, which account for about 1.8 GtC/yr. Future rates of CO_2 increase in the atmosphere are uncertain because the rates of release and the rates at which CO_2 will be taken up by the ocean and the land biota are not known with great precision. However, projections of the future increase of CO_2 have been made by carbon cycle and climate models based on scenarios concerning the use of fossil energy sources. For example, most climate modelers employ business-as-usual scenarios for future greenhouse gas amounts.

Extensive numerical experiments using a one-dimensional climate model have been performed to investigate the effects of an increase in CO_2 on temperature, beginning with the pioneering work of Manabe and Wetherald (1967). These equilibrium experiments were generally performed by doubling a fixed CO_2 concentration (e.g., from 300 to 600 ppmv). Various feedbacks involving moisture that are associated with CO_2 doubling have been investigated by Hansen *et al.* (1981) using a one-dimensional model. Based on a fixed relative humidity, a lapse rate of 6.5 K km^{-1}, and a prescribed cloud altitude, the equilibrium surface temperature, T_s, increases. This is because rising temperatures in the atmosphere and at the surface increase the water vapor concentration in the atmosphere, which in turn traps more thermal IR radiation and, to a lesser degree, absorbs more solar radiation. The feedback due to increasing water vapor through higher temperatures is positive. If, however, a moist adiabatic lapse rate is used instead of a fixed lapse rate, T_s is less sensitive to radiative perturbations as more moisture is added to the atmosphere because of the decreasing lapse rate. Subsequently, the temperature differences between the top of the convective

region and the surface are reduced, and ΔT_s decreases. Thus, the feedback due to the use of a moist adiabatic lapse rate is negative. The surface temperature change, ΔT_s, is sensitive to the assumption of fixed versus varying cloud altitude. The assumption that clouds move to a higher altitude as temperature increases results in a greater ΔT_s than for the case in which cloud altitude is assumed fixed. Clouds that move aloft enhance the trapping of thermal IR emission from the surface by water vapor and exert a positive feedback. However, hypothetical black clouds were used in the early one-dimensional climate models, and thus cloud radiative properties were not properly accounted for.

The exchange of latent and sensible heat and radiative fluxes through the planetary boundary layer provides an important source of ocean–atmospheric interaction and hence plays a significant role in climate perturbations. Through hydrological cycle feedbacks, surface warming due to doubled CO_2 may be divided into three processes. The first two processes involve direct surface and atmospheric heating due to the greenhouse effect of doubled CO_2. The third process is related to the interactions among the ocean surface temperature, the hydrological cycle, and the tropospheric convective adjustment. Surface warming due to the first two processes enhances the evaporation of water vapor into the troposphere, which indirectly amplifies surface warming via the latent heat release within the troposphere and increases the tropospheric absolute humidity that, in turn, increases downward IR emission to the surface. Based on one-dimensional model calculations, the third process is most significant in the feedback process (Ramanathan, 1981). Overall, the one-dimensional models show a surface temperature increase of about 2–3°C in the doubling of CO_2 concentration experiments.

8.4.2   Ozone and Other Greenhouse Gases

8.4.2.1   OZONE

As presented in Section 3.2.2, ozone is produced primarily in the middle and upper stratosphere by three-body collisions involving molecular oxygen, O_2, its atomic form, O, and a third body. Atomic oxygen is produced by the photolysis of O_2 due to the absorption of solar fluxes with wavelengths shorter than 2423 Å. The destruction of O_3 is mainly due to the catalytic effects of various free radical species, including nitrogen oxides, chlorine oxides, and hydrogen oxides. The excited atomic oxygen in the ^1D state, produced by the photodissociation of ozone from solar wavelengths shorter than 3100 Å, is essential for the production of these radicals. Ozone absorbs solar radiation in the UV Hartley and Huggins bands and in the visible Chappuis band, as discussed in Section 3.2.1. The majority of the heating at altitudes above $\sim$45 km is due to absorption in the Hartley band, while the heating below $\sim$30 km is due to absorption in the Chappuis band. In the infrared, O_3 exhibits a number of vibrational–rotational bands. The 9.6 μm band is most important because of its location in the window region.

The total column ozone concentrations have been decreasing over much of the globe at a rate of about 4–5% per decade at midlatitude in both hemispheres, according to ground-based and satellite observations. As indicated in Section 3.2.2, a deep

ozone hole has developed over Antarctica since the mid-1970s, with a decrease of as much as 50% or more during austral spring. This ozone hole is caused by heterogeneous chemical reactions between chlorine species derived from CFCs and polar stratospheric ice clouds. In addition, the average stratospheric ozone depletion over the past decade appears to be associated with the injection of sulfur dioxide from the Mt. Pinatubo volcanic eruption in 1991. A reduction in stratospheric O_3 could lower the temperature in the region in which it occurs because of the reduction of UV absorption. At the same time, however, two competing effects on tropospheric and surface temperatures are produced. First, more solar UV and visible fluxes could reach the troposphere and surface, leading to a warming effect. On the other hand, the greenhouse effect, due to the trapping of IR fluxes by O_3, could be reduced and, in turn, cooling could result. The competition of these two effects determines whether the troposphere is warmed or cooled due to the decrease in O_3.

Based on one-dimensional climate models, a reduction in the concentration of O_3 would lead to a decrease in atmospheric and surface temperatures. Temperature perturbations due to other radiative forcings, such as a doubling of CO_2, may lead to changes in the reaction rates involving ozone. This in turn affects the ozone concentration. Increases in the CO_2 concentration reduce stratospheric temperatures, resulting in an increase in the total ozone due to a slowing down of the ozone loss reaction rate. The uncertainty involved in these chemical reaction rates, however, is very large.

Contrary to stratospheric ozone, tropospheric ozone, about 10% of the total column ozone, has increased in the Northern Hemisphere over the past three decades. This increase has been closely associated with gaseous exchanges between the surface and the atmosphere, regional gaseous emissions, and aerosol and cloud processes. For example, anthropogenic sources of hydrocarbons, methane, nitrous oxide, and carbon monoxide enhance the formation of tropospheric ozone. Also, cumulus convection and precipitation, as well as heterogeneous chemical processes occurring within clouds, have a profound effect on tropospheric chemistry involving ozone. In the upper troposphere, the injection of nitric oxide from aircraft can produce excess ozone. The greenhouse effect of the increased tropospheric ozone, associated with its IR emission at the 9.6 μm band, appears to be rather important (see Fig. 8.12).

8.4.2.2   METHANE

As described in Section 4.2.1, methane exhibits an absorption band at $\sim$1400 cm^{-1} (7.6 μm) and is a greenhouse gas. Atmospheric CH_4 has been increasing since the beginning of the 19th century. The current levels of methane are about 1.7–1.8 ppmv and are the highest that have been observed, including in the ice-core records that go back to 1850. The primary natural source of CH_4 appears to be microbial decay of organic matter under anoxic conditions in wetlands. Anthropogenic sources include rice production, termites, bacterial decay in landfills, leakages from the use of fossil fuels and natural gas, and biomass burning, the sum of which may be twice as great as that from natural sources. The atmospheric concentration of CH_4 is controlled by its reaction with hydroxyl radicals in the troposphere via the reaction $CH_4 + OH \rightarrow H_2O + CH_3$. This reaction is largely the result of water vapor in the troposphere. The CH_4 growth rate has declined in recent years, probably in part because of the increase
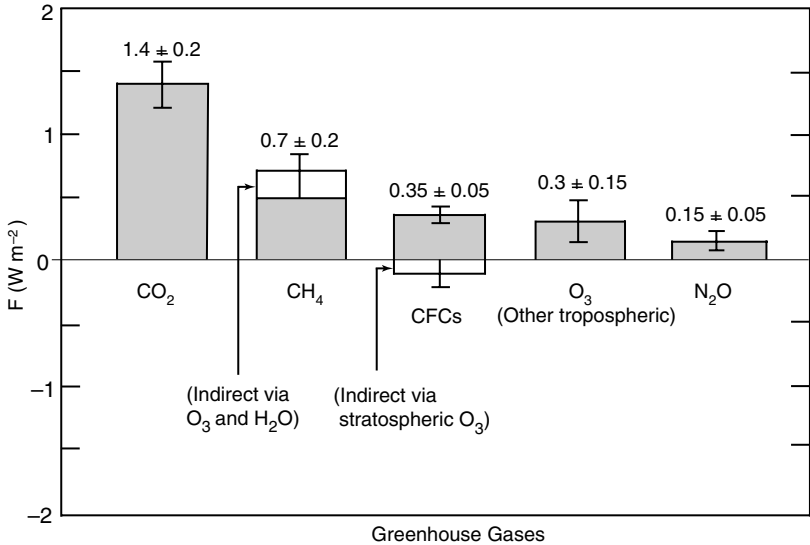
Figure 8.12   Estimated climatic radiative forcings between 1850 and 2000 produced by CO_2, CH_4, CFCs, O_3, and N_2O. Indirect effects associated with CH_4 and CFCs, as well as uncertainties, are also displayed (data taken from Hansen *et al.,* 2000).

in chemical emissions such as CO that affect the production of OH, the primary sink for CH_4.

Based on one-dimensional climate models, a doubling of CH_4 from 1.7 to 3.4 ppmv would lead to an increase in the surface temperature of $\sim$0.2–0.4°C (Wang *et al.,* 1986). Moreover, changes in the CH_4 concentration may affect the global O_3 distribution through reactions with OH and other trace gases. An increase in CH_4 may lead to an increase in O_3 and H_2O via the net reaction $CH_4 + 4O_2 \rightarrow CH_2O + H_2O + 2O_3$, if enough NO is present (Crutzen, 1983). The process is a positive feedback to the radiative forcing involving CH_4.

8.4.2.3   NITROUS OXIDE

N_2O exhibits complex IR absorption bands located at 7.9 μm that overlap with the CH_4 bands. The mixing ratio of N_2O is $\sim$0.3 ppmv and it has a long atmospheric lifetime of $\sim$120 years. The principal atmospheric source of N_2O is denitrification in natural and agricultural biomass. The oceans are also a significant source of N_2O. Measurements indicate a global increase in the N_2O concentration at a rate of $\sim$0.2% per year, attributable to the increase in fossil fuel combustion and fertilizer denitrification. Based on a one-dimensional radiative–convective model, the surface temperature is estimated to increase by $\sim$0.3 to 0.4°C as a result of a doubling of N_2O from 0.3 to 0.6 ppmv (Wang *et al.,* 1986). Dissociation of N_2O by excited oxygen atoms, $O(^1D)$, is the major source of nitrogen oxides ($NO_x = NO, NO_2$) [see Eqs. (3.2.13)–(3.2.14)]. NO_x are important in determining the distribution of both tropospheric and stratospheric O_3. At $\sim$25 km, the net effect of NO_x additions to the stratosphere will be to

lower the O_3 concentration. However, below $\sim$25 km in the stratosphere, NO_x protect ozone from destruction (Crutzen, 1983).

8.4.2.4   HALOCARBONS

As discussed in Section 4.2.1, chlorofluorocarbons (CFCs) are important greenhouse gases because of their absorption in the 10 μm window region. CFCs comprise a family of compounds containing chlorine and fluorine in various combinations. Their sources are generally manufactured. Most of them are chemically inert and have long atmospheric lifetimes on the order of 100 years. The most prevalent CFCs are CFC-11 ($CFCl_3$) and CFC-12 (CF_2Cl_2). Chlorine can be released in the form of Cl_x, via, e.g., $CFCl_3 + h\tilde{v} \rightarrow CFCl_2 + Cl$, which can erode the ozone layer via the catalytic destruction reactions described in Eqs. (3.2.13)–(3.2.14). The growth rate of atmospheric CFCs produced by refrigerants and spray-can propellants noted in the 1970s, particularly CFC-11 and CFC-12, has been reduced as a result of production restrictions imposed by the Montreal Protocol and its Amendments (IPCC, 1996). Most CFCs will be phased out if production agreements are followed.

The direct climatic effects of CFCs and other bromocarbons are exerted through their absorption of IR radiation in the atmospheric 10 μm window (Ramanathan *et al.*, 1987). The indirect effects are associated with the destruction of ozone by the chlorine and bromine released through photodissociation. There is an overall negative feedback due to this indirect radiative forcing, as noted in Subsection 8.4.2.1.

Finally, we note that carbon monoxide does not exhibit a significant IR absorption band. However, it is a climatically important gas because of its chemical reactions involving O_3, CO_2, and CH_4 in the troposphere. The anthropogenic component of CO production is associated with transportation, industrial fossil fuel combustion, deforestation, and biomass burning, as well as modification of CH_4 source via the principal reaction $CO + OH \rightarrow CO_2 + H$. This reaction leads to the reduction of OH concentrations, which, in turn, affects the lifetimes of a number of other gases, principally CH_4 and SO_2.

Figure 8.12 summarizes the radiative forcings of the principal greenhouse gases, including direct and indirect effects, as estimated by Hansen *et al.* (2000). Direct radiative forcing is defined as the gain or loss of radiative flux in the earth–atmosphere system produced by the addition or reduction of a specific greenhouse gas in the system. Indirect radiative forcing is that generated by the increase or decrease of one greenhouse gas (e.g., O_3) caused by the addition of another greenhouse gas (e.g., CH_4). Climatic radiative forcing by CO_2 is the largest, with a value of about 1.4 W m^{-2}, produced by the increase from the preindustrial concentration of about 278 ppmv to the current level of about 365 ppmv. Forcing by CH_4, including the indirect effects via O_3 and H_2O, is half as large as that of CO_2. The total forcing by all non-CO_2 greenhouse gases combined is comparable to that of CO_2.

8.4.3   Radiation Feedback Consideration

In Section 8.4.1, we discussed various feedbacks involving water vapor that are associated with greenhouse perturbations. The H_2O concentration varies substantially

with space and time. In the troposphere it is determined by the local hydrological cycle of evaporation, condensation, and precipitation, and by large-scale transport processes. From the analysis of measured H_2O concentrations across the globe, the mean annual relative humidity, RH, as a function of height, is fairly constant, i.e.,

$$\text{RH}(z) = \frac{q(z)}{q_s(z)} \cong \text{const}, \tag{8.4.1}$$

where q is the specific humidity and q_s is its saturation value. The saturation of specific humidity can be expressed in terms of the saturation vapor pressure as follows:

$$q_s = \frac{\varepsilon e_s}{p}, \tag{8.4.2}$$

where the ratio of the molecular weight of water vapor to that of dry air, ε, is 0.622, and p is the air pressure. Based on the conservation of the Gibbs function during phase changes, the differential changes of the saturation vapor pressure with respect to temperature are governed by the Clausius–Clapeyron equation in the form

$$\frac{de_s}{dT} = \frac{L}{R_v} \frac{e_s}{T^2}, \quad \text{or} \quad d \ln e_s = \left(\frac{L}{R_v T} \right) d \ln T, \tag{8.4.3}$$

where L is the latent heat involving vapor and water, and R_v is the gas constant for water vapor. By integration and empirical fitting, we have

$$e_s(\text{mb}) = 6.11 \exp \left[\frac{a(T - T_o)}{T - b} \right], \tag{8.4.4}$$

where $T_o = 273.16$ K; $a = 21.874$ and $b = 7.66$ if $T \leq 273.16$ K; and $a = 17.269$ and $b = 35.86$ if $T > 273.16$ K.

As a result of the increasing greenhouse gases discussed above, both surface and tropospheric temperatures increase. It follows that the amount of tropospheric water vapor in saturated air increases. Consequently, since the relative humidity in the troposphere is fairly constant, the amount of absolute water vapor increases there and traps more IR flux emitted from the surface and lower atmosphere through rotational and vibrational line spectra. This leads to a further increase in the surface temperature, a positive feedback mentioned in Section 8.4.1. The degree and extent of this positive feedback must be determined from appropriate climate models.

One final note is in order regarding the surface temperature, T_s, and emitted IR flux at the top of the atmosphere, $F_{ir}(z_\infty)$. Although their relationship is complex from the perspective of radiative transfer and convection in the atmosphere, to the extent that the temperature profiles have more or less the same shape at all latitudes, and that the IR fluxes are dependent on temperature at all levels, we may, to a good approximation, relate these two in a linear form as follows:

$$F_{ir}(z_\infty) \cong a + bT_s, \tag{8.4.5}$$

where the empirical coefficients $a \cong 216$ W m^{-2} and $b = 1.58$ W m^{-2} C^{-1} are determined from the analysis of climatological records of zonal surface temperature,

cloud cover, and satellite-observed IR fluxes (see Section 8.5.2 for further discussion). If we use the climatological mean annual surface temperature of 15°C, we obtain an IR flux of 239.7 W m^{-2}. This value is about the same as $Q(1 - r)$, if we use a solar constant of 1366 W m^{-2} (i.e., $Q \cong 342$ W m^{-2}) and a global albedo of 30%. Thus, Eq. (8.4.5) satisfies the requirement of radiative equilibrium at TOA. It also accounts for the convective nature of the lower atmosphere, since the surface temperature in this equation represents the value corresponding to radiative–convective equilibrium.

From Eq. (8.4.5), we may study the sensitivity of surface temperature with respect to IR flux. By differentiation, we have

$$\frac{\Delta T_s}{\Delta F_{ir}} = \frac{1}{b} \cong 0.63°\text{C/(W m}^{-2}).$$ (8.4.6)

Radiative forcing calculations for the doubling of the CO_2 concentration from 300 to 600 ppmv reveal an increase in F_{ir} of about 4 W m^{-2}, which translates to an increase in the surface temperature of about 2.52°C.

8.4.4   Aerosols and Radiation

In Section 5.1, we discussed in some detail the size, shape, and type of atmospheric aerosols produced by both natural and anthropogenic processes. Aerosol particles in the atmosphere can affect the radiation balance of the earth–atmosphere system by reflecting sunlight back into space, by absorbing sunlight, and by absorbing and emitting IR radiation. Aerosols are usually considered to be important for their influence on solar radiation in which the significance of scattering relative to absorption is determined by the aerosols' chemical composition and particle size distribution and shape. However, if the size of aerosol particles is comparable to the thermal infrared wavelength, particularly in the 10 μm window region, they can function to trap the infrared radiation and produce the greenhouse effect. All the recent radiative forcing studies involving aerosols essentially focus on their impact on solar radiation, referred to as the *direct effect*. In Section 5.4, we presented extinction coefficients, single-scattering albedos, asymmetry factors, and phase functions for representative aerosol particles. Aerosols can also act as condensation and/or ice nuclei, which can modify the optical properties, amount, and evolution of clouds, referred to as the *indirect effect*. Climatic perturbations caused by this effect are largely unknown at this point and are a subject of contemporary research and development.

For climatic forcing investigations, aerosols are classified according to their location in the troposphere and the stratosphere. Stratospheric aerosols are usually associated with volcanic eruptions. Volcanic dust and particles formed via photochemical reactions involving sulfur dioxide are generally smaller than 1 μm with small absorption in the visible. Consequently, their effects are primarily associated with the reflection of sunlight leading to cooling of the surface. The radiative effects of aerosols on the temperature of a planet depend not only on their optical depth, visible absorptance, and average size, but also on the variation of these properties with time. Aerosol particles resulting from volcanic activity can reach the stratosphere, where
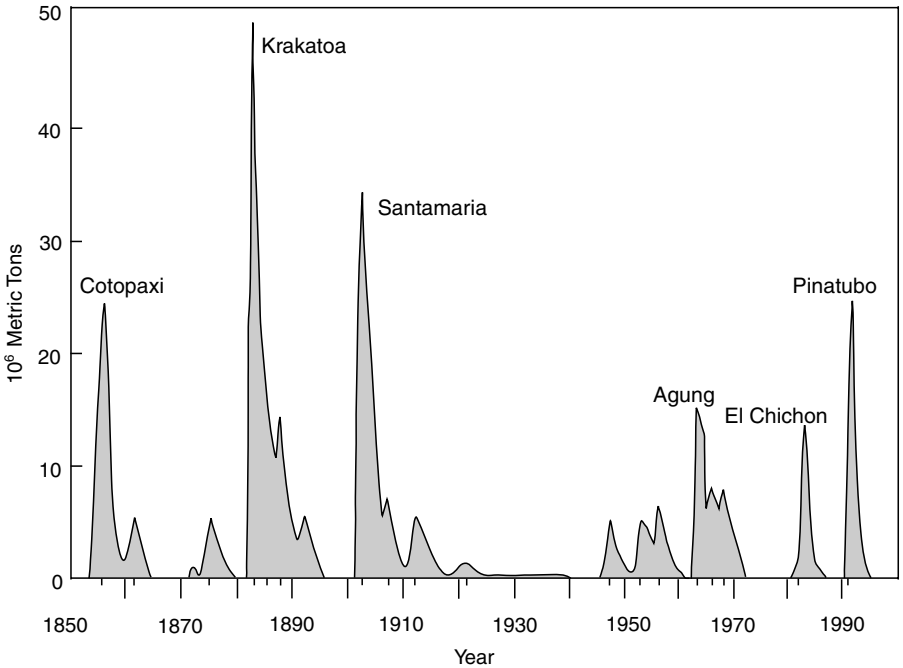
Figure 8.13   Estimated chronology of annual average global aerosol loading by volcanic activity from 1850 to 2000. Major volcanic episodes are indicated in the diagram (data taken from *Study of Man's Impact on Climate,* 1971; Sato *et al.,* 1993).

they are transported by general circulation around the globe over many months or years and can cause anomalous weather.

Figure 8.13 displays the chronological series of major volcanic eruptions and the estimated annual average global aerosol loading dating back to 1850. The period from 1850 to 1920 was characterized by frequent eruptions of climatic importance. From 1920 to about 1960 was a period of reduced aerosol loading. Since 1960, volcanic aerosol loading has increased. The most recent episode of Mt. Pinatubo's volcanic eruption in 1991 initiated weather disturbances on a major global scale for about a year. The solar occultation technique (see Section 7.2.3) has been employed to infer the extinction properties for Mt. Pinatubo aerosols in the stratosphere. Sato *et al.* (1993) estimated the visible optical depth of volcanic aerosols and showed that the global mean value associated with the Mt. Pinatubo eruption is about 0.1, which could produce a negative radiative forcing of about -3 to $-4\,\mathrm{W\,m^{-2}}$. Volcanic activity could be important in interpreting some of the interdecadal variation in surface temperature.

Tropospheric aerosols are produced by both natural and anthropogenic processes. Tropospheric aerosols undergo chemical and physical transformation, particularly in clouds, and are removed primarily by precipitation. They have typical residence times of a few days. As a result of their short lifetime, aerosols in the lower troposphere are distributed inhomogeneously with a maximum concentration close to their source

regions, such as deserts, and industrial and biomass combustion areas. Most aerosols with anthropogenic sources are found in the lower troposphere, below about 2 km. Radiative forcing by anthropogenic aerosols has been recognized as an important contributor to climate change. Anthropogenic sulfate aerosols have been shown to have a direct negative forcing because of their solar reflecting properties. Carbonaceous aerosols are also mainly anthropogenic and are composed of two components: black carbon and organic carbon. Organic carbon from anthropogenic activities is similar to sulfate aerosols in terms of its scattering properties. However, black carbon absorbs significant solar radiation and increases the amount of radiation absorbed by the atmosphere. The presence of black carbon in aerosols may lead to a net heating of the atmosphere. Because of the regional character of sulfate and carbonaceous aerosols, a global model is required to investigate their overall radiative and climatic impact. This subject will be further elaborated upon in Section 8.6.3.

As presented here, the possible response of the global mean temperature field to an increase in aerosol concentrations is far more difficult to estimate than in the cases of changes in CO_2 concentration and the solar constant for the following reasons. First, the optical properties of various aerosol species in terms of the real and imaginary parts of the refractive index, with respect to wavelength, have not been quantified completely. Second, unlike the increase of CO_2, which appears to be uniform around the globe, increases in atmospheric aerosols are likely to be regional. Third, there are questions concerning particle shape and particle size distribution as a function of altitude.

Despite all of these uncertainties, one-dimensional radiative-convective climate models have been used to clarify our understanding of how the perturbation caused by a specific aerosol might influence the radiation field of the earth–atmosphere system and its temperature structure. Charlock and Sellers (1980) have carried out a comprehensive study of the effects of aerosols on temperature perturbation using a one-dimensional climate model. Aerosol optical depth and single-scattering albedo in the visible are used as the two basic parameters in the perturbation studies. The former parameter is an indicator of the attenuation power of aerosols, while the latter represents their relative strength of scattering and absorption. For aerosols with weak absorption, surface temperature decreases as the optical depth increases because of the domination of backscattering. For aerosols with strong absorption, however, warming could occur as the optical depth increases.

Hansen *et al.* (1978) investigated temperature perturbation due to the increase of volcanic dust caused by the explosive eruption of Mount Agung in Indonesia in 1963 based on a one-dimensional climate model. Results showed that after the Agung eruption, average tropospheric temperatures decreased by a few tenths of a degree within a time scale on the order of about 1 year, in agreement with those observed. This cooling effect is caused by the highly reflective properties of the sulfuric acid composition in volcanic aerosols, which block out solar radiation. As a consequence, their addition tends to decrease the amount of solar flux absorbed by the earth–atmosphere system. Hansen *et al.* (1992) conducted a similar simulation involving the eruption of Mt. Pinatubo in 1991 and showed surface cooling for about a year. The potential global atmospheric and climatic consequences of nuclear war

have also been investigated using the one-dimensional models that were developed to study the effects of volcanic eruption (Turco *et al.,* 1983). Significant hemispherical attenuation of solar radiation and subfreezing land temperatures may result from the fine dust particles produced by high-yield nuclear surface bursts and the smoke from city and forest fires ignited by airbursts, a concept referred to as *nuclear winter theory*.

Perhaps the most significant and uncertain role that aerosols play in climate is through their interactions with clouds, the indirect effect. Some aerosols are effective cloud condensation nuclei, which can affect the droplet size distribution and, hence, the radiative properties of clouds and precipitation processes. In addition, the incursion of stratospheric aerosols into the upper troposphere is also a potential source of alteration in the radiative properties of cirrus clouds. Reducing the uncertainty introduced by the indirect aerosol effects on radiative forcing is an important task in climate studies that requires the successful combination of satellite observations and chemical transport modeling.

8.4.5　Cloud Radiative Forcing

Numerical experiments using one-dimensional climate models have been carried out to aid in the understanding of the role clouds play in climate. The effects of high, middle, and low clouds on the equilibrium temperature assuming 100% cloud cover are shown in Fig. 8.14. Nonblack high clouds often produce a warming effect in the troposphere and low stratosphere due to the combined effects of the transmission of solar fluxes and emission of IR fluxes from the clouds. Middle and low clouds produce a significant cooling in the atmosphere and at the surface due to significant reflection by the clouds. The degree of the warming and cooling that results depends on the radiative properties and positions of the clouds used in the model calculations.

In the calculations, representative low and middle clouds (assumed to be black) have solar albedos of 76.9 and 82.5%, respectively, while high clouds have a solar albedo of 10.5% with a nonblack emissivity of 47.5%. Because optically thin high clouds are largely composed of ice crystals, whereas optically thick middle and low clouds primarily contain water droplets, it appears appropriate to suggest that ice clouds are greenhouse elements and that the presence of water clouds would exert a significant solar albedo effect. Clouds thus exert two competing effects on the radiation field of the earth–atmosphere system. On the one hand, they reflect a significant portion of the incoming solar flux, and on the other, they trap the outgoing thermal IR fluxes emitted from the atmosphere below the clouds and from the surface. The competition between the solar albedo and IR greenhouse effects determines whether the surface will undergo cooling or warming.

8.4.5.1　Cloud Position and Cover

The first issue concerning the role of clouds in greenhouse perturbations is the possible variation in cloud position and cover. If a formation of high clouds were to rise higher in the atmosphere, there would be a positive feedback because of
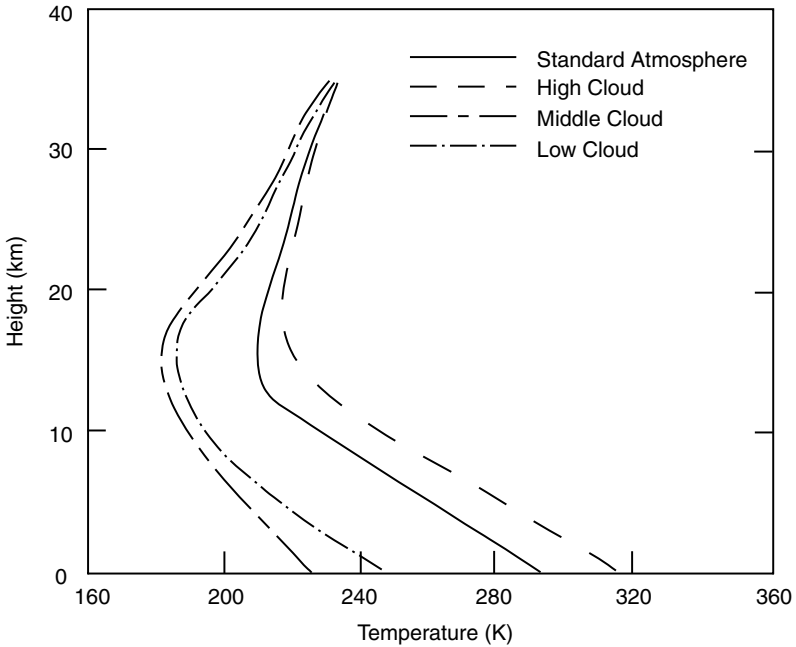
Figure 8.14 Effects of high, middle, and low clouds on atmospheric temperatures in a radiative–convective model. The solid curve is the temperature profile of the standard atmosphere (data taken from Liou and Ou, 1983).

the enhanced downward IR flux. A positive feedback would also be evident if high cloud cover increased because of greenhouse perturbations. The reverse would be true if middle and low cloud covers increased as a result of greenhouse warming. The prediction of cloud cover and position based on physical principles is a difficult task and its success, even with comprehensive climate models, is limited at this point. This difficulty is also associated with our limited ability to determine cloud cover and cloud position from observations. Because of the uncertainties and limitations of the remote sensing of clouds from satellite radiometers, at this point we do not have sufficient cloud data to correlate with the greenhouse warming that has occurred so far. Section 8.6.2 provides further discussion of this subject.

8.4.5.2 CLOUD MICROPHYSICS

The second factor that determines the role clouds play in global climate and greenhouse warming is related to cloud microphysics in terms of their liquid/ice water content and particle size. Some evidence exists, based on aircraft observations, that an increase in temperature leads to an increase in cloud liquid water content (LWC) (Feigelson, 1981; Matveev, 1984). The temperature dependence of ice water content (IWC) is also evident from aircraft measurements of midlatitude cirrus clouds (Heymsfield and Platt, 1984). Shown in Fig. 8.15 is LWC/IWC as a function of
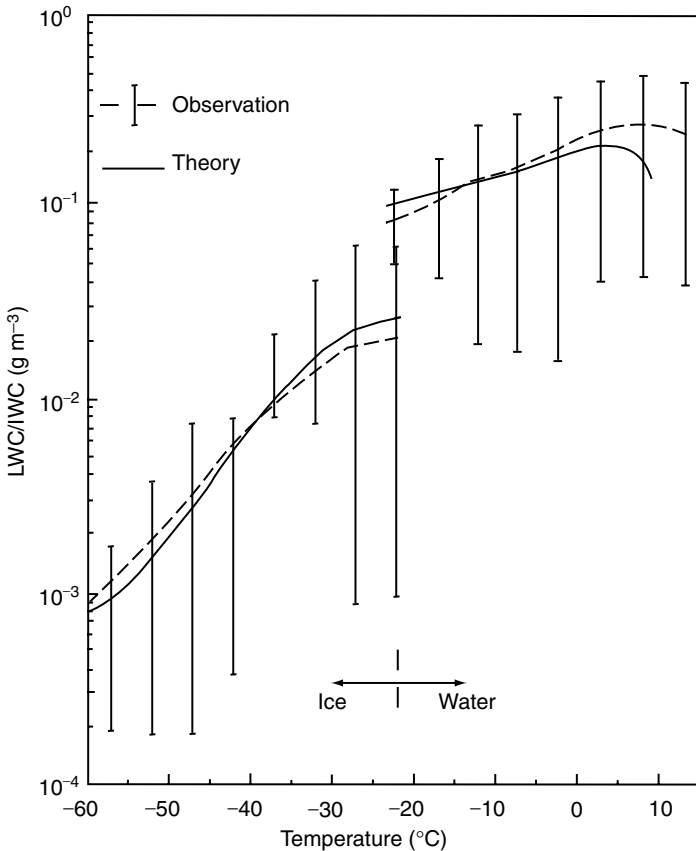
Figure 8.15   Observed and model computed liquid water content (LWC) and ice water content (IWC) as functions of temperature. The observed LWC and IWC data are based on statistical averages of aircraft measurements presented by Matveev (1984) and Heymsfield and Platt (1984), respectively. Theoretical results are derived from diffusion and accretion models.

temperature from $10°C$ to $-60°C$, a range covering the terrestrial atmosphere. Although observed values show large standard deviations, Fig. 8.15 nevertheless reveals the temperature-dependent trend in terms of the mean value, as denoted by the dashed lines. Theoretical results based on diffusion and accretion processes are also illustrated in the diagram for comparison purposes. Based on both the fundamental cloud physics principle and available observations, then, it is clear that as temperature increases, the LWC/IWC of clouds increases as well. On the basis of aircraft microphysics measurements, Heymsfield and Platt (1984) illustrated that there is a distinct correlation between ice-crystal size distribution and temperature. Ice crystals are smaller (larger) at colder (warmer) temperatures. In particular, the colder cloud top primarily contains pristine ice crystals, whereas the warmer cloud bottom is composed of larger irregular ice particles (see Fig. 5.3).

One-dimensional model studies show that if the temperature increases in the region of low clouds, which are composed entirely of water droplets, these clouds will reflect more solar radiation because of greater LWC (the solar albedo effect), leading to a negative feedback (Charlock, 1982; Somerville and Remer, 1984). For high cirrus clouds containing primarily nonspherical ice crystals, a one-dimensional radiative–convective model study illustrates, however, that the balance of solar albedo versus greenhouse effects depends not only on the IWC, but also on the ice crystal size (Ou and Liou, 1995).

The radiative properties of clouds are functions of both the vertical water content and the size spectrum of cloud particles. Because cloud particles scatter an amount of light in proportion to their cross-sectional area, we may define a physical parameter, referred to as the mean effective radius, a_e, in the case of spherical water droplets, or the mean effective size, D_e, in the case of nonspherical ice crystals [Eqs. (7.3.13a) and (7.3.15a)]. From the analysis presented in Section 7.3.5, the visible optical depth is then given by [Eqs. (7.3.13d) and (7.3.15c)]

$$\tau(\text{water}) \cong \frac{3}{2\rho_l}\text{LWP}/a_e,$$

$$\tau(\text{ice}) \cong \text{IWP}(c + b/D_e), \tag{8.4.7}$$

where LWP (IWP) = LWC (IWC) $\cdot \Delta z$, with Δz the cloud thickness. Thus, the optical depth of a cloud is a function of both the LWP (IWP) and mean effective radius (size). For the same LWP or IWP, clouds containing smaller particles would have larger optical depths.

Figure 8.16 shows the cloud albedo covering the solar spectrum (0.2–5 μm) as a function of LWP/IWP and a_e/D_e computed from a multiple scattering program (Liou *et al.,* 1998) involving representative spherical water droplets for water clouds and representative hexagonal ice crystals for ice clouds. For a given a_e/D_e, cloud albedo increases with increasing LWP/IWP. But for a given LWP/IWP, a cloud with a smaller a_e/D_e reflects more solar radiation because of its larger effective optical depth as noted above. The relationship between cloud albedo and a_e/D_e is fundamental to the discussion of the indirect aerosol effect in cloud and climate studies.

8.4.5.3   AEROSOLS/CLOUDS AND PRECIPITATION

Cloud cover and LWC are potentially related to precipitation, which is produced by the autoconversion of cloud droplets, where the time constant is proportional to the droplet size distribution. Variations in cloud particle size would affect cloud albedo, but variations in precipitation would affect the washout of cloud condensation nuclei (CCN), a type of aerosol particle, and thus could alter the formation of the cloud particle size spectrum.

Both theory and experiments indicate that the cloud droplet concentration is approximately proportional to the level of CCN, and that the size of the droplets decreases as the CCN in clouds increase. Over land, numerous observations have indicated that pollution increases the number of CCN leading to more small cloud droplets per unit
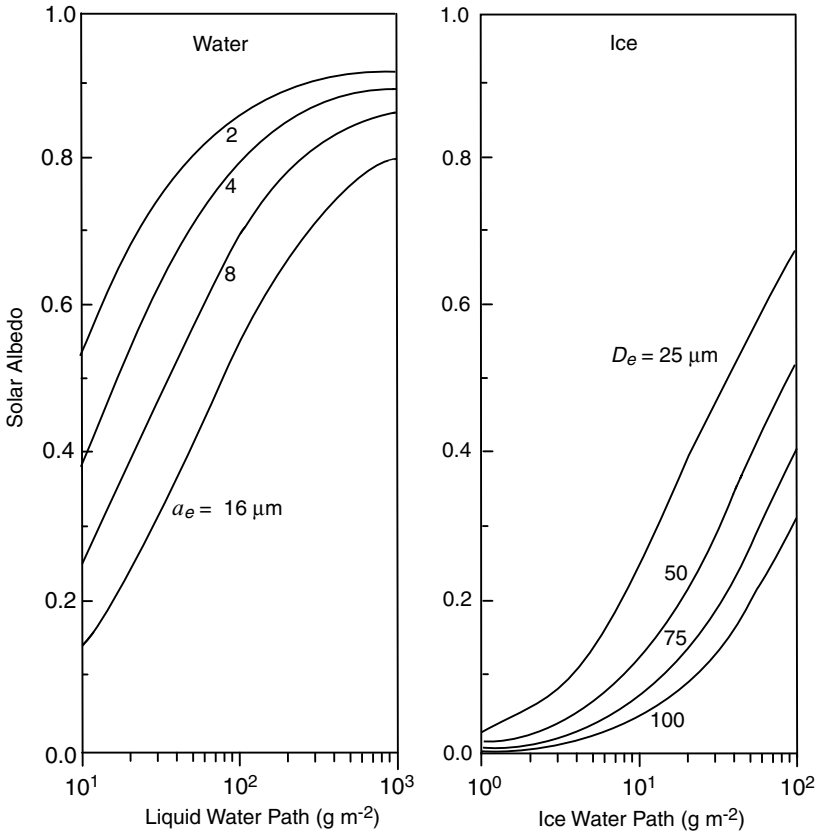
Figure 8.16   Broadband solar albedo of water and ice clouds as a function of the liquid water path (LWP)/ice water path (IWP) and the mean effective water droplet radius (a_e)/ice crystal size (D_e). The range of these values covers representative water and ice cloud conditions.

volume (Twomey *et al.,* 1984). This effect has also been observed from the ship stack effluents that modify the radiative properties of shallow stratus clouds over the oceans (Coakley *et al.*, 1987).

The major source of CCN over the oceans is non-sea-salt sulfate, which is produced from the emission of dimethylsulfide (DMS) by marine organisms. The highest rate of DMS emission into the atmosphere is associated with the warmest, most saline, and most intensely illuminated regions of the oceans (Charlson *et al.,* 1987). An increase in the surface temperature associated with greenhouse warming could cause an increase in DMS emission and, hence, in CCN. The consequence of smaller cloud droplets as a result of the additional production of CCN could lead to brighter clouds, an indirect solar albedo effect that is the functional opposite of greenhouse warming.

A one-dimensional cloud-precipitation-climate model has been developed by Liou and Ou (1989) to investigate the potential link between the perturbed cloud particle size distribution produced by greenhouse effects and climate perturbations. If the

perturbed mean cloud particle radii produced by greenhouse effects or air pollution are smaller than the climatological mean value, precipitation could decrease, leading to increases in cloud LWC. Thus, the solar albedo effect would outweigh the infrared greenhouse effect. A reduction of the mean droplet radius by 0.5 μm could cool the atmosphere to such a degree as to offset the warming produced by CO_2 doubling. Leaitch *et al.* (1992) observed a reduction in droplet radii of about 1 μm in eastern North America as a result of anthropogenic pollution.

A number of model studies have shown that more precipitation could be triggered by the warmer temperature fields produced by greenhouse warming. Precipitation is considered to be the primary mechanism for the removal of atmospheric aerosols, including CCN. Thus, the increase in precipitation could cause a reduction in the number of CCN. The low concentration of efficient CCN could, in turn, lead to larger cloud droplets and a reduction of the cloud albedo. It follows that this could provide a possible mechanism for a positive feedback to greenhouse warming. The potential positive feedback associated with particle size in cloud-climate feedback problems was first noted by Liou and Ou (1989) and Albrecht (1989).

The role of clouds in climate and climatic perturbations due to external radiative forcing is indeed intricate and multidimensional. The results derived from one-dimensional models shed some light on the potential interaction and feedback involving cloud height, cloud LWC, and cloud particle size associated with temperature increases produced by greenhouse warming. However, one-dimensional models do not account for horizontal variations in cloud parameters. In view of the fact that clouds are three-dimensional in nature, their role in climate and global climate change must be resolved from models that can simulate both vertical and horizontal cloud structure and composition based on fundamental physical principles. Further discussion of the role of clouds and radiation in climate is presented in Section 8.6.2.

8.5   Radiation in Energy Balance Climate Models

In the previous section dealing with the earth's radiation budget, we noted that, on average, there is a radiation excess in the tropical region and a radiation deficit in the middle and high latitudes (Fig. 8.2). Significant temperature and water vapor gradients also occur in the atmosphere near the surface. Thus, there must be poleward as well as upward energy transfers in order to produce an overall energy balance. Energy exchanges at TOA are solely due to radiative processes. However, within the earth-atmosphere system, these exchanges involve a number of mechanisms of which radiative energy transfer is only one component.

8.5.1   Energy Budget of the Atmosphere and the Surface

8.5.1.1   ATMOSPHERE AND OCEANS

Based on the energy conservation principle, the zonal mean rate of energy storage per unit area in the atmosphere must be the residual of the net surface flux, the net
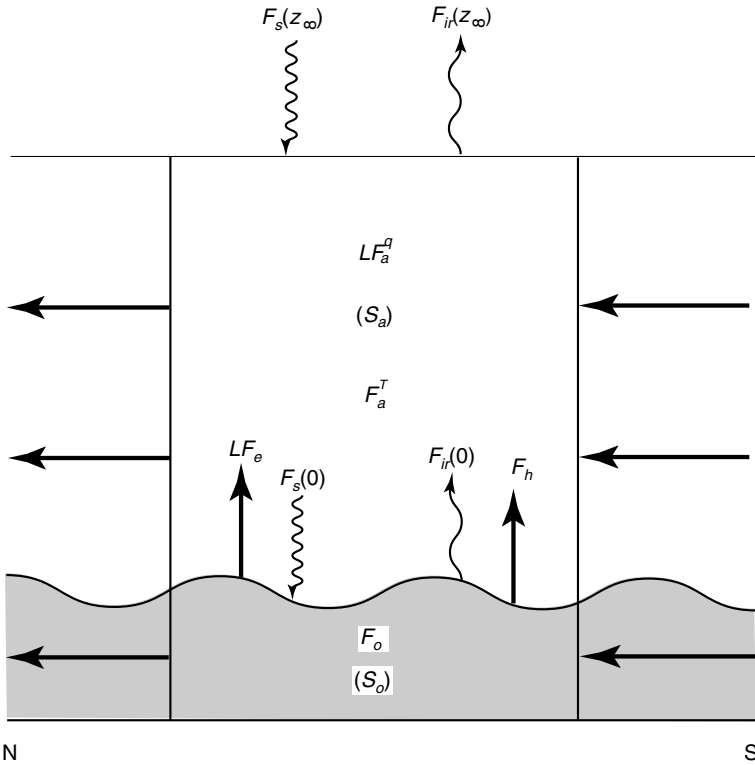
Figure 8.17   Energy components of the earth and the atmosphere, where $F_s(z_\infty)$ and $F_{ir}(z_\infty)$ denote the absorbed solar flux and the emitted IR flux at TOA; $F_s(0)$ and $F_{ir}(0)$ are the absorbed solar flux and the emitted IR flux at the surface; LF_a^q and F_a^T are the divergence of meridional transports of latent and sensible heat; F_o is the divergence of oceanic transports of sensible heat; LF_e and F_h are the vertical fluxes of latent and sensible heat, respectively; and S_a and S_o are the zonal mean rate of energy storage per unit area for the atmosphere and the oceans, respectively.

TOA radiative flux $F(z_\infty)$, and the divergence of atmospheric transport of heat F_a, given by

$$S_a = [F_v(0) + F(0)] - F_a - F(z_\infty), \tag{8.5.1}$$

where $F_v(0)$ denotes the surface vertical flux associated with the transport of sensible (F_h) and latent heat (LF_e) fluxes [see Eq. (8.5.5)]. The atmospheric transport term also consists of two terms involving sensible (F_a^T) and latent heat (LF_a^q) components. Equation (8.5.1) is the basic energy balance equation for an atmospheric column in which the total heat storage rate is balanced by the surface input minus the sum of the divergence of the poleward transport of heat and the net radiative flux at TOA, as displayed in Fig. 8.17. The net radiative flux is the difference between the absorbed solar flux F_s and the emitted IR flux F_{ir}.

At the surface, the energy balance is governed by the rate of energy storage, the net surface flux, and the divergence of the transport of sensible heat below the earth's surface. The heat capacity of the continental surface is generally very small, as in the cases of snow and ice. Thus, on a large scale, the principal heat exchange must occur between the atmosphere and the oceans. Analogous to Eq. (8.5.1) for atmospheric storage, the zonal mean rate of energy storage per unit area for the oceans may be expressed by

$$S_o = -[F_v(0) + F(0)] - F_o, \qquad (8.5.2)$$

where the first term represents the net surface flux between the radiative and turbulent fluxes and F_o is the divergence of the oceanic transport of sensible heat. The zonal mean rate of energy storage per unit area in the atmosphere varies with the seasons from 10 to 20 W m^{-2}, but the annual mean value is about zero. For the oceans, seasonal variations can be as large as 100 W m^{-2}.

The divergence of the atmospheric transport of sensible and latent heat is shown in Fig. 8.18a. Positive and negative values represent the divergence and convergence of heat, respectively. The atmospheric latent heat component has a minimum between about 10°N and 10°S because of excess precipitation in the tropical convective zone. Minima between about 40° and 60° in both hemispheres are a result of the surplus precipitation produced by cyclonic storm activity. Maxima patterns are evident in the subtropics between about 20° and 30° in both hemispheres, where evaporation rates are highest. In the northern subtropics evaporation is less pronounced because of the presence of large desert areas. The atmospheric sensible heat component shows a large maximum in the tropics associated with higher temperatures and Hadley circulation. Small maxima at 40°N and S are related to the transport of heat by eddies and mean meridional circulation. On an annual basis, about 20 W m^{-2} of sensible and latent heat is lost between 40°S and 40°N, while about 50 to 70 W m^{-2} is gained poleward of 60°. In the middle and high latitudes, sensible and latent heat is largely transported by transient eddies. The mean circulation contribution to heat transport is most pronounced in the tropics. For annual mean conditions, the divergence of the oceanic transport of sensible heat may be estimated from Eq. (8.5.1) as a residual, if the atmospheric heat storage, the atmospheric divergence of sensible and latent heat, and the net radiative flux at TOA are given. About 40 W m^{-2} of heat is transported by ocean currents out of the tropics, and about 25 W m^{-2} of heat is transported by ocean currents into latitudes poleward of 40°. Since there are no oceans south of about 70°S, the energy convergence of heat in the oceans should disappear there. Finally, the sum of the divergence of atmospheric and oceanic transports of heat must be balanced by the net radiative flux at TOA on an annual basis (see also Fig. 8.2).

To obtain the total horizontal flux (energy/time) for the entire atmosphere for each component displayed in Fig. 8.18a, we may perform an integration of the horizontal flux F_h over the latitudinal cross section to obtain

$$f_a = 2\pi a_e \cos\varphi \int_0^{z_\infty} F_h \, dz, \qquad (8.5.3a)$$
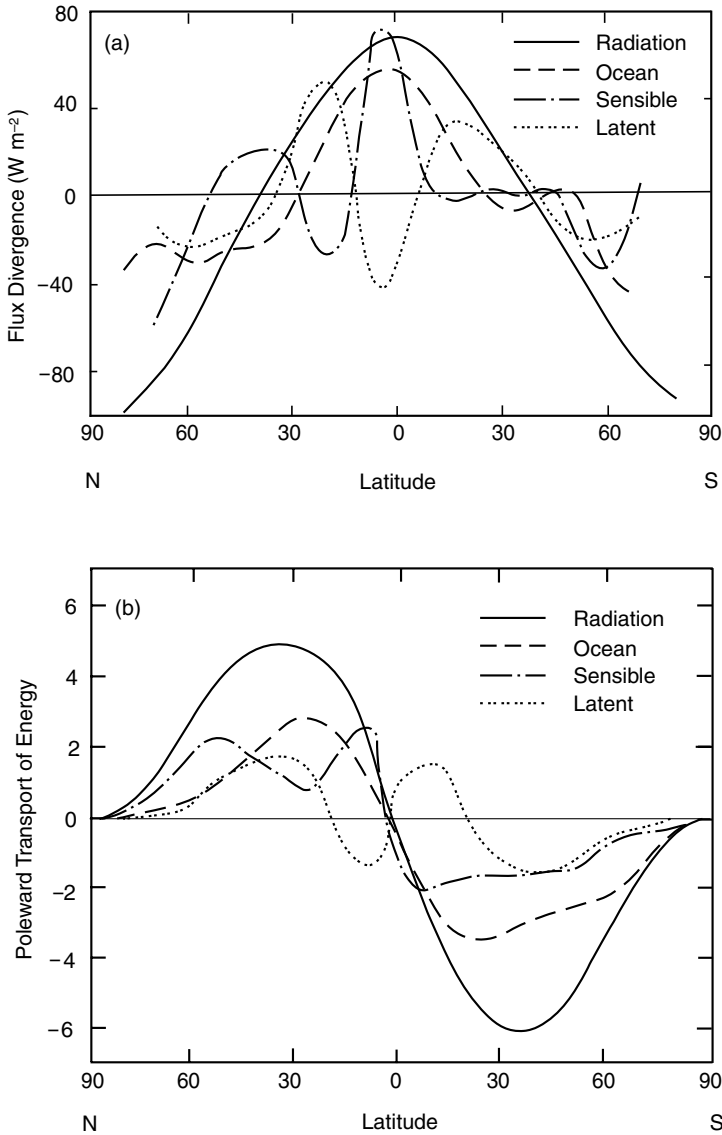
Figure 8.18 (a) Divergence of atmospheric transports of sensible and latent heat, and oceanic transports of sensible heat. Also shown is the net radiative flux density at TOA. (b) Meridional profiles of the poleward transports of sensible and latent heat in the atmosphere, and of sensible heat in the oceans. Also shown is the net radiative flux. All the terms are in units of 10^{15} W (data taken from Oort and Peixóto, 1983).

where a_e is the radius of the earth and φ is the latitude. Figure 8.18b shows the meridional profiles of the poleward transports of sensible and latent heat in the atmosphere, the poleward transport of sensible heat in the oceans, and the net radiative flux (energy/time) at TOA. The oceanic transports dominate in low latitudes with maximum poleward transports of about 3×10^{15} W near 25°N and about -3.5×10^{15} W near 20°S. Latent heat is transported both toward the equator and toward the poles from about 20°N and S, where the evaporation maxima are located. The transports of sensible heat in the atmosphere show double maxima at about 10° and 50° in both hemispheres with values of about 2×10^{15} W. In lower latitudes between 20°N and 20°S, the atmospheric transports of sensible and latent heat largely cancel each other out. As a result, the total atmospheric transport of energy is more important in the middle and high latitudes.

8.5.1.2   SURFACE ENERGY BUDGET

The net surface flux defined in Eq. (8.5.1) consists of the vertical flux and the net radiative flux, which can be divided into its solar and IR components. A large fraction of the solar flux at TOA is transmitted through the clear atmosphere, as is evident in Fig. 3.9. In cloudy conditions, the solar flux available at the surface is largely dependent on the solar zenith angle and cloud optical depth, which is a function of cloud liquid/ice water content and particle size distribution, as illustrated in Fig. 3.19. In reference to Eq. (8.2.30), let the absorbed portion of the solar flux be denoted by

$$F_s(0) = F_s^{\downarrow}(0)(1 - r_s). \tag{8.5.3b}$$

The net IR flux at the surface is the difference between the upward IR flux emitted by the surface and the downward IR flux from the atmosphere reaching the surface $F_{ir}^{\downarrow}(0)$. We may then let the net emitted IR flux be denoted by

$$F_{ir}(0) = \varepsilon_s \sigma T_s^4 - \varepsilon_s F_{ir}^{\downarrow}(0). \tag{8.5.4}$$

The vertical transport of energy from the surface to the atmosphere consists of the flux of sensible heat from the surface F_h and the evaporative water loss from the surface F_e. Thus, we write

$$F_v(0) = LF_e + F_h, \tag{8.5.5}$$

where L is the coefficient for the latent heat of evaporation.

On the basis of the energy conservation principle, the net flux of energy absorbed by the surface must equal the rate at which the surface is storing energy S_s, so that

$$S_s = F_s(0) - F_{ir}(0) - F_v(0)$$
$$= F_s^{\downarrow}(0)(1 - r_s) - \varepsilon_s \sigma T_s^4 + \varepsilon_s F_{ir}^{\downarrow}(0) - LF_e - F_h. \tag{8.5.6}$$

Over the oceans, we have $S_s = S_o + F_o$, from Eq. (8.5.2). Over land and on a microscopic scale, S_s may be expressed in terms of the rate of heat conduction into the underlying surface. Equation (8.5.6) represents the basic surface energy equation.
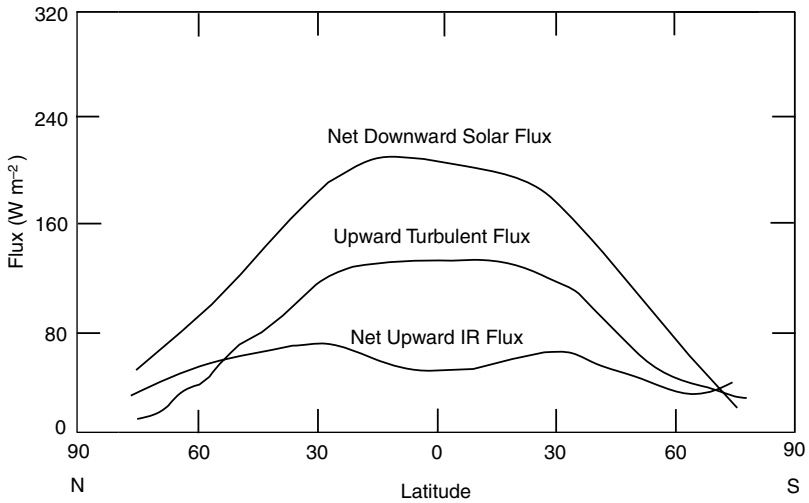
Figure 8.19   Latitudinal distributions of estimated mean annual net solar flux, net IR flux, and total turbulent flux at the earth's surface (data taken from Ou *et al.,* 1989; Budyko, 1982; Sellers, 1965).

Local variations such as the melting of snow and ice have not been considered here. The assumption that $S_s = 0$ has been frequently used in weather prediction and climate models because of its simplicity. This assumption is approximately correct for land surfaces, averaged over 24 hours, but would lead to large errors in determining surface temperature over a diurnal cycle. If $S_s = 0$, surface temperature can be directly computed from Eq. (8.5.6), provided that all the radiative and turbulent fluxes are known.

The surface albedo is dependent on the type of surface, as well as on the solar zenith angle and the wavelength. For water surfaces, the albedo generally ranges from 6 to 9%. The albedo varies significantly over land surfaces with values ranging from 10 to 40%. The albedos of snow and ice are usually greater than about 40%. The dependence of vegetation albedo on solar wavelength is significant (see Fig. 7.7). The thermal IR emissivities of water and land surfaces are normally between 90 and 95%. It is usually assumed that the earth's surfaces are approximately black in infrared radiative transfer calculations. Exceptions include snow and some sand surfaces whose emissivities are wavelength dependent and could be less than 90%. The transport of sensible heat and water vapor from surfaces is governed by turbulent motion and is generally expressed by the vertical gradients of temperature and specific humidity near the surface based on the mixing length theory (Brutsaert, 1982).

Latitudinal distributions of annual net solar and IR fluxes and the total turbulent flux at the earth's surface are displayed in Fig. 8.19. The annual net surface solar flux decreases poleward, resulting in a slight hemispheric asymmetry. Latitudinal distributions of annual net surface IR flux show double maxima at about 35° in both hemispheres; these maxima are associated with subtropical highs. The minimum

pattern in the tropics is due to the large downward IR fluxes produced by the extensive cloud cover in the ITCZ. This pattern is similar to the dip that occurs in the emitted IR flux at TOA. Estimates of the vertical turbulent fluxes of sensible and latent heat can be made by using available climatological data, as shown in Fig. 8.19. The surface turbulent flux decreases from about 100–150 W m^{-2} in the tropics to near zero at the poles. There is a large amount of uncertainty in the estimate of vertical flux in the tropics. A large fraction of the vertical transport takes place through cumulus convection, which occurs in only about 3–4% of the total tropical area. For this reason, it is very difficult to quantify vertical transport based on data obtained from a limited number of observational points. On an annual basis, the sum of the turbulent flux and the net IR flux at the surface is approximately equal to the solar flux absorbed by the surface.

8.5.2   Radiative Forcing in Energy Balance Climate Models

The zonal mean rates of energy storage per unit area in the atmosphere and the oceans may be related to the local time rate of change of surface temperature, as a first approximation, in the form

$$S_a + S_o = c(\varphi)\frac{\partial T}{\partial t}, \tag{8.5.7}$$

where the surface temperature is denoted by T in this section, c is referred to as the *thermal inertia coefficient,* and φ is the latitude. On eliminating the net surface flux from Eqs. (8.5.1) and (8.5.2) and letting the total atmospheric and oceanic transports of energy be

$$R = F_a + F_o, \tag{8.5.8}$$

we have

$$-c(x)\frac{\partial T(t, x)}{\partial t} = -[F_s(t, x) - F_{ir}(t, x)] + R(t, x), \tag{8.5.9}$$

where $x = \sin\varphi$. This is the basic equation for the one-dimensional energy balance climate model. To seek a solution for surface temperature, it is necessary to relate both the infrared and horizontal fluxes to surface temperature and to express the absorbed solar flux as a function of surface albedo.

We may approach the determination of the equilibrium surface temperature using the steady-state condition to obtain

$$F_s(x) = F_{ir}(x) + R(x). \tag{8.5.10}$$

In an infinitesimal latitude belt under equilibrium, the absorbed solar flux must be balanced by the emitted infrared flux and the net horizontal flux divergence, including sensible and latent heat transports by atmospheric motions and sensible heat transports by ocean currents (see Fig. 8.18). In the following, we present two approaches for the evaluation of surface temperature associated with solar constant perturbations.

8.5.2.1   LINEAR HEATING APPROACH

On the basis of the monthly mean values of radiation flux density at the top of the atmosphere for 260 stations, Budyko (1969) developed an empirical formula relating the outgoing infrared flux density, the surface temperature, and the fractional cloud cover in the form

$$F_{ir}(x) = a_1 + b_1 T(x) - [a_2 + b_2 T(x)]\eta, \qquad (8.5.11)$$

where, for simplicity, the surface temperature T_s is set as T, as noted above, η is the fractional cloud cover, and a_1, b_1, a_2, and b_2 are empirical constants. The influence of the deviation of cloudiness from its mean global value on the temperature is normally neglected because of the intricate relation of clouds with the radiation field and surface albedo. With a fractional cloud cover η of 0.5, Eq. (8.5.11) may be rewritten in the form

$$F_{ir}(x) = a + bT(x), \qquad (8.5.12)$$

where the coefficients a and b are defined in Eq. (8.4.5). As pointed out in Section 8.4.3, the physical explanation of the linear relation between the outgoing infrared flux and the surface temperature is that, since the temperature profiles have more or less the same shape at all latitudes, infrared cooling, which depends on the temperature at all levels, may be expressed in terms of the surface temperature.

The incoming solar flux density may be expressed by

$$F_s(x) = Qs(x)[1 - r(x)] = Qs(x)A(x), \qquad (8.5.13)$$

where $Q = S/4$, S is the solar constant, r is the planetary albedo, which is allowed to vary in terms of temperature, A is the solar flux absorbed by the earth–atmosphere system, and $s(x)$ is the normalized mean annual distribution of insolation at each latitude. $Qs(x)$ can be obtained from the values presented in Fig. 2.8.

Moreover, to relate the surface temperature distribution and the horizontal heat transfer in the atmosphere and hydrosphere, Budyko derived a simple linear equation by comparing the observed mean latitudinal values of $R(x)$ with the difference of the annual mean temperature at a given latitude and the global mean temperature $\bar{T}$, given by

$$R(x) = c[T(x) - \bar{T}], \qquad (8.5.14)$$

with the empirical constant $c \cong 3.75 \ \text{W m}^{-2} \ \text{C}^{-1}$.

Substituting Eqs. (8.5.12)–(8.5.13) into Eq. (8.5.10) leads to

$$Qs(x)A(x) - [a + bT(x)] = c[T(x) - \bar{T}]. \qquad (8.5.15a)$$

The surface temperature is then given by

$$T(x) = \frac{Qs(x)A(x) - a + c\bar{T}}{c + b}. \qquad (8.5.15b)$$

Furthermore, over a climatological time scale, the earth–atmosphere system as a whole should be in radiative equilibrium so that

$$Q\bar{A} - (a + b\bar{T}) = 0, \tag{8.5.16}$$

where the global surface temperature, global absorptivity, and global albedo are defined by

$$\bar{T} = \int_0^1 T(x)\,dx, \qquad \bar{A} = 1 - \bar{r} = \int_0^1 s(x)A(x)\,dx. \tag{8.5.17a}$$

Consequently,

$$\bar{T} = (Q\bar{A} - a)/b. \tag{8.5.17b}$$

At this point, the latitudinally dependent surface temperature may be computed as a function of x for given $s(x)$ and $A(x)$. The planetary albedo is usually defined by a simple step function depending on whether or not an ice sheet exists. By letting x_i represent the sine of latitude φ_i of the ice line, the absorptivity or albedo can be expressed by (Budyko, 1969; Sellers, 1969)

$$A(x, x_i) = 1 - r(x, x_i) = \begin{cases} A_1, & x > x_i \\ A_2, & x < x_i. \end{cases} \tag{8.5.18}$$

Exercise 8.5 requires computation of the temperature as a function of x.

   To study the model's sensitivity to solar constant perturbations, a linear analysis of the global surface temperature, solar constant, and global albedo may be carried out by setting $\bar{T} = \bar{T}_0 + \Delta\bar{T}$, $Q = Q_0 + \Delta Q$, and $\bar{r} = \bar{r}_0 + \Delta\bar{r}$, where $\bar{T}_0$, Q_0, and $\bar{r}_0$ denote the mean values for the present climate condition. It follows that

$$\begin{aligned} T(x) = \quad &\frac{1}{c+b}\left\{ Q_0 s(x)[1 - r(x)]\left(1 + \frac{\Delta Q}{Q_0}\right) - a + c\bar{T}_0 \right. \\ &\left. + \frac{cQ_0}{b}\left[\frac{\Delta Q}{Q_0}(1 - \bar{r}_0 - \Delta\bar{r}) - \Delta\bar{r}\right] \right\}. \end{aligned} \tag{8.5.19}$$

Thus, the effects of a change in the solar constant on the earth's mean annual surface temperature can be investigated in a manner that includes the coupling effect of the changing global albedo, an indication of the change in a glaciated area. In his original work, Budyko expressed $\Delta\bar{r}$ in terms of an empirical function in the form $0.3\,l(x, \bar{x}_i)s^*(x_i)$, where $\bar{x}_i (= 0.95)$ represents the present ice-line position, l is the ratio of the change in the ice-covered area to the total area of the northern hemisphere, and s^* is the ratio of the mean solar flux in the zone of the additional ice area to the mean solar flux for the entire hemisphere.
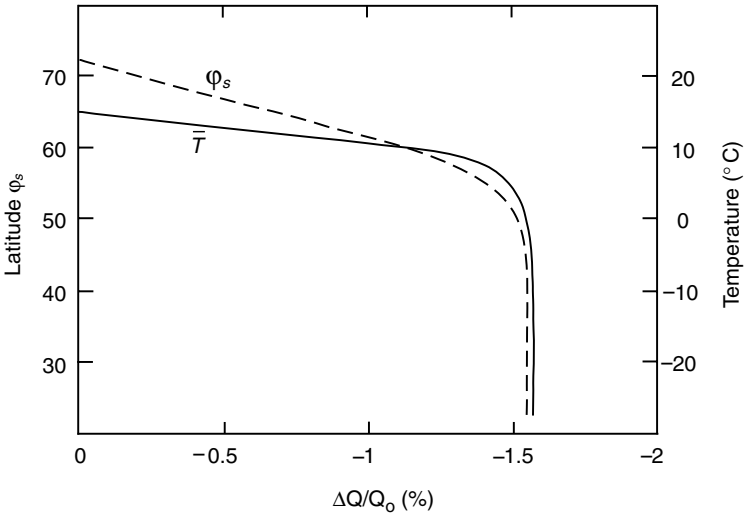
Figure 8.20 Dependence of the global surface temperature and the latitude of glaciation on changes in the solar constant (data taken from Budyko, 1969). If this change is reduced by about 1.6%, the ice line reaches a latitude of about 50°N and advances southward to the equator without any further reduction in the solar constant. This is referred to as Budyko's theory of the runaway ice-covered earth, due to the ice–albedo feedback. Additional theoretical calculations show that a substantial decrease in the solar constant on the order of 10% is required to cause an ice-covered earth.

Using Eq. (8.5.19), calculations can be carried out to investigate the position of the ice line for different values of $\Delta Q/Q_0$. Shown in Fig. 8.20 are the latitude φ_s, corresponding to the ice-line position, and the global temperature $\bar{T}$ as functions of $\Delta Q/Q_0$. In these calculations, the mean value for the solar constant used was 1353 W m^{-2}, and the global albedo was set at 0.33. Based on climatological data, the ice-line temperature $T(x_i)$ was assumed to be $-10°$C. The step functions for absorptivity were $A_1 = 0.38$, $A_2 = 0.68$, and $A_{12} = 0.5$ at the ice line. With a 1% decrease in the incoming solar flux, the global surface temperature is reduced by about 5°C. Further, a 1.5% decrease in the incoming solar flux reduces the global surface temperature by about 9°C. The response to these decreases in temperature is a southward advance of glaciation by about 18° of latitude, corresponding to the advance of quarternary glaciation. Based on these calculations, when $\Delta Q/Q_0$ is reduced by about 1.6%, the ice line reaches a latitude of about 50°N, at which point the global surface temperature decreases to several tens of degrees below zero. As a result, the ice sheet begins to advance continuously southward to the equator with no further reduction in the solar flux required. Note that with a constant global mean albedo (i.e., $\Delta \bar{r} = 0$), a 1% decrease in $\Delta Q/Q_0$ lowers the global surface temperature by only 1.2–1.5°C. The significance of the ice–albedo coupling effect is quite apparent. The preceding process describes Budyko's theory of the runaway ice-covered earth, which is based on the ice–albedo feedback. A decrease in the solar constant produces surface cooling, which leads to an increase in the ice-covered area. In

turn, the surface albedo increases, resulting in a reduction in the absorption of solar flux.

8.5.2.2   DIFFUSION APPROACH

A more rigorous approach to parameterizing horizontal heat transport can be constructed by utilizing the diffusion concept in a large-scale sense. From Eq. (8.5.8), we have

$$R = LF_a^q + F_a^T + F_o = F_s - F_{ir}. \tag{8.5.20}$$

The three large-scale transport terms may be expressed by a linear function of surface temperature differences in terms of a second-order equation in ΔT.

The second-order equation in ΔT for horizontal heat transport is, in essence, the diffusion approximation (Held and Suarez, 1974; North, 1975; Ghil, 1976). A thermal diffusion form for horizontal heat transport, $-D\nabla^2 T$, was adopted with D being an empirical coefficient to be determined by fitting the present climate. Thus, all the transport processes are parameterized with a single coefficient, analogous to an eddy diffusion approach to dispersion by macroturbulence in the entire geofluid system. Using the spherical coordinates for the Laplace operator [see Eq. (5.2.29)] in the context of a one-dimensional latitudinal variation, we have

$$-D\nabla^2 T = \frac{-D}{a_e^2 \sin\theta} \frac{d}{d\theta}\left(\sin\theta \frac{dT}{d\theta}\right) = \frac{-D}{a_e^2} \frac{d}{dx}(1-x^2)\frac{d}{dx}T(x), \tag{8.5.21}$$

where the polar angle $\theta = 90° - \varphi$ and a_e is the radius of the earth. Letting $D' = D/a_e^2$, we find from Eq. (8.5.10) that

$$D'\frac{d}{dx}(1-x^2)\frac{d}{dx}T(x) = F_{ir}(x) - Qs(x)A(x, x_i). \tag{8.5.22}$$

Since F_{ir} and T are linearly related through Eq. (8.5.12), we may rewrite Eq. (8.5.22) in the form

$$\left[\frac{d}{dx}(1-x^2)\frac{d}{dx} - \frac{1}{D''}\right]F(x) = \frac{Q}{D''}s(x)A(x, x_i), \tag{8.5.23}$$

with $D'' = D'/b$, and we let $F_{ir} = F$ for the convenience of presentation. We must now specify the ice-sheet edge x_i. It is generally assumed that if $T(x) < T_i$, ice will be present, whereas if $T(x) > T_i$, there will be no ice. In terms of infrared flux, we may set $F(x_i) = F_i$. As noted previously, the ice-line temperature T_i is normally assumed to be $-10°C$.

For a mean annual model with symmetric hemispheres, the boundary condition must be that there is no heat flux transport at the poles or across the equator, i.e., $\nabla F(x)$ or $\nabla T(x) = 0$ at $x = 1$ and 0, respectively. Thus,

$$(1-x^2)^{1/2}\frac{d}{dx}F(x)\bigg|_{x=0} = (1-x^2)^{1/2}\frac{d}{dx}F(x)\bigg|_{x=1} = 0. \tag{8.5.24}$$

To obtain a solution we may expand $F(x)$ in terms of the Legendre polynomials in the form

$$F(x) = \sum_{n=\text{even}} F_n P_n(x), \tag{8.5.25}$$

where only even terms are used because $F(x)$ is an even function of x in the mean annual case, i.e., symmetric between two hemispheres, and F_n represent the unknown coefficients to be determined. Since the Legendre polynomials are the eigenfunctions of the spherical diffusion equation, defined in Eq. (5.2.42) (for $\ell = 0$), we have

$$\frac{d}{dx}(1 - x^2)\frac{d}{dx}P_n(x) = -n(n+1)P_n(x). \tag{8.5.26}$$

Moreover, $(1 - x^2)^{1/2}dP_n(x)/dx = 0$ for $x = 0$, and 1 when $n = $ even. It follows that the imposed boundary conditions described in Eq. (8.5.24) are satisfied by the expansion.

Substituting Eq. (8.5.25) into Eq. (8.5.23) and utilizing the orthogonal property of P_n (see Appendix E), we find

$$F_n = QH_n(x_i)/[1 + n(n+1)D''], \tag{8.5.27a}$$

where

$$H_n(x_i) = (2n+1)\int_0^1 s(x)A(x, x_i)P_n(x)\,dx, \tag{8.5.27b}$$

which can be evaluated from the known values for $s(x)$ and $A(x, x_i)$. The final step in completing the solution is to determine the diffusion transport coefficient D''. This can be done empirically by varying D'' in Eq. (8.5.27a) to satisfy the present climate conditions, i.e., $\bar{x}_i = 0.95$, $F(\bar{x}_i) = F_i$, and $Q = Q(\bar{x}_i) = Q_0$. The solution can then be used to investigate the ice-line position as a function of Q. Using Eqs. (8.5.25) and (8.5.27a) and letting $x = x_i$, we obtain

$$Q(x_i) = F_i\left[\sum_{n=\text{even}}\frac{H_n(x_i)P_n(x_i)}{1 + n(n+1)D''}\right]^{-1}. \tag{8.5.28}$$

The normalized mean annual distribution of insolation $s(x)$ can be fitted by Legendre polynomial expansions. With a degree of accuracy of about 2%, it is given by

$$s(x) = \sum_{n=\text{even}} s_n P_n(x) \approx 1 + s_2 P_2(x), \tag{8.5.29}$$

where $s_2 = 0.482$. Based on observed data, the absorptivity of ice-free latitudes may also be fitted by Legendre polynomial expansions in the form

$$A_2(x, x_s) = d_0 + d_2 P_2(x), \quad x < x_s, \tag{8.5.30}$$

with $d_0 = 0.697$ and $d_2 = -0.0779$. The absorptivity over ice or snow with 50% cloud cover is assumed to be 0.38 as defined previously.

The results derived from the preceding diffusion model show the multiple branch nature of the solution. A southward advance of glaciation requires a decrease in the solar constant of about 8 to 9% (North, 1975). After the ice line reaches ∼45–50°N, its southward advance continues even though the incoming solar flux increases. This result is basically in agreement with Budyko's theory, although the decrease in the solar constant that is needed to produce an ice-covered earth is substantially greater. Finally, it is noted that the diffusion models with nonlinear coefficients, such that D is proportional to dT/dx, produce virtually the same results and conclusions (Held and Suarez, 1974; North *et al.,* 1981).

The simple climate models just described are not in reasonable agreement with the nearly isothermal surface temperatures observed within 30° of the equator, as argued by Lindzen and Farrell (1977). To introduce tropical transport, referred to as Hadley cell transport, into simple climate models, they introduced an empirical adjustment to include a heat-flux term. This heat flux approached zero for latitudes greater than a prescribed latitude $\varphi_h(\sim 25°\text{N})$. For Budyko's linear heating model, $Qs(x)A(x, x_i)$ was replaced by its average over the region $0 \leq \varphi \leq \varphi_h$. Lindzen and Farrell's model showed that at 25° latitude, identified as the Hadley stability ledge, a reduction in the solar constant did not significantly alter the ice-line position. It was not until the solar constant reduction reached about 15 to 20% that the glaciation advanced continuously southward. From 25° to about 60°, identified as the moderately strong stability region, a southward advance of glaciation was related to the decrease of the solar constant. This model illustrated that the reduction in the solar constant necessary for an ice-covered earth is much greater than Budyko's predicted value of 1.6%.

8.5.3   Solar Insolation Perturbation

A number of external factors have been proposed as the major causes of the earth's climatic variations during the Pleistocene period, about 600,000 years ago. This period is the first epoch of the Quaternary period in the Cenozoic era, characterized by the spreading and recession of continental ice sheets and by the appearance of the modern human. Variations in the output of the sun, seasonal and latitudinal distributions of incoming solar radiation due to the earth's orbital changes, the volcanic dust content of the atmosphere, and the distribution of carbon dioxide between the atmosphere and the oceans have all been hypothesized to have caused the fluctuations of the great ice sheets that were characteristic of this period. The orbital theory of climate change originally described by Milankovitch (1941) appears to receive considerable support based on climatological and geological records. In Section 2.2, we detailed the variation in solar insolation as a function of the earth's orbital parameters: eccentricity, oblique angle, and the longitude of the perihelion (precession).

Based on measurements of the oxygen isotopic composition of planktonic foraminifera from deep-sea sediment cores taken in the southern Indian Ocean, Hays *et al.* (1976) have reconstructed the climatic record of the earth up to ∼500,000 years ago. Summer sea-surface temperatures at the core site have been estimated based on statistical analyses of radiolarian assemblages. Spectral analyses of paleoclimatic
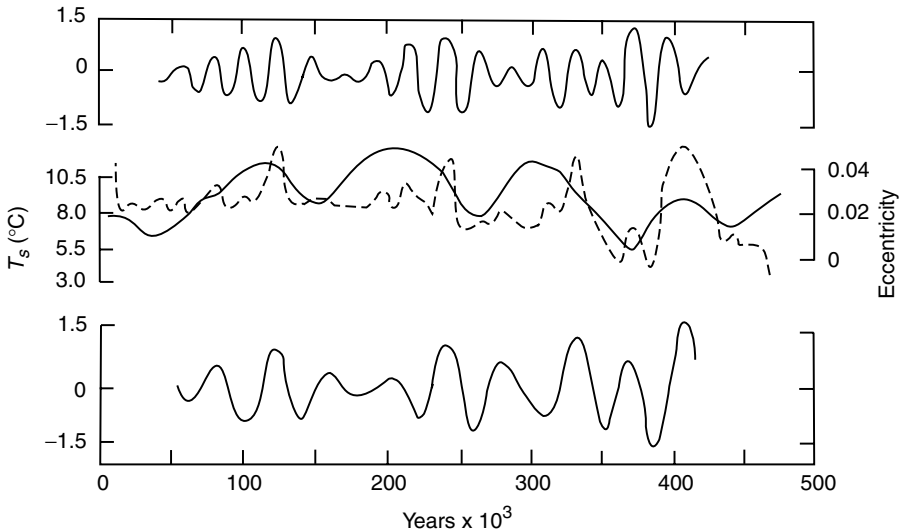
Figure 8.21   Variation in eccentricity and surface temperature over the past 500,000 years. Dashed line in the center shows variations in the estimated sea surface temperature T_s. Solid line denotes the orbital eccentricity. Upper and lower curves are the 23,000- and 40,000-year frequency components extracted from T_s based on a statistical filter method (data taken from Hays *et al.,* 1976).

time series indicate significant peaks in the frequencies at which the earth's orbital parameters are known to vary. These observations are of fundamental importance to our understanding of climatic sensitivity. Figure 8.21 shows that the principal periods of climatic variation (~100,000, 42,000, and 23,000 years) correspond to periods of orbital changes in eccentricity (~105,000 years), obliquity (~41,000 years) and precession (~23,000 and 19,000 years).

Down-core variations in $\delta^{18}O$ reflect changes in oceanic isotopic composition, which are considered to be caused primarily by the waxing and waning of the great Pleistocene ice sheets. Measurements of the ratio of ^{18}O and ^{16}O have been reported with respect to an international standard as $\delta^{18}O$ in parts per thousand. Employing observations of $\delta^{18}O$ in five deep-sea cores, Imbrie *et al.* (1984) have developed a geological time series for the past 780,000 years to evaluate the orbital theory of the Pleistocene ice ages and found that the orbital and isotopic signals are strikingly coherent in the 41,000- and 22,000-year components.

One-dimensional energy balance climate models have been used in an attempt to explain glacial cycles in terms of solar insolation changes associated with the earth's orbit. The fractional changes of mean annual global insolation are approximately given by $\Delta e^2/2$, as defined in Eq. (2.2.26). Since the eccentricity is less than about 0.07, the resulting changes in solar insolation are less than ~0.2%, which corresponds to a change of about 0.4°C in the global mean surface temperature based on a one-dimensional energy balance model (North *et al.,* 1981). This change in surface temperature is an order of magnitude smaller than the changes reported by Hays *et al.*

(1976), cited earlier. It has been speculated that the glacial maximum of 18,000 years ago was due to a variation in the declination of the sun from about $22.2°$ to its present value of $23.45°$. This change resulted in less insolation at the poles and more at the equator. Based on the annual energy balance model, this obliquity change causes only a 1–$2°$ equatorward shift in the ice line, much smaller than the approximately $15°$ shift reconstructed by the CLIMAP Project Members (1976).

The small responses of the annual energy-balance climate model to orbital parameters have led to the suggestion that seasonal variations must be incorporated in the model to produce more accurate orbital solar forcings. Suarez and Held (1979) have developed a seasonal energy balance model to investigate orbital forcings. Significant discrepancies, however, exist between the model's present climate and the actual record. Moreover, the response of the surface temperature to orbital parameters in the model is fairly linear and cannot be used to explain the strong surface-temperature variations that are present in the spectral analyses of the paleoclimatic record over the last 100,000 years. Although physical factors such as oceanic heat transports, variations in cloudiness, and atmospheric transports of sensible and latent heat could be accounted for more comprehensively in one- or two-dimensional energy-balance climate models, it is not evident that such simplified models, with their inherent limitations, would be physically adequate to successfully reproduce the paleoclimatic record. However, through the use of global climate models, some success has been achieved in understanding how past glacial climates were maintained and in estimating the relative significance of the known feedback mechanisms involving the changes in ice sheets, land albedo, and atmospheric CO_2 in forcing the temperature changes that existed 18,000 years ago (Broccoli and Manabe, 1987). Simulations of paleoclimate records require an efficient physically based global model that couples the atmosphere and the oceans, as well as a computer that is capable of carrying out long-term calculations.

8.6   Radiation in Global Climate Models

8.6.1   An Introduction to General Circulation Modeling

Radiative processes directly influence the dynamics and thermodynamics of the atmosphere through the generation of radiative heating/cooling rates, as well as through net radiative fluxes available at the surface. In Fig. 8.2 we showed the radiation balance at the top of the atmosphere in terms of the zonally averaged patterns for the absorbed solar and emitted infrared fluxes observed from satellites. As illustrated in Fig. 8.2, there is a gain of radiative energy between $40°N$ and $40°S$, whereas there are losses in the polar regions. This pattern is largely caused by the sharp decrease in insolation during the winter season and the high surface albedo in the polar regions. In addition, we note that the outgoing infrared flux is only slightly latitudinally dependent, owing to the larger amount of atmospheric water vapor and the higher and colder clouds in the tropics.

This pattern of radiative energy excess and deficit generates the equator-to-pole temperature gradient and, subsequently, produces a growing store of zonal mean available potential energy. In the equatorial region, warm air expands upward and creates a poleward pressure gradient force at the upper altitudes, where air flows poleward from the equator. In the upper levels, air cools and sinks in the subtropical high-pressure belts located at about 30° and returns to the equator at the surface. Kinetic energy is generated as a result of the work done by the horizontal pressure gradient force. This thermally driven circulation between the equator and subtropics is referred to as the *Hadley cell*. As a result of the earth's rotation, air flowing toward the equator at the surface is deflected to the west and creates easterly trade winds. In the upper level of the Hadley cell, the Coriolis deflection of poleward-moving air generates westerly winds.

Similar thermally driven circulation is found in the polar regions. Cold air shrinks downward, producing an equator-directed pressure gradient force and motion in the lower altitudes. The sinking motion over the poles results in airflow in the upper level toward the pole and into the low-pressure belts located at about 60°. Thus, a Hadley circulation develops between the poles and the subpolar low-pressure regions. Here, the effect of the Coriolis force is the same; easterly winds are produced at the surface, while westerly winds are generated aloft. In the Hadley circulation, the atmosphere may be regarded as an engine that absorbs net heat from a high-temperature reservoir and releases this heat to a low-temperature reservoir. The temperature differences generate the available potential energy, which in turn is partly converted to kinetic energy to overcome the effect of friction.

Poleward zonal thermal winds at the upper altitudes become unstable in the middle latitudes, an effect referred to as *baroclinic instability* (Charney, 1947). A reverse circulation is generated in which warm air sinks in the subtropical highs and cold air rises in the subpolar lows in such a manner that westerly winds prevail at all levels. The meridional circulation in this region cannot be explained by the direct heating and cooling effects, as in the Hadley circulation, and cannot generate kinetic energy. The maintenance of westerlies in the middle latitudes is explained by the continuous transfer of angular momentum from the tropics, influenced by large-scale wave disturbances. The baroclinic waves transport heat poleward and intensify until heat transport is balanced by the radiation deficit in the polar regions.

The general circulation of the atmosphere described above is governed by a set of physical and mathematical equations defining the horizontal velocity fields, the first law of thermodynamics, the requirement of air-flow continuity, the large-scale hydrostatic balance, and the thermodynamic state. This system of equations consists of six unknowns: the two horizontal velocity components, the large-scale vertical velocity, density, pressure, and temperature. Additional equations can be introduced to include moisture variation based on the conservation of water vapor, and liquid and ice water content. This system of equations constitutes the so-called *general circulation model* (GCM).

To describe a GCM, representation in the vertical direction must be introduced. There are several alternative ways of treating the vertical coordinate in numerical

models. Because of the hydrostatic nature of large-scale motion, the height coordinate can be converted to the pressure coordinate to simplify the continuity equation and the density variation that occurs in other equations. However, the pressure coordinate encounters difficulties over a mountain, since a particular constant-pressure surface may intersect this mountain at certain times but not others. For this reason, the σ-coordinate system, introduced by Phillips (1957), has been used to remove the difficulty in the representation of lower-boundary conditions. The term $\sigma = p/p_s$, where p_s is the surface pressure; $\sigma = 0$ when $p = 0$; and $\sigma = 1$ when $p = p_s$. The surface pressure in the σ-coordinate follows the terrain height at a given point in time and hence becomes a variable. While the σ-coordinate removes the problem in the representation of the lower-boundary conditions, it generates another problem in the computation of the pressure gradient force term, which is a small quantity but splits into two terms with large values. Numerical methods have been developed to circumvent this problem by interpolating between σ and p coordinates. Since the observed data are normally available in the p-coordinate, numerical interpolations are also required for the data corresponding to the initial field. For verification purposes, transformation of the predicted variables in the σ-coordinate from the model to the p-coordinate is required.

Appropriate numerical methods and initial conditions are required to solve the velocity components, temperature, humidity variables, and a term referred to as the *geopotential height,* the product of the height and gravitational acceleration, as functions of time and space, defined in spherical coordinates, based on a set of nonlinear partial differential equations. The evolution of the atmospheric flow patterns may be determined by integrating the prediction equations beginning at some initial time. Because of the nature of the nonlinearity in these partial and ordinary differential equations, analytic solutions are not practical, and numerical methods for their solution must be used. In the design of numerical methods, appropriate time and spatial differencing schemes are required to ensure computational stability. In the finite-difference approach, any dependent variable is approximated in terms of its values at discrete points in space and time (Arakawa, 1966). An alternative approach is the spectral method, which represents the flow field in space as a finite series of smooth and orthogonal functions. The prediction equations are then expressed in terms of the expansion coefficients, which are dependent on time.

An initial condition consisting of three-dimensional distributions of dependent variables is required to perform the numerical integration of a prediction model. Initial conditions for weather prediction are based on global observations at a specific time from which interpolations of the atmospheric variables are performed to model grid points by means of the objective analysis. Because of the initial imbalance of pressure and wind fields due to the presence of gravity waves, the initial data must be adjusted by numerical methods in order to eliminate meteorological noise. The prediction of large-scale geopotential height fields for a time period of about 5 days is largely influenced by the initial conditions and the data initialization procedure. Observational uncertainty in the initial state grows during model prediction because of the random turbulent nature of atmospheric motion. There is an inherent time limit
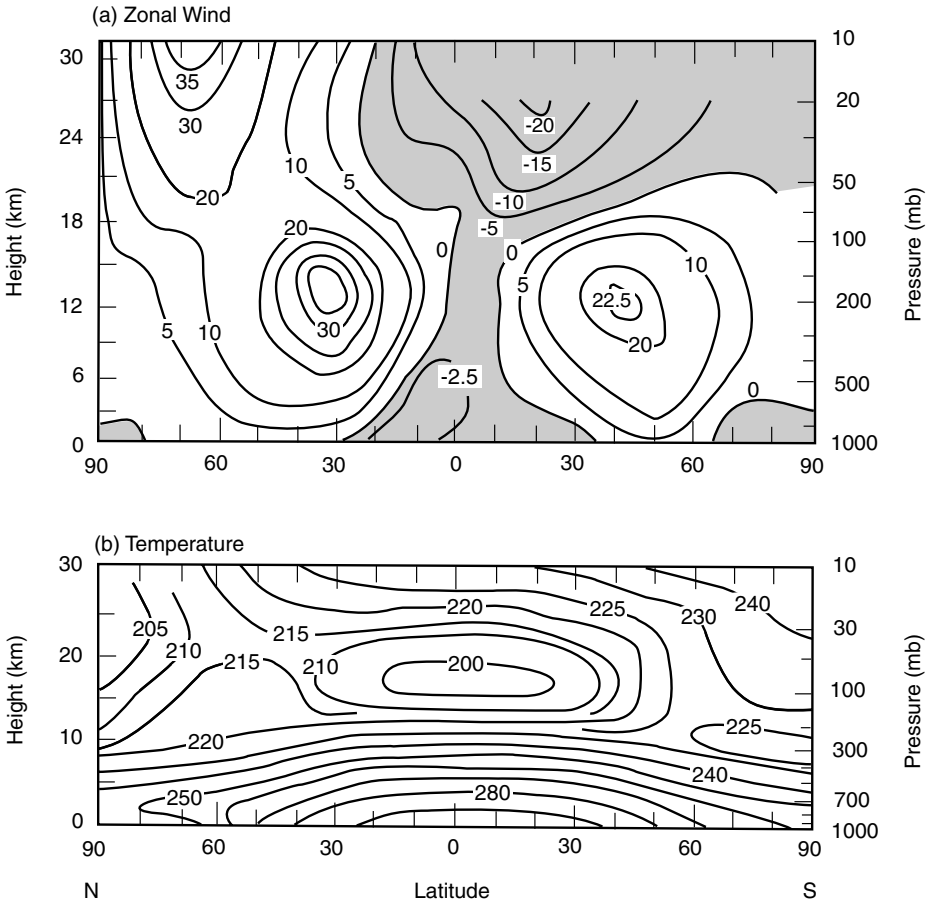
Figure 8.22   Observed (a) zonal mean wind (m s^{-1}) and (b) temperature (K) in height (pressure)–latitude cross section for December, January, and February. Negative regions of winds are shaded (data taken from Newell *et al.,* 1974).

of about a few weeks for predictability of the atmospheric state by means of a numerical integration of atmospheric models (Lorenz, 1969). This predictability limit is associated with the limited ability of atmospheric models to predict the day-to-day evolution of the atmosphere in a deterministic sense. It is feasible, however, for models to predict the statistical properties of the atmosphere. Models may be "tuned" to the mean or equilibrium state of the atmosphere, that is, the present climate, in terms of general circulation, temperature, radiative balance, cloud, and other pertinent patterns, and are referred to as *climate models*. Figure 8.22 displays the zonal mean wind and temperature for climatological winter conditions. Climatological radiation budgets were presented in Section 8.2.2, while cloud-cover climatology was illustrated in Fig. 7.21.

Through the process of tuning the model, all the relevant coefficients and/or computational schemes are adjusted to present climate conditions. When climate is perturbed by external forcings, such as the increase of greenhouse gases, model coefficients (e.g., drag coefficient) and computation schemes (e.g., cloud-cover formation) may no longer be representative of the perturbed climate. Thus, it is critically important to understand the physical processes involved in the model and to incorporate physical feedbacks into the model.

The oceans cover about 70% of the earth's surface. Through their fluid motion and high heat capacity, as well as their ecosystems, the oceans play an essential role in shaping the earth's climate and its variability. Wind stress at the ocean surface drives the large-scale circulation in the upper levels of the oceans. Evaporation from the ocean surface provides the water vapor needed for cloud formation and the resulting latent heat release. Ocean circulation redistributes heat and fresh water around the globe. The oceans store heat and absorb solar visible radiation at the surface, contributing substantially to the variability of climate on many time scales. The ocean thermohaline circulation also allows water from the surface to be carried into the deep ocean and may sequester heat away from the atmosphere for periods of thousands of years. The oceans absorb CO_2 and other trace gases as well as aerosol particles and exchange them with the atmosphere, thereby altering ocean circulation and climate variability. It is clear that the modeling of climate and climate change must include an adequate description of oceanic processes and the coupling between the oceans and the atmosphere.

Figure 8.23 presents a schematic description of the principal components of the physical and mathematical definitions and feedbacks of a GCM for climate studies, including the contribution from the oceans, with reference to the transfer of solar and IR radiation in the atmosphere and their exchange with the earth's surface. In the following, we present the relative significance of radiative processes in global climate models, particularly those concerning the roles of clouds and aerosols.

8.6.2 Cloud Radiative Forcing in Global Climate Models

In Sections 8.2.3 and 8.4.5, we pointed out the importance of clouds in climate from the perspective of the radiation budget of the earth–atmosphere system as well as their role in greenhouse warming perturbations. The estimates of warming due to doubled CO_2 produced by a number of numerical simulations of GCMs have shown surface temperature increases from about 1.3 to 5.4 K. Much of the difference in results has been attributed to the representation of clouds in the models. Early GCMs that were used for predicting climate change included fixed cloud covers estimated from climatology. Variations in clouds that may be forced by a climate perturbation and any feedback due to these cloud variabilities cannot be simulated in a fixed cloud experiment. Moreover, the variability of the stimulated present climate condition and the effect of the diurnal cycle may also be affected in this type of experiment. Interactive cloud parameterizations developed for climate models involve the direct use of relative humidity for the formation of cloud cover and the incorporation of a prognostic cloud water variable. The latter scheme allows the interaction and feedback
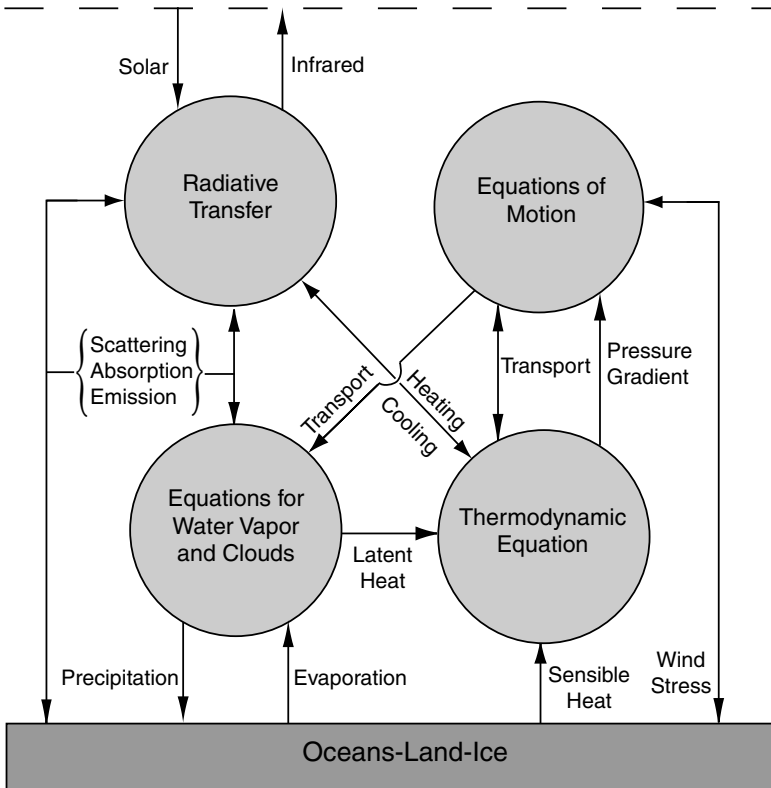
Figure 8.23   Principal components of the physical and mathematical definitions and interactions of a general circulation model (GCM) for climate simulations, particularly in reference to radiative transfer in the earth-atmosphere system. See also Fig. 8.17.

of the radiative properties of clouds to be made dependent on the predicted cloud water path, a more physically based approach to understanding the role of clouds and radiation in climate and climate change. In the following, we present representative cloud simulations in GCMs with a particular emphasis on cloud radiative forcing.

8.6.2.1   INTERNAL RADIATIVE FORCING

Clouds are produced by upward motion and the hydrological cycle and are highly interactive with the dynamic system. Atmospheric and surface temperatures as well as humidity are affected by the presence of clouds. As illustrated in Section 8.2.3, cloud radiative forcing information has been extracted from ERB data. The degree of cloud radiative forcing is determined by internal interactions and feedbacks and represents a measure of the instantaneous impact of clouds on radiative fluxes and heating rates. Thus, cloud radiative forcing is a useful diagnostic parameter for climate models.

The cloud radiative forcing discussed previously is in reference to the radiative fluxes at TOA. It is important, however, to recognize that, in principle, there are an

infinite number of cloud profiles in the atmosphere that can give a consistent ERB, and that the general circulation of the atmosphere is not directly related to the ERB but to radiative heating within the atmosphere. A realistic definition and understanding of cloud radiative forcing must begin with a consideration of radiative heating fields. Analogous to the definitions of ERB cloud radiative forcing, cloud radiative forcing with respect to internal solar and IR radiative heating can be defined as follows:

$$\dot{C}_{s,ir} = \left(\frac{\partial T}{\partial t}\right)_{s,ir} - \left(\frac{\partial T}{\partial t}\right)_{s,ir}^{cl}. \tag{8.6.1}$$

The presence of clouds produces strong IR cooling at cloud tops and significant heating at cloud bases, while the intensity of these cooling/heating configurations is dependent on the position and optical depth of the clouds (Section 4.7).

Internal cloud radiative forcing can be best understood from a zonally averaged profile. Figure 8.24 shows the latitudinal distribution of zonal mean cloud radiative forcing computed from the difference between atmospheric (including clouds) and clear-sky heating rates for a perpetual January simulation based on a climate model (Randall *et al.,* 1989). Zonally averaged solar radiative forcing is rather small: a 0.4 K day^{-1} warming in the upper tropical troposphere and weak negative values at low levels in the tropics and the Northern Hemisphere, in association with a reduction of gaseous absorption in the shadows of upper-level clouds. On the other hand, IR radiative forcing shows pronounced features, including a strong cooling associated with low-level cloud tops in the midlatitudes, a maximum warming of about 2 K day^{-1} at the 7 km level in the tropics, and a maximum cooling aloft of about 1 K day^{-1} at 16 km. The latter two features constitute a radiative dipole pattern that signifies the existence of optically thick convective anvil clouds. Anvils destabilize the upper troposphere because of cloud-top cooling, while warming in the middle troposphere tends to suppress shallow convection. The cloud-induced radiative destabilization of deep convection suggests that cloud IR radiative forcing has a positive feedback to the processes that generate high clouds. Owing to the domination of IR radiative cooling, the pattern of cloud net radiative forcing resembles that of the IR component.

8.6.2.2   GREENHOUSE WARMING AND CLOUD COVER FEEDBACK

Global cloud cover has been routinely mapped from satellite visible and IR radiometers. In particular, the ISCCP program has produced useful cloud climatology for climate studies (Section 7.4.5). No discernible changes in global cloud cover have been reported during the past 20 years. Cloud cover can exert a large influence on climate, as shown by the results of one-dimensional radiative–convective models. The presence of clouds reduces the solar flux available to the earth and the atmosphere for absorption, known as the solar albedo effect, but at the same time it enhances the trapping of the outgoing IR fluxes, the greenhouse effect. Cloud-cover sensitivity experiments using global climate models have been performed by numerous researchers to understand the impact of clouds on the greenhouse warming produced by the increase of carbon dioxide. As an example, we show cloud-cover feedback processes in a climate model based on comprehensive numerical experiments carried out by
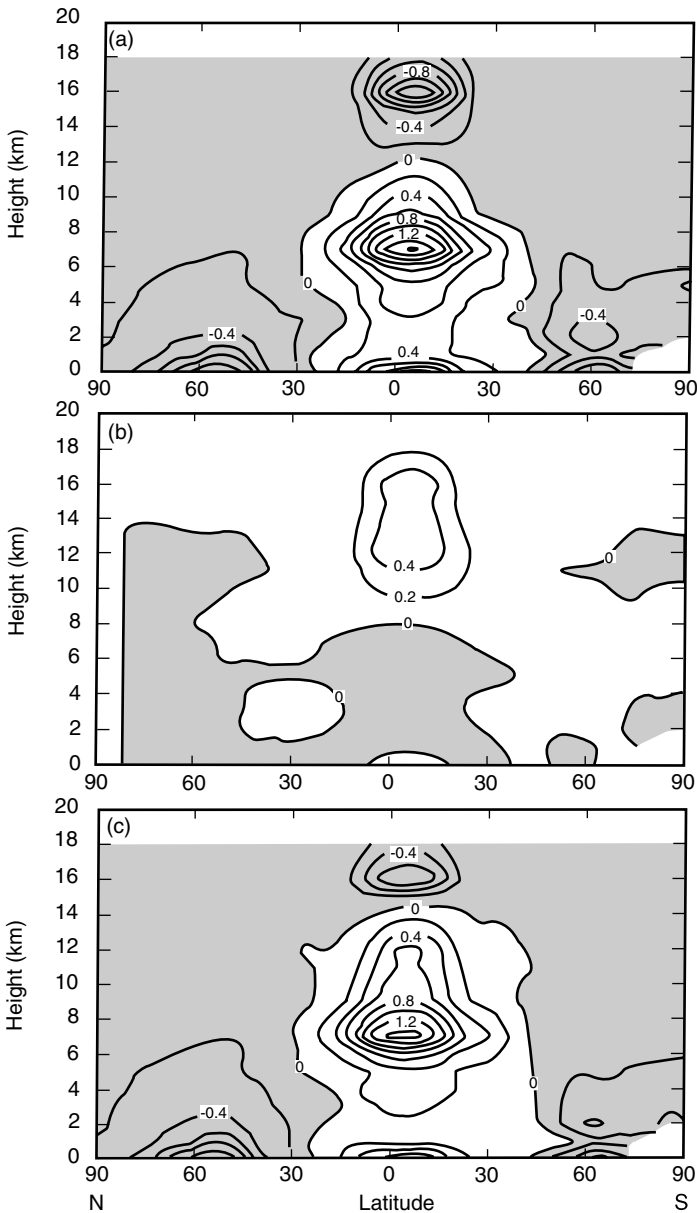
Figure 8.24 Latitudinal distribution of the zonal mean cross section for (a) cloud longwave (IR) radiative forcing; (b) cloud shortwave (solar) radiative forcing; and (c) cloud net radiative forcing. The results are obtained from the perpetual January control simulation; the contour interval is 0.2 K day^{-1} (data taken from Randall *et al.,* 1989).

Wetherald and Manabe (1988). In order to facilitate the interpretation and analysis of the results from numerical experiments, a simple scheme for the computation of cloud cover was used in these experiments. At each grid point, clouds were placed in the layer where the relative humidity exceeded 99%; otherwise, no clouds were predicted. This simple scheme gives a global integral of total cloud amount approximately equal to 50%. Clouds may occur in a single (thin) layer or in multiple contiguous vertical layers (thick). The radiative properties of these clouds were prescribed. Two experiments were carried out: one prescribed the cloud cover, while the other predicted the distribution of cloud cover. The latter experiment incorporated the interaction between cloud cover and radiative transfer in the atmosphere in a simple way.

In response to the increase in CO_2 concentration, cloud cover increases around the tropopause but decreases in the upper troposphere as shown in Fig. 8.25. The increase of high-level clouds reduces the outgoing IR flux from the top of the model atmosphere. At most latitudes a substantial reduction in cloud cover in the upper troposphere increases the available solar flux in the lower atmosphere and hence reduces the solar albedo effect. The change in the distribution of high clouds also increases the heating associated with the absorption of solar flux. The latitudinal profiles of the cloud-induced changes in the outgoing IR fluxes show a positive feedback to the increase in CO_2 concentration. The increase in the net incoming solar flux coupled with the reduction of the outgoing IR flux, due to the change in cloud cover, leads to a significant positive feedback to greenhouse warming.

The cloud-cover feedback associated with radiative forcing depends on the performance of a model with respect to the formation of clouds, which, in turn, is a function of such factors as the vertical layer, numerical method, parameterization of moist convection and cloud formation, and the radiative properties of clouds. A cloud-cover feedback study can, therefore, be viewed as a study of the possible mechanisms involved in a model setting rather than a quantitative assessment of the influence of cloud cover on the sensitivity of climate. To reduce a large amount of the uncertainty in the estimate of climate sensitivity, improvements must be made in the treatment of cloud feedback processes in climate models. A model must be capable of reproducing the horizontal distributions of cloud cover and radiative fluxes that have been derived from satellite observations, such as the results from ISCCP and ERB. Moreover, as demonstrated from one-dimensional climate models, cloud LWC and mean effective particle size, two parameters that are dependent on temperature and other relevant variables, have significant feedbacks to greenhouse perturbations, a subject of discussion in the following subsection.

8.6.2.3 GREENHOUSE WARMING AND CLOUD LIQUID/ICE WATER CONTENT FEEDBACK

Effects of the representation of clouds determined from different parameterization schemes on climate and climate perturbations have been investigated by Senior and Mitchell (1993), utilizing a low-resolution version of a GCM. This model included a prognostic cloud water variable and incorporated interactive cloud optical properties. The model also incorporated parameterization approaches for the production of
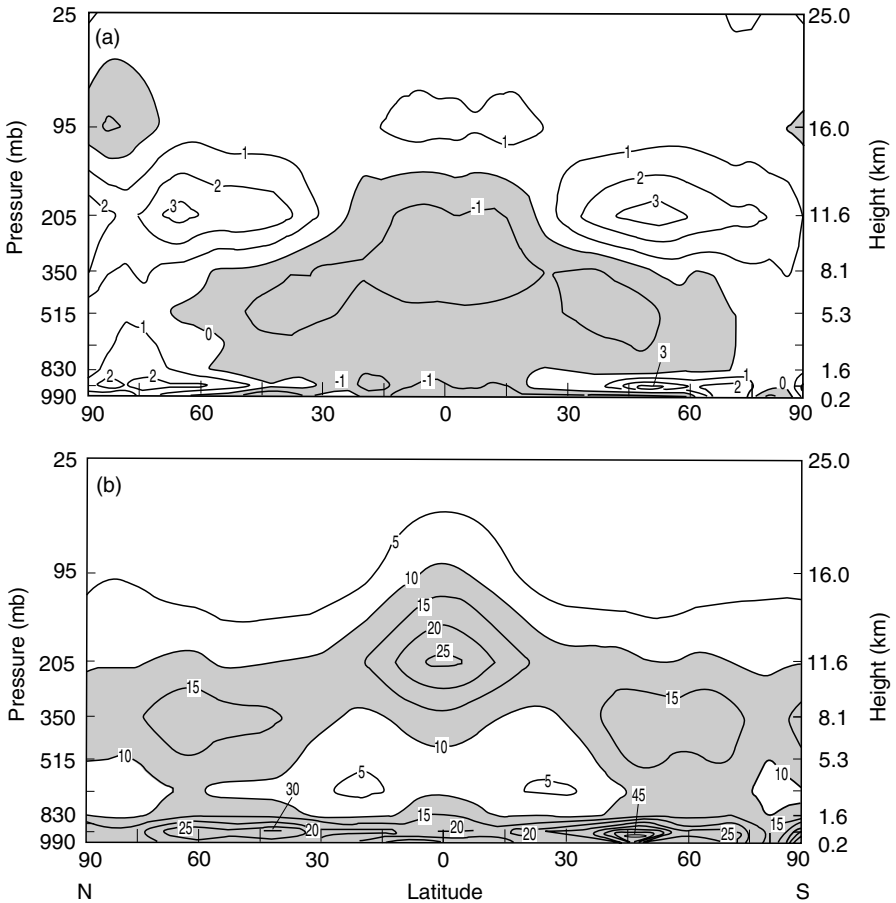
Figure 8.25 Height-latitude cross section for (a) CO_2-induced change of zonal mean cloud amount (%); and (b) zonal mean cloud amount (%) obtained from a GCM experiment (data taken from Wetherald and Manabe, 1988).

precipitation from water and ice clouds in which the empirical coefficients in the parameterization equations were taken from microphysics measurements (Heymsfield and Donner, 1990). In addition, cloud radiative properties as a function of liquid/ice water path were expressed in terms of reflectivity, transmissivity, and absorptivity using the values presented by Liou and Wittman (1979). Effects of cloud particle size, which was not a prognostic variable in the model, on the cloud radiative properties were not accounted for, however.

The sensitivity of cloud representations in the model to a doubling of atmospheric CO_2 was investigated over perpetual summer (June, July, August) and winter (December, January, February) conditions for a 10-year period after reaching equilibrium. Figure 8.26 shows the height-latitude cross section of equilibrium changes
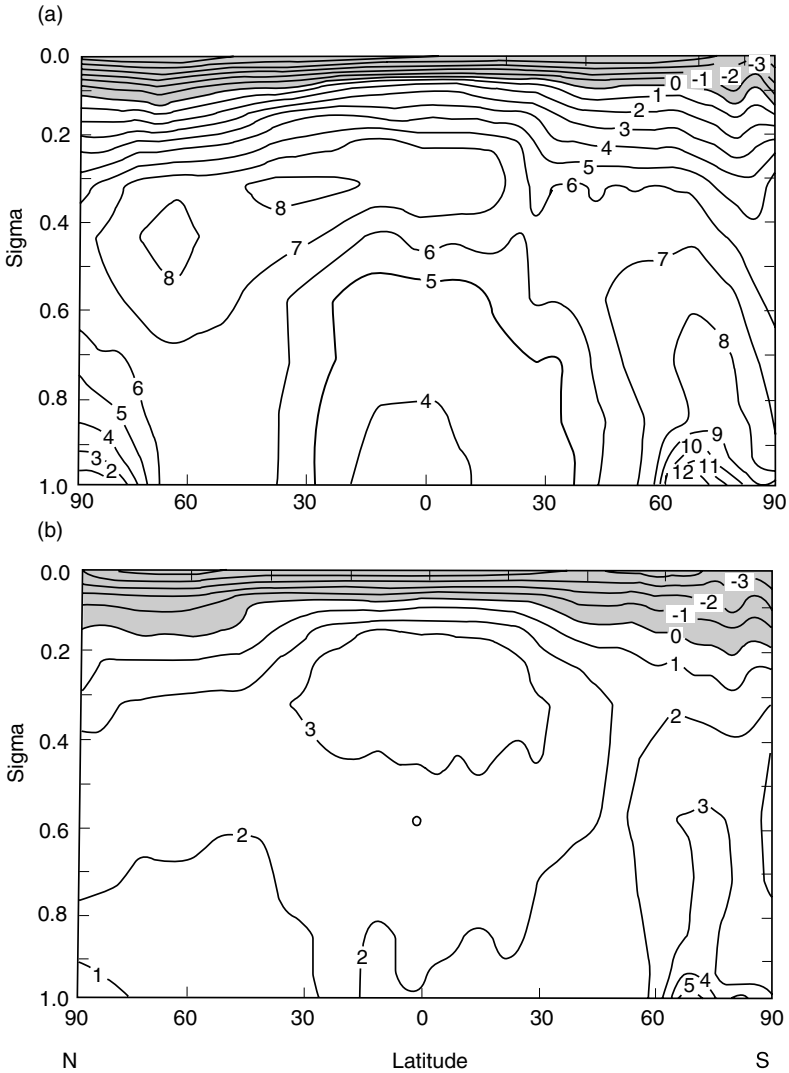
Figure 8.26 Height (σ-coordinate)–latitude cross section of equilibrium temperature changes in a 10-year summer simulation of doubling CO_2. Contours in every 1 K and reductions are shaded. (a) Cloud parametrization using an RH method; and (b) an interactive cloud water prognostic equation and radiative transfer feedback (data taken from Senior and Mitchell, 1993).

of temperature based on two cloud formation schemes. One employed the prognostic cloud formation coupled with the interactive radiative properties described above. The other used the relative humidity (RH) scheme in which the fractional amount of layer clouds is predicted from a quadratic dependence on RH. Cloud amount increased from zero at an assigned threshold level to full cover at saturation. Similar to the

experiment performed by Wetherald and Manabe (1988) described in the preceding subsection, the global mean total cloud cover was around 50%, while the radiative properties of high, middle, and low clouds determined from the RH scheme were fixed. The two experiments produced a large warming in the upper tropical troposphere, an amplification of the surface warming in high latitudes in winter, and a substantial stratospheric cooling of up to 5°C, in agreement with most model simulations incorporating enhanced CO_2. However, the prognostic cloud formation model with interactive cloud optical properties differed from the RH scheme in both the magnitude and distribution of the warming. The maximum warming of the upper troposphere was reduced from 8 K in the RH model to 3 K in the interactive cloud formation scheme. There was also a large reduction in the lower-level midlatitude warming from over 6 K to 2 K between the two cloud schemes. The pattern was similar in the summer simulation. The global mean annual change in surface temperature was reduced from 5.4 K in the RH model to 1.9 K in the interactive cloud formation scheme. When the interactive cloud formation was included, the resultant reduction in warming was due to a negative feedback produced from the change in cloud thermodynamic phase from ice to water and the associated prevalence of the solar albedo effect.

8.6.2.4   CLOUD PARTICLE SIZE FEEDBACK

Studies of the effects of possible changes in cloud particle size associated with air pollution and/or global warming using global climate models are still at an embryonic stage. In Eq. (8.4.7), we showed that cloud particle size directly affects cloud optical depth and hence, cloud albedo. Modeling of the formation of water-droplet and ice-crystal size spectra and the associated precipitation process based on physical principles is an involved and intricate computational task, considered to be a luxury for global climate models under present computer power limitations. However, parameterization approaches have been used to incorporate cloud particle size in climate models based on an assumed aerosol size distribution and large-scale variables. Uncertainty in the radiative forcing produced by the aerosol-cloud indirect effect estimated from climate models is so large that the confidence level in these estimates has been noted to be extremely low (IPCC, 1996; Hansen *et al.,* 2000). Reduction of this uncertainty requires a long-term global observational program for simultaneous measurements of aerosol and cloud particle sizes, coupled with independent validation.

8.6.3   Direct Radiative Forcing: Aerosols and Contrails

The subject of cloud radiative forcing is concerned with the effects of clouds on the radiation field of the earth–atmosphere system indirectly induced by external forcings such as the increase in anthropogenic greenhouse gases and/or particulates, which has been referred to as the indirect effect. However, two particle types of anthropogenic origin can directly exert a radiative forcing effect in climate systems through their scattering and absorption processes: aerosols generated in the planetary boundary layer and contrails produced in the upper troposphere and lower stratosphere.

8.6.3.1 AEROSOLS

In Section 8.4.4, we discussed the interaction of aerosols with solar radiation and pointed out that the radiative forcing produced by anthropogenic tropospheric aerosols has been recognized as a significant contributor to climate change. Two types of aerosols have been identified to be of specific importance in climate studies: sulfate and carbonaceous aerosols. The latter aerosols result from biomass burning as well as from fossil fuel production.

A number of GCM experiments have been conducted to investigate the impact of sulfate and carbonaceous aerosols on climate radiative forcing. For this purpose, a GCM must be coupled with an atmospheric chemical transport model, which is used to compute aerosol concentration in a separate manner. The computation of aerosol concentration for radiative transfer requires its inventory from available observations as an initial condition, including the estimation of carbonaceous aerosols produced from biomass burning, black and organic carbons generated from fossil fuel emissions, and sulfur emission. The radiative properties of aerosols are a function of their size distribution and refractive index, which is determined by their chemical composition. The chemical state and size of an aerosol are also dependent on RH. To model the RH dependence, the fundamental Köhler equation associated with diffusion growth can be used, but a size distribution must be assumed.

In the following, we present a GCM experiment conducted by Penner *et al.* (1998) to illustrate the relative significance of various aerosol types in the context of climate radiative forcing. In the experiment, these authors assumed that sulfate was present as $(NH_4)_2SO_4$ in sulfate aerosols, that smoke particles were made up of a variety of water-soluble salts and other elements, that biomass aerosols behaved similarly to an aerosol having 30% of $(NH_4)_2SO_4$ by mass, and that black and organic carbon particles were hydrophobic. Having determined the size distribution and chemical composition as a function of RH, single-scattering and absorption calculations for the aforementioned aerosol types assumed to be spherical in shape can then be carried out by using the Lorenz–Mie theory detailed in Section 5.2. A chemical transport model was employed to calculate aerosol concentrations every 6 hours. The results were then incorporated into the GCM to compute the radiative forcings at both TOA and the surface, defined as the difference between the net solar flux with and without the contributions by anthropogenic aerosols. In the course of model integration, the cloud distribution remained unchanged in both experiments, implying that the indirect forcing via cloud formation was not accounted for.

Figure 8.27 illustrates the global annual average radiative forcing at TOA for various aerosol types determined from the coupled GCM and chemical transport model. For the aerosols produced from biomass burning, two negative forcing values are presented because of the assumed size distributions. Negative forcing implies that more solar flux is reflected back to space, thereby reducing the absorption by aerosols in the earth–atmosphere system. The computed annual average forcings produced by fossil fuel black carbon (BC) alone and a combination of fossil/urban BC and organic carbon (OC) are 0.2 and 0.16 W m^{-2}, respectively. These forcings are primarily associated with emissions in North America, Europe, and Asia. The positive forcings
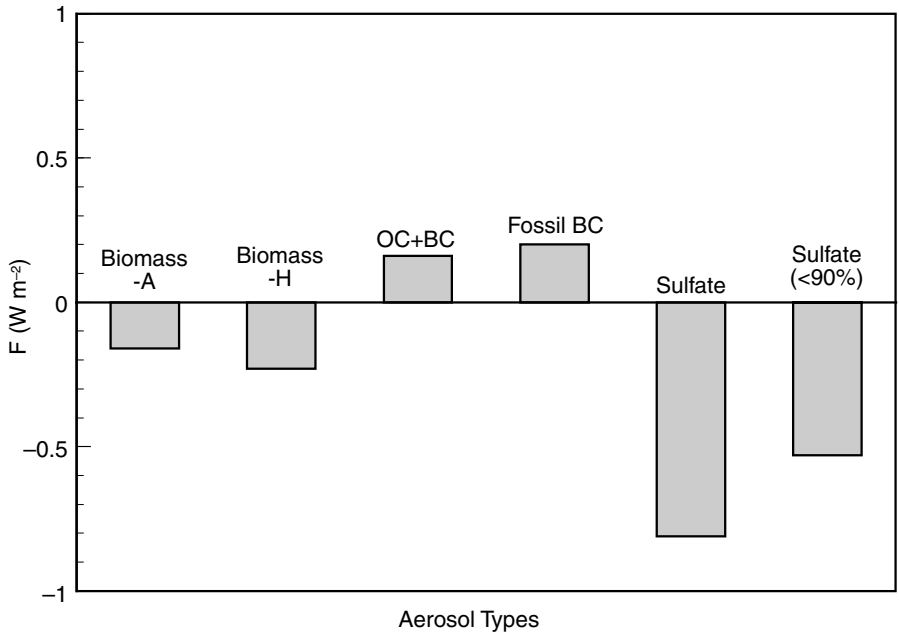
Figure 8.27 Global average radiative forcing at TOA for various aerosol types estimated from a GCM coupled with a chemical transport model. Biomass-A and -H denote the use of different assumed size distributions, OC denotes organic carbon, BC denotes black carbon, and sulfate (<90%) represents the case where the particle size and optical properties remain fixed at RH = 90% if RH > 90% (data taken from Penner *et al.*, 1998).

are produced by the domination of absorption due to BC. In the case of sulfate, a larger negative forcing of -0.81 W m^{-2} is shown because of the negligible absorption of $(NH_4)_2SO_4$ and water in the solar spectrum. When the RH dependence was capped at 90%, the forcing decreased by 50% because of the aerosol size effect.

On the basis of the preceding discussion, it is clear that direct climate forcing by anthropogenic aerosols in the troposphere is subject to a large degree of uncertainty in view of the numerous assumptions involved in the calculation of the spectral single-scattering and absorption properties of various aerosol types, as well as the treatment of aerosol transport, which is regional in character. At present, we do not know the precise sign of the current trend of aerosol forcing because such information would require knowledge of the trends of different aerosol compositions. Direct aerosol forcing largely depends on the aerosol's single-scattering albedo, and hence, on the amount of absorbing constituents. Indirect aerosol forcing depends on the effectiveness of aerosols in forming cloud particles and the consequence of cloud and precipitation processes, a subject of great complexity and uncertainty. In summary, climate forcing by anthropogenic aerosols may present the largest source of uncertainty about future climate change, the solution of which must rely on accurate long-term global observations of aerosol size and composition with appropriate validation.

8.6.3.2 CONTRAILS

The upper level ice crystal clouds produced by jet aircraft are known as contrails, or condensation trails as noted by Appleman (1953). Contrails are visible line clouds resulting from water vapor emissions that form behind aircraft flying in sufficiently cold air. It has been suggested that water droplets can form on the soot and sulfuric acid particles emitted from aircraft and/or background particles, and can then freeze to become ice particles. Persistent contrails often develop into more extensive contrail cirrus, particularly in ice-supersaturated air masses, in which the ice supersaturation is generally too small to allow cirrus clouds to form naturally. Contrails may enhance the extension of the natural cirrus cover in adjacent areas where RH is too low for the spontaneous nucleation of ice crystals, an indirect effect that has not been quantified at this point. Contrails can affect the atmospheric moisture budget by scavenging large cloud particles from the upper troposphere and by seeding clouds at lower altitudes.

In an analysis of cirrus cloud cover in Salt Lake City based on surface observations, Liou *et al.* (1990c) found that a substantial increase in cirrus clouds occurred in about 1965, coinciding with a sharp increase in domestic jet fuel consumption in the mid-1960s. Increased cirrus cloudiness has also been detected in climate data from stations in the midwestern and northwestern United States that are located beneath the major upper-tropospheric flight paths. A comprehensive analysis of jet aircraft contrails over the United States and Europe using satellite infrared imagery has been carried out (see, e.g., Minnis *et al.,* 1998; IPCC, 1999). Figure 8.28 displays a high-resolution IR image from NOAA 14, showing a variety of contrails in forms of spreading, diffusion, and thinning, associated with patchy cirrus.

The growing concern over the impact of increasing jet aircraft activities on the environment and climate led the Intergovernmental Panel on Climate Change to produce a report on the subject of aviation and global atmosphere (IPCC, 1999). Discussions of the issues related to contrails, cirrus clouds, and climate radiative forcing are detailed in Chapter 3 of the report. It has been estimated that aircraft line-shaped contrails cover about 0.1% of the earth's surface on an annually averaged basis, but with much higher values in local regions. Projections of contrail cover have also been made based on the rate of growth in aviation fuel consumption.

Radiative forcing of contrails depends not only on their cover, but also on their ice-crystal size distribution and shape. The ice-crystal images collected by the optical probe and replicator system from aircraft platforms during a number of field experiments show that contrails predominantly consist of bullet rosettes, columns, and plates with sizes ranging from about 1 μm to about 100 μm. The solar albedo and thermal emission of contrails computed from radiative transfer models reveal their significant dependence on both mean effective ice-crystal size and shape. Despite a large degree of uncertainty regarding the cover and ice-crystal size and shape of contrails, estimations of their globally and annually averaged radiative forcing have been made (IPCC, 1999). For subsonic aircraft emissions in 1992, a positive radiative forcing of 0.02 W m^{-2} with an uncertainty of more than a factor of 2 was estimated. It is much smaller than those produced by tropospheric aerosols. Nevertheless, the future projection of air traffic shows that the direct climatic effects of contrails could
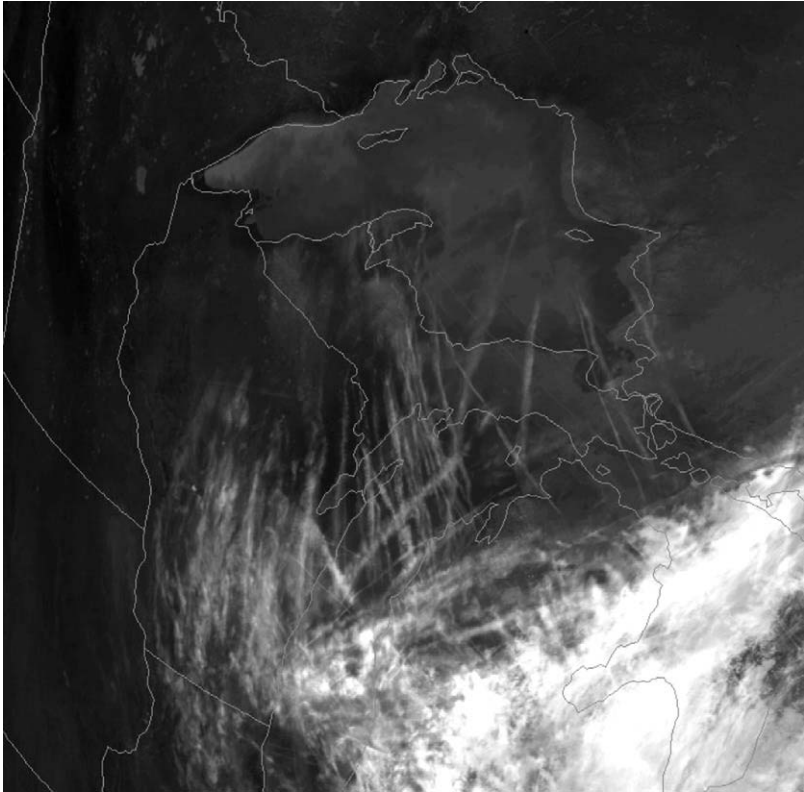
Figure 8.28   High-resolution thermal infrared image over the Great Lakes area of the United States from the NOAA-14 satellite at 2050 UTC October 2000, showing a number of north–south oriented contrails associated with patchy cirrus (courtesy of Patrick Minnis of the NASA Langley Research Center).

be on the same order as some tropospheric aerosol types. Moreover, it is noted that contrails have a significant regional character.

Perhaps the most significant contrail impact on climate is through their indirect effect on cirrus cloud formation. To reduce substantial uncertainties in contrail-cirrus radiative forcing estimations, a long-term observation program is needed to separate the cirrus clouds that are generated through contrails and other substances emitted from aircraft from those that occur naturally. The radiative and climatic impact of cirrus clouds have been articulated in a number of review articles (Liou, 1986; Sassen, 1997; and IPCC, 1999). The subject of the remote sensing of cirrus clouds from satellites was presented in Section 7.3.5.

8.6.4   Radiation in El Niño–Southern Oscillation

The importance of clouds and their radiative properties in shaping climate and very long-term climate trends, such as those due to increasing greenhouse gases and

aerosols, has long been well known. However, their impact on interannual to decadal time-scale variability is not as generally recognized. In the following, we present an example of the impact of clouds and their radiative effects on a simulation of the El Niño–Southern Oscillation (ENSO) based on a global climate model coupling the atmosphere and the oceans developed at the University of California, Los Angeles (UCLA).

El Niño is a climatic feature associated with the interannual oceanic warming of the eastern tropical Pacific off the coasts of Peru and Ecuador. In 1982–83, El Niño was linked to the disastrous droughts on the continents of southeastern Asia and Australia, unusual weather patterns over North and South America, and the weakening of monsoons over India. The Southern Oscillation involves a seesaw effect in the surface pressure across the tropical Pacific Ocean. Bjerknes (1969) discovered that continual Southern Oscillation is both the cause and the consequence of continually changing sea surface temperature (SST) patterns and that El Niño, associated with warm SSTs of the eastern tropical Pacific, is but one phase of a cycle. Analogous to the seasonal cycle representing an oscillation between winter and summer, the Southern Oscillation is a fluctuation between El Niño and a complementary state referred to as La Niña. While the seasonal cycle is forced by the variability of solar insolation associated with the earth's obliquity, the Southern Oscillation corresponds to a natural mode of interannual oscillation of the coupled ocean and atmosphere between the El Niño and La Niña states associated with SST anomalies. It is clear that the successful simulation and prediction of ENSO requires a global climate model that incorporates dynamic and physical couplings of the atmosphere and the oceans.

The UCLA coupled atmosphere–ocean general circulation model that was used to study the effect of clouds and their radiative properties on SST simulations consisted of an atmospheric GCM (AGCM) and an oceanic GCM (OGCM) that were coupled in the following way. The surface wind stress and heat flux were calculated hourly by the AGCM and its daily averages were transferred to the OGCM. At the same time, the SST was calculated hourly by the OGCM and its value at the time of coupling was passed to the AGCM. The AGCM had a horizontal resolution of $4°$ latitude by $5°$ longitude and 15 layers in the vertical. The OGCM domain was from $30°$S to $50°$N and from $130°$E to $70°$W with 27 vertical layers, had a constant depth at 4150 m, and covered a longitudinal resolution of $1°$ with varying latitudinal resolution. Figure 8.29a displays the climatological SST patterns simulated by the UCLA coupled model with the incorporation of the formation of stratus clouds over the Peruvian coast areas and of a larger high cloud emissivity (Yu and Mechoso, 1999). The interannual SST variabilities in terms of the amplitude, frequency, and the associated atmospheric and oceanic structures closely resemble those of the observed ENSO episodes. However, without the inclusion of proper representations of stratus clouds and high-cloud emissivity, the simulated SSTs are much lower (Fig. 8.29b) than those displayed in Fig. 8.29a, and their variabilities are about four times weaker. The inclusion of appropriate high-cloud emissivity produces a warmer upper troposphere, leading to a reduction of the temperature gradient below high clouds and the subsequent reduction of cumulus activity and surface evaporation in the tropical Pacific region. The introduction of

(a)

(b)
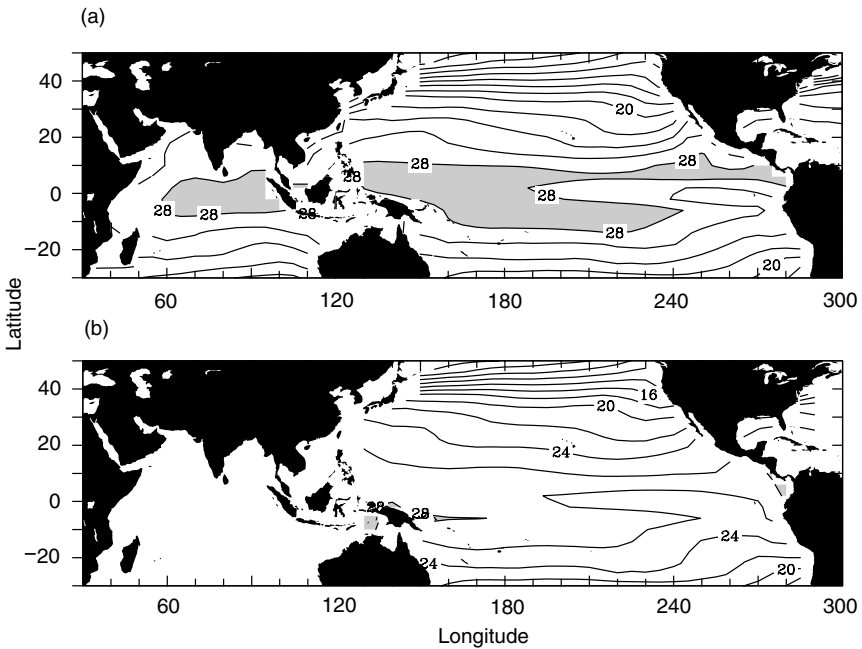
Figure 8.29 Climatological SST (°C) patterns simulated by the UCLA coupled atmosphere–ocean general circulation model: (a) a version with a suitable incorporation of high-cloud emissivity and the formation of low-level Peruvian stratus clouds, and (b) an earlier version without a specific inclusion of these two cloud effects. The shaded area in (a) shows the SST higher than 28°C including the Indian ocean (data provided by Jin-Yi Yu of the University of California, Los Angeles).

the low-level Peruvian stratus clouds decreases the amount of solar flux reaching the surface and, hence, the surface temperature locally, allowing a better simulation of the SST pattern in the central equatorial Pacific.

In light of the foregoing discussion, it appears evident that the interaction and feedback of clouds and their radiative properties are important physical processes governing the interannual SST variability. Further numerical experiments employing internally consistent and physically based cloud formation and radiative transfer parameterizations are required to explore the issues concerning the role of cloud and radiation processes in the context of interannual and decadal time-scale variabilities, particularly when these variabilities are coupled with greenhouse warming.

Exercises

8.1 Assume that the atmosphere acts as a single isothermal layer with a temperature T_a that transmits solar radiation but absorbs all thermal infrared radiation. Show that the global surface temperature $T = (2T_a)^{1/4}$. Let the global albedo be 30%, and the solar constant be 1366 W m^{-2}. What is the global surface temperature?

8.2 The mean global surface temperature is only about 15°C. The mean global absorptivity of solar radiation by the atmosphere is about 0.2. Use the global albedo and solar constant given in Exercise 8.1 and compute the mean global emissivity and temperature of the atmosphere. Repeat the calculation if the solar constant decreases by 1%.

8.3 Let the global reflectivities of the atmosphere and the surface be $\bar{r}$ and r_s, respectively. Consider the multiple reflection between the surface and the atmosphere and show that the global albedo of the earth–atmosphere system is given by

$$\bar{r} + (1 - \bar{A} - \bar{r})^2 r_s/(1 - r_s\bar{r}).$$

8.4 In Eq. (8.4.5), express the IR flux based on the Stefan–Boltzmann law and use a constant emissivity for the earth–atmosphere system. What would be the sensitivity of surface temperature to this IR flux and why is this sensitivity much smaller than that presented in Eq. (8.4.6)?

8.5 (a) The normalized mean annual distribution of insolation is approximately given by Eq. (8.5.29), and the albedo is given by the step function

$$r(x, x_i) = \begin{cases} 0.62, & x > 0.95 \\ 0.32, & x < 0.95. \end{cases}$$

Compute and plot the latitudinal surface temperature as a function of x from Eq. (8.5.15b).

(b) Also compute the temperature at the ice line assuming an albedo of 0.5. Show that the solution of the ice-line position x_i is quadratic. Plot x_i as a function of Q/Q_0 from 0.97 to 1.2.

8.6 (a) Derive Eq. (8.5.19) by means of a linear perturbation analysis.

(b) Show that the area covering the earth poleward of latitude φ is given by $2\pi a_e^2(1 - x)$, where a_e is the radius.

(c) Let $\Delta\bar{r} = 0.3s^*(x)\ell(x, \bar{x}_i)$, where ℓ is the ratio of the change in the ice-covered area to the total area of the Northern Hemisphere, and

$$s^*(x) = \int_{x_i}^{0.95} s(x)\,dx.$$

Show that the solution of x_i is given by a fourth-order polynomial equation. Compute and plot x_i as a function of $\Delta Q/Q_0$.

8.7 From Eq. (8.5.27a), we find

$$F_0 = QH_0(x_i), \qquad F_2 = QH_2(x_i)/(6D'' + 1).$$

Based on the two-mode approximation, we also have from Eq. (8.5.25)

$$F(x) = F_0 + F_2 P_2(x).$$

By fitting the present climate conditions, i.e., $\bar{x}_i = 0.95, 4Q_0 = 1366$ W m^{-2}, find the empirical coefficient D''.
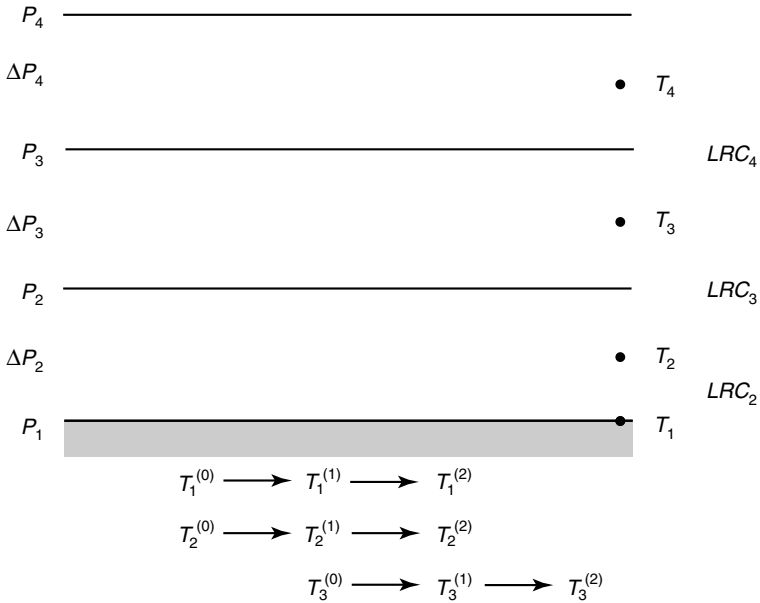
Figure 8.30   Graphic representation of the temperature iteration in the convective adjustment scheme. The superscript (n) $(n = 0, 1, 2, \ldots)$ denotes the iterative step. The subscripts 1 and $(2, 3, \ldots)$ denote the surface and atmospheric layers, respectively. The term LRC_k $(k = 2, 3, \ldots)$ represents the critical temperature difference.

8.8 Based on Eqs. (8.3.19)–(8.3.21) and in reference to Fig. 8.30, we may construct an iterative procedure for the computation of a temperature profile by employing the following convective adjustment scheme:

(a) First, compute $T_k^{(0)}$, the temperature for each atmospheric layer, for $k = 2, 3, 4, \ldots, N$ by using

$$T_k^{(0)} = T_k^{\chi} + \left(\frac{\partial T}{\partial t} \right)^{\chi} \Delta t, \qquad (8.E.1)$$

where the superscript $(n)(n = 0, 1, 2, \ldots)$ denotes the iterative step during the χth time step, and T_k is the temperature at the kth layer. The subscript $k = 1$ denotes the surface, and $k = N$ is the top level. The meanings of superscript (n) and subscript k are graphically displayed in Fig. 8.30.

(b) Second, calculate the radiative equilibrium temperature for the surface $(k = 1)$ by using

$$T_1^{(0)} = \left(\frac{F_{s,1} + F_{ir,1}^{\chi}}{\sigma} \right)^{1/4}, \qquad (8.E.2)$$

where $F_{s,1}$ and $F_{ir,1}^{\chi}$ are the net solar flux and downward longwave flux at the surface, respectively, and σ is the Stefan–Boltzmann constant. To a

good approximation, the net solar flux in the atmosphere may be computed independently of the temperature.

(c) Third, compute the critical temperature difference for each layer as follows:

$$\text{LRC}_k = (z_k - z_{k-1})\gamma_c, \tag{8.E.3}$$

where z_k is the altitude of the kth level and γ_c is the critical lapse rate.

(d) Fourth, calculate $T_1^{(1)}$ and $T_2^{(1)}$ such that they satisfy the energy balance:

$$\frac{C_p}{g}\Delta p_2\big(T_2^{(1)} - T_2^{(0)}\big) = \sigma\,\Delta t\big[\big(T_1^{(0)}\big)^4 - \big(T_1^{(1)}\big)^4\big], \tag{8.E.4a}$$

$$T_2^{(1)} = T_1^{(1)} - \text{LRC}_2, \tag{8.E.4b}$$

where $\Delta p_2 = p_1 - p_2$. The term on the right-hand side of Eq. (8.E.4a) represents the net upward radiative flux. After eliminating $T_1^{(1)}$ in Eqs. (8.E.4a) and (8.E.4b), $T_2^{(1)}$ can be solved from the following equations employing an iterative procedure (e.g., Newton's method) as follows:

$$\sigma\,\Delta t\big(T_2^{(1)} + \text{LRC}_2\big)^4 + \frac{C_p}{g}\Delta p_2 T_2^{(1)} - \frac{C_p}{g}\Delta p_2 T_2^{(0)} - \sigma\,\Delta t\big(T_1^{(0)}\big)^4 = 0. \tag{8.E.5}$$

(e) Fifth, if $T_2^{(1)} - T_3^{(0)} > \text{LRC}_3$, that is, if the layer is unstable, $T_1^{(2)}$, $T_2^{(2)}$, and $T_3^{(1)}$ can be simultaneously computed so that they satisfy the following relationships:

$$\frac{C_p}{g}\big[\Delta p_2\big(T_2^{(2)} - T_2^{(1)}\big) + \Delta p_3\big(T_3^{(1)} - T_3^{(0)}\big)\big] = \sigma\,\Delta t\big[\big(T_1^{(1)}\big)^4 - \big(T_1^{(2)}\big)^4\big], \tag{8.E.6a}$$

$$T_3^{(1)} = T_2^{(2)} - \text{LRC}_3, \tag{8.E.6b}$$

$$T_1^{(2)} = T_2^{(2)} + \text{LRC}_2, \tag{8.E.6c}$$

where $\Delta p_3 = p_2 - p_3$. These equations can also be solved by an iterative method. If $T_2^{(1)} - T_3^{(0)} \le \text{LRC}_3$, the layer is stable. We can then determine $T_2^{(2)}$ and $T_3^{(1)}$ by setting

$$T_2^{(2)} = T_2^{(1)}, \qquad T_3^{(1)} = T_3^{(0)}. \tag{8.E.7}$$

(f) Sixth, for levels $k = 3, 4, 5, \ldots, N$, it is necessary to find $T_{k+1}^{(1)}$ and $T_k^{(2)}$. If $T_k^{(1)} - T_{k+1}^{(0)} > \text{LRC}_{k+1}$, we may use

$$\frac{C_p}{g}\big[\Delta p_{k+1}\big(T_{k+1}^{(1)} - T_{k+1}^{(0)}\big) + \Delta p_k\big(T_k^{(2)} - T_k^{(1)}\big)\big] = 0, \tag{8.E.8a}$$

and

$$T_k^{(2)} - T_{k+1}^{(1)} = \text{LRC}_{k+1}, \quad k \ge 3. \tag{8.E.8b}$$

Table 8.1

Radiative Heating Rate, Temperature, and Air Density Profiles

Height (km)	Temperature (K)	Solar (K day^{-1})	IR (K day^{-1})	Air density (g m^{-3})
0.2	303.7	0.74	−2.07	1.20×10^3
0.7	295.3	0.43	−1.10	1.13×10^3
1.1	287.3	0.42	−0.77	1.09×10^3
1.6	278.8	1.04	−3.09	1.05×10^3
2.4	269.8	0.92	−2.29	0.95×10^3
3.5	259.2	0.93	−1.95	8.6×10^2
4.8	246.9	0.69	−1.52	7.4×10^2
6.3	234.3	0.50	−1.41	6.3×10^2
8.1	224.4	0.31	−0.89	5.0×10^2
10.4	222.9	0.21	−0.01	3.8×10^2
13.9	216.4	0.23	−0.30	2.3×10^2
19.8	219.4	0.80	−0.85	8.9×10^1
27.3	227.1	1.49	−1.29	2.9×10^1
35.2	234.1	3.41	−4.52	8.5×10^0
43.2	255.3	9.41	−10.15	3.0×10^0
52.1	268.5	10.84	−7.89	1.0×10^0
61.0	239.8	3.26	−4.62	0.9×10^0

Substituting Eq. (8.E.8b) into Eq. (8.E.8a) and solving for $T_k^{(2)}$ leads to

$$T_k^{(2)} = \frac{\Delta p_{k+1}\left(T_{k+1}^{(0)} + \mathrm{LRC}_{k+1}\right) + \Delta p_k T_k^{(1)}}{\Delta p_k + \Delta p_{k+1}}. \qquad (8.E.8c)$$

$T_{k+1}^{(1)}$ can then be solved by substituting $T_k^{(2)}$ back into Eq. (8.E.8b). If $T_k^{(1)} - T_{k+1}^{(0)} < \mathrm{LRC}_{k+1}$, we set

$$T_k^{(2)} = T_k^{(1)}, \qquad T_{k+1}^{(1)} = T_{k+1}^{(0)}. \qquad (8.E.9)$$

(g) Seventh, replace $T_k^{(2)}$ by $T_k^{(0)}$ for $k = 1, 2, 3 \dots, N$.

(h) Finally, repeat steps (d) through (g) until all layers of the supercritical lapse rate are eliminated.

Letting the critical lapse rate be 6.5 K km^{-1}, compute the temperature profile corresponding to the radiative heating profile listed in Table 8.1. In the calculations, let the net solar and IR fluxes at the surface be 159.1 and 386.9 W m^{-2}, respectively, and the surface air density be 1.221×10^3 g m^{-3}.

Suggested Reading

Budyko, M. I. (1982). *The Earth's Climate: Past and Future*. Academic Press, New York. Chapter 3 provides a discussion of the energy balance of the earth's surface and some results from energy balance models.

Hartmann, D. L. (1994). *Global Physical Climatology*. Academic Press, New York. Chapters 10, 11, and 12 present concise discussions of climate models and natural and anthropogenic climate changes.

IPCC (1996). *Climate Change 1995: The Science of Climate Change*. J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell (eds.). Cambridge University Press, Cambridge, U.K. Chapter 2 gives a comprehensive discussion of radiative forcing due to various greenhouse gases and aerosols.

Liou, K. N. (1992). *Radiation and Cloud Processes in the Atmosphere. Theory, Observation, and Modeling*. Oxford University Press, New York. Chapters 6 and 7 provide basic discussions of radiation budgets and climatic perturbations due to greenhouse gases, aerosols, and clouds.

Peixóto, J. P., and Oort, A. H. (1992). *Physics of Climate*. American Institute of Physics, New York. Chapter 17 presents various types of climate models.

Trenberth, K. E., ed. (1992). *Climate System Modeling*. Cambridge University Press, Cambridge, U.K. Chapters 1, 7, 10, and 20, authored by S. S. Schneider, R. D. Turco, J. T. Kiehl, and W. M. Washington, respectively, contain useful information pertaining to the role of radiation in climate and climate modeling.

This Page Intentionally Left Blank

Appendix A: Derivation
of the Planck Function

In accordance with Boltzmann statistics, if N_0 denotes the number of oscillators in any given energy state, then the number N in a state having energy higher by an amount ε is given by

$$N = N_0 e^{-\varepsilon/KT}, \tag{A.1}$$

where K is Boltzmann's constant and T is the absolute temperature. On the basis of Planck's first postulation, an oscillator cannot have just any energy, but rather only energies given by Eq. (1.2.1). Thus, the possible values of E must be 0, $h\tilde{v}$, $2h\tilde{v}$, and so on. If the number of oscillators with zero energy is N_0, then by virtue of Eq. (A.1), the number with energy $h\tilde{v}$ is $N_0 e^{-hv/KT}$, the number with energy $2h\tilde{v}$ is $N_0 e^{-2h\tilde{v}/KT}$, and so on. The total number of oscillators with frequency $\tilde{v}$ for all states is therefore

$$\begin{aligned} N &= N_0 + N_0 e^{-h\tilde{v}/KT} + N_0 e^{-2h\tilde{v}/KT} + \cdots \\ &= N_0[1 + e^{-h\tilde{v}/KT} + (e^{-h\tilde{v}/KT})^2 + \cdots] \\ &\cong N_0/(1 - e^{-h\tilde{v}/KT}). \end{aligned} \tag{A.2}$$

The total energy of these oscillators may be obtained by multiplying each term in Eq. (A.2) by the appropriate energy level:

$$\begin{aligned} E &= 0 \cdot N_0 + h\tilde{v} \cdot N_0 e^{-h\tilde{v}/KT} + 2h\tilde{v} \cdot N_0 e^{-2h\tilde{v}/KT} + 3h\tilde{v} \cdot N_0 e^{-3h\tilde{v}/KT} + \cdots \\ &= h\tilde{v}N_0 e^{-h\tilde{v}/KT}[1 + 2e^{-h\tilde{v}/KT} + 3(e^{-h\tilde{v}/KT})^2 + \cdots] \\ &\cong h\tilde{v}N_0 e^{-h\tilde{v}/KT}/(1 - e^{-h\tilde{v}/KT})^2. \end{aligned} \tag{A.3}$$

The average energy per oscillator is then given by

$$\frac{E}{N} = \frac{N_0 h\tilde{v}e^{-h\tilde{v}/KT}/(1 - e^{-h\tilde{v}/KT})^2}{N_0/(1 - e^{-h\tilde{v}/KT})} = h\tilde{v}/(e^{h\tilde{v}/KT} - 1). \tag{A.4}$$

According to Planck's second postulation, the quanta of energy are emitted only when an oscillator changes from one to another of its quantized energy states. The average emitted energy of a group of oscillators is, therefore, given by Eq. (A.4), which is the factor appearing in Planck's formula.

To obtain the Planck function, we let $u_{\tilde{\nu}}$ denote the monochromatic energy density, i.e., the energy per unit volume per unit frequency interval in a cavity with temperature T. With this definition, we write

$$u_{\tilde{\nu}} = Ah\tilde{\nu}/(e^{h\tilde{\nu}/KT} - 1), \tag{A.5}$$

where A is a constant to be determined. In accordance with the principle of equipartition of energy, the energy density in a cavity is given by the classical Rayleigh–Jeans formula:

$$u_{\tilde{\nu}} = (8\pi\tilde{\nu}^2/c^3)KT. \tag{A.6}$$

This formula is valid when the temperature T is high and the frequency $\tilde{\nu}$ is small. By letting $h\tilde{\nu}/KT \to 0$ in Eq. (A.5), we find $A = 8\pi\tilde{\nu}^2/c^3$. Thus, the monochromatic energy density is

$$u_{\tilde{\nu}} = \frac{8\pi h\tilde{\nu}^3}{c^3(e^{h\tilde{\nu}/KT} - 1)}. \tag{A.7}$$

For blackbody radiation, the emitted photons travel in all directions (4π solid angle) at the speed of light c. Thus, the emitted intensity (or radiance) in a cavity with a temperature T in units of energy/area/time/sr/frequency may be expressed by

$$B_{\tilde{\nu}}(T) = u_{\tilde{\nu}}c/4\pi. \tag{A.8}$$

On substituting Eq. (A.7) into (A.8), we obtain the Planck function in the form

$$B_{\tilde{\nu}}(T) = \frac{2h\tilde{\nu}^3}{c^2(e^{h\tilde{\nu}/KT} - 1)}, \tag{A.9}$$

as stated in Eq. (1.2.3).

Appendix B: The Schrödinger Wave Equation

According to wave mechanics, the behavior of a corpuscle is described by the Schrödinger wave equation in the form

$$\frac{h}{2\pi i}\frac{\partial \psi}{\partial t} = \frac{h^2}{8\pi^2 m}\nabla^2\psi - E_p\psi, \tag{B.1}$$

where ψ denotes the wave function, m is the mass, E_p is the potential energy of the corpuscle, which is a function of position only, and $\nabla^2 = \partial^2/\partial^2 x + \partial^2/\partial^2 y + \partial^2/\partial^2 z$. The quantum-mechanical Hamiltonian operator is defined as

$$H = -\frac{h^2}{8\pi^2 m}\nabla^2 + E_p. \tag{B.2}$$

Thus, we have

$$\frac{h}{2\pi i}\frac{\partial \psi}{\partial t} = -H\psi. \tag{B.3}$$

We may express the wave function as the product of two variables such that

$$\psi(q, t) = \varphi(q)\phi(t), \tag{B.4}$$

where q represents the particle position or coordinate. It follows that

$$\frac{h}{2\pi i}\varphi(q)\frac{d\phi(t)}{dt} = -\phi(t)H\varphi(q). \tag{B.5a}$$

Dividing both sides by $\varphi(q)\,\phi(t)$ yields

$$\frac{h}{2\pi i}\frac{1}{\phi(t)}\frac{d\phi(t)}{dt} = -\frac{1}{\varphi(q)}H\varphi(q) = -E. \tag{B.5b}$$

The left-hand side is a function of t only, but the right-hand side is a function of position q only. It follows that each must be equal to the same constant, which has the dimension of energy. This constant is identified with the total energy of the

particle and is denoted by $-E$. Thus, we obtain two equations from Eq. (B.5b) as follows:

$$\frac{h^2}{8\pi^2 m}\nabla^2\varphi + (E - E_p)\varphi = 0, \tag{B.6}$$

$$\frac{d\phi}{dt} = -\frac{2\pi i}{h}E\phi. \tag{B.7}$$

The solution of the time-dependent equation upon ignoring a constant of integration is

$$\phi = \exp(-2\pi i E t/h). \tag{B.8}$$

Thus, the wave function from Eq. (B.4) is given by

$$\psi(q,t) = \varphi(q)\exp(-2\pi i E t/h), \tag{B.9}$$

where the amplitude of the wave function φ is soluble from Eq. (B.6) only for certain specific values of E, referred to as the eigenvalues. A particular function that satisfies Eq. (B.6) is called the eigenfunction of the problem. The eigenfunctions and eigenvalues depend on the form of the potential energy function E_p of the corpuscle.

For a hydrogen atom in which one electron moves under the Coulomb attraction of the nucleus, the potential energy is given by

$$E_p = -e^2/r, \tag{B.10}$$

where e is the charge of the electron and r is the distance between the electron and the nucleus. Equation (B.6) can then be solved and the eigenvalues of this case are defined by

$$E_n = -R_H hc/n^2, \tag{B.11}$$

where $R_H = 1.097 \times 10^5$ cm^{-1} is the Rydberg constant. See also Eq. (1.3.3).

Appendix C: Spherical Geometry

In reference to Fig. C.1, we define

$$\overline{CD} = \overline{CO}\tan\theta', \qquad \overline{OD} = \overline{CO}\sec\theta',$$
$$\overline{CE} = \overline{CO}\tan\theta, \qquad \overline{OE} = \overline{CO}\sec\theta, \qquad (C.1)$$

where $\overline{CD}$ and $\overline{CE}$ are the tangent lines of the arcs CA and CB, respectively. For the triangle $\triangle CDE$, we find

$$\overline{DE^2} = \overline{CD^2} + \overline{CE^2} - 2\overline{CE}\,\overline{CD}\cos DCE. \qquad (C.2)$$

For the triangle $\triangle ODE$, we have

$$\overline{DE^2} = \overline{OD^2} + \overline{OE^2} - 2\overline{OD}\,\overline{OE}\cos DOE. \qquad (C.3)$$

Upon substituting Eq. (C.1) into Eqs. (C.2) and (C.3), we obtain

$$\overline{DE^2} = \overline{CO^2}[\tan^2\theta' + \tan^2\theta - 2\tan\theta'\tan\theta\cos(\phi - \phi')], \qquad (C.4)$$

$$\overline{DE^2} = \overline{CO^2}[\sec^2\theta' + \sec^2\theta - 2\sec\theta'\sec\theta\cos\Theta]. \qquad (C.5)$$
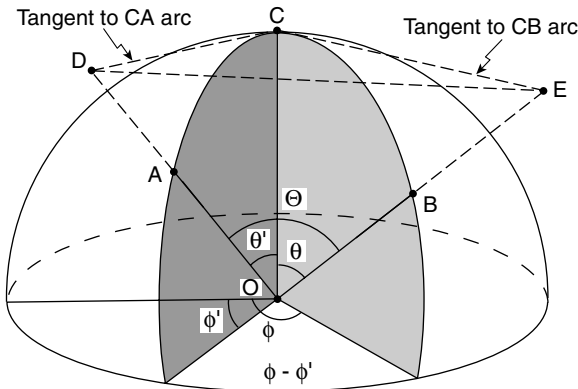
Figure C.1   Relationship between the scattering angle, zenith angle, and azimuthal angle in spherical coordinates.

It follows that

$$\tan^2 \theta' + \tan^2 \theta - 2 \tan \theta' \tan \theta \cos(\phi - \phi')$$
$$= \sec^2 \theta' + \sec^2 \theta - 2 \sec \theta' \sec \theta \cos \Theta. \tag{C.6}$$

But $\sec^2 \theta - \tan^2 \theta = 1$, so Eq. (C.6) becomes

$$2 - 2 \sec \theta' \sec \theta \cos \Theta = -2 \tan \theta' \tan \theta \cos(\phi - \phi'). \tag{C.7}$$

Thus, we have

$$\cos \Theta = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi')$$
$$= \mu \mu' + (1 - \mu^2)^{1/2}(1 - \mu'^2)^{1/2} \cos(\phi - \phi'), \tag{C.8}$$

where $\mu = \cos \theta$ and $\mu' = \cos \theta'$.

Appendix D: Complex Index of Refraction, Dispersion of Light, and Lorentz–Lorenz Formula

Within a dielectric, positive and negative charges are impelled to move in opposite directions by an applied electric field. As a result, electric dipoles are generated. The product of charges and the separation distance of positive and negative charges is called the dipole moment, which, when divided by the unit volume, is referred to as polarization $\mathbf{P}$. The displacement vector $\mathbf{D}$ (charge per area) within a dielectric is defined by

$$\mathbf{D} = \varepsilon\mathbf{E} = \mathbf{E} + 4\pi\mathbf{P}, \tag{D.1}$$

where ε is the permittivity of the medium. Thus,

$$\varepsilon = 1 + 4\pi\mathbf{P}\cdot\mathbf{E}/E^2. \tag{D.2}$$

The velocity of light in terms of ε and the permeability μ is given by

$$c = \sqrt{\frac{1}{\mu\varepsilon}}. \tag{D.3}$$

The permeability μ in air or water is nearly equal to the permeability μ_0 in vacuum, i.e., $\mu \approx \mu_0$. The index of refraction is defined as the ratio of the velocity of light in vacuum to that in the medium and may be expressed by

$$m = \frac{c_0}{c} \approx \sqrt{\varepsilon} = \sqrt{1 + \frac{4\pi\mathbf{P}\cdot\mathbf{E}}{E^2}}. \tag{D.4}$$

But the polarization vector for N dipoles is [see Eq. (3.3.1)]

$$\mathbf{P} = N\alpha\mathbf{E}. \tag{D.5}$$

Inserting Eq. (D.5) into Eq. (D.4) leads to

$$m^2 = 1 + 4\pi N\alpha. \tag{D.6}$$

Now, we have to find the polarizibility in terms of frequency. On the basis of the definition of a polarization vector, we have

$$\mathbf{P} = N e\mathbf{r}, \tag{D.7}$$

where e is the charge of an electron, and $\mathbf{r}$ represents the vector distance. Combining Eqs. (D.5) and (D.7), we find

$$\alpha \mathbf{E} = e\mathbf{r}. \tag{D.8}$$

Further, from the Lorentz force equation, the force generated by the electric and magnetic fields is given by

$$\mathbf{F} = e[\mathbf{E} + (\mu/c)\mathbf{v} \times \mathbf{H}], \tag{D.9}$$

where $\mathbf{v}$ denotes the velocity of an electron, which is very small compared to the velocity of light. Hence, the force produced by the magnetic field may be neglected. The force in the vibrating system in terms of the displacement r is due to (1) the acceleration of the electron; (2) the damping force, which carries away energy when the vibrating electrons emit electromagnetic waves, and which is proportional to the velocity of the electrons; and (3) the restoring force of the vibration, which is proportional to the distance r. From Newton's second law, we find

$$\frac{\mathbf{F}}{m_e} = \frac{e\mathbf{E}}{m_e} = \frac{d^2\mathbf{r}}{dt^2} + \gamma \frac{d\mathbf{r}}{dt} + \xi\mathbf{r}, \tag{D.10}$$

where γ and ξ are the damping and restoring coefficients, respectively, and m_e is the mass of the electron. In scalar form, we write

$$\frac{d^2r}{dt^2} + \gamma \frac{dr}{dt} + \xi r = \frac{eE}{m_e}. \tag{D.11}$$

The homogeneous solution of this second-order differential equation is given by

$$r = r_0 e^{-i\omega t} = r_0 e^{-i2\pi \tilde{v} t}. \tag{D.12}$$

Substituting Eq. (D.12) into Eq. (D.11), we obtain

$$[(\xi - 4\pi^2 \tilde{v}^2) - i2\pi \tilde{v}\gamma]r = eE/m_e. \tag{D.13}$$

The natural (or resonant) frequency is defined by $\tilde{v}_0 = \sqrt{\xi}/2\pi$. Thus, we find

$$\alpha = \frac{er}{E} = \frac{e^2}{m_e} \frac{1}{4\pi^2(\tilde{v}_0^2 - \tilde{v}^2) - i2\pi\gamma\tilde{v}}$$

$$= \frac{e^2}{m_e} \left[\frac{\tilde{v}_0^2 - \tilde{v}^2}{4\pi^2(\tilde{v}_0^2 - \tilde{v}^2)^2 + \gamma^2\tilde{v}^2} + \frac{i}{2\pi} \frac{\gamma\tilde{v}}{4\pi^2(\tilde{v}_0^2 - \tilde{v}^2)^2 + \gamma^2\tilde{v}^2} \right]. \tag{D.14}$$

Let the real and imaginary parts of the index of refraction be m_r and m_i, respectively, so that the index of refraction is defined by

$$m = m_r + im_i. \tag{D.15}$$

From Eq. (D.6), we can then show that

$$m_r^2 - m_i^2 = 1 + \frac{4\pi N e^2}{m_e} \frac{\tilde{v}_0^2 - \tilde{v}^2}{4\pi^2 (\tilde{v}_0^2 - \tilde{v}^2)^2 + \gamma^2 \tilde{v}^2}, \tag{D.16a}$$

$$2 m_r m_i = \frac{2 N e^2}{m_e} \frac{\gamma \tilde{v}}{4\pi^2 (\tilde{v}_0^2 - \tilde{v}^2)^2 + \gamma^2 \tilde{v}^2}. \tag{D.16b}$$

For air, $m_r \approx 1$ and $m_i \ll (m_r - 1)$. Also, in the neighborhood of the resonant frequency $(\tilde{v}^2 - \tilde{v}_0^2) = (\tilde{v}_0 + \tilde{v}) \cdot (\tilde{v} - \tilde{v}_0) \cong 2\tilde{v}_0 (\tilde{v} - \tilde{v}_0)$. Further, the half-width of the natural broadening depends on the damping and is given in the form $\alpha_N = \gamma/4\pi$, while the line strength S is $\pi N e^2/(m_e c)$. Thus, we obtain the real part

$$m_r - 1 = -\frac{N e^2}{4\pi m_e \tilde{v}_0} \frac{\tilde{v} - \tilde{v}_0}{(\tilde{v} - \tilde{v}_0)^2 + \alpha_N^2}, \tag{D.17}$$

and the absorption coefficient (Born and Wolf, 1975)

$$k_{\tilde{v}} = \frac{4\pi \tilde{v}_0 m_i}{c} = \frac{S}{\pi} \frac{\alpha_N}{(\tilde{v} - \tilde{v}_0)^2 + \alpha_N^2}. \tag{D.18}$$

Equation (D.18) is the Lorentz profile discussed in Section 1.3.

Shown in Fig. D.1 is the dependence of $(m_r - 1)$ and $k_{\tilde{v}}$ on the frequency. The value of $(m_r - 1)$ increases as the frequency increases when $(\tilde{v}_0 - \alpha_N) > \tilde{v}$. This mode is
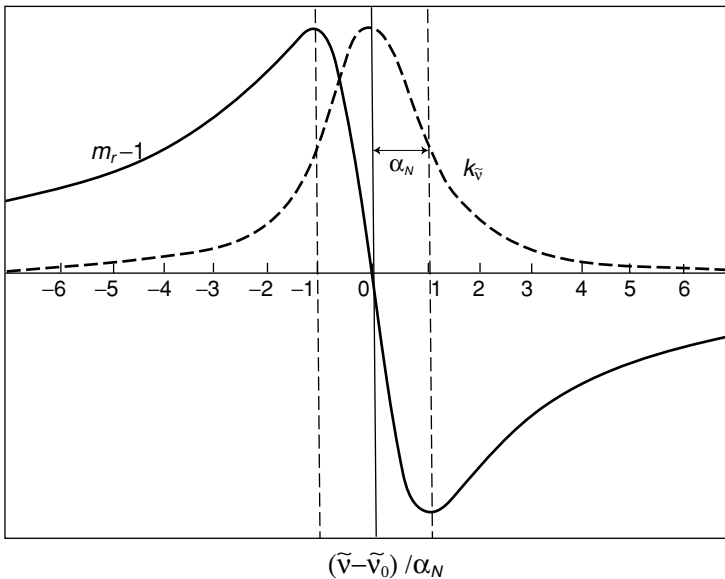
Figure D.1   Real and imaginary parts of the complex index of refraction as functions of the frequency.

referred to as *normal dispersion* under which light is dispersed by a prism into its component colors. For the region $(\tilde{\nu}_0 + \alpha_N) > \tilde{\nu} > (\tilde{\nu}_0 - \alpha_N)$, $(m_r - 1)$ decreases with increasing frequency and is called *anomalous dispersion*. For the range $\tilde{\nu} > (\tilde{\nu}_0 + \alpha_N)$, normal dispersion takes place again, but $(m_r - 1)$ is smaller than unity.

In this appendix we also wish to prove Eq. (3.3.16). We consider a dielectric placed between the plates of a parallel-plate condenser without the end effect. Moreover, we consider an individual molecule constituting this dielectric and draw a sphere with radius a about this molecule. The molecule is, therefore, affected by the fields caused by (1) the charges of the surfaces of the condenser plates; (2) the surface charge on the dielectric facing the condenser plates; (3) the surface charge on the spherical boundary of radius a; and (4) the charges of molecules (other than the one under consideration) contained within the sphere. For items (1) and (2), the electric field produced by these charges is

$$\mathbf{E}_1 + \mathbf{E}_2 = (\mathbf{E} + 4\pi\mathbf{P}) - 4\pi\mathbf{P} = \mathbf{E}. \tag{D.19}$$

For item (3), the electric field, which is produced by the polarization charge presented on the inside of the sphere, is given by

$$dE_3 = \frac{4\pi P \cos\theta \, dA}{4\pi a^2}, \tag{D.20}$$

where $P\cos\theta$ represents the component of the polarization vector in the direction of the electric field vector, and the differential area $dA = a^2 \sin\theta \, d\theta \, d\phi \times \cos\theta$. Thus,

$$E_3 = \int_0^{2\pi} \int_0^{\pi} \frac{4\pi P \cos\theta}{4\pi a^2} a^2 \sin\theta \cos\theta \, d\theta \, d\phi = \frac{4\pi P}{3}. \tag{D.21}$$

For item (4), it turns out that $E_4 = 0$. Thus, the effective electric field is

$$\mathbf{E}' = \mathbf{E} + 4\pi\mathbf{P}/3. \tag{D.22}$$

However, from Eq. (D.5), we have

$$\mathbf{P} = \alpha N \mathbf{E}' = \alpha N(\mathbf{E} + 4\pi\mathbf{P}/3). \tag{D.23}$$

It follows that

$$\mathbf{P} = \alpha N \mathbf{E}/(1 - 4\pi\alpha N/3). \tag{D.24}$$

Thus, from the definition of the index of refraction in Eq. (D.4), we find

$$m^2 = 1 + 4\pi\alpha N/(1 - 4\pi\alpha N/3). \tag{D.25}$$

Rearranging the terms, we obtain the Lorentz–Lorenz formula in the form

$$\alpha = \frac{3}{4\pi N} \frac{m^2 - 1}{m^2 + 2}. \tag{D.26}$$

Appendix E: Properties of the Legendre Polynomials and Addition Theorem

As indicated in Eqs. (5.2.42) and (5.2.43), the solution of the second-order differential equation

$$(1 - \mu^2)\frac{d^2 y}{d\mu^2} - 2\mu\frac{dy}{d\mu} + \left[\ell(\ell+1) - \frac{m^2}{1-\mu^2}\right] y = 0 \qquad (E.1)$$

is given by

$$y(\mu) = P_\ell^m(\mu) = \frac{(1-\mu^2)^{m/2}}{2^\ell \ell!} \frac{d^{m+\ell}}{d\mu^{m+\ell}}(\mu^2 - 1), \qquad (E.2)$$

where $\mu = \cos\theta$. When $m = 0$, $P_\ell^0(\mu) = P_\ell(\mu)$ are the Legendre polynomials. From Eq. (E.2) we have

$$P_\ell^m(\mu) = (1 - \mu^2)^{m/2}\frac{d^m P_\ell(\mu)}{d\mu^m}. \qquad (E.3)$$

The associated Legendre polynomials satisfy the orthogonal properties

$$\int_{-1}^{1} P_\ell^m(\mu)P_k^m(\mu)\,d\mu = \begin{cases} 0, & \ell \neq k, \\ \dfrac{2}{2\ell+1}\dfrac{(\ell+m)!}{(\ell-m)!}, & \ell = k, \end{cases} \qquad (E.4)$$

$$\int_{-1}^{1} P_\ell^m(\mu)P_\ell^n(\mu)\frac{d\mu}{1-\mu^2} = \begin{cases} 0, & m \neq n, \\ \dfrac{1}{m}\dfrac{(\ell+m)!}{(\ell-m)!}, & m = n. \end{cases} \qquad (E.5)$$

Some useful recurrence relations in conjunction with light scattering and radiative transfer are

$$\frac{dP_\ell^m}{d\theta} = -\sqrt{1-\mu^2}\frac{dP_\ell^m}{d\mu} = \frac{1}{2}\big[(\ell-m+1)(\ell+m)P_\ell^m - P_\ell^{m+1}\big], \qquad (E.6)$$

$$(2\ell+1)\mu P_\ell^m = (\ell+m)P_{\ell-1}^m + (\ell-m+1)P_{\ell+1}^m, \qquad (E.7)$$

$$(2\ell+1)(1-\mu^2)^{1/2}P_\ell^m = \big(P_{\ell+1}^{m+1} - P_{\ell-1}^{m+1}\big). \qquad (E.8)$$

A number of low-order associated Legendre and Legendre polynomials are

$$P_1^1(\mu) = (1 - \mu^2)^{1/2}, \qquad\qquad P_2^1(\mu) = 3\mu(1 - \mu^2)^{1/2},$$

$$P_3^1(\mu) = \tfrac{3}{2}(5\mu^2 - 1)(1 - \mu^2)^{1/2}, \quad P_2^2(\mu) = 3(1 - \mu^2),$$

$$P_3^2(\mu) = 15\mu(1 - \mu^2), \qquad\qquad P_3^3(\mu) = 15(1 - \mu^2)^{3/2}, \qquad \text{(E.9)}$$

$$P_0(\mu) = 1, \qquad\qquad\qquad P_1(\mu) = \mu,$$

$$P_2(\mu) = \tfrac{1}{2}(3\mu^2 - 1), \qquad\qquad P_3(\mu) = \tfrac{1}{2}(5\mu^3 - 3\mu),$$

$$P_4(\mu) = \tfrac{1}{8}(35\mu^4 - 30\mu^2 + 3). \qquad\qquad \text{(E.10)}$$

Below we present the addition theorem for the Legendre polynomials. Let $g(\mu, \phi)$ be an arbitrary function on the surface of a sphere where this function and all of its first and second derivatives are continuous. Then $g(\mu, \phi)$ may be represented by an absolutely convergent series of surface harmonics as follows:

$$g(\mu, \phi) = \sum_{\ell=0}^{\infty} \left[a_{\ell 0} P_\ell(\mu) + \sum_{m=1}^{\ell} (a_{\ell m} \cos m\phi + b_{\ell m} \sin m\phi) P_\ell^m(\mu) \right]. \quad \text{(E.11)}$$

The coefficients can be determined by

$$a_{\ell 0} = \frac{2\ell + 1}{4\pi} \int_0^{2\pi} \int_{-1}^1 g(\mu, \phi) P_\ell(\mu) \, d\mu \, d\phi, \qquad\qquad \text{(E.12)}$$

$$a_{\ell m} = \frac{(2\ell + 1)(\ell - m)!}{2\pi(\ell + m)!} \int_0^{2\pi} \int_{-1}^1 g(\mu, \phi) P_\ell^m(\mu) \cos m\phi \, d\mu \, d\phi, \quad \text{(E.13)}$$

$$b_{\ell m} = \frac{(2\ell + 1)(\ell - m)!}{2\pi(\ell + m)!} \int_0^{2\pi} \int_{-1}^1 g(\mu, \phi) P_\ell^m(\mu) \sin m\phi \, d\mu \, d\phi. \quad \text{(E.14)}$$

We note that

$$\int_{-1}^1 P_\ell^m(\mu) P_k^m(\mu) \, d\mu = \begin{cases} 0, & \ell \neq k \\ \dfrac{2(\ell + m)!}{(2\ell + 1)(\ell - m)!}, & \ell = k, \end{cases} \qquad \text{(E.15)}$$

$$\int_0^{2\pi} \cos m\phi \cos n\phi \, d\phi = \begin{cases} 0, & m \neq n \\ \pi, & m = n, \end{cases} \qquad\qquad \text{(E.16)}$$

and also that $P_\ell(1) = 1$, and $P_\ell^m(1) = 0$. Thus, we write

$$[g(\mu, \phi)]_{\mu=1} = \sum_{\ell=0}^{\infty} a_{\ell 0} = \frac{1}{4\pi} \sum_{\ell=0}^{\infty} (2\ell + 1) \int_0^{2\pi} \int_{-1}^1 g(\mu, \phi) P_\ell(\mu) \, d\mu \, d\phi. \quad \text{(E.17)}$$

We may now define the surface harmonic function in the form

$$Y_\ell(\mu, \phi) = \sum_{m=0}^{\ell} (a_{\ell m} \cos m\phi + b_{\ell m} \sin m\phi) P_\ell^m(\mu). \tag{E.18}$$

Let $Y_\ell(\mu, \phi)$ of order ℓ be $g(\mu, \phi)$, and by virtue of Eq. (E.17), we find

$$[Y_\ell(\mu, \phi)]_{\mu=1} = \frac{2\ell + 1}{4\pi} \int_0^{2\pi} \int_{-1}^1 Y_\ell(\mu, \phi) P_\ell(\mu) \, d\mu \, d\phi. \tag{E.19}$$

From the scattering geometry, we have

$$\cos \Theta = \mu\mu' + (1 - \mu^2)^{1/2}(1 - \mu'^2)^{1/2} \cos(\phi - \phi'). \tag{E.20}$$

Thus, we may let

$$P_\ell(\cos \Theta) = \sum_{m=0}^{\ell} (c_m \cos m\phi + d_m \sin m\phi) P_\ell^m(\mu)$$

$$= \frac{c_0}{2} P_\ell(\mu) + \sum_{m=1}^{\ell} (c_m \cos m\phi + d_m \sin m\phi) P_\ell^m(\mu). \tag{E.21}$$

Using the orthogonal properties denoted in Eqs. (E.15) and (E.16), we find

$$\int_0^{2\pi} \int_{-1}^1 P_\ell(\cos \Theta) P_\ell^m(\mu) \cos m\phi \, d\mu \, d\phi = \frac{2\pi(\ell + m)!}{(2\ell + 1)(\ell - m)!} c_m. \tag{E.22}$$

By letting $P_\ell^m(\mu) \cos m\phi = Y_\ell(\mu, \phi)$, and using Eq. (E.19), Eq. (E.22) becomes

$$\int_0^{2\pi} \int_{-1}^1 P_\ell(\cos \Theta) \left[P_\ell^m(\mu) \cos m\phi \right] d\mu \, d\phi = \frac{4\pi}{2\ell + 1} \left[P_\ell^m(\mu) \cos m\phi \right]_{\cos \Theta = 1}$$

$$= \frac{4\pi}{2\ell + 1} P_\ell^m(\mu') \cos m\phi'. \tag{E.23}$$

Note that $\cos \Theta = 1$ and $\Theta = 0$, so we have $\mu = \mu'$, and $\phi = \phi'$. It follows from Eq. (E.21) that

$$c_m = \frac{2(\ell - m)!}{(\ell + m)!} P_\ell^m(\mu') \cos m\phi'. \tag{E.24}$$

In a similar manner, we find

$$d_m = \frac{2(\ell - m)!}{(\ell + m)!} P_\ell^m(\mu') \sin m\phi'. \tag{E.25}$$

Thus, from Eqs. (E.24), (E.25), and (E.21), we obtain

$$P_\ell(\cos \Theta) = P_\ell(\mu) P_\ell(\mu') + 2 \sum_{m=1}^{\ell} \frac{(\ell - m)!}{(\ell + m)!} P_\ell^m(\mu) P_\ell^m(\mu') \cos m(\phi' - \phi). \tag{E.26}$$

Appendix F: Some Useful Constants

Acceleration of gravity (at sea level and 45° latitude)	$g = 9.80616$ m sec^{-2}
Angular velocity of rotation of the earth	$\omega = 7.27221 \times 10^{-5}$ rad sec^{-1}
Avogadro's number	$N_0 = 6.02297 \times 10^{23}$ molecule mol^{-1}
Boltzmann's constant	$K = 1.38062 \times 10^{-23}$ J K^{-1}
Density of air at standard pressure and temperature	$\rho = 1.273 \times 10^{-3}$ g cm$^{-3} = 1.273$ kg m^{-3}
Density of ice (0°C)	$\rho_i = 0.917$ g cm$^{-3} = 0.917 \times 10^3$ kg m^{-3}
Density of liquid water (4°C)	$\rho_\ell = 1$ g cm$^{-3} = 1 \times 10^3$ kg m^{-3}
Electron charge	$e = 1.60219 \times 10^{-19}$ C (coulomb, mks)
Gravitational constant	$G = 6.673 \times 10^{-11}$ N m^2 kg^{-2}
Latent heat of condensation at 0°C	$L_c = 2.5 \times 10^6$ J kg^{-1}
Loschmidt's number (at standard temperature and pressure)	$n_0 = 2.68719 \times 10^{25}$ molecule m^{-3}
Mass of an electron	$m_e = 9.10956 \times 10^{-31}$ kg
Mass of the earth	$M_e = 5.988 \times 10^{24}$ kg
Mean distance between the earth and the sun	$r_0 = 1.49598 \times 10^{11}$ m
Mean radius of the earth	$a_e = 6.37120 \times 10^6$ m
Mean radius of the sun (visible disk)	$a_s = 6.96000 \times 10^8$ m
Molecular weight of dry air	$M = 28.97$ g mol^{-1}
Permeability of vacuum	$\mu_0 = 12.56637 \times 10^{-7}$ kg m C^{-2} (mks)
	$= 1$ Gaussian unit (cgs)
Permittivity of vacuum	$\varepsilon_0 = 8.85419 \times 10^{-12}$ C kg^{-1} m^{-3} sec^2 (mks)
	$= 1$ Gaussian unit (cgs)
Planck's constant	$h = 6.62620 \times 10^{-34}$ J sec
Saturation vapor pressure (0°C)	$e_0 = 6.1078$ mb $= 0.61078$ k Pa
Solar constant	$S \cong 1366$ W m^{-2} (J sec^{-1} m^{-2})
Specific heat of air at constant pressure	$C_p = 10.04 \times 10^2$ m^2 sec^{-2} K^{-1}
Specific heat of air at constant volume	$C_v = 7.17 \times 10^2$ m^2 sec^{-2} K^{-1}
Standard pressure	$p_0 = 1013.25$ mb $= 101.325$ k Pa
Standard temperature	$T_0 = 273.16$ K
Stefan-Boltzmann constant	$\sigma = 5.66961 \times 10^{-8}$ J m^{-2} sec^{-1} K^{-4}
Universal gas constant	$R^* = 8.31432$ J mol^{-1} K^{-1}
Velocity of light	$c = 2.99792458 \times 10^8$ m sec^{-1}
Wien's displacement constant	$\alpha = 0.2897 \times 10^{-2}$ m K

Appendix G: Standard Atmospheric Profiles[a]

Altitude (km)	Pressure (mb)	Temperature (K)	Density (g m^{-3})	Water Vapor (g m^{-3})	Ozone (g m^{-3})
0	1.013×10^3	288.1	1.225×10^3	5.9×10^0	5.4×10^{-5}
1	8.986×10^2	281.6	1.111×10^3	4.2×10^0	5.4×10^{-5}
2	7.950×10^2	275.1	1.007×10^3	2.9×10^0	5.4×10^{-5}
3	7.012×10^2	268.7	9.093×10^2	1.8×10^0	5.0×10^{-5}
4	6.166×10^2	262.2	8.193×10^2	1.1×10^0	4.6×10^{-5}
5	5.405×10^2	255.7	7.364×10^2	6.4×10^{-1}	4.5×10^{-5}
6	4.722×10^2	249.2	6.601×10^2	3.8×10^{-1}	4.5×10^{-5}
7	4.111×10^2	242.7	5.900×10^2	2.1×10^{-1}	4.8×10^{-5}
8	3.565×10^2	236.2	5.258×10^2	1.2×10^{-1}	5.2×10^{-5}
9	3.080×10^2	229.7	4.671×10^2	4.6×10^{-2}	7.1×10^{-5}
10	2.650×10^2	223.2	4.135×10^2	1.8×10^{-2}	9.0×10^{-5}
11	2.270×10^2	216.8	3.648×10^2	8.2×10^{-3}	1.3×10^{-4}
12	1.940×10^2	216.6	3.119×10^2	3.7×10^{-3}	1.6×10^{-4}
13	1.658×10^2	216.6	2.666×10^2	1.8×10^{-3}	1.7×10^{-4}
14	1.417×10^2	216.6	2.279×10^2	8.4×10^{-4}	1.9×10^{-4}
15	1.211×10^2	216.6	1.948×10^2	7.2×10^{-4}	2.1×10^{-4}
16	1.035×10^2	216.6	1.665×10^2	6.1×10^{-4}	2.3×10^{-4}
17	8.850×10^1	216.6	1.423×10^2	5.2×10^{-4}	2.8×10^{-4}
18	7.565×10^1	216.6	1.216×10^2	4.4×10^{-4}	3.2×10^{-4}
19	6.467×10^1	216.6	1.040×10^2	4.4×10^{-4}	3.5×10^{-4}
20	5.529×10^1	216.6	8.891×10^1	4.4×10^{-4}	3.8×10^{-4}
21	4.729×10^1	217.6	7.572×10^1	4.8×10^{-4}	3.8×10^{-4}
22	4.047×10^1	218.6	6.451×10^1	5.2×10^{-4}	3.9×10^{-4}
23	3.467×10^1	219.6	5.500×10^1	5.7×10^{-4}	3.8×10^{-4}
24	2.972×10^1	220.6	4.694×10^1	6.1×10^{-4}	3.6×10^{-4}
25	2.549×10^1	221.6	4.008×10^1	6.6×10^{-4}	3.4×10^{-4}
30	1.197×10^1	226.5	1.841×10^1	3.8×10^{-4}	2.0×10^{-4}
35	5.746×10^0	236.5	8.463×10^0	1.6×10^{-4}	1.1×10^{-4}
40	2.871×10^0	250.4	3.996×10^0	6.7×10^{-5}	4.9×10^{-5}
45	1.491×10^0	264.2	1.966×10^0	3.2×10^{-5}	1.7×10^{-5}
50	7.978×10^{-1}	270.6	1.027×10^0	1.2×10^{-5}	4.0×10^{-6}
75	5.520×10^{-2}	219.7	8.754×10^{-2}	1.5×10^{-7}	8.6×10^{-8}
100	3.008×10^{-4}	210.0	4.989×10^{-4}	1.0×10^{-9}	4.3×10^{-11}

[a] Source: U.S. Standard Atmosphere, 1976.

Appendix H: Answers to Selected Exercises

Chapter 1

1.3 $B_\nu(T) = 2h\nu^3 c^2/(e^{hc\nu/KT} - 1)$

1.4   Show that $(5 - x) = 5 \cdot e^{-x}$, where $x = hc/(K\lambda T)$, and find x

1.5   Insert $\lambda_m = \alpha/T$ into the Planck function

1.6 $\sim$300 K

1.7 7.52×10^{-23} J sec^{-1} m^{-2} sr^{-1}/μm; 81.2×10^{-3} J sec^{-1} m^{-2} sr^{-1}/(cm^{-1}); 8.57×10^{-17} J sec^{-1} m^{-2} sr^{-1}/Hz

1.8 5.22×10^2 J sec^{-1} m^{-2}; 4.96×10^2 J sec^{-1} m^{-2}; 9.36 μm

1.9   Both cases are about 0.4%

1.11 (b) $n = 1, 2$; $\lambda_{12} = 1216$ Å

1.12 1 J $= 5.0345 \times 10^{22}$ cm^{-1}

1.13 Use Planck's relation, $\Delta E = h\nu c$

1.14 Note: $\displaystyle\int e^{ax} \cos bx \, dx = \frac{e^{ax}(a \cos bx + b \sin bx)}{a^2 + b^2}$

1.15 Note: $\displaystyle\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \pi$; $\displaystyle\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$

1.17 (a) $\displaystyle\lim_{x,y \to \infty} K(x, y) = \frac{1}{\pi} \frac{y}{x^2 + y^2}$;

     (b) Define a delta function $f(x - t) = \dfrac{1}{\pi} \dfrac{y}{y^2 + (x - t)^2}$

1.18 (a) $n_1 b_{12} = n_2 b_{21}$; (b) $n_1 C_{12} u_{\tilde{\nu}} = n_2(C_{21} u_{\tilde{\nu}} + A_{21})$;

     (c) $n_1(b_{12} + C_{12} u_{\tilde{\nu}}) = n_2(b_{21} + C_{21} u_{\tilde{\nu}} + A_{21})$

1.19 $\beta_\lambda = 0.1, 0.5$ m^{-1}; $\tau = 1, 5$

1.20 x(aerosol) $= 48.9$ km, x(dense fog) $= 0.039$ km

1.21 $I_\lambda(s) = I_\lambda(0)(1 - R_\lambda)^2 T_\lambda/(1 - R_\lambda^2 T_\lambda^2)$

1.23 $F_\nu(\tau = 0) = \pi B_\nu(T_s) 2 E_3(\tau_*) + \pi B_\nu(T)[1 - 2E_3(\tau_*)]$

Chapter 2

2.1 At the poles, $\varepsilon = \delta$; at $60°$, $\varepsilon = 30° + \delta$; at $45°N$ and equinox, $2H = 12$ hr; at $45°N$ and at the solstice, $2H = 15.44$ hr

2.2 Use the equation for an ellipse involving the radius, the true anomaly, and eccentricity

2.4 (b) 102 min; (c) 35,865 km

2.5 5754 K

2.6 1.52991×10^{22} J

2.7 4.51×10^{-10}

2.8 (a) 1.14×10^{9} J sec^{-1}; (b) 1.82×10^{-2} J sec^{-1}

2.10 $T(\text{Venus}) = 225.8$ K

2.11 70.85 K (use the solid angle from Ex. 1.2)

2.12 4.2 K

2.13 (a) 556.60 W m^{-2}; (b) 434.81 W m^{-2} (daily mean)

2.14 Note: $\sin(-\varphi) = -\sin \varphi; \sin(-\varepsilon) = -\sin \varepsilon$

2.15 Note: $e \ll 1$ and $(1 - x)^{-2} = 1 + 2x - \cdots$

2.17 Note: Solar irradiance at TOA, $F_\lambda = \Omega B_\lambda(T)$, where Ω is a solid angle

2.18 $F_{0\lambda} = 22.31$ W m^{-2}, $T_\lambda = 0.68465$

Chapter 3

3.1 Note that the Boltzmann constant $K = MR$, where M is the molecular weight and R is the gas constant for air.

3.3 (a) Use Beer's law for the change of solar flux due to absorption from the outer edge of the atmosphere to the point P and (b) take an exponential variation of the density in the vicinity of the point P.

3.5 $[O]^2 = J_2[O_2]/(K_{11}[M] + K'_{11})$

3.6 $A_{\bar{\nu}}(u/\mu_0) = 2\sqrt{Su\alpha/\mu_0}$

3.7 $\sigma_s(0.7\,\mu\text{m}) = 1.71 \times 10^{-27}$ cm^2

3.8 $\tau(0.5\,\mu\text{m}) \cong 0.15$

3.9 At $z = 10$ km; $\rho = 0.414 \times 10^{-3}$ g cm^{-3}, $(m_r - 1) \times 10^{-3} = 0.099$

3.10 $F \cong F_0 l \cdot 1.37 \times 10^{-11}$ (at 10 km)

3.11 For $\lambda = 10$ cm; $\beta_\pi \cong 9.76 \times 10^{-10}$ km^{-1}, 8.66×10^{-10} km^{-1}

3.12 The total deviation from the original direction is $\theta' = 2(\theta_i - \theta_t) + 2(p - 1) \cdot (\pi/2 - \theta_t)$, where $p = 1$ denotes two refractions and $p \geq 2$ denotes internal

reflections. The first rainbow $(p = 2)$ is located at the $137°$ scattering angle.

3.13 The deviation from the original ray is $\theta' = (\theta_i - \theta_t) + (\theta'_i - \theta'_t)$. Note that $\theta'_i = \theta_t, \theta'_t = \theta_t$ and $\theta_t = A/2$. The $22°$ and $46°$ halos are defined by $A = 60°$ and $90°$, respectively.

3.14 At about $24.1°$ from the sun

3.15 An aerosol particle size of about 0.45 to $0.48 \, \mu m$ (use blue and red light in the calculations).

3.17 Note that for $\tilde{\omega} = 1$, $k = 0$ and $\gamma_1 = \gamma_2$

3.18 $\chi = -[(\gamma_1 - \gamma_2)(1 - 2\gamma_3) + 1/\mu_0]\tilde{\omega} F_\odot$

Chapter 4

4.2 $x \to 0$, $L(x) \cong x$; $x \to \infty$, $L(x) \cong 2\sqrt{x/2\pi}$ (prove)

4.3 $f(k) = \dfrac{1}{k}\dfrac{1}{\pi}\left(\dfrac{2k\delta}{S}\coth\beta - \dfrac{k^2\delta^2}{S^2} - 1\right)^{-1/2}$

4.4 (d) $f(k) = L^{-1}(e^{-au}) = \delta(k - a)$;

$f(k) = L^{-1}(\exp[-(au)^{1/2}]) = \dfrac{1}{2}\left(\dfrac{a}{\pi}\right)^{1/2} k^{-3/2} \exp(-a/4k)$

4.5 $\dfrac{\overline{W}}{\delta} = \dfrac{2}{\pi}(b_{\bar{\nu}}y)^{1/2}\tan^{-1}\left(\dfrac{\delta}{\pi \nu_0}(b_{\bar{\nu}}y)^{1/2}\right) + \dfrac{2\nu_0 C_y}{1 + \nu_0 C y(\delta/\nu_0)}$, $y = a_{\bar{\nu}}u$

4.6 Note: $du = q \, dp/g$

4.7 $T_{\bar{\nu}} = F_{\bar{\nu}}/F_{0,\bar{\nu}}$

4.8 To obtain $c_{\bar{\nu}}$ and $d_{\bar{\nu}}$, use strong- and weak-line approximations

4.9 Note: $\displaystyle\int_0^\infty e^{-a^2x^2}\cos bx \, dx = \dfrac{\sqrt{\pi}}{2a}e^{-b^2/4a^2}$; $\displaystyle\int_0^\infty \dfrac{\cos ax}{1 + x^2}\, dx = \dfrac{\pi}{2}e^{-a}$

4.10 Note: $\displaystyle\int_0^\infty\int_{k''}^\infty \to \int_0^\infty\int_0^k$

4.11 Let $M = I^\uparrow + I^\downarrow$, and $N = I^\uparrow - I^\downarrow$

Chapter 5

5.1 Perform $mk \, \nabla \times \mathbf{N}_\psi = \nabla \times \nabla \times \mathbf{M}_\psi$

5.2 Take a trial solution: $rv^i = \dfrac{1}{k}\displaystyle\sum_{n=1}^\infty \beta_n \, \psi_n(kr) \, P_n^1 (\cos\theta)\sin\phi$, and determine β_n

5.3 (a) Note: $\nabla \times \nabla\psi = 0$ and $\nabla \cdot \nabla \times \mathbf{A} = 0$;

(b) $\mathbf{M}_\psi = \nabla \times (\mathbf{a}_z\psi) = \dfrac{1}{r}\dfrac{\partial\psi}{\partial\phi}\mathbf{a}_r - \dfrac{\partial\psi}{\partial r}\mathbf{a}_\phi$

5.4 Let $\psi(r, \phi, z) = e^{-i\omega t} R(r)\Phi(\phi)Z(z)$

5.5   See Eq. (5.4.24)

5.6 $\theta_i = \cos^{-1} \sqrt{1/(1 + m^2)}$

5.7   (a) Red outside and violet inside, $a \cong 3.5\,\mu$m;
      (b) Primary cloudbow, 137.78° (red) and 139.49° (violet);
      (c) 22° halo $(A = 60°)$, $\Delta\theta = 0.76°$; 46° halo $(A = 90°)$, $\Delta\theta = 2.17°$

5.8   Let $m \cos\theta_t = u + iv$ in Eq. (5.3.23a)

5.9   See also Eq. (5.2.113)

5.10 (c) $\phi_i = A/2$, $\phi_i = (A + \theta'_h)/2$, where θ'_h is the minimum deviation angle projected on the horizontal plane, and $\sin\theta'/2 = \cos\varepsilon_i \sin\theta'_h/2$ (prove) where ε_i is the elevation angle; θ (red) $= 24.54°$

5.14 The average cross section of randomly oriented hexagonal crystals is given by
$$\overline{G} = \frac{6}{\pi} \int_0^{\pi/6} \int_0^{\pi/2} G(\alpha, \beta) \cos\alpha \, d\alpha \, d\beta$$

Chapter 6

6.1 $\tilde{\omega}(\tilde{\nu}) = (1 + x^2)/(2 + x^2)$, where $x = (\nu - \nu_0)/\alpha$, $R(\mu, \mu_0)$
$$= \frac{\tilde{\omega}}{4(\mu + \mu_0)} \times \frac{(1 + \sqrt{3}\mu)(1 + \sqrt{3}\mu_0)}{[1 + \mu\sqrt{3(1-\tilde{\omega})}][1 + \mu_0\sqrt{3(1-\tilde{\omega})}]}$$

6.2 $r(\mu_0) = F^{\uparrow}(\mu_0)/\mu_0 F_{\odot}$; $\tilde{\omega} = 0.8$, $r(\mu_0 = 1) = 0.32$, $\bar{r} = 0.35$

6.3   Follow the procedures outlined in Section 6.3.2

6.4   Note: $1 + x + x^2 + \cdots = 1/(1 - x)$

6.6   Let $F_{\odot} = \pi$ and $\phi_0 = 0°$. For $\tau = 0.1$, $I(\mu = 1, \phi = 0°) = 0.028$

6.7 $I_2(0, \mu, \phi)$
$$= \frac{\mu_0 F_{\odot} \tilde{\omega}^2}{4} \int_0^{2\pi} \int_{-1}^{1} P(\mu, \phi; \mu'\phi')P(\mu', \phi', -\mu_0, \phi_0)g(\mu, \mu_0, \mu')\,d\mu'\,d\phi',$$
where $g(\mu, \mu_0, \mu') = \dfrac{1}{4(\mu_0 + \mu')} \left[\dfrac{\mu_0}{\mu_0 + \mu} \left\{ 1 - \exp\left[-\tau_1\left(\dfrac{1}{\mu_0} + \dfrac{1}{\mu}\right)\right] \right\} \right.$
$$+ \frac{\mu'}{\mu - \mu'}\left\{ \exp\left[-\tau_1\left(\frac{1}{\mu_0} + \frac{1}{\mu'}\right)\right]\right.$$
$$\left.\left. - \exp\left[-\tau_1\left(\frac{1}{\mu_0} + \frac{1}{\mu}\right)\right]\right\} \right], \mu \neq \mu'$$

6.8 $I^{\uparrow}(0, \mu_1) = \dfrac{1}{2}\left\{ \dfrac{\mu_0}{\mu_1}\left[2S^- - 2b\dfrac{\mu_0}{\mu_1}(S^+ + S^-)\right] + \dfrac{K\mu_1}{2b} + H \right\}$, where

$$K = -\frac{2\mu_0}{\mu_1}\left[\left(S^- - \frac{b\mu_0}{\mu_1}\right)e^{-\tau_1/\mu_0} + S^+ + \frac{b\mu_0}{\mu_1}(S^+ + S^-)\right] \bigg/ \left(\frac{\tau_1 + \mu_1}{b}\right),$$

$$H = \frac{2\mu_0}{\mu_1}\left[S^+ + \frac{b\mu_0}{\mu_1}(S^+ + S^-)\right] + \frac{K\mu_1}{2b}$$

6.12 $(E_l/a_l)^2 + (E_r/a_r)^2 - 2(E_l/a_l)(E_r/a_r)\cos\delta = \sin^2\delta$

6.15 (b) Let $I_0 = 1$, $\left[2 - \frac{1}{2}\,0\,\frac{1}{2}\right]$, $I = 2$, $P = \frac{1}{2\sqrt{2}}\%$, $I_r = \frac{5}{4}$; (c)$\chi = 0°$, $\beta = 22.5°$;
(d)$\frac{1}{4}[(4 - \sqrt{2})\,(2\sqrt{2} - 1)\,0\,(2\sqrt{2} - 1)]$, $\frac{1}{4}[(4 - \sqrt{2})\,(1 - 2\sqrt{2})\,0\,(1 - 2\sqrt{2})]$;
(e)$\frac{1}{4}[(4 + \sqrt{2})\,(2\sqrt{2} - 3)\,0\,(1 + 2\sqrt{2})]$ for right-hand polarization

6.17 $\nabla^2 I_0^0 - k^2\beta_e^2 I_0^0 = -\chi\beta_e^2 e^{-\tau_s}$; where $\chi = 3\tilde{\omega}F_\odot(1 + g - \tilde{\omega}g)/4\pi$

Chapter 7

7.1 (b) $\nu^* = 3.13$

7.2 (a) $g(k) = -\frac{12}{k^2}e^{-k} + \left(\frac{24}{k^2} - \frac{8}{k} + 2\right)\left[\frac{1}{k^2} - \left(\frac{1}{k} + \frac{1}{k^2}\right)e^{-k}\right]$, $i = 10$, $k = 5$,
$g(k) = 0.0490$; (b) $i = 10$, $k = 5$, $g(k) = 0.0489$

7.3

Second difference, $\mathbf{H} = \begin{bmatrix} 1 & -2 & 1 & & & & & & 0 \\ -2 & 5 & -4 & 1 \\ 1 & -4 & 6 & -4 & 1 \\ & 1 & -4 & 6 & -4 & 1 \\ & & & & \cdots \\ & & & & & 1 & -4 & 5 & -2 \\ 0 & & & & & & 1 & -2 & 1 \end{bmatrix}$

7.5 $\tau \sim \text{LWC}^{2/3}N^{1/3}$

Chapter 8

8.1 302 K

8.2 $\bar{\varepsilon} = 0.88$, $T_a = 250\,\text{K}$; $\bar{\varepsilon} = 0.9$, $T_a = 249\,\text{K}$

8.3 Note: $1 + y + y^2 + \cdots = 1/(1 - y)$

8.4 $\partial T_s/\partial F_{ir} = T_s/4F_{ir}$

8.5 $T(x_i = 0.95) \cong -7°C$; $x = [a - b/(Q/Q_0)]^{1/2}$, where $a \cong 1.71646$ and
$b \cong 0.81396$

8.6 (c) $x_i^4 - ax_i^3 - bx_i^2 + cx_i - d + [e + f(\triangle Q/Q_0)]/(1 + \triangle Q/Q_0) = 0$,
where a, b, c, d, e, and f are certain coefficients

8.7 $F_2(6D'' + 1) = QH_2(x_i)$, find D''

References

Albrecht, B. A. (1989). Aerosols, cloud microphysics, and fractional cloudiness. *Science* **245,** 1227–1230.

Ambartzumian, V. A. (1936). The effect of absorption lines on the radiative equilibrium of the outer layers of the stars. *Publ. Observ. Astonom.* **6,** Univ. Leningrad, Leningrad 7.

Ambartzumian, V. A. (1942). A new method for computing light scattering in turbid media. *Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz.* **3,** 97–104.

Ambartzumian, V. A. (1958). "Theoretical Astrophysics." Pergamon Press, New York.

Anderson, G. P., Picard, R. H., and Chetwynd, J. H. (1995). Proceedings of the 17th Annual Review Conference on Atmospheric Transmission Models, Special Report No. 274, 332 pp., Phillips Laboratory Geophysics Directorate, MA.

Anderson, S. M., and Mauersberger, K. (1992). Laser measurements of ozone absorption cross-sections in the Chappuis band. *Geophys. Res. Lett.* **19,** 933–936.

Appleman, H. (1953). The formation of exhaust condensation trails by jet aircraft. *Bull. Am. Meteorol. Soc.* **34,** 14–20.

Arakawa, A. (1966). Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comp. Phys.* **1,** 119–143.

Arking, A., and Grossman, K. (1972). The influence of line shape and band structure on temperatures in planetary atmospheres. *J. Atmos. Sci.* **29,** 937–949.

Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *Philosoph. Mag.* **41,** 237–275.

Asano, S. (1975). On the discrete-ordinates method for the radiative transfer. *J. Meteorol. Soc. Jpn.* **53,** 92–95.

Bacon, N. J, Swanson, B. D., Baker, M. B., and Davis, E. J. (1998). Laboratory studies of light scattering by single levitated ice crystals. *Proc. Conf. Cloud Phys., Am. Meteorol. Soc.,* Everett, WA, pp. 427–428.

Baker, J. (1990). "Planet Earth: The View from Space." Harvard Univ. Press, Cambridge, MA.

Barkey, B., Liou, K. N., Takano, Y., Gellermann, W., and Sokolsky, P. (1999). An analog light scattering experiment of hexagonal icelike particles. II. Experimental and theoretical results. *J. Atmos. Sci.* **56,** 613–625.

Bellman, R. E., Kalaba, R. E., and Prestrud, M. C. (1963). "Invariant Imbedding and Radiative Transfer in Slabs of Finite Thickness." Elsevier, New York.

Berenger, J. P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *J. Comput. Phys.* **114,** 185–200.

Berger, A. L. (1978). Long-term variations of daily insolation and quaternary climatic changes. *J. Atmos. Sci.* **35,** 2362–2367.

Berger, A. L. (1988). Milankovich theory and climate. *Rev. Geophys.* **26,** 624–657.

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Res.* **97,** 163–172.

Bohr, N. (1913). On the constitution of atoms and molecules. *Phil. Mag.* **(6), 26,** 1–25.

Born, M., and Wolf, E. (1975). "Principles of Optics." Pergamon, Oxford.

Brasseur, G., and Simon, P. C. (1981). Stratospheric chemical and thermal response to long-term variability in solar UV irradiance. *J. Geophys. Res.* **86,** 7343–7362.

Brasseur, G., and Solomon, S. (1986). "Aeronomy of the Middle Atmosphere," 2nd Ed., Reidel, Dordrecht, Netherlands.

Broccoli, A. J., and Manabe, S. (1987). The influence of continental ice, atmospheric CO_2, and land albedo on the climate of the last glacial maximum. *Climate Dyn.* **1,** 87–99.

Browell, E. V. (1994). Remote sensing of trace gases from satellites and aircraft. *In* "The Chemistry of the Atmosphere: Its Impact on Global Change" (J. G. Calvert, ed.), pp. 121–134. Blackwell Scientific Publications, London.

Brutsaert, W. (1982). "Evaporation into the Atmosphere." Reidel, Dordrecht, Netherlands.

Brutsaert, W., Hsu, A. Y., and Schmugge, T. J. (1993). Parameterization of surface heat fluxes above forest with satellite thermal sensing and boundary-layer sounding. *J. Appl. Meteorol.* **32,** 909–917.

Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the earth. *Tellus,* **21,** 611–619.

Budyko, M. I. (1982). "The Earth's Climate: Past and Future." Academic Press, New York.

Budyko, M. I. (1986). "The Evolution of the Biosphere." Reidel, Dordrecht, Netherlands.

Cahalan, R., Ridgway, W., and Wiscombe, W. (1994). Independent pixel and Monte Carlo estimates of stratocumulus albedo. *J. Atmos. Sci.* **51,** 3776–3790.

Cai, Q., and Liou, K. N. (1982). Polarized light scattering by hexagonal ice crystals: Theory. *Appl. Opt.* **21,** 3569–3580.

Chahine, M. T. (1970). Inverse problems in radiative transfer: Determination of atmospheric parameters. *J. Atmos. Sci.* **27,** 960–967.

Chahine, M. T., Evans, N. L., Gilbert, V., and Haskins, R. D. (1984). Requirements for a passive IR advanced moisture and temperature sounder. *Appl.Opt.* **23,** 979–989.

Chamberlain, J. W., and Hunten, D. M. (1987). "Theory of Planetary Atmospheres: An Introduction to Their Physics and Chemistry." Academic Press, New York.

Chandrasekhar, S. (1950). "Radiative Transfer." Oxford Univ. Press, Oxford.

Chapman, S. (1930). A theory of upper-atmospheric ozone. *Mem. Roy. Meteorol. Soc.* **3,** 103–125.

Chapman, S. (1931). The absorption and dissociation or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth. *Proc. Phys. Soc.* **43,** 26–45.

Charlock, T. P. (1982). Cloud optical feedback and climatic stability in a radiative–convective model. *Tellus* **34,** 245–254.

Charlock, T. P., and Alberta, T. L. (1996). The CERES/ARM/GEWEX Experiment (CAGEX) for the retrieval of radiative fluxes with satellite data. *Bull Am. Meteorol. Soc.* **77,** 2673–2683.

Charlock, T. P., and Sellers, W. D. (1980). Aerosol effects on climate: Calculations with time-dependent and steady-state radiative convective models. *J. Atmos. Sci.* **37,** 1327–1341.

Charlson, R. J., Lovelock, J. E., Andreae, M. O., and Warren, S. G. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326,** 655–661.

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.* **4,** 135–162.

Chepfer, H., Brogniez, G., and Fouquart, Y. (1998). Cirrus clouds' microphysical properties deduced from POLDER observations. *J. Quant. Spectrosc. Radiat. Transfer.* **60,** 375–390.

Chou, M. D., and Arking, A. (1980). Computation of infrared cooling rates in the water vapor bands. *J. Atmos. Sci.* **37,** 855–867.

Chou, M. D., Kratz, D. P., and Ridgway, W. (1991). IR radiation parameterizations in numerical climate models. *J. Climate* **4,** 424–437.

CLIMAP Project Members (1976). The surface of the ice-age earth. *Science* **191,** 1131–1137.

Coakley, J. A., Bernstein, J. R., and Durkee, P. A. (1987). Effect of ship track effluents on cloud reflectivity. *Science* **237,** 1020–1021.

Coffeen, D. L., (1979). Polarization and scattering characteristics in the atmosphere of Earth, Venus, and Jupiter. *J. Opt. Soc. Am.* **69,** 1051–1064.

Collis, R. T. H. (1969). Lidar. *Adv. Geophys.* **13,** 113–139.

Conrath, B. J., Hanel, R. A., Kunde, V. G., and Prabhakara, C. (1970). The infrared interferometer experiment on Nimbus 3. *J. Geophys. Res.* **75,** 5831–5857.

Coulson, K. L. (1975). "Solar and Terrestrial Radiation." Academic Press, New York.

Coulson, K. L. (1983). Effects of the El Chichon volcanic cloud in the stratosphere on the polarization of light from the sky. *Appl. Opt.* **22,** 1036–1050.

Cox, C., and Munk, W. (1954). Measurement of the roughness of the sea surface from photographs of the sun's glitter. *J. Opt. Soc. Am.* **44,** 838–850.

Crutzen, P. J. (1970). The influence of nitrogen oxides on the atmospheric ozone content. *Quart. J. Roy. Meteorol. Soc.* **96,** 320–325.

Crutzen, P. J. (1983). Atmospheric interactions: Homogeneous gas reactions of C, N, and S containing compounds. *In* "The Major Biogeophysical Cycles and Their Interactions" (B. Bolin and R. B. Cook, eds.), pp. 67–114. Wiley, New York.

Curtis, A. R. (1952). Contribution to a discussion of "A statistical model for water vapor absorption," by R. M. Goody. *Quart. J. Roy. Meteorol. Soc.* **78,** 638–640.

Dahlback, A., and Stamnes, K. (1991). A new spherical model for computing the radiation field available for photolysis and heating at twilight. *Planet. Space Sci.* **59,** 671–683.

d'Almeida, G. A., Koepke, P., and Shettle, E. P. (1991). "Atmospheric Aerosols." A. Deepak Publishing, Hampton, VA.

Dave, J. V. (1978). Effect of aerosols on the estimate of total ozone in an atmospheric column from the measurements of its ultraviolet radiance. *J. Atmos. Sci.* **35,** 899–911.

Dickinson, R. E. (1973). Method of parameterization for infrared cooling between altitudes of 30 and 70 kilometers. *J. Geophys. Res.* **78,** 4451–4457.

Diner, D. J., Beckert, J. C., Reilly, T. H., *et al.* (1998). Multi-angle Imaging SpectroRadiometer (MISR) Instrument description and experiment overview. *IEEE Trans. Geosci. Remote Sens.* **36,** 1072–1087.

Dobson, G. M. (1957). Observers' handbook for the ozone spectrometer. *Ann. Int. Geophys. Year* **5,** 46–89.

Eddington, A. S. (1916). On the radiative equilibrium of the stars. *Mon. Not. Roy. Astronom. Soc.* **77,** 16–35.

Eddy, J. A. (1977). Historical evidence for the existence of the solar cycle. *In* "The Solar Output and Its Variation" (O.R. White, ed.), pp. 51–71, Colorado Assoc. Univ. Press, Boulder, CO.

Elsasser, W. M. (1938). Mean absorption and equivalent absorption coefficient of a band spectrum. *Phys. Rev.* **54,** 126–129.

Elsasser, W. M. (1942). "Heat Transfer by Infrared Radiation in the Atmosphere." *Harvard Meteorol. Studies,* **6,** Harvard Univ. Press, Cambridge, MA.

Elsasser, W. M., and Culbertson, M. F. (1960). Atmospheric radiation tables. *Meteorol. Monogr.* **4,** 1–43.

Evans, K. F. (1993). Two-dimensional radiative transfer in cloudy atmospheres: The spherical harmonic spatial grid method. *J. Atmos. Sci.* **50,** 3111–3124.

Farman, J. C., Gardiner, B. G., and Shanklin, J. D. (1985). Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature* **315,** 207–208.

Feigelson, E. M. (1981). "Radiation in Cloudy Atmosphere." Section of Meteorology and Atmospheric Physics of Soviet Geophysical Committees, Hydrometeorology Press (IZD).

Freeman, K. P., and Liou, K. N. (1979). Climatic effects of cirrus clouds. *Adv. Geophy.* **21,** 231–287.

Fresnel, A. J. (1823). Mémoire sur la diffraction de la lumiere. *In* "Oeuvres complétes d'Augustin Fresnel" (H. Senarmont, E. Verdet, and L. Fresnel, eds.), **1,** pp. 247–767. Paris, 1866–1870.

Freudenthaler, V., Homburg, F., and Jäger, H. (1996). Optical parameters of contrail from lidar measurements: Linear depolarization. *Geophy. Res. Lett.* **23,** 3715–3718.

Fu, Q., and Liou, K. N. (1992). On the correlated *k*-distribution method for radiative transfer in nonhomogeneous atmospheres. *J. Atmos. Sci.* **49,** 2139–2156.

Fu, Q., and Liou, K. N. (1993). Parameterization of the radiative properties of cirrus clouds. *J. Atmos. Sci.* **50,** 2008–2025.

Fu, Q., Liou, K. N., Cribb, M. C., Charlock, T. P., and Grossman, A. (1997). Multiple scattering parameterization in thermal infrared radiative transfer. *J. Atmos. Sci.* **54,** 2799–2812.

Gao, B. C., and Kaufman, Y. J. (1995). Selection of the 1.375-μm MODIS channel for remote sensing of cirrus clouds and stratospheric aerosols from space. *J. Atmos. Sci.* **52,** 4231–4237.

Gasiewski, A. J. (1993). Microwave radiative transfer in hydrometeors. In "Atmospheric Remote Sensing by Microwave Radiometry" (M.A. Janssen, ed.), pp 91–144. Wiley, New York.

Ghil, M. (1976). Climate stability for a Sellers-type model. *J. Atmos. Sci.* **33,** 3–20.

Gille, J. C., and House, F. B. (1971). On the inversion of limb radiance measurements. I: Temperature and thickness. *J. Atmos. Sci.* **28,** 1427–1442.

Gille, J. C., Bailey, P. L., Massie, S. T., *et al.* (1996). Accuracy and precision of cryogenic limb array etalon spectrometer (CLAES) temperature retrieval. *J. Geophys. Res.* **101,** 9583–9601.

Godson, W. L. (1953). The evaluation of infrared radiative fluxes due to atmospheric water vapor. *Quart. J. Roy. Meteorol. Soc.* **79,** 367–379.

Goody, R. M. (1952). A statistical model for water vapor absorption. *Quart. J. Roy. Meteorol. Soc.* **78,** 165–169.

Goody, R. M. (1964). "Atmospheric Radiation I: Theoretical Basis." Oxford Univ. Press, London.

Goody, R. M. (1995). "Principles of Atmospheric Physics and Chemistry." Oxford Univ. Press, New York.

Goody, R. M. and Yung, Y. L. (1989). "Atmospheric Radiation. Theoretical Basis," 2nd ed. Oxford Univ. Press, New York.

Goody, R. M., West, R., Chen, L., and Crisp, D. (1989). The correlated-*k* method for radiation calculations in nonhomogeneous atmospheres. *J. Quant. Spectrosc. Radiat. Transfer* **42,** 539–550.

Greenler, R. (1980). "Rainbows, Halos, and Glories." Cambridge Univ. Press, Cambridge, U.K.

Grenfell, T. C., and Warren, S. G. (1999). Representation of a nonspherical ice particle by a collection of independent spheres for scattering and absorption of radiation. *J. Geophys. Res.* **104,** 31697–31709.

Griggs, M. (1968). Absorption coefficients of ozone in the ultraviolet and visible regions. *J. Chem. Phys.* **49,** 857–859.

Grody, N. C. (1976). Remote sensing of atmospheric water content from satellites using microwave radiometry. *IEEE Trans. Antennas Propag.* **AP 24,** 155–162.

Grody, N. C. (1993). Remote sensing of the atmosphere from satellites using microwave radiometry. *In* "Atmospheric Remote Sensing by Microwave Radiometry" (M. A. Janssen, ed.), pp 91–144. Wiley, New York.

Hansen, J. E. (1971). Multiple scattering of polarized light in planetary atmospheres. Part II. Sunlight reflected by terrestrial water clouds. *J. Atmos. Sci.* **28,** 1400–1426.

Hansen, J. E., and Hovenier, J. W. (1974). Interpretation of the polarization of Venus. *J. Atmos. Sci.* **31,** 1137–1160.

Hansen, J. E., and Pollack, J. B. (1970). Near-infrared light scattering by terrestrial clouds. *J. Atmos. Sci.* **27,** 265–281.

Hansen, J. E., and Travis, L. D. (1974). Light scattering in planetary atmospheres. *Space Sci. Rev.* **16,** 527–610.

Hansen, J. E., Wang, W. C., and Lacis, A. (1978). Mount Agung eruption provides test of a global climate perturbation. *Science* **199,** 1065–1068.

Hansen, J. E., Johnson, D., Lacis, A., Lebedeff, S., Lee, P., Rind, D., and Russell, G. (1981). Climate impact of increasing atmospheric carbon dioxide. *Science* **213,** 957–966.

Hansen, J. E., Lacis, A., Ruedy, R., and Sato, M. (1992). Potential climate impact of Mt. Pinatubo eruption. *Geophys. Res. Lett.* **19,** 215–218.

Hansen, J. E., Sato, M., Ruedy, R., Lacis, A., and Oinas, V. (2000). Global warming in the twenty-first century: An alternative scenario. *Proc. Natl. Acad. Sci. USA* **97,** 9875–9880.

Harrison, E. F., Minnis, P., Barkstrom, B. R., Ramanathan, V., Cess, R. D., and Gibson, G. G. (1990). Seasonal variation of cloud radiative forcing derived from the Earth Radiation Budget Experiment. *J. Geophys. Res.* **95,** 18,687–18,703.

Hartmann, D. L. (1994). "Global Physical Climatology." Academic Press, New York.

Hays, J. D., Imbrie, J., and Shackleton, N. J. (1976). Variations in the earth's orbit: Pacemaker of the ice ages. *Science* **194,** 1121–1132.

Heintzenberg, J., and Charlson, R. J. (1996). Design and application of the integrating nephelometer: A review. *J. Atmos. Oceanic Technol.* **13,** 987–1000.

Held, I. M., and Suarez, M. J. (1974). Simple albedo feedback models of the icecaps. *Tellus* **26,** 613–629.

Henderson-Sellers, A., and Wilson, M. F. (1983). Surface albedo data for climate modeling. *Rev. Geophys. Space Phys.* **21,** 1743–1778.

Henyey, L. C., and Greenstein J. L. (1941). Diffuse radiation in the galaxy. *Astrophys. J.* **93,** 70–83.

Herman, B. M., Ben-David, A., and Thorne, K. J. (1994). Numerical technique for solving the radiative transfer equation for a spherical-shell atmosphere. *Appl. Opt.* **33,** 1760–1770.

Herman, B. M., Caudill, T., Flittner, D., and Liou, K. N. (1995). Comments on "Solar heating rates: The importance of spherical geometry." *J. Atmos. Sci.* **52,** 380–382.

Hertz, H. (1889). Über Strahlen elecktrischen Kraft. *Wiedem. Ann. Phys. Chem.* **36,** 769–783.

Herzberg, G. (1945). "Molecular Spectra and Molecular Structure." Van Nostrand Reinhold, Princeton, NJ.

Herzberg, L. (1965). Solar optical radiation and its role in upper atmospheric processes. *In* "Physics of the Earth's Upper Atmosphere" (C. O. Hines, I. Paglis, R. Hartz, and J. A. Fejer, eds.), pp. 31–45. Prentice Hall, New Jersey.

Heymsfield, A. J., and Donner, L. J. (1990). A scheme for predicting ice cloud water contents in a general circulation model. *J. Atmos. Sci.* **47,** 1865–1877.

Heymsfield, A. J., and Iaquinta, J. (2000). Cirrus crystal terminal velocities. *J. Atmos. Sci.* **57,** 916–938.

Heymsfield, A. J., and Platt, C. M. R. (1984). A parameterization of the particle size spectrum of ice clouds in terms of the ambient temperature and the ice water content. *J. Atmos. Sci.* **41,** 846–855.

Holland, A. C., and Gagne, G. (1970). The scattering of polarized light by polydisperse systems of irregular particles. *Appl. Opt.* **9,** 1113–1121.

Houghton, J. T., Taylor, F. W., and Rodgers, C. D. (1984). "Remote Sounding of Atmospheres." Cambridge Univ. Press, Cambridge, U.K.

Hovenier, J. W. (1969). Symmetry relations for scattering of polarized light in a slab of randomly oriented particles. *J. Atmos. Sci.* **26,** 488–499.

Hovenier, J. W. (2000). Measuring scattering matrix of small particles at optical wavelengths. *In* "Light Scattering by Nonspherical Particles" (M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, eds.), pp. 335–365. Academic Press, San Diego.

Hoyt, D. V., and Schatten, K. H. (1997). "The Role of the Sun in Climate Change." Oxford Univ. Press, New York.

Huang, R., and Liou, K. N. (1983). Polarized microwave radiation transfer in precipitating cloudy atmospheres: Applications to window frequencies. *J. Geophys. Res.* **88,** 3885–3893.

Huang, R., and Liou, K. N. (1984). Remote sounding of the cirrus cloud optical depth and temperature from 3.7 and 11 micrometer windows. *Adv. Atmos. Sci.* **1,** 150–164.

Humphreys, W. J. (1954). "Physics of the Air." Dover, New York.

Imbrie, J., Hays, J. D., Martinson, D. G., McIntyre, A., Mix, A. G., Morley, J. J., Pisias, N. G., Prell, W. L., and Shackleton, N. J. (1984). The orbital theory of Pleistocene climate: Support from a revised chronology of marine δ^{18}O record. *In* "Milankovitch and Climate, Part 1" (A. L. Berger, J. Imbrie, J. Hays, G. Kukla, and B. Saltzman, eds.), pp. 269–305. Reidel, Boston.

IMSL User's Manual (1987). "Math/Library." IMSL, Houston, Texas, pp. 1151.

Inn, E. C., and Tanaka, Y. (1953). Absorption coefficient of ozone in the ultraviolet and visible regions. *J. Opt. Soc. Am.* **43,** 870–873.

Inoue, T. (1985). On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bispectral measurements in the 10 micron window region. *J. Meteorol. Soc. Jpn.* **63,** 88–99.

IPCC (1996). "Climate Change 1995: The Science of Climate Change." (J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg and K. Maskell, eds.). Cambridge Univ. Press, Cambridge, U.K.

IPCC (1999). "Aviation and the Global Atmosphere" (J. E. Penner, D. H. Lister, D. J. Griggs, D. J. Dokken, and M. McFarland, eds.), Chapter 3, "Aviation-Produced Aerosols and Cloudiness" (D. W. Fahey, and U. Schumann, eds.), pp. 63–120. Cambridge Univ. Press, Cambridge, U.K.

Jackson, J. D. (1975). "Classical Electrodynamics." Wiley, New York.

Jacobowitz, H., Tighe, R. J., and Nimbus 7 ERB experiment Team (1984). The earth radiation budget derived from the Nimbus 7 ERB experiment. *J. Geophys. Res.* **89,** 4997–5010.

Janssen, M. A., ed. (1993). "Atmospheric Remote Sensing by Microwave Radiometry." Wiley, New York.

Jastrow, R., and Thompson, M. H. (1984). "Astronomy: Fundamentals and Frontiers." 4th ed., Wiley, New York.

Johnston, H. S. (1971). Reduction of stratospheric ozone by nitrogen oxide catalysis from SST exhaust. *Science* **173,** 517–522.

Joseph, J. H., Wiscombe, W. J., and Weinman, J. A. (1976). The delta-Eddington approximation for radiative flux transfer. *J. Atmos. Sci.* **33,** 2452–2459.

Junge, C. E. (1963). "Air Chemistry and Radioactivity." Academic Press, New York.

Kaplan, L. D. (1959). Inference of atmospheric structure from remote radiation measurements. *J. Opt. Soc. Am.* **49,** 1004–1007.

Kerker, M. (1969). "The Scattering of Light and Other Electromagnetic Radiation." Academic Press, San Diego.

Kidder, S. Q., and Vonder Haar, T. H. (1995). "Satellite Meteorology." Academic Press, San Diego.

King, J. I. F. (1956). The radiative heat transfer of planet Earth. *In* "Scientific Uses of Earth Satellites," pp. 133–136. Univ. of Michigan Press, Ann Arbor.

King, M. D., Byrne, D. M., Herman, B. M., and Reagan, J. A. (1978). Aerosol size distributions obtained by inversion of spectral optical depth measurements. *J. Atmos. Sci.* **35,** 2153–2167.

King, M. D., Kaufman, Y. J., Menzel, W. P., and Tanre, D. (1992). Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectroradiometer. *IEEE Trans. Geosci. Remote Sens.* **30,** 2–27.

King, M. D., Tsay, S. C., Platnick, S. E., Wang, M., and Liou, K. N. (1997). Cloud retrieval algorithm for MODIS: Optical thickness, effective particle radius, and thermodynamic phase. MODIS Algorithm Theoretical Basis Document No. ATBD-MOD-05, MOD06-Cloud Product. Goddard Space Flight Center.

Kirchhoff, G. (1860). Über das Verhältnis zwischen dem Emissionsvermögen und dem Absorptionsvermögen der Körper für Wärme und Licht. *Ann. Phys. Chem.* **109,** 275–301.

Kondratyev, K. Ya. (1969). "Radiation in the Atmosphere" (International Geophysics Series; Vol. 12). Academic Press, New York.

Korolev, A. V., Issac, G. A., and Hallett, J. (1999). Ice particle habits in Arctic clouds. *Geophys. Res. Lett.* **26,** 1299–1302.

Kourganoff, V. (1952). "Basic Methods in Transfer Problems." Clarendon Press, London.

Kunde, V. G., Conrath, B. J., Hanel, R. A., Maguire, W. C., Prabhakara, C., and Solomonson, V. V. (1974). The Nimbus IV infrared spectroscopy experiment. 2. Comparison of observed and theoretical radiances from 425–1450 cm^{-1}. *J. Geophys. Res.* **79,** 777–784.

Lacis, A. A., and Oinas, V. (1991). A description of the correlated k-distribution method for modeling non-grey gaseous absorption, thermal emission and multiple scattering in vertically inhomogeneous atmospheres. *J. Geophys. Res.* **96,** 9027–9063.

Lacis, A., Wang, W. C., and Hansen, J. F. (1979). Correlated k-distribution method for radiative transfer in climate models: Application to effect of cirrus clouds on climate. *NASA Conf. Publ.* **207b,** 309–314.

Laplace, P. S. de (1796). "Exposition du systéme du monde." Imprimerie Cercle-Social, Paris.

Leaitch, W. R., Isaac, G. A., Strapp, J. W., Banic, C. M., and Wiebe, H. A. (1992). The relationship between cloud droplet number concentrations and anthropogenic pollution: Observations and climatic implications. *J. Geophys. Res.* **97,** 2463–2474.

Lean, J. (1987). Solar ultraviolet irradiance variations: A review. *J. Geophys. Res.* **92,** 839–868.

Lean, J., and Rind, D. (1998). Climate forcing by changing solar radiation. *J. Climate* **11,** 3069–3094.

Lenoble, J. (1993). "Atmospheric Radiative Transfer." A. Deepak Publishing, Hampton, VA.

Lenoble, J., and Sekera, Z. (1961). Equation of radiative transfer in a planetary spherical atmosphere. *Proc. Nat. Acad. Sci. USA* **47,** 372–378.

Leovy, C. B. (1969). Energetics of the middle atmosphere. *Adv. Geophys.* **13,** 191–221.

Lhermittee, R. (1990). Attenuation and scattering of millimeter wavelength radiation by clouds and precipitation. *J. Atmos. Oceanic Technol.* **7,** 464–479.

Li, Z., Cribb, M., and Trishchenko, A. (2002). Impact of surface inhomogeneity on the closure of solar radiative transfer and a method of estimating surface areal-mean albedo from downwelling flux measurements. *J. Geophy. Res.* (in press).

Lindzen, R. S., and Farrell, B. (1977). Some realistic modifications of simple climate models. *J. Atmos. Sci.* **34,** 1487–1501.

Liou, K. N. (1973a). A numerical experiment on Chandrasekhar's discrete-ordinates method for radiative transfer: Application to cloudy and hazy atmospheres. *J. Atmos. Sci.* **30,** 1303–1326.

Liou, K. N. (1973b). Transfer of solar irradiance through cirrus cloud layers. *J. Geophys. Res.* **78,** 1409–1418.

Liou, K. N. (1974). Analytic two-stream and four-stream solutions for radiative transfer. *J. Atmos. Sci.* **31,** 1473–1475.

Liou, K. N. (1975). Applications of the discrete-ordinates method for radiative transfer to inhomogeneous aerosol atmospheres. *J. Geophys. Res.* **80,** 3434–3440.

Liou, K. N. (1986). The influence of cirrus on weather and climate process: A global perspective. *Mon. Weather Rev.* **114,** 1167–1199.

Liou, K. N. (1992). "Radiation and Cloud Processes in the Atmosphere. Theory, Observation, and Modeling." Oxford Univ. Press, New York.

Liou, K. N., and Duff, A. D. (1979). Atmospheric liquid water content derived from parameterization of Nimbus 6 scanning microwave spectrometer data. *J. Appl. Meteorol.* **18,** 99–103.

Liou K. N., and Hansen J. E. (1971). Intensity and polarization for single scattering by polydisperse spheres: A comparison of ray optics and Mie theory. *J. Atmos. Sci.* **28,** 995–1004.

Liou, K. N., and Lahore, H. (1974). Laser sensing of cloud composition: A backscattered depolarization technique. *J. Appl. Meteorol.* **13,** 257–263.

Liou, K. N., and Ou, S. C. (1983). Theory of equilibrium temperatures in radiative-turbulent atmospheres. *J. Atmos. Sci.* **40,** 214–229.

Liou, K. N., and Ou, S. C. (1989). The role of cloud microphysical processes in climate: An assessment from a one-dimensional perspective. *J. Geophys. Res.* **94,** 8599–8607.

Liou, K. N., and Rao, N. (1996). Radiative transfer in cirrus clouds. Part IV: On cloud geometry, inhomogeneity, and absorption. *J. Atmos. Sci.* **53,** 3046–3065.

Liou, K. N., and Takano, Y. (1994). Light scattering by nonspherical particles: Remote sensing and climatic implications. *Atmos. Res.* **31,** 271–298.

Liou, K. N., and Wittman, G. D. (1979). Parameterization of the radiative properties of clouds. *J. Atmos. Sci.* **36,** 1261–1273.

Liou, K. N., and Xue, Y. (1988). Exploration of the remote sounding of infrared cooling rates due to water vapor. *Meteorol. Atmos. Phys.* **38,** 131–139.

Liou, K. N., Aufderhaar, G. C., and Nipko, P. T. (1981). Effects of clouds and precipitation on the temperature profile retrieval for the DMSP SSM/T microwave sounders. *J. Appl. Meteorol.* **20,** 821–825.

Liou, K. N., Fu, Q., and Ackerman, T. P. (1988). A simple formulation of the delta-four-stream approximation for radiative transfer parameterizations. *J. Atmos. Sci.* **45,** 1940–1947.

Liou, K. N., Takano, Y., Ou, S. C., Heymsfield, A., and Kreiss, W. (1990a). Infrared transmission through cirrus clouds: A radiative model for target detection. *Appl. Opt.* **29,** 1886–1896.

Liou, K. N., Ou, S. C., Takano, Y., Valero, F. P. J., and Ackerman, T. P. (1990b). Remote sounding of the tropical cirrus cloud temperature and optical depth using 6.5 and 10.6-micron radiometers during STEP. *J. Appl. Meteorol.* **29,** 716–726.

Liou, K. N., Ou, S. C., and Koenig, G. (1990c). An investigation of the climatic effect of contrail cirrus. *In* "Air Traffic and the Environment-Background, Tendencies and Potential Global Atmospheric Effects" (U. Schumann, ed.), pp. 154–169. Springer-Verlag, Berlin.

Liou, K. N., Yang, P., Takano, Y., Sassen, K., Charlock, T. P., and Arnott, W. P. (1998). On the radiative properties of contrail cirrus. *Geophys. Res. Lett.* **25,** 1161–1164.

Liou, K. N., Takano, Y., and Yang, P. (2000). Light scattering and radiative transfer in ice crystal clouds: Applications to climate research. *In* "Light Scattering by Nonspherical Particles" (M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, eds.), pp. 417–449. Academic Press, San Diego.

Livingston, W. C., Wallace, L. and White, O. R. (1988). Spectrum line intensity as a surrogate for solar irradiance variations. *Science* **240,** 1765–1767.

Logan, J. A., Prather, M. J., Wofsy, S. C., and McElroy, M. B. (1978). Atmospheric chemistry: Response to human influence. *Phil. Trans. Roy. Soc. London Ser. A* **290,** 187–234.

Lopez-Puertas, M., Rodrigo, R., Molina, A., and Taylor, F. W. (1986). A non-LTE radiative transfer model for infrared bands in the middle atmosphere. I. Theoretical basis and application to CO_2 15 μm bands. *J. Atmos. Terr. Phys.* **48,** 729–748.

Lorentz, H. A. (1906). The absorption and emission of lines of gaseous bodies. *In* "H.A. Lorentz Collected Papers" (The Hague, 1934–1939), **3,** 215–238.

Lorenz, E. N. (1969). The predictability of a flow which possesses many scales of motion. *Tellus* **21,** 289–307.

Lorenz, L. V. (1890). Lysbevaegelsen i og uder en plane lysbolger belyst kulge. *Vidensk. Selk. Skr.,* **6,** 1–62.

Lyot, B. (1929). Recherches sur la polarisation de la lumière des planètes et de quelques substances terrestres. *Ann. Observ. Paris (Meudon)* **8,** 161 pp. [Available in English as NASA TTF-187, 1964.]

Mace, G., Ackerman, T. P., Minnis, P., and Young, D. F. (1998). Cirrus layer microphysical properties derived from surface-based millimeter radar and infrared interferometer data. *J. Geophys. Res.* **103,** 23207–23216.

Malkmus, W. (1967). Random Lorentz band model with exponential-tailed S^{-1} line intensity distribution function. *J. Opt. Soc. Am.* **57,** 323–329.

Manabe, S., and Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of the relative humidity. *J. Atmos. Sci.* **24,** 241–259.

Marchuk, G. I., Mikhailov, G. A., Nazaraliev, M. A., Darbinjan, R. A., Kargin, B. A., and Elepov, B. S. (1980). "The Monte Carlo Methods in Atmospheric Optics." Springer-Verlag, Berlin.

Marshall, J. S., and Palmer, W. (1948). The distribution of raindrops with size. *J. Meteorol.* **5,** 165–166.

Martin, P. G. (1974). Interstellar polarization from a medium with changing grain alignment. *Astrophy. J.* **187,** 461–472.

Mateer, C. L., Heath, D. F., and Krueger, A. J. (1971). Estimation of total ozone from satellite measurements of backscattered ultraviolet earth radiance. *J. Atmos. Sci.* **28,** 1307–1311.

Matveev, L. T. (1984). "Cloud Dynamics." Kluwer Group, Dordrecht, Netherlands.

McCartney, E. J. (1983). "Absorption and Emission by Atmospheric Gases." Wiley, New York.

McClain, E. P., Pichel, W. G., and Walton, C. C. (1985). Comparative performance of AVHRR-based multichannel sea surface temperatures. *J. Geophy. Res.* **90,** 11587–11601.

McCormick, M. P., Hamill, P., Pepin, T. J., Chu, W. P., Swissler, T. J., and McMaster, L. R. (1979). Satellite studies of the stratospheric aerosol. *Bull. Am. Meteorol. Soc.* **60,** 1038–1046.

McMillin, L. M., and Dean, C. (1982). Evaluation of a new operational technique for producing clear radiances. *J. Appl. Meteorol.* **21,** 1005–1014.

Meador, W. E., and Weaver, W. R. (1980). Two stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement. *J. Atmos. Sci.* **37,** 630–643.

Meeks, M. L., and Lilley, A. E. (1963). The microwave spectrum of oxygen in the Earth's atmosphere. *J. Geophys. Res.* **68,** 1683–1703.

Menzel, D. H. (1966). "Selected Papers on the Transfer of Radiation." Dover, New York.

Mie, G. (1908). Beigrade zur Optick trüber Medien, speziell kolloidaler Metallösungen. *Ann. Physik.* **25,** 377–445.

Milankovitch, M. (1941). "Kanon der Erdbestrahlung und seine Anwendung auf des Eiszeitproblem." Königlich Serbische Akademie, Belgrade. [English translation by the Israel Program for Scientific Translation, published by the U.S. Department of Commerce and the National Science Foundation.]

Minnis, P., Liou, K. N., and Takano, Y. (1993). Inference of cirrus cloud properties using satellite-observed visible and infrared radiances. Part I: Parameterization of radiance field. *J. Atmos. Sci.* **50,** 1279–1304.

Minnis, P., Young, D. F., Garber, D. P., Nguyen, L., Smith, W. L. Jr., and Palikonda, R. (1998). Transformation of contrails into cirrus during SUCCESS. *Geophys. Res. Lett.* **25,** 1157–1160.

Mishchenko, M. I. (1991). Extinction and polarization of transmitted light by partially aligned nonspherical grains. *Astrophys. J.* **367,** 561–574.

Mishchenko, M. I., Hovenier, J. W., and Travis, L. D., eds. (2000). "Light Scattering by Nonspherical Particles." Academic Press, San Diego.

Molina, L. T., and Molina, M. J. (1986). Absolute absorption cross section of ozone in the 185- to 350-nm wavelength range. *J. Geophys. Res.* **91,** 14501–14508.

Molina, M. J., and Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes: chlorine atom catalyzed destruction of ozone. *Nature* **249,** 810–812.

Möller, F. (1943). "Das Strahlungsdiagramm." Reichsamt für Wetterdienst (Luftwaffe), Berlin.

Mugnai, A., and Smith, E. A. (1988). Radiative transfer to space through a precipitating cloud at multiple microwave frequencies. Part I. Model description. *J. Appl. Meteorol.* **27,** 1055–1073.

National Research Council (1994). "Solar Influences on Global Change." National Academy Press, Washington, D.C.

National Research Council (1996). "Aerosol Radiative Forcing and Climate Change." National Academy Press, Washington, D.C.

Neftel, A., Moor, E., Oeschger, H., and Stauffer, B. (1985). Evidence from polar ice cores for the increase in atmospheric CO_2 in the past two centuries. *Nature* **315,** 45–47.

Newell, R. E., Kidson, J. W., Vincent, D. G., and Boer, G. J. (1974). "The General Circulation of the Tropical Atmosphere." Vol. 1, MIT, Cambridge, MA.

North, G. R. (1975). Theory of energy-balance climatic models. *J. Atmos. Sci.* **32,** 2033–2043.

North, G. R., Cahalan, R. F., and Coakley, J. A. (1981). Energy balance climate models. *Rev. Geophys. Space Phys.* **19,** 91–121.

Nussenzveig, H. M. (1979). Complex angular momentum theory of the rainbow and the glory. *J. Opt. Soc. Am.* **69,** 1068–1079.

Oort, A. H., and Péixoto, J. P. (1983). Global angular momentum and energy balance requirements from observations. *Adv. Geophys.* **25,** 355–490.

Ou, S. C., and Liou, K. N. (1982). Generalization of the spherical harmonic method to radiative transfer in multi-dimensional space. *J. Quant. Spectrosc. Radiat. Transfer* **28,** 271–288.

Ou, S. C., and Liou, K. N. (1995). Ice microphysics and climatic temperature perturbations. *Atmos. Res.* **35,** 127–138.

Ou, S. C., Liou, K. N., and Liou, W. J. (1989). The seasonal cycle of the global zonally averaged energy balance. *Theor. Appl. Climatol.* **40,** 9–23.

Ou, S. C., Liou, K. N., Takano, Y., Rao, N. X., Fu, Q., Heymsfield, A. J., Miloshevich, L. M., Baum, B., and Kinne, S. A. (1995). Remote sounding of cirrus cloud optical depths and ice crystal sizes from AVHRR data: Verification using FIRE II IFO measurements. *J. Atmos. Sci.* **52,** 4143–4158.

Ou, S. C., Liou, K. N., King, M. D., and Tsay, S. C. (1999). Remote sensing of cirrus cloud parameters based on a 0.63–3.7 μm radiance correlation technique applied to AVHRR data. *Geophys. Res. Lett.* **26,** 2437–2440.

Peebles, G. H., and Plesset, M. S. (1951). Transmission of gamma rays through large thicknesses of heavy materials. *Phys. Rev.* **81,** 430–439.

Peixóto, J. P., and Oort A. H. (1992). "Physics of Climate." American Institute of Physics, New York.

Penner, J. E., Chuang, C. C., and Grant, K. (1998). Climate forcing by carbonaceous and sulfate aerosols. *Climate Dyn.* **14,** 839–851.

Penner, S. S. (1959). "Quantitative Molecular Spectroscopy and Gas Emissivities." Addison-Wesley, Reading, MA.

Perrin, F. (1942). Polarization of light scattered by isotropic opalescent media. *J. Chem. Phys.* **10,** 415–426.

Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* **9,** 84–97.

Phillips, N. A. (1957). A coordinate system having some special advantages for numerical forecasting. *J. Meteorol.* **14,** 184–185.

Pinker, R. T., Laszlo, I., Whitlock, C. H., and Charlock, T. P. (1995). Radiative flux opens new window on climate research. *EOS* **76,** No. 15., April 11 issue.

Planck, M. (1901). Ueber das Gesetz der Energieverteilung im Normalspectrum. *Ann. Physik* **4,** 553–563.

Plass, G. N., and Kattawar, G. W. (1968). Monte Carlo calculations of light scattering from clouds. *Appl. Opt.* **7,** 415–419.

Platt, C. M., Abshire, N. L., and McNice, G. T. (1978). Some microphysical properties of an ice cloud from lidar observations of horizontally oriented crystals. *J. Appl. Meteorol.* **17,** 1220–1224.

Prabhakara, C., Kratz, D. P., Yoo, J. M., Dalu, G. and Vernekar, A. (1993). Optically thin cirrus clouds: Radiative impact on the warm pool. *J. Quant. Spectrosc. Radiat. Transfer* **49,** 467–483.

Pruppacher, H. R., and Pitter, R. L. (1971). A semi-empirical determination of the shape of cloud and raindrops. *J. Atmos. Sci.* **28,** 86–94.

Ramanathan, V. (1981). The role of ocean-atmosphere interactions in the CO_2 climate problem. *J. Atmos. Sci.* **38,** 918–930.

Ramanathan, V., Callis, L., Cess, R., Hansen, J., Isaksen, I., Kuhn, W., Lacis, A., Luther, F., Mahlman, J., Reck, R., and Schlesinger, M. (1987). Climate-chemical interactions and effects of changing atmospheric trace gases. *Rev. Geophys.* **25,** 1441–1482.

Randall, D. A., Harshvardhan, Dazlich, D. A., and Corsetti, T. G. (1989). Interactions among radiation, convection, and large-scale dynamics in a general circulation model. *J. Atmos. Sci.* **46,** 1943–1970.

Rayleigh, Lord (1871). On the light from the sky, its polarization and colour. *Phil. Mag.* **41,** 107–120, 274–279.

Reagan, J., Thomason, L., and Herman, B. (1986). Assessment of atmospheric limitations on the determination of the solar spectral constant from ground-based spectroradiometer measurements. *IEEE Trans. Geosci. Remote Sens.* **24,** 258–266.

Rigone, J. L., and Stogryn, A. P. (1977). Data processing for the DMSP Microwave Radiometer System. *11th Inter. Symp. Remote Sensing of the Environment,* Univ. of Michigan, pp. 1599–1608.

Rodgers, C. D., and Walshaw, C. D. (1966). The computation of infra-red cooling rate in planetary atmospheres. *Quart. J. Roy. Meteorol. Soc.* **92,** 67–92.

Rolland, P., and Liou, K. N. (1998). Remote sensing of optical and microphysical properties of cirrus clouds using MODIS channels. *Proceeding Cirrus Topical Meeting.* Opt. Soc. Am., Baltimore, pp. 17–19.

Rolland, P., Liou, K. N., King, M. D., Tsay, S. C., and McFarquhar, G. M. (2000). Remote sensing of optical and microphysical properties of cirrus clouds using Moderate-Resolution Imaging Spectroradiometer channels: Methodology and sensitivity to physical parameters. *J. Geophys. Res.* **105,** 11721–11738.

Rossow, W. B., and Schiffer, R. A. (1999). Advances in understanding clouds from ISCCP. *Bull. Am. Meteorol. Soc.* **80,** 2261–2288.

Rothman, L. S., Rinsland, C. P., Goldman, A., *et al.* (1998). The HITRAN molecular spectroscopic database and HAWKS (HITRAN Atmospheric Workstation): 1996 edition. *J. Quant. Spectrosc. Radiat. Transfer* **60,** 665–710.

Russell, P. B., Livingston, J. M., Dutton, E. G., *et al.* (1993). Pinatubo and pre-Pinatubo optical depth spectra: Mauna Loa measurements, comparisons, inferred particle size distributions, radiative effects, and relationship to lidar data. *J. Geophys. Res.* **98,** 22969–22985.

Sassen, K. (1976). Polarization diversity lidar returns from virga and precipitation: Anomalies and the bright band analogy. *J. Appl. Meteorol.* **15,** 292–300.

Sassen, K. (1991). The polarization lidar technique for cloud research: A review and current assessment. *Bull. Am. Meteorol. Soc.* **72,** 1848–1886.

Sassen, K. (1997). Contrail-cirrus and their potential for regional climate change. *Bull. Am. Meteorol. Soc.* **78,** 1885–1903.

Sassen, K., and Hsueh, C. (1998). Contrail properties derived from high-resolution polarization lidar studies during SUCCESS. *Geophys. Res. Lett.* **25,** 1165–1168.

Sassen, K., and Liou, K. N. (1979). Scattering of polarized laser light by water droplet, mixed-phase and ice crystal clouds. I. Angular scattering patterns. *J. Atmos. Sci.* **36,** 838–851.

Sato, M., Hansen, J. E., McCormick, M. P., and Pollack, J. B. (1993). Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98,** 22987–22994.

Schelkunoff, S. A. (1943). "Electromagnetic Theory." Van Nostrand, New York.

Schiffer, R. A., and Rossow, W. B. (1983). The international satellite cloud climatology project (ISCCP): The first project of the world climate research programme. *Bull. Am. Meteorol. Soc.* **64,** 779–784.

Schotland, R. M. (1969). Some aspects of remote atmospheric sensing by laser radar. *In* "Atmospheric Exploration by Remote Probes," pp. 179–200. Committee on Atmospheric Sciences, Nat. Acad. Sci–Nat. Res. Council, Washington, D.C.

Schrödinger, E. (1926). Quantiscerung als Eigenwertproblem. *Ann. Physik* **79,** 361–376, 489–527.

Schuster, A. (1905). Radiation through a foggy atmosphere. *Astrophys. J.* **21,** 1–22.

Schwarzschild, K. (1906). On the equilibrium of the sun's atmosphere. *Nach. K. Gesell, Wiss. Göttingen,* Math-Phys. Klasse **195**, pp. 41–53. *In* "Selected Papers on the Transfer of Radiation" (D. H. Menzel, ed.). Dover, New York.

Schwarzschild, K. (1914). Diffusion and absorption in the Sun's atmosphere. *Sitzungsberichte der Königlichen Preussichen Akademie der Wissenschaften,* pp. 1183–1200. In "Selected Papers on the Transfer of Radiation" (D. H. Menzel, ed.). Dover, New York.

Sellers, W. D. (1965). "Physical Climatology." Univ. of Chicago Press, Chicago.

Sellers, W. D. (1969). A global climatic model based on the energy balance of the earth–atmosphere system. *J. Appl. Meteorol.* **8,** 392–400.

Senior, C. A., and Mitchell, J. F. (1993). Carbon dioxide and climate: The impact of cloud parameterization. *J. Climate* **6,** 393–418.

Smith, W. L. (1970). Iterative solution of the radiative transfer equation for the temperature and absorbing gas profile of an atmosphere. *Appl. Opt.* **9,** 1993–1999.

Smith, W. L. (1991). Atmospheric soundings from satellites-false expectation or the key to improved weather prediction. *Quart. J. Roy. Meteorol. Soc.* **117,** 267–297.

Smith, W. L., and Platt, C. M. (1978). Comparison of satellite-deduced cloud heights with indications from radiosonde and ground-based laser measurements. *J. Appl. Meteorol.* **17,** 1796–1802.

Smith, W. L., Revercomb, H. L., Knuteson, R. O., Best, F. A., Dedecker, R., Howell, H. B., and Woolf, H. M. (1995). Cirrus cloud properties derived from high spectral resolution infrared spectrometer during FIRE II. Part I: The High resolution Interferometer Sounder (HIS) Systems. *J. Atmos. Sci.* **52,** 4238–4245.

Smith, W. L., Ackerman, S., Revercomb, H., Huang, H., DeSlover, D. H., Feltz, W., Gumley, L., and Collard, A. (1998). Infrared spectral absorption of nearly invisible cirrus clouds. *Geophys. Res. Lett.* **25,** 1137–1140.

Sobolev, V. V. (1975). "Light Scattering in Planetary Atmospheres." Translated by W. M. Irvine. Pergamon Press, Oxford.

Solomon, S. (1999). Stratospheric ozone depletion: A review of concepts and history. *Rev. Geophys.* **37,** 275–316.

Solomon, S., Portmann, R. W., Sanders, R. W., and Daniel, J. S. (1998). Absorption of solar radiation by water vapor, oxygen, and related collision pairs in the Earth's atmosphere. *J. Geophys. Res.* **103,** 3847–3858.

Somerville, R., and Remer, L. (1984). Cloud optical thickness feedbacks in the CO_2 climate problem. *J. Geophys. Res.* **89,** 9668–9672.

Spänkuch, D., and Döhler, W. (1985). Radiative properties of cirrus clouds in the middle ir derived from Fourier spectrometer measurements from space. *Z. Meteorol.* **6,** 314–324.

Spencer, R. W., Goodman, H. M., and Hood, R. E. (1989). Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. *J. Atmos. Ocean. Tech.* **6,** 254–273.

Staelin, D. H., Kunzi, K. F., Pettyjohn, R. L., Poon, R. K. L., and Wilcox, R. W. (1976). Remote sensing of atmospheric water vapor and liquid water with the Nimbus 5 microwave spectrometer. *J. Appl. Meteorol.* **15,** 1204–1214.

Staley, D. O., and Jurica, G. M. (1970). Flux emissivity tables for water vapor, carbon dioxide and ozone. *J. Appl. Meteorol.* **9,** 365–372.

Staley, D. O., and Jurica, G. M. (1972). Effective atmospheric emissivity under clear skies. *J. Appl. Meteorol.* **11,** 349–356.

Stamnes, K., and Dale, H. (1981). A new look at the discrete-ordinate method for radiative transfer calculations in anisotropically scattering atmospheres. II. Intensity computations. *J. Atmos. Sci.* **38,** 2696–2706.

Stephens, G. L. (1994). "Remote Sensing of the Lower Atmosphere." Oxford Univ. Press, New York.

Stephens, G. L., Campbell, G. G., and Vonder Haar, T. H. (1981). Earth radiation budgets. *J. Geophys. Res.* **86,** 9739–9760.

Stokes, G. G. (1852). On the composition and resolution of streams of polarized light from different sources. *Trans. Cambridge Philos. Soc.* **9,** 399–423.

Stokes, G. G. (1862). On the intensity of the light reflected from or transmitted through a pile of plates. *Proc. Roy. Soc. London* **11,** 545–556.

Stratton, J. A. (1941). "Electromagnetic Theory." McGraw-Hill, New York.

Study of Man's Impact on Climate (1971). "Inadvertent Climate Modification." MIT Press, Boston.

Suarez, M. J., and Held, I. M. (1979). The sensitivity of an energy balance climate model to variations in the orbital parameters. *J. Geophys. Res.* **84,** 4825–4836.

Suomi, V. E. (1958). The radiation balance of the earth from a satellite. *Ann. Int. Geophys. Year* **6,** 331–340.

Taflove, A. (1995). "Computational Electrodynamics: The Finite-Difference Time Domain Method." Artech House, Boston.

Takano, Y., and Asano, S. (1983). Fraunhofer diffraction by ice crystals suspended in the atmosphere. *J. Meteorol. Soc. Jpn.* **61,** 289–300.

Takano, Y., and Liou, K. N. (1989a). Solar radiative transfer in cirrus clouds. I. Single-scattering and optical properties of hexagonal ice crystals. *J. Atmos. Sci.* **46,** 3–19.

Takano, Y., and Liou, K. N. (1989b). Solar radiative transfer in cirrus clouds. II. Theory and computation of multiple scattering in an anisotropic medium. *J. Atmos. Sci.* **46,** 20–36.

Takano, Y., and Liou, K. N. (1993). Transfer of polarized infrared radiation in optically anisotropic media: Application to horizontally oriented ice crystals. *J. Opt. Soc. Am. A* **10,** 1243–1256.

Takano, Y., and Liou, K. N. (1995). Radiative transfer in cirrus clouds. III. Light scattering by irregular ice crystals. *J. Atmos. Sci.* **52,** 818–837.

Thekaekara, M. P. (1976). Solar irradiance: Total and spectral and its possible variations. *Appl. Opt.* **15,** 915–920.

Thomas, G. E., and Stamnes, K. (1999). "Radiative Transfer in the Atmosphere and Ocean." Cambridge Univ. Press, Cambridge, U.K.

Toon, O. B., and Turco, R. (1991). Polar stratospheric clouds and ozone depletion. *Sci. Am.* **264,** 68–74.

Toon, O. B., McKay, C. P., and Ackerman, T. P. (1989). Rapid calculation of radiative heating rates and photodissociation rates in inhomogeneous multiple scattering atmospheres. *J. Geophys. Res.* **94,** 16287–16301.

Trenberth, K. E., ed. (1992). "Climate System Modeling." Cambridge Univ. Press, Cambridge, U.K.

Turco, R. P., Toon, O. B., Ackerman, T. P., Pollack, J. B., and Sagan, C. (1983). Nuclear winter: Global consequences of multiple nuclear explosions. *Science* **222,** 1283–1292.

Twomey, S. (1963). On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature. *J. Assoc. Comput. Mach.* **10,** 97–101.

Twomey, S. (1977). "Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements." Elsevier, New York.

Twomey, S., and Seton, K. J. (1980). Inference of gross microphysical properties of clouds from spectral reflectance measurements. *J. Atmos. Sci.* **37,** 1065–1069.

Twomey, S. A., Piepgrass, M., and Wolfe, T. L. (1984). An assessment of the impact of pollution on global cloud albedo. *Tellus* **36B,** 356–366.

U.S. Standard Atmosphere (1976). NOAA-S/T76-1562. Supt. of Documents, U.S. Govt. Printing Office, Washington, D.C.

van de Hulst, H. C. (1945). Theory of absorption lines in the atmosphere of the earth. *Ann. Astrophys.* **8,** 1–10.

van de Hulst, H. C. (1957). "Light Scattering by Small Particles." Wiley, New York.

van de Hulst, H. C. (1980). "Multiple Light Scattering. Tables, Formulas, and Applications," Vols. 1 and 2. Academic Press, New York.

Vigroux, E. (1953). Contribution à l'étude expérimentale de l'absorption de l'ozone. *Ann. Phys.* **8,** 709–762.

Volkovitsky, O. A., Pavlova, L. N., and Petrushin, A. G. (1980). Scattering of light by ice crystals. *Izv. Acad. Sci. USSR, Atmos. Oceanic Phys.* (Engl. Transl.), **16,** 98–102.

Volten, H., Jalava, J., Lumme, K., de Haan, J., Vassen, V., and Hovenier, J. W. (1999). Laboratory measurements and T-matrix calculations of the scattering matrix of rutile particles in water. *Appl. Opt.* **38,** 5232–5240.

Vonder Haar, T. H., and Suomi, V. E. (1971). Measurement of the earth's radiation budget from satellites during a five-year period. *J. Atmos. Sci.* **28,** 305–314.

Walshaw, C. D., and Rogers, C. D. (1963). The effect of the Curtis-Godson approximation on the accuracy of radiative heating-rate calculations. *Quart. J. Roy. Meteorol. Soc.* **89,** 122–130.

Wandinger, U. (1998). Multiple-scattering influence on extinction- and backscatter-coefficient measurements with Raman and high-spectral-resolution lidars. *Appl. Opt.* **37,** 417–427.

Wang, P. (1982). Mathematical description of the shape of conical hydrometeors. *J. Atmos. Sci.* **39,** 2615–2622.

Wang, W. C., Wuebbles, D. J., Washington, W. M., Issacs, R. G., and Molnar, G. (1986). Trace gases and other potential perturbations to global climate. *Rev. Geophys.* **24,** 110–140.

Waterman, P. C. (1971). Symmetry, unitarity, and geometry in electromagnetic scattering. *Phys. Rev. D* **3,** 825–839.

Waters, J. W., Kunzi, K. F., Pettyjohn, R. L., Poon, R. K. L., and Staelin, D. H. (1975). Remote sensing of atmospheric temperature profiles with the Nimbus 5 microwave spectrometer. *J. Atmos. Sci.* **32,** 1953–1969.

Watson, G. N. (1944). "A Treatise on the Theory of Bessel Functions," 2nd ed., Cambridge Univ. Press, London and New York.

Weickmann, H. K. (1948). Die Eisphase in der Atmosphere. Royal Aircraft Establishment, 96 pp.

Weinman, J. (1976). Effects of multiple scattering on light pulses reflected by turbid atmospheres. *J. Atmos. Sci.* **33,** 1763–1771.

Wendling, P., Wendling, R., and Weickmann, H. K. (1979). Scattering of solar radiation by hexagonal ice crystals. *Appl. Opt.* **18,** 2663–2671.

Wetherald, R. T., and Manabe, S. (1988). Cloud feedback processes in a general circulation model. *J. Atmos. Sci.* **45,** 1397–1415.

Whittaker, E. T., and Watson, G. N. (1940). "Modern Analysis." 4th ed., Cambridge Univ. Press, London and New York.

Wilheit, T. T., and Chang, A. T. (1977). A satellite technique for quantitatively mapping rainfall rates over the oceans. *J. Appl. Meteorol.* **16,** 551–559.

Wiscombe, W. J. (1980). Improved Mie Scattering Algorithms. *Appl. Opt.* **19,** 1505–1509.

Woodbury, G. E., and McCormick, M. P. (1986). Zonal and geographyxical distributions of cirrus clouds determined from SAGE data. *J. Geophys. Res.* **91,** 2775–2785.

Wylie, D. P., Menzel, W. P., Woolf, H. M., and Strabala, K. I. (1994). Four years of global cirrus cloud statistics using HIRS. *J. Climate* **7,** 1972–1986.

Yamamoto, G. (1952). On a radiation chart. *Sci. Rep. Tohoku Univ. Ser. 5 Geophys.* **4,** 9–23.

Yang, P., and Liou, K. N. (1995). Light scattering by hexagonal ice crystals: Comparison of finite-difference time domain and geometric optics models. *J. Opt. Soc. Am. A* **12,** 162–176.

Yang P., and Liou, K. N. (1996a). Geometric-optics-integral-equation method for light scattering by nonspherical ice crystals. *Appl. Opt.* **35,** 6568–6584.

Yang, P., and Liou, K. N. (1996b). Finite-difference time domain method for light scattering by small ice crystals in three-dimensional space. *J. Opt. Soc. Am. A* **13,** 2072–2085.

Yang, P., and Liou, K. N. (2000). Finite-difference time domain method for light scattering by nonspherical and inhomogeneous particles. *In* "Light scattering by Nonspherical Particles." (M. I. Mishchenko, J. W. Hovenier, and L. D. Travis, eds.), pp. 173–221. Academic Press, San Diego.

Yee, S. K. (1966). Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antennas Propag.* **14,** 302–307.

Yu, J. Y., and Mechoso, C. R. (1999). Links between annual variation of Peruvian stratocumulus clouds and of SST in the eastern equatorial Pacific. *J. Climate* **12,** 3305–3318.

Zerull, R. H. (1976). Scattering measurements of dielectric and absorbing nonspherical particles. *Beitr. Phys. Atmos.* **49,** 168–188.

This Page Intentionally Left Blank

Index

This Page Intentionally Left Blank

International Geophysics Series

EDITED BY

RENATA DMOWSKA
Division of Engineering and Applied Sciences
Harvard University
Cambridge, Massachusetts

JAMES R. HOLTON
Department of Atmospheric Sciences
University of Washington
Seattle, Washington

H. THOMAS ROSSBY
Graduate School of Oceanography
University of Rhode Island
Kingston, Rhode Island

*Out of Print.

*Out of Print.

*Out of Print.