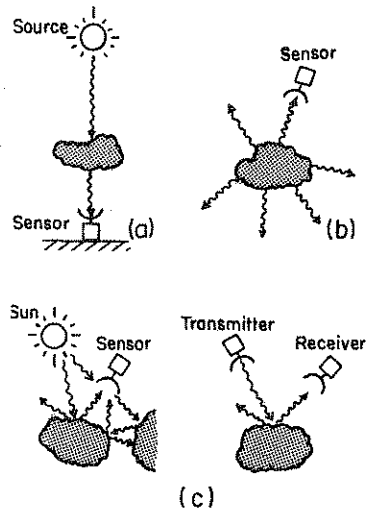

Figure 1.1 Three basic types of radiative properties used in electromagnetic remote sensing topics discussed in this book: (a) extinction, (b) emission, (c) scattering of either an incoherent or coherent source.

sensing in a number of different ways. For example, *passive* sensing relies on measuring natural levels of radiation, such as the radiation emitted from the Earth's surface or sunlight reflected by the atmosphere. On the other hand, *active* sensing makes use of electromagnetic energy transmitted from a specially chosen source to some "object" or "target," and then monitoring the interactions between this object and the radiation. The target may be a naturally occurring fluctuation in the atmosphere or a fixed target at a set distance from the transmitter.

Emphasis is placed on understanding the interactions between electromagnetic radiation and the atmosphere; these interactions provide a convenient way of categorizing the topics discussed in this book. The remote sensing topics are discussed in terms of

- extinction methods
- emission methods
- scattering methods

according to the different types of radiative processes (extinction,

scattering, absorption–emission) that provide the basis for the approach under consideration. Figure 1.1 is a schematic view of these processes as they might be used in certain remote sensing applications. In the first case, the radiation from a given, known source is observed, and the amount of radiation attenuated or lost from the beam as it is transmitted to the sensor is used to determine the amount of matter along the path. This loss can either occur through absorption or scattering, or from a combination of both. A large number of remote sensing methods actually rely on this kind of approach (see Chapters 6 and 7). Radiation can also be emitted by the object and the second category deals with measurements of infrared and microwave emission which are used to obtain information about the thermal structure of the atmosphere and other properties (Chapter 7). The third category of remote sensing utilizes the properties of scattering.[1] In the example shown in Fig. 1.1c, the medium is illuminated either by a source of incoherent radiation, such as sunlight or infrared radiation emitted from the Earth's surface or by an artificial coherent source, like that transmitted by a laser source (as in the case of lidar – light detection and ranging) or a source of microwave radiation (as in the case of radar – radio detection and ranging). This radiation is then scattered by an object and received by a sensor optimally tuned to detect the signal.

1.2 The Nature of Inverse Problems

Problems of remote sensing fall into a category referred to as *inverse problems*. The nature of these inverse problems is portrayed in Fig. 1.2 in the form of an analogy and can be posed in the following way. How well can we infer the dragon from observations of its tracks? Clearly this line of inquiry is fruitless without some a priori knowledge of dragons and the types of tracks they leave. Therefore a central theme of the first part of this book is the understanding of how electromagnetic radiation interacts with matter since these are the tracks we exploit in remote sensing applications to infer information about the atmosphere.

[1] Scattering means elastic scattering, for which the frequency of light is unaltered. Another class of scattering is also relevant to remote sensing: Inelastic scattering results in a shift of the frequency of the incident light and the desired properties of the atmosphere are related to this frequency shift. Examples of these methods are Raman and fluorescent scattering and neither are dealt with in this book.

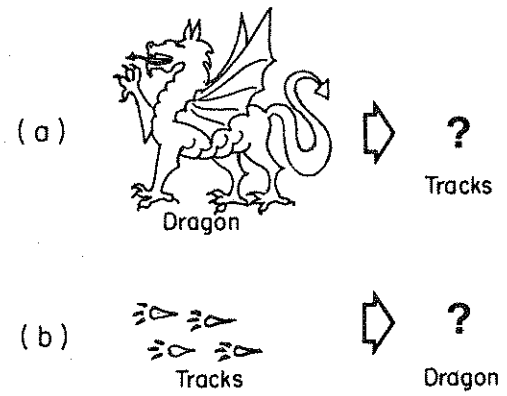

Figure 1.2 (a) The direct problem: Describe the tracks of a dragon. (b) The inverse problem: Describe a dragon from its tracks (from Bohren and Huffman, 1983.)

The theory of inverse problems is referred to as *retrieval or inversion theory*; atmospheric remote sensing is only one of a number of disciplines for which retrieval theory is an important topic. Detailed discussion of the general inversion problem and methods of solution by interactive, nonlinear and statistical approaches is considered beyond the scope of this book and the reader is referred to an introductory treatment of this subject provided by Twomey (1977) and others cited later in the notes.

Before leaving inversion theory, a simple example relevant to later topics will now be described. Suppose that an absorbing gas in the atmosphere emits radiation to a satellite in a way that depends on its temperature (which is true, as we see later in Chapter 2). Suppose also that the wavelength of the emission is related to a single level (this is not true, as we learn later) so that we obtain a relationship between wavelength and temperature as sketched in Fig. 1.3. Knowing how the intensity of radiation, the wavelength of the radiation, and the temperature of the emitting gas are related provides us with a way of estimating the temperature as a function of wavelength and thus altitude.

In reality, the actual situation is unfortunately more complex than the ideal one just described. The radiation at one wavelength

does not come from a single height but is distributed over a broad layer of the atmosphere. This results in a blurring or a departure from a one–to–one correspondence between wavelength and height as supposed for the ideal case illustrated in Fig. 1.3b.

This example is typical of many of the inversion problems in remote sensing where a set of measurements (in this case emission as a function of wavelength) is influenced by all values of the unknown distribution (i.e., the vertical distribution of temperature throughout the layer that contributes to the emission). We represent this blurring effect in the following way. Let the sought–after distribution be $f(x)$ (which may be the temperature profile as a function of altitude z), and let $K_i(x)$ be the relative contribution curve for a wavelength λ_i. The interval between x and $x + \Delta x$ contributes to the measurement of the ith channel of a radiometer the amount $f(x)K_i(x)\Delta x$. The total measured radiation in this channel is

$$g_i = \int_a^b K_i(x) f(x) dx \tag{1.1}$$

where the limits of the integral depend on the details of the problem at hand. This equation is known as the *Fredholm integral equation of the first kind* because the limits of the integral are fixed and because $f(x)$ appears only in the integrand. The function $K_i(x)$ is known as the "kernel" or "kernel function."

We will see in later chapters how different remote sensing problems reduce to (1.1) or a form similar to (1.1) and its solution requires inversion to obtain the distribution $f(x)$. This is an unfortunate circumstance since the solutions to (1.1) suffer a number of difficulties, including non–existence, non–uniqueness, and instability. Non–existence is usually not an issue for most practical problems since we measure g_i and $f(x)$ exists through physical considerations. Non–uniqueness is related to the blurring effect discussed earlier. The practical consequence of this is that there exist several functions $f(x)$ which produce the same function g_i. This problem tends to be overcome to a certain extent by restricting the class of admissible solutions to physically realizable ones. In this way, a priori information is introduced into the retrieval scheme.

The major difficulty in solving (1.1), at least for most practical problems, concerns the problem of instability which arises, for example, from errors in the observations g_i. With a small error ϵ_i in

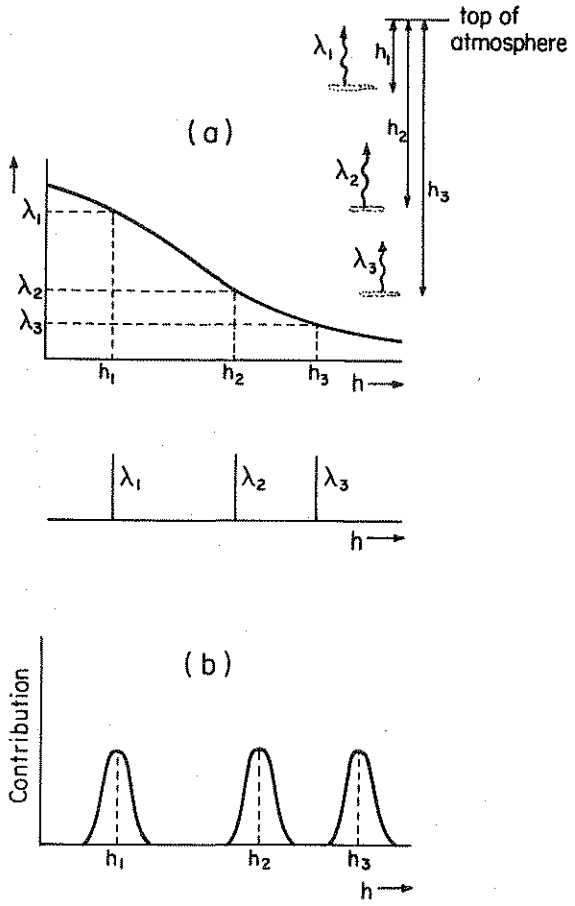

Figure 1.3 (a) Schematic diagram for a satellite-based atmospheric temperature profile measurement. The contribution to the radiation measured at the selected wavelength arises from a discrete level. (b) A version of the contribution function of Fig. 1.3a corresponding to a more practical case where the measured radiation originates from a range of different levels in unequal proportions (from Twomey, 1977).

g_i, (1.1) becomes

$$g_i + \epsilon_i = \int_a^b K_i(x) f(x) dx \tag{1.2}$$

where ϵ_i produces an arbitrarily large change in $f(x)$ and the ultimate success of any retrieval largely depends on the accuracy of the measurement g_i and on the shape of $K_i(x)$. A simple example that underlines this important point will now be presented.

Excursus: The Ill-Posed Nature of the Inversion Problem

Let us illustrate the nature of the instability using the problem of temperature retrieval. We can write the transfer equation of infrared radiation in the form of (1.1) (this equation and its interpretation is given in Chapter 7). The integral term of (7.32) can be expressed as

$$I_i = \int_{z_s}^{z_o} \mathcal{B}_i(z') \mathcal{W}_i(z', \infty) dz' \tag{1.3}$$

where I_i is the measured intensity of the radiation, $\mathcal{W}_i$ is the kernel function (for this particular problem we call this the weighting function, and a more complete discussion of this function is given in Chapter 7), and $\mathcal{B}_i$ is the distribution function containing temperature information. We use the subscript i to refer to the measurement of a particular satellite radiometer channel and assume there are $i = 1, \ldots, M$ spectral measurements.

Consider a simple example of the solution to (1.3). Suppose we have only two observations (i.e., $M = 2$), then we might discretize the integral in (1.3) as

$$\begin{aligned} \mathcal{B}_1 \mathcal{W}_{1,1} + \mathcal{B}_2 \mathcal{W}_{1,2} &= I_1 \\ \mathcal{B}_1 \mathcal{W}_{2,1} + \mathcal{B}_2 \mathcal{W}_{2,2} &= I_2 \end{aligned} \tag{1.4}$$

where $\mathcal{W}_{i,j}$ is the weighting function of the jth layer for the ith channel. Let us suppose for the sake of illustration that the weighting functions have the numerical values $\mathcal{W}_{1,1} = 1, \mathcal{W}_{1,2} = 1, \mathcal{W}_{2,1} = 2$, and $\mathcal{W}_{2,2} = 2.000001$, and that the intensities are $I_1 = 2$ and $I_2 = 4.000001$. We then obtain

$$\mathcal{B}_1 = 1$$

$$\mathcal{B}_2 = 1$$

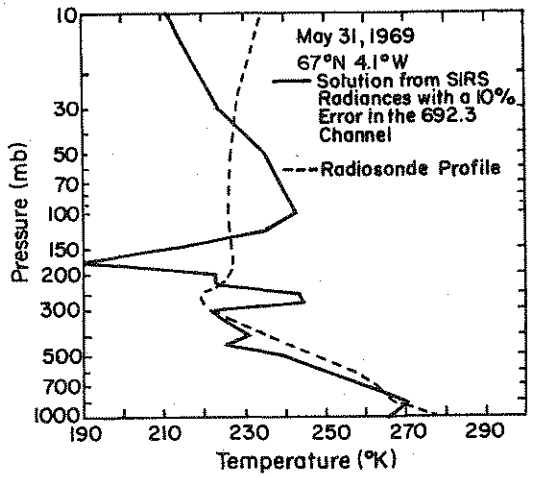

Figure 1.4 Comparison of radiosonde and a retrieved temperature profile when a spurious error of 10% is added to one channel (from Smith, 1972).

as a solution to (1.4). Suppose that there is a small uncertainty in one of the measured intensities, I_2, such that the value $I_2 = 4$ is recorded instead of the value 4.000001. Then the solution to (1.4) is

$$\mathcal{B}_1 = 2$$
$$\mathcal{B}_2 = 0$$

This is a dramatic change to the solution and nicely illustrates the problem of instability. In reality, matters become even worse as the number of measurements (i.e., as M) increases. Figure 1.4 provides an example of how random errors propagate in an actual temperature retrieval. The retrieved temperature profile obtained using infrared measurements from a radiometer when a spurious error of 10% is added to a single instrument channel is shown in this diagram. This point is explored further in one of the projects introduced in Appendix 2.

To summarize, it is important to remember for later discussion that the difficulties associated with inversion theory can be reduced if:

1. We make instruments as accurate (noise–free) as possible.
2. We select channels (i.e., wavelengths) so that the kernel functions are as sharp as possible.
3. We develop methods of solution that are stable (i.e., provide minimum distortion of the true solution) in the presence of unavoidable measurement noise.

1.3 The Chemical Composition of the Atmosphere

Table 1.1 lists the relative abundance of various species in the Earth's atmosphere. The concentrations are given as *mixing ratios* by volume, the unit commonly used by atmospheric scientists, and is identical to the chemists' *mole fraction*. In the table, the mixing ratios are given as fractions, but parts per million (p.p.m or p.p.m.v) are commonly used for minor constituents. The quoted mixing ratios are actually averages for the lower atmosphere. Figure 1.5 shows representative vertical profiles of the species listed in Table 1.1, for average, midlatitude conditions.

Concentrations of atmospheric gases are subject to chemical and photochemical alteration. Molecular oxygen, for example, decomposes into atoms above 90 km. Methane and nitrous oxide are unstable in the stratosphere. Relevant to remote sensing topics addressed later is the fact that both molecular oxygen and carbon dioxide are uniformly mixed below about 100 km.

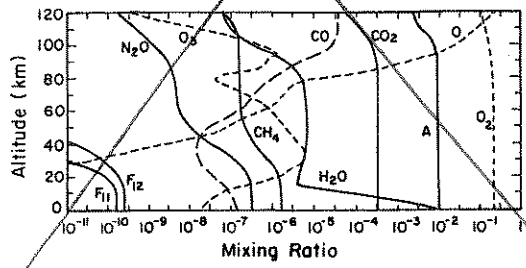

Figure 1.5 Vertical profiles of mixing ratios of selected species in the atmosphere (from Goody and Yung, 1989).

mosphere, what is the temperature of the surface if the observed intensity is 0.98×10^4 ergs cm^{-2} μm^{-1} sr^{-1}?

2.12. What is the ratio of the spectral radiances of black bodies at 300 K and 6000 K at (a) 1 GHz, (b) 1000 GHz, (c) 1 μm, and (d) 0.1 μm?

2.13. Show that, for a black body, the wavelength at which $\mathcal{B}_\nu$ is maximum is about 1.76 times greater than the wavelength at which $\mathcal{B}_\lambda$ is maximum at the same temperature.

2.14. Derive the Rayleigh–Jeans distribution (2.49) from (2.47).

2.15. Find the wavelength at which the incoming solar irradiance at the top of the Earth's atmosphere is equal to the outgoing terrestrial irradiance. Assume the sun and Earth to be emitting as blackbodies at 6000 K and 255 K, respectively. The radius of the Earth and the Earth–sun distance are given in Problem 6.3.

3

Microscopic Interactions — Atomic and Molecular Absorption

Our ability to interpret measurements and subsequently infer information about the atmosphere from them could not be possible without some knowledge of the way that electromagnetic radiation interacts with matter and, in turn, not possible without some understanding of the fundamental properties of matter itself. It is certainly beyond the scope of this book to provide a thorough discussion of such a basic topic. The focus of this and the next two chapters concerns the properties of matter that are most relevant in its interaction with radiation. The specific topics discussed in this chapter apply to the microscopic scale (i.e., how radiation interacts with matter on the atomic and molecular level) and primarily concentrate on the absorption of radiation by gases in the Earth's atmosphere.

3.1 The Atomic Absorption Spectrum

3.1.1 The Bright Line Spectrum

The concept of a blackbody was introduced in Chapter 2. It was shown how the emission by a solid body produces a spectrum that is continuous in wavelength and, temperature aside, independent of the nature of matter. However, this is hardly the complete story. Almost two centuries before Planck's description of blackbody radiation, it was noted how vaporization of certain volatile metals produces distinct colors in flames and how these colors clearly relate to the type of metals being vaporized. In 1752 Thomas Melvill studied the color of the flame after its light was passed through a prism and discovered that the spectrum was not continuous like that of the sun or of the radiation emitted from cavities; rather, it existed as a series of distinct bright lines, the spectral locations of which varied according to the particular substance placed in the flame. This work, along with an instrument developed in the early part of the nineteenth century (the *spectroscope*) to observe the spectrum, heralded the modern science of *spectroscopy*. Analyzed in the spectroscope, the spectra of

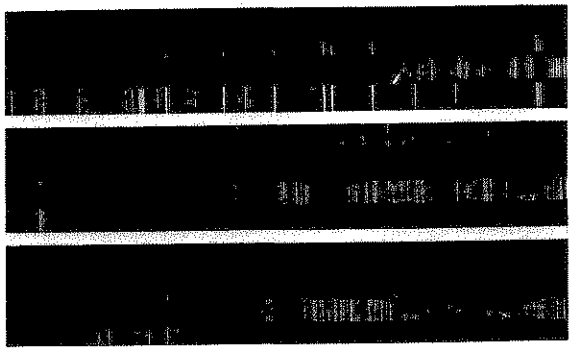

Figure 3.1 The bright line spectrum of the vaporized element iron is shown with the spectrum of the sun. The wavelength region is from 300 to 330 nm, in the ultraviolet. The solar spectrum is in the center of each strip, and the emission spectrum of iron lies above and below it. The bright lines of iron occur at the same wavelengths as some of the dark lines in the solar spectrum (Sobel, 1987).

flames (and electric discharges in gases) produce a line spectrum unique to the particular chemical element being vaporized.

Explanation of the line spectra phenomenon posed a challenge to theorists perhaps as significant as that posed by blackbody radiation. In some ways the challenge was even greater because, unlike the continuous emission from hot solid objects, the line spectrum was directly dependent on the nature of the object being observed. What transpired in the earlier part of this century with the experiments of Rutherford and the insights of Bohr was the development of modern quantum theory and a revolution that shook the foundations of both physics and scientific philosophy thereafter.

Bohr provided us with an early explanation of line spectra like that shown in Fig. 3.1. He proposed a planetary model of the atoms in which an electron moved around the nucleus in discrete orbits, with the position of the orbit relative to the nucleus representing discrete or quantized energy levels. As long as the electron remains in a fixed energy state it can emit no radiation since its energy does not change. The electron, however, can spontaneously fall from one level to a lower level losing energy in the form of radiation. A photon

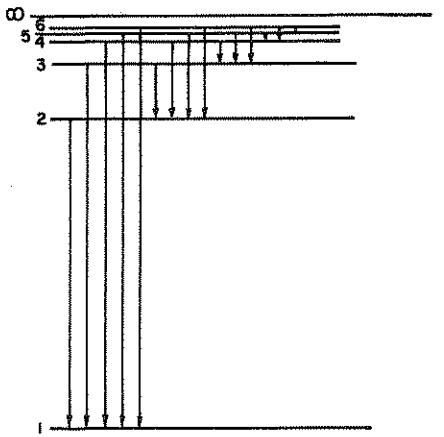

Figure 3.2 Energy levels of a hydrogen atom according to the Bohr theory. The first six levels are shown and drawn to scale in the sense that distance between levels is proportional to energy. Level 1 is the atom's ground state. Light is emitted whenever an atom makes a transition from a higher state to a lower one and the frequency of light is proportional to the energy difference. Higher levels are closer together. The line labeled by (∞) is the energy the electron would have to be barely able to escape the nucleus.

of light is therefore created with a frequency that is determined by the change in energy state of the atom,

$$h\nu = E_2 - E_1 = \Delta E \tag{3.1}$$

where E_2 and E_1 are the upper and lower energy states. Thus isolated atoms emit light at only certain definite frequencies as observed in the line spectra of Fig. 3.1. Moreover, the energy loss occurs by transitions to several intermediate levels such as that shown schematically in Fig. 3.2 for the simple hydrogen atom. The result is a spectrum of several lines, many of which are separated by only small frequency shifts.

To understand how these lines occur, we need to reconsider the Bohr model of an atom. Despite the success of the Bohr theory, it could not explain a number of fundamental issues underlying the

quantum theory of Planck and Einstein. For example, why are states of oscillators and atoms quantized? How can this quantum or photon possess both particle and wavelike properties? We learn from quantum theory that the position of an electron cannot be precisely located at any particular point in space. As a result, the position of an electron must be treated as some sort of probability of existing at a particular point. It is here that the concept of a wave emerges as some kind of distribution in space, whereas a particle is thought of as a fixed point. Since it is fundamentally impossible to observe the position of the electron, the Bohr concept of an electronic orbit has no physical meaning. [1] The new theory replaces an orbit with a probability distribution, the highest probability being concentrated in the region of the Bohr orbit.

Our inherent inability to describe the motions of electrons in a classical way is stated in terms of the *uncertainty principle* of Heisenberg

$$\Delta E \approx \frac{\hbar}{\Delta t} \tag{3.2}$$

where $\hbar = h/2\pi$. This formula expresses the impossibility of determining both the energy states of an atom or molecule and the lifetime of these states with unlimited accuracy. According to (3.1), this limitation also means that the spectral lines are not sharply defined but are smeared over a finite frequency interval,

$$\Delta\nu = \frac{\Delta E}{h} \tag{3.3}$$

which prescribes the minimum possible line width. This broadening about the position of the line is referred to as *natural broadening.*

3.1.2 The Absorption Line Spectrum

One way to excite an atom to a higher energy level is to use light as a source of energy. Let the frequency of the light be such that the quantum energy $h\nu$ precisely equals the difference in energy between the ground state and some higher state of the atom. The quantum is absorbed and an electron jumps to a higher state; this is the reverse process of atomic emission. If the original light source has

[1] There is another inconsistency with the circularly orbiting electron model of Bohr. This motion requires an accelerating electron which, by definition, must radiate and fall to a lower orbit.

a continuous distribution of frequencies, like the radiation from the sun, then only those frequencies that correspond to transitions of that atom are absorbed. A spectrum of dark lines is produced on viewing the light source through a cell containing a gas of atoms and the positions of these lines correspond to the absorbed frequencies. We refer to such a spectrum as the ***line absorption spectrum.***

The key to our understanding of the underlying physics of absorption lines of an atom came when it was discovered that the dark absorption lines occur at the same frequencies as the bright line spectra of a glowing gas composed of the same atoms (as shown in the example of Fig. 3.1). In Bohr's quantized atom, we found a simple explanation of both bright line spectra and absorption spectra. The energy levels are characteristic of each element. In an excited gas, such as in a discharge tube, the atoms are in higher levels and fall to lower levels, emitting light. On illumination, the same kind of atoms absorb light undergoing transitions from lower to higher levels.

Absorption and emission spectra are of profound importance to remote sensing. These absorption lines are a form of spectral fingerprint which can be used to identify the composition of the atmosphere of distant stars and planets. For instance, we can determine that the sun contains vapors of sodium, calcium, and iron, among other elements. Using the same techniques to identify molecules, we learn that the atmosphere of Mars is predominantly composed of carbon dioxide and further that the atmosphere of Titan has significant concentrations of methane.

3.2 Molecular Absorption Spectra

The absorption spectrum of a molecule is substantially more complex than is that of an atom. Both transitions between the energy states of the atoms that make up the molecule and transitions between energy states associated with movements of the atoms themselves are possible.

Since the energy required to produce a transition from a lower to a higher state is inversely proportional to the wavelength of the photon, the types of mechanisms that induce absorption also depend on the wavelength of the absorbed photon. These mechanisms must induce either a magnetic or an electric effect that can be influenced by electromagnetic radiation. Mechanisms responding fastest occur at the shortest wavelengths whereas the more sluggish mechanisms produce absorption at longer wavelengths. We can use this wave-

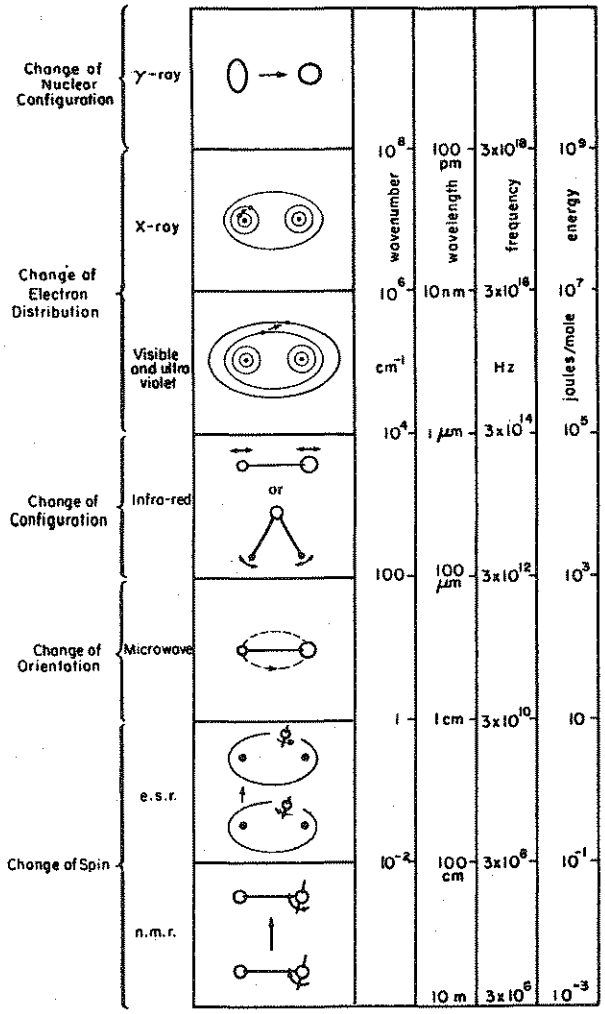

Figure 3.3 The electromagnetic spectrum and the possible types of interactions between photons and a molecule or atom (Banwell, 1983).

length dependence as a convenient classification of the absorption mechanisms as shown in Fig. 3.3 although the dividing boundaries are by no means precise.

- In the radio frequency regime, the absorption is associated with the nucleons and electrons which we consider to be tiny charged particles that spin producing tiny magnetic dipoles. The reversal of this dipole due to spin reversal interacts with the magnetic field at frequencies in the range 3×10^6 to 3×10^{10} Hz.
- In the visible and ultraviolet region excitation of valence electrons results in moving electric charges in the molecule. Changes in the electric dipole give rise to a spectrum by its interaction with the oscillating electric field of radiation. These electronic transitions occur within the individual atoms of molecules and dominate the visible and ultraviolet portions of the electromagnetic spectrum. At even shorter wavelengths, photons can actually disrupt the absorbing molecule by *photodissociation* or even produce *photo-ionization* of individual atoms.
- Absorption by molecules in the mid- and near infrared occur by vibration (although a mixture of vibrations and rotations are usually induced at these frequencies). Induction of vibrations requires more energy than rotations and thus takes place at higher frequencies corresponding to infrared wavelengths between about 0.7 μm and about 20 μm.
- In the microwave and far infrared, the molecule undergoes a rotation like that depicted in Fig. 3.4a and the component of the dipole in a given direction fluctuates in a regular fashion as shown in the lower part of Fig. 3.4a. These fluctuations are more sluggish than are the fluctuations associated with vibrations or the fluctuations associated with electronic transitions. Rotational lines generally occur in bands at the longer infrared wavelengths beyond about 20 μm extending into the microwave spectral region where individual rotational lines can be resolved.

As a consequence of the vibrational–rotational transitions, absorption lines are spread into bands containing many lines (as illustrated in Fig. 3.4b) that are used, either individually or as a group, to fingerprint molecules in the same way that atomic spectral lines fingerprint atoms. It is the vibrational–rotational absorption spectrum of molecules that is largely of interest to topics discussed in this book.

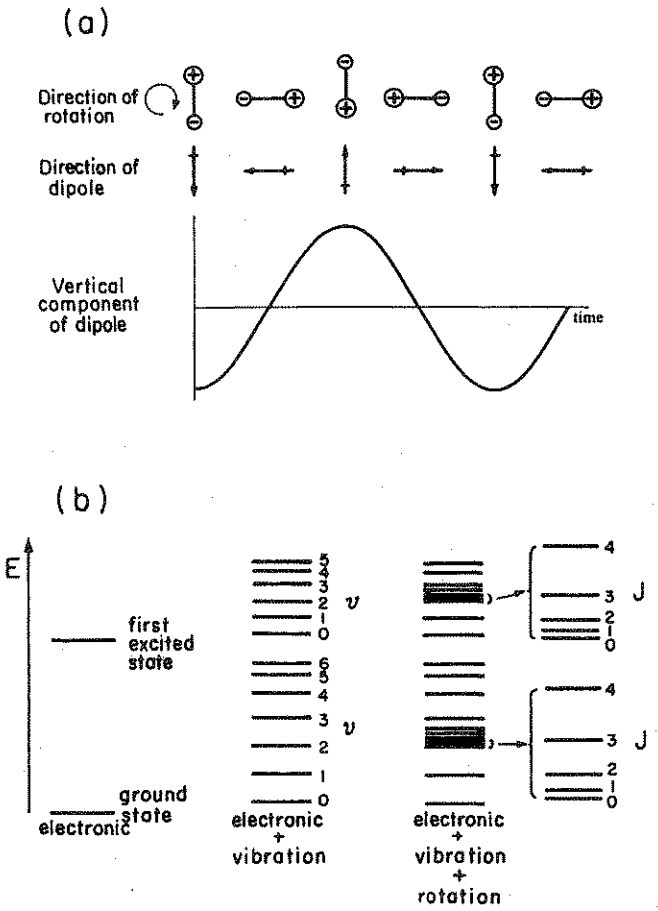

Figure 3.4 (a) The rotation of a simple diatomic molecule showing the fluctuation in the dipole moment measured in a particular direction (Banwell, 1983). (b) Molecular absorption spectra actually consist of closely spaced lines due to rotational and vibrational transitions. J and v refer to the quantum numbers associated with the rotational and vibrational transitions, respectively. This diagram shows how these transitions are superimposed on electronic states.

3.2.1 Molecular Bonding and the Molecular Dipole Moment

To appreciate the nature of molecular absorption, it is useful to reflect on the characteristics of the forces that bond atoms to form molecules and to reflect on the properties of molecules themselves which give rise to the absorption spectra we observe. The *covalent bond* is the prevalent type of bond that forms radiatively active molecules of the Earth's atmosphere. The bond is considered to be the result of the electrostatic forces that arise from the sharing of electrons between the atoms that compose the molecules. In circulating among the atoms, the electrons spend more of their time between atoms than on the outside of the molecule. This lopsided circulation produces a net attractive electrostatic force bonding atoms into molecules.

The basic importance of the oscillating electric dipole to our view of the way radiation interacts with matter was emphasized in Chapter 2 as well as in relation to Fig. 3.3. For the interactions discussed here, a photon can only be absorbed to produce vibrational and rotational movements in the molecule if the charge around a molecule is separated in some way to produce a dipole moment (Section 2.1). How can an electrically neutral molecule then possess such a dipole moment and hence be radiatively active? To understand how this is so, we consider the molecule as a charge distribution which can be described by a probability function in the same way that atoms are considered as a distribution of charge. If the geometric center of this distribution happens to coincide with the center of mass, as it will for a symmetric homonuclear molecule like N_2, then the electric dipole moment is 0. Such molecules are said to be *homopolar* because they lack a permanent dipole moment. Therefore, most *homonuclear* molecules such as N_2 and H_2, when isolated from each other, are also homopolar and are radiatively inactive to infrared radiation[2].

[2] Under conditions of high pressure, such as those that occur in the atmospheres of Jupiter and Saturn, the charge distribution of homonuclear molecules is distorted sufficiently by frequent collisions with other molecules that a dipole moment is produced. The resulting spectrum produces a pressure-induced absorption band. Since the collision time between molecules is very brief, the absorption lines are correspondingly broad. This general characteristic is evident in the emission spectra between 300 and 800 cm^{-1}, which will be presented in Fig. 3.27 for Saturn. This feature is unimportant for the topics of this book.

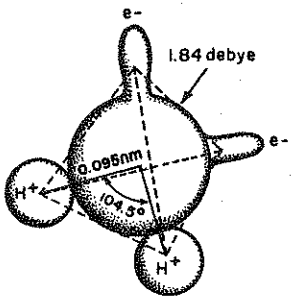

Figure 3.5 Schematic of the atomic configuration and electronic orbitals of the water molecule, with an oxygen atom located at the center and hydrogen atoms at an angle of 105 degrees. Dipolar character comes from protons at H+ positions and unshared electrons at e– locations; the direction of the dipole moment is along the symmetry axis. The atoms and the electronic orbitals have tetrahedral symmetry.

Molecules that possess a permanent electric dipole moment are said to be *heteropolar*. This dipole moment arises from the particular bonding configuration of the molecule. An important example of a heteropolar molecule is the H_2O molecule. We see by studying the molecule how a dipole moment is established through the geometric arrangement of atoms in the molecule. The specific structure of the water molecule, highlighted in Fig. 3.5, shows how the bonding of hydrogen atoms around the oxygen atom help produce a permanent dipole moment from the charge separation of the protons at the $H+$ positions and from the unshared electrons as shown in the diagram.

Molecular oxygen is a special case of a homonuclear molecule that is radiatively active at those wavelengths which we normally associate with rotational transitions. Rotational transitions arise by virtue of the permanent magnetic dipole of the molecule resulting from the presence of an unshared electron in the outer orbit of each oxygen atom. As we shall see, this magnetic dipole gives rise to absorption bands in the microwave region that are exploited by the MSU for temperature sounding of the atmosphere.

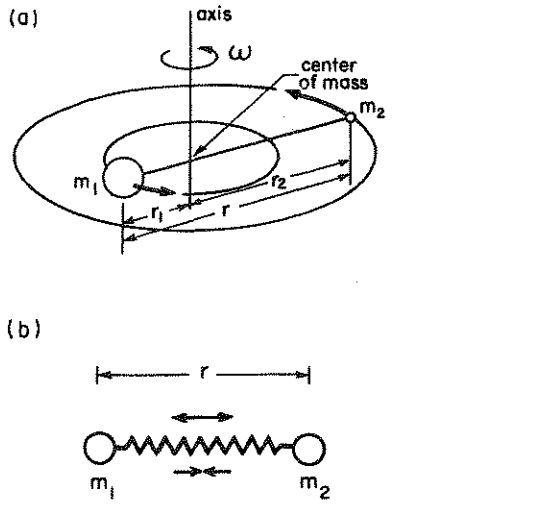

Figure 3.6 (a) A simple analog to a diatomic molecule that rotates about its center of mass. (b) A vibrating diatomic molecule can be thought of as two masses held together by a spring. Vibrations occur when the spring is distorted from its equilibrium position.

3.2.2 Vibration and Rotation Spectra of Simple Diatomic Molecules: Illustrative Examples

Simple mechanical analogs to a rotating–vibrating diatomic molecule are now described to provide an elementary understanding of these vibrational–rotational absorption spectra.

Rotating Molecules

First we invoke a simple mechanical model of a two body rigid rotator, as shown in Fig. 3.6a, as an analog of an absorbing diatomic molecule. An elementary treatment of rigid body rotation tells us that the energy of such a system rotating with an angular velocity ω is

$$E = \frac{1}{2}\mathcal{I}\omega^2 = \frac{\mathcal{L}^2}{2\mathcal{I}} \tag{3.4}$$

where $\mathcal{L}$ is the angular momentum and $\mathcal{I}$ is the moment of inertia,

$$\mathcal{I} = m_1 r_1^2 + m_2 r_2^2 \tag{3.6}$$

With the definition of the center of mass as

$$m_1 r_1 = m_2 r_2 \tag{3.7}$$

then $\mathcal{I}$ follows as

$$\mathcal{I} = \frac{m_1 m_2}{m_1 + m_2}(r_1 + r_2)^2 = m' r^2 \tag{3.8}$$

where m' is the reduced mass of the molecule and r is the distance between the two atoms. Schroedinger's equation shows that the angular momentum allowed to the rigid diatomic molecule has discrete values given by $\mathcal{L} = \sqrt{J(J+1)}\hbar$ where J is the *rotational quantum number* ($J = 0, 1, 2, ...$). When this expression is used for the angular momentum factor in (3.4), we find that the energy of the Jth rotational state of the molecule is

$$E_J = \frac{J(J+1)\hbar^2}{2\mathcal{I}} \tag{3.9}$$

and the change in energy associated with a transition from this state to $(J+1)$ is

$$\Delta E = E_{J+1} - E_J = h\nu \tag{3.10}$$

It follows from a combination of these expressions that the spectral position of an absorption line of a diatomic molecule is defined by

$$\nu = \frac{\hbar}{2\pi m' r^2}(J+1) = 2B(J+1) \tag{3.11}$$

where $B = \hbar/4\pi\mathcal{I}$ is referred to as the rotational constant. Since this is a factor that contains the moment of inertia of each molecule, it is a factor that is specific to that molecule. Equation (3.11) predicts absorption lines as shown in Fig. 3.7a that are equally spaced by a frequency interval $2B$. From (3.9) and (3.11), the moment of inertia of a molecule can be deduced from its rotational spectra and given the masses of its constituent atoms, the interatomic separation r can be calculated.

For more complex molecules, their rotational energy may be generally specified in terms of three principal moments of inertia and three sets of rotational quantum numbers. Molecules can be divided into four types of rotators based on their geometric structure and resulting moments of inertia:

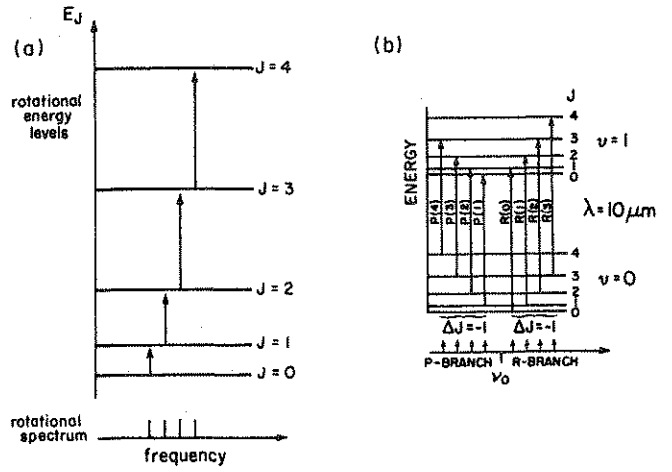

Figure 3.7 (a) Energy levels and the resulting spectrum of molecular rotation. (b) Vibrational–rotational transitions between the lowest vibrational and rotational levels.

- **Linear molecules** (CO_2,N_2O,C_2H_2, all diatomic molecules): For these molecules, the moment of inertia along the internuclear axis is zero and the other two principal moments of inertia about axes perpendicular to the internuclear axis are equal. Rotational energy for these molecules is characterized by a single rotational constant and absorption spectra resemble the simple spectra predicted by our rigid rotator model shown earlier.
- **Symmetric top molecules** (NH_3,CH_3Cl,CF_3Cl): These are nonlinear molecules possessing two moments of inertia that are equal. Thus two rotational constants define the rotational energy states of the molecule.
- **Spherical symmetric top molecules**: This special case is of interest because one of the important trace gases in the atmosphere, namely methane (CH_4), is a classic example of a spherical top rotator. A spherical top is one in which all three principal moments of inertia are equal.

• **Asymmetric top molecules** (H_2O,O_3): These molecules possess three different moments of inertia and are thus characterized by three rotational constants and three sets of rotational quantum numbers. The rotational absorption spectra are therefore quite complex.

Vibrating Molecules

When sufficiently excited, polar molecules vibrate as well as rotate. Vibrational transitions in diatomic molecules can also be studied via the simple mechanical model shown in Fig. 3.6b. In this case, two atoms can be thought of as being held together by a spring that permits vibrations along the line connecting the two atoms. In a nonvibrating state, the atoms are separated by some distance r, which is determined by the tension in the spring. A distortion of the "spring" occurs by absorption to displace one atom with respect to another and the force,

$$F = -\mathrm{k}\,(r' - r) \tag{3.12}$$

acts to restore the molecule to its original undisturbed state. This sets up an oscillation at a frequency

$$\nu' = \frac{1}{2\pi}\sqrt{\frac{\mathrm{k}}{m'}} \tag{3.13}$$

which follows from the classic theory of a harmonic oscillator where k is the spring constant and m' is the reduced mass of the system. Now quantum theory predicts that the frequency of a harmonic oscillator is quantized according to

$$\nu = \left(v + \frac{1}{2}\right)\frac{1}{2\pi}\sqrt{\frac{\mathrm{k}}{m'}} = \left(v + \frac{1}{2}\right)\nu' \tag{3.14}$$

where v is the *vibrational quantum number.* The energies corresponding to these frequencies are

$$E_v = h\nu = (v + 1/2)h\nu' \tag{3.15}$$

The energy required for a vibrational transition is larger than that required for a rotational transition. Vibrations, however, are typically accompanied by rotations so the rotating molecule is not exactly like a rigid rotator. We learn from quantum mechanics,

however, that only certain types of vibrations and rotations are permitted together. These are defined by selection rules which, for the diatomic molecule (or a longitudinal polyatomic molecule like the carbon dioxide molecule), the transition $\Delta v = \pm 1$ occurs simultaneously with a $\Delta J = \pm 1$ transition. This selection rule produces pairs of transitions of the form shown in Fig. 3.7b. As a rule, each vibrational transition frequency is split up into a series of spectral lines with mutual separations that approximately correspond to the respective rotational constant. In Fig. 3.7b, the vibrational transition from $v = 0$ to $v = 1$ is shown. Two branches of rotation lines result for this vibrational transition: one for $\Delta J = +1$, which is referred to as the *R branch*, and the other for $\Delta J = -1$, the *P branch.*

3.2.3 Absorption by Two Triatomic Gases

Because of the dependence of both rotations and vibrations on m' we can conclude that it is possible for molecules with the same approximate total mass to possess very different absorption spectra (Problem 3.5). These spectra are determined not only by the mass of the molecule but also by the distribution of mass within the molecule.

In additional to the effects of the distribution of mass, the rotational and vibrational absorption spectra of polyatomic molecules are much more complex than are the spectra of diatomic molecules because of the higher degrees of freedom of both vibrational and rotational motions. The absorption spectra of the CO_2 molecule is highly relevant to atmospheric remote sensing. The CO_2 molecule vibrates in four different modes, two of which are energetically equivalent. These modes are referred to as *the symmetric stretch mode*, the *asymmetric stretch mode*, and the *bending mode* which has two equivalent modes of vibration. The dipole moment of the symmetric stretch mode is plainly 0 throughout the whole motion (Fig. 3.8a) and this vibration is radiatively inactive. The asymmetric stretch produces a periodic alteration of the dipole moment and this mode is "infrared active" as is the bending mode (Figs. 3.8b and c). The bending mode actually permits $\Delta v = \pm 1, \Delta J = 0$ transitions. These transitions then produce a large absorption peak centered on the fundamental frequency of the oscillator. The absorption is strong at these frequencies is referred to as the Q branch and is especially strong because all the $\Delta v = 1$ transitions accumulate for all available J-levels.

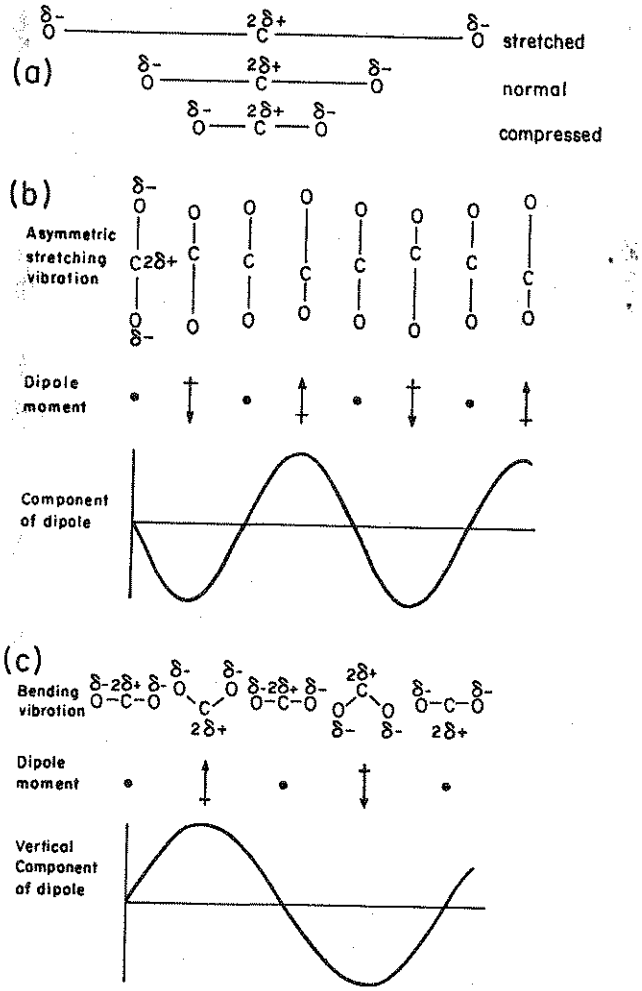

Fig. 3.8 (a) The symmetric stretching of a vibrating CO_2 molecule. (b) The asymmetric stretching of the CO_2 molecule showing the fluctuating dipole moment. (c) The bending motion of the carbon dioxide molecule and its associated dipole fluctuation (Banwell, 1983).

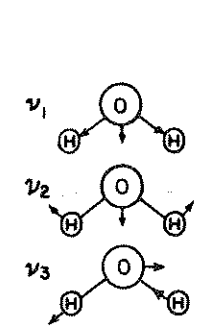

Fig. 3.8 (Cont.) (d) Vibrational modes of a water-vapor molecule.

Another molecule of considerable importance to the study of the Earth's atmosphere is the H_2O molecule. Since this molecule is not linearly arranged like the CO_2 molecule, geometrically different modes of vibration and rotation occur. Figure 3.8d indicates the three modes of vibration of the water molecule. Superimposed on these three modes are the rotational modes around three axes of rotation. The spectra arising from the multiplicity of vibration–rotation transitions are accordingly complex producing irregular absorption spectra composed of bands of thousands of lines.

Excursus: The Infrared Carbon Dioxide Laser

The infrared CO_2 gas laser exploits vibration–rotation transitions of the CO_2 molecule. This particular laser is successfully used in laser studies of the atmosphere (Chapter 8). Lasers (light amplification by stimulated emission of radiation) possess four extraordinary properties compared to ordinary light: coherence, monochromaticity, directionality, and intensity. That laser light is emitted in an extremely narrow range of directions is a consequence of the geometry of the laser tube. The property of monochromaticity is a consequence of quantum mechanics and the special property of coherence is a property of *stimulated emission*. Stimulated emission is an idea that dates back to Einstein in 1917. Bohr's theory of the atom explains two types of transitions; excitation from a lower state to a higher one by absorption of a photon and the reverse process which results in

emission. Einstein argued for a third type of transition, stimulated emission, in which an atom or molecule in an upper state is forced to emit radiation and decay to a lower state by the presence in its vicinity of a photon of the same frequency. The original photon, together with that from the stimulated emission, proceed to stimulate emission in other molecules. Provided enough molecules exist with populated upper states, a kind of chain reaction occurs.

Absorption, spontaneous emission, and stimulated emission occur simultaneously. Whether more photons are emitted than absorbed depends on whether more molecules exist with populated upper states relative to the lower state. The natural population distribution is for most atoms to be in the ground state, fewer in the first excited state, fewer still in the next state, and so on. Increasing temperature increases the populations in the upper states, but the ordering stays the same. Gas lasers generally require the existence of a *population inversion*, one that has the upper states of the molecule populated more than lower states.

We can understand the basic operation of a CO_2 laser by considering Fig. 3.9. The typical medium for the laser is a gas mixture of nitrogen (N_2), helium (He), and CO_2 in an approximate ratio of 6:3:1. The actual laser transition occurs as one of several possible P or R transitions (Fig. 3.9) located in the 9–11 μm region. One of the keys to laser action is for the upper level (in this case the 001 vibration state of the CO_2 molecule) to exist long enough to allow sufficient accumulation of molecules within this state, thus promoting stimulated emission. Some method is required to elevate the CO_2 molecule to the excited state (referred to as pumping of the laser). Nitrogen is used as an intermediary here as the 001 state of the CO_2 molecule very nearly coincides with the first vibrational state of the N_2 molecule. The nitrogen molecule is excited to this vibrational state through collisions with electrons which are initiated by an electric discharge in the tube. The N_2 molecule, pumped to this state, is unable to radiate to its ground state (remember this molecule is homonuclear and has no dipole moment). The vibrational energy of this molecule is thus readily transferred to the CO_2 molecule via collisions which in turn excite a vibration in the CO_2 molecule to the 001 and higher states. The CO_2 molecule then undergoes a vibration–rotation transition to a lower vibration–rotation state (such as the 100 vibration state depicted in Fig. 3.9) via stimulated emission. Collisions among the CO_2 molecules transfer vibrational energy from a molecule in this excited state to a CO_2

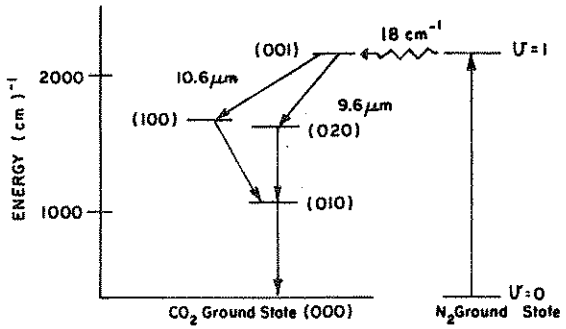

Figure 3.9 The energy levels of the CO_2 molecule relevant to the operation of an infrared CO_2 gas laser. In the example shown the CO_2 gas lasers at 10.6 μm for the $(001) \rightarrow (100)$ transition.

molecule in the ground state. This results in two molecules in the intermediate (010) state. Thus, a bottleneck occurs at (010). Lasing requires that this level be depopulated rapidly to maintain the cycle of pumping and stimulated emission. This is achieved by He atoms. The vibrational energy of the CO_2 molecules in this bottleneck state is transferred to translational energy of the lighter He atoms thus increasing the temperature of the gas mixture which is then mitigated by circulating cool water over the laser tube.

Laser light is intense because of both the multiplicative effect of stimulated emission and the fact that the gas is enclosed between two parallel mirrors at either end of the tube so that photons bounce back and forth producing even more new photons. The wavelength of the stimulated photon is precisely that of the photon that produced it so a beam of high monochromaticity results. One of the mirrors at the end of the tube is not completely reflecting; it transmits some fraction of the incident light and a beam emerges from the tube. This beam, transmitted to the atmosphere and received via appropriated collecting optics, constitutes the technique of lidar remote sensing.

3.3 Line Shapes

Discussion so far has focused only on those factors that determine the spectral positions of the absorption lines. However, the amount of radiation absorbed by molecules also depends on the strength

of the absorption line. In fact, there are three main properties of an absorbing line that mathematically define molecular absorption. These properties are the central position of the line in the spectrum (ν_o), the strength of the line (S), and the shape or profile of the line.

The shape of a line characterizes its spectral "fine" structure or its "fuzziness." The *shape factor* $f(\nu - \nu_o)$ represents this property and provides an estimate of the relative absorption at a frequency which is displaced by $\nu - \nu_o$ from the line center. By convention, the shape factor is defined such that

$$\int_{-\infty}^{\infty} f(\nu - \nu_o) d\nu = 1 \tag{3.16}$$

The broadening of absorption lines is a consequence of a number of factors. It was mentioned earlier that there is a natural line broadening because the energy levels are not sharp. Since the lifetime of a molecule in a particular energy state cannot be longer than the natural lifetime of the upper energy state of the molecule (which is determined from quantum mechanical considerations), then this limit defines the lower limit to the width of a spectral line. This is the *natural line width* and is represented as

$$\alpha_N = \frac{\Delta\nu}{2} \tag{3.17}$$

where $\Delta\nu$ is obtained from (3.3) given an estimate of the lifetime in the upper state. Natural broadening is only a significant mechanism in the upper atmosphere (in the mesosphere and upper stratosphere) where collisions are relatively infrequent. In the lower atmosphere typical lifetimes for upper states of vibration–rotation transitions are much shorter than the natural lifetimes because of the effects of collisions with other molecules.

3.3.1 Pressure Broadening

The main forms of absorption line broadening relevant to the transmission of electromagnetic radiation through the gases of the Earth's lower atmosphere arise from both the individual motions of the molecules and their collisions with other molecules. Broadening by collisions is a complex subject and no exact theory exists. We consider this process only in a simple and somewhat heuristic way here. In this view, we consider that collisions stop the oscillator momentarily after which it starts instantaneously again with a phase that

has no relationship to the phase of the oscillator just prior to the collision. In the case of a rotating molecule, this means that orientation of the molecule after collision is random (a so–called strong collision). When the collisions are frequent enough, as they are in the lower atmosphere, then a kind of equilibrium is reached between the thermal energy associated with molecular motions and the electrical energy of the dipole oscillations. This equilibrium is referred to as *local thermodynamic equilibrium* and under this equilibrium population of the various energy levels closely obey the Boltzmann energy distribution and depend on the temperature as a result.

The phase shift model just described is generally a good approximation to the collision broadened line shape. The simplest description of this collision broadening is the *Lorentz line shape*,

$$f_L(\nu - \nu_o) = \frac{\alpha_L/\pi}{(\nu - \nu_o)^2 + \alpha_L^2} \tag{3.19}$$

where $\alpha_L = (2\pi\bar{t})^{-1}$ and $\bar{t}$ is the mean time between collisions. α_L is referred to as the *Lorentz half–width* and is a measure of the frequency shift that corresponds to the half power point of the line. When typical kinetic theory is adopted to describe molecular collisions, it can be shown that

$$\alpha_L \approx \alpha_{L,s}(p/p_s)(T_s/T)^{1/2} \tag{3.20}$$

where $p_s = 1000$ mb, $T_s = 273$ K, and $\alpha_{L,s}$ is the half–width value at these "standard" temperatures and pressures. For most gases of interest, the values of $\alpha_{L,s}$ generally fall in the range 0.01–0.1 cm^{-1}. For example, a typical value of a CO_2 line is $\alpha_{L,s} \approx 0.07$ cm^{-1}. Figure 3.10 provides an illustration of the Lorentz line profile at pressures of 1, 0.5 and 0.25 bar. The regions far from the line center are referred to as *line wings* and it is relevant to note how the absorption in the wings of the lines increases with increasing pressure in contrast to the absorption in the center of the line which decreases with increasing pressure.

The relationship expressed by (3.20) bears profoundly on the transmission of radiation in the Earth's atmosphere. It states that the Lorentz line width is proportional to atmospheric pressure. Since pressure varies by about three orders of magnitude from the surface to 50 km, so then does α_L. The effect of this pressure variation has been confirmed by experiments. The dependence of α_L on temperature is both less important, since T does not vary to the same

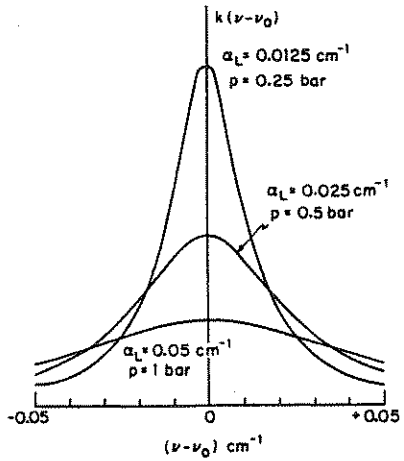

Figure 3.10 Lorentz profiles for three pressures. A line width of 0.05 cm^{-1} at a pressure of 1 bar is typical for vibration–rotation bands (from Goody and Yung, 1989). Note how the dependence of line shape on pressure changes as ν varies from line center to line wings.

extent through the atmosphere, and less well understood. Temperature dependence varies with the particular transition and is often stronger than the $T^{-1/2}$ relationship quoted. Fortunately, the temperature dependence on line shape is usually of secondary importance compared to the temperature dependence of the line strength as discussed below.

The pressure dependence of absorption has a profound effect on the remote sounding of the atmosphere, as suggested by reference to Fig. 3.11. The left part of the figure schematically presents atmospheric pressure as a function of altitude with three different levels highlighted for discussion. The right-hand portion of the diagram shows the pressure–broadened absorption line such as might occur by a transition in the CO_2 molecule at each of these levels. Consider a downward–looking radiometer that receives radiation through a filter which has a spectral width narrower than the line width and is capable of detecting radiation emitted at the three neighboring spectral regions indicated. For a line of suitable strength, when the filter position is close to the line center, the radiation detected by

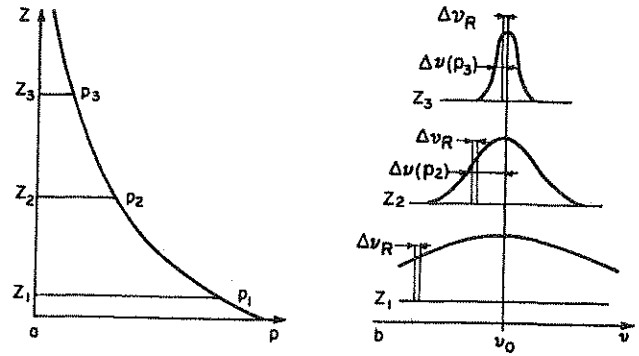

Figure 3.11 The atmospheric pressure as a function of height (left) and the variation of line shape (right) at three different heights. The position of the filter response of a hypothetical radiometer is indicated by $\Delta\nu_R$ relative to this line.

the instrument originates mainly from the higher levels of the troposphere. Further out from the line center, the radiation detected by the instrument arrives from lower down, perhaps in the middle levels of the atmosphere. Still further out into the wings of the line, little of the radiation emitted from the broadened line is absorbed by the atmosphere above owing to the narrowing of the absorption line at higher altitudes. Thus, the radiation detected at these wavelengths originates largely from emission in the lower atmosphere. In this example, we see how the variation of the absorption line width with pressure provides an intuitive explanation of the weighting functions discussed in relation to temperature sounding in Chapter 7.

The Lorentz model of an absorption line is generally accurate for pressure–broadened lines at levels typical of the middle stratosphere (around 10 mb) and below. There are, however, some limitations to this model. For instance, there are significant variations in α_L from gas to gas, from band to band, and even between lines in the same band. The line width is also a function of the type of colliding molecule, or the *broadening gas*. The principal broadening gases in the Earth's atmosphere are naturally N_2 and O_2 since these are most abundant. When these gases act to broaden the absorption lines of other gases, like H_2O and CO_2, then the line is said to undergo

foreign broadening. Broadening by like molecules, such as a water molecule by a water molecule, is referred to as *self broadening* and α_L is usually larger than that in foreign broadening.

Departures in the Lorentz line shape occur in the line wings (i.e., where $|\nu - \nu_o| \gg \alpha_L$). This is unfortunate because these portions of the lines tend to dominate radiative transfer in the relatively transparent regions of the absorption spectrum (the so called "windows" of the spectrum discussed later). Departures in Lorentz lines in the far wings are very difficult to measure and are a major concern both to studies of atmospheric radiative transfer and to various problems of remote sensing which seek to exploit these transparent portions of the absorption spectrum.

Variations of the actual line shape from the simple Lorentz model are also important for those spectral regions where the width of rotational lines is comparable to the central frequency of the line. This situation occurs in the microwave region where the simple phase shift model is no longer strictly appropriate. At these wavelengths, the van Vleck–Weisskopf function

$$f_{VW}(\nu - \nu_o) = \frac{1}{\pi}\left(\frac{\nu}{\nu_o}\right)^2 \left[\frac{\alpha_L}{(\nu-\nu_o)^2 + \alpha_L^2} + \frac{\alpha_L}{(\nu+\nu_o)^2 + \alpha_L^2}\right]$$

is a better approximation to laboratory measurements. Figure 3.12 is a comparison of the measured and modeled water vapor absorption line centered at 22 GHz which is asymmetrical about the central frequency of the line. The different models do depart from observations in ways that are significant especially since the absorption by this particular line is important to atmospheric remote sensing of water vapor as discussed in Chapter 7.

Excursus: Vertical Distributions of Chlorine Monoxide

The shape of an absorption line relevant to paths of varying pressure are not represented by a pure Lorentz line profile. The profile may be thought of as a superposition of several Lorentz lines at different pressures. For example, the line shape for the vertical path that varies from p_1 to p_3 in Fig. 3.11 may be thought of as a kind of superposition of the three Lorenz line profiles drawn. The net effect is a line that is narrow in the center and broad in the wings relative to a pure Lorentz line defined for a single pressure.

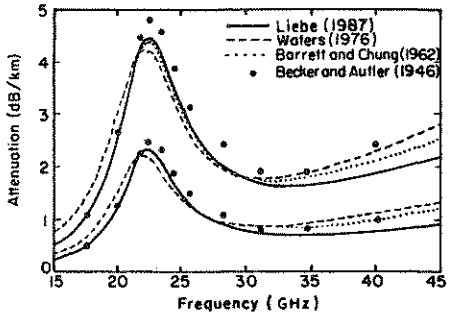

Figure 3.12 The absorption coefficient of the 22 GHz water vapor line in air (10 g H_2O m^{-3}). Comparison between measurement and the van Vleck–Weisskopf line shape is shown (after Walter, 1992a).

An important advantage of microwave spectroscopy over the spectroscopic measurements using optical systems (like those discussed at the end of this chapter) is that individual absorption lines in the microwave emission spectrum can actually be resolved. This provides the possibility of exploiting the shape information of the line to arrive at a qualitative vertical profile of the absorbing gas. An example of this approach is given in De Zafra et al. (1989) who report spectroscopic measurements of the rotational line emission of chlorine monoxide (ClO). In this study, the authors invert the emission line profile, which is just a superposition of line emissions at different pressures, to gain quantitative information on the vertical profile of ClO. The existence of elevated levels of ClO in the stratosphere over the south pole is directly implicated in the formation of the Antarctic ozone hole. Figure 3.13a presents two emission spectra measured across the 512–MHz bandwidth of the instrument. The upper spectrum is for a period four hours after sunrise and the lower spectrum represents emission at night. Ozone lines appear at 110 and 179 MHz and the ClO line is located at 278 MHz (1.1 m wavelength). Figure 3.13b shows spectra obtained over a five–day period, averaged to produce spectra at two–hour intervals after having the ozone line absorptions removed. It is important to note that the ClO line spectra shows a broadening of the daytime profiles (spectra labeled as d,e, and f) compared to the night profiles (traces i, j). The interpretation of these results is that a low level peak concentration

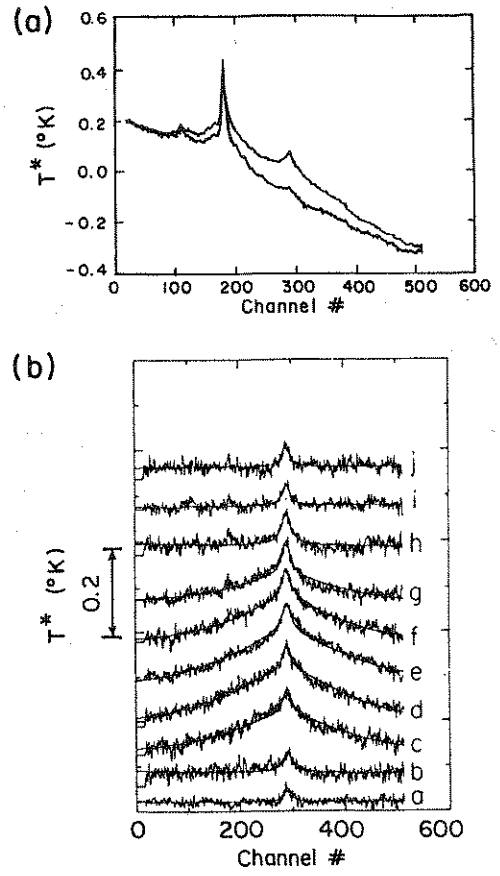

Figure 3.13 (a) Raw spectral data in 512–MHz spectral band pass showing the ClO line at 278 MHz for day (upper trace) and night (lower trace). The ordinate is in terms of brightness temperature difference between day and night and corresponds to observations in the zenith. (b) ClO spectra corresponding to two hourly intervals during the course of 24 hours showing diurnal behavior of the emission. Data represent an average over five days. Trace a, 2–0 hours before sunrise; trace b, 0–2 hours after sunrise, and so on. The midday interval is represented

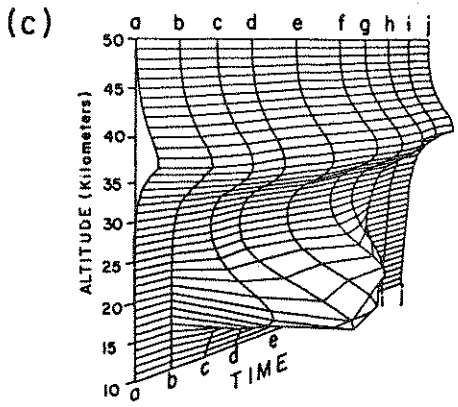

Figure 3.13 (Cont.) (c) Time evolution of the vertical profile of ClO mixing ratio obtained from a deconvolution algorithm applied to the spectra shown in (b) (from De Zafra et al., 1989).

of ClO exists during the day as well as the stratospheric peak that persists throughout both the day and night (Fig. 3.13c).

3.3.2 Doppler Broadening

Even without pressure and natural broadening effects, finite line widths arise simply due to the motions of the molecules themselves. The notion of the Doppler frequency shift was previously encountered in the discussion of wave propagation where it was discovered that the frequency shift depends on whether the dipole is moving toward or away from the observer. It is this shift in frequency that gives rise to *Doppler broadening*. However, molecules of a gas move rapidly with a range of velocities. Some are sluggish in their motion, while others move rapidly, and many move with velocities between these extremes. Experiments show that the velocity distribution of molecules obeys the Maxwellian distribution. As a result of the distributed velocities, the Doppler shifts are correspondingly represented by the same distribution function from which the *Doppler line profile* is derived

$$f_D(\nu - \nu_o) = (\pi\alpha_D)^{-1/2} \exp[-(\nu - \nu_o)^2/\alpha_D^2] \qquad (3.21)$$

where $\alpha_D = u_m\nu_o/c, u_m = (2k_BT/m_a)^{1/2}$, and m_a is the molecular

mass; u_m is essentially the root–mean–square molecular velocity and for CO_2 at a temperature of 250 K, $\alpha_D \approx 7 \times 10^{-4}$ cm^{-1}. We note from this broadening formula that absorption lines of heavy molecules broaden less than do light molecules. This fact is actually exploited in high spectral resolution lidar measurements to separate aerosol and molecular backscattering, an aspect discussed further in Chapter 8.

The Doppler broadened line has a Gaussian shape centered at ν_o and this shape is compared to the Lorentz profile in Fig. 3.14a. According to definition, the Doppler half width depends directly on the resonant frequency and on $T^{1/2}$. Thus accurate measurements of line width provide a way of deducing the temperature of the absorbing molecule provided it can be established that Doppler broadening is the predominant effect. This temperature dependence is an important tool for mesospheric and upper stratospheric sounding.

The relative importance of Doppler broadening compared to pressure broadening can be appreciated in terms of the ratio

$$\frac{\alpha_D}{\alpha_L} \approx 10^{-12} \frac{\nu_o}{p} \tag{3.22}$$

where ν_o is in Hertz and p is the pressure expressed in millibars. Derivation of (3.22) is left as an exercise for the interested reader (Problem 3.6). This ratio is an approximation where the numerical factor is deduced for an averaged weight molecule at a temperature of $T \approx 300$K. Figure 3.14b shows the variation of line width of an oxygen and carbon dioxide molecule, expressed in frequency units, as a function of height in the atmosphere. Due to the dependence of the Doppler width on line center frequency, Doppler broadening of the O_2 line at 2.5 mm wavelength (or about 118 GHz) is two orders of magnitude smaller than that of a CO_2 line at 15 μm. As a consequence, the transition between regions dominated by pressure broadening and regions dominated by Doppler broadening occurs approximately 40 km higher up in the atmosphere for O_2 at 2.5 mm.

In the transition region where both Doppler and Lorentz broadening are important, that is where $\alpha_D \approx \alpha_L$, the shape factor may be represented as a combination of both line shapes according to a function referred to as *Voigt line shape*. This particular line shape will not be discussed further.

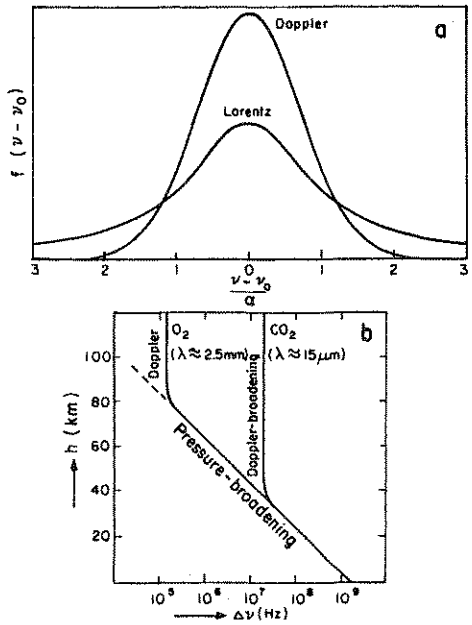

Figure 3.14 (a) A comparison of the Doppler and Lorentz line shapes. (b) Approximate relationship between atmospheric height and the linewidth for a microwave line of O_2 and an infrared line of CO_2 (from Elachi, 1987).

3.4 Absorption Coefficients and Transmission Functions

3.4.1 The Absorption Coefficient

The three factors that characterize line absorption — position, strength, and shape — combine to define the single parameter

$$k_\nu = Sf(\nu - \nu_o) \tag{3.23}$$

which is known as the *absorption coefficient*. This coefficient is an important parameter in characterizing the attenuation of electromagnetic radiation as it propagates through an absorbing gas. The factor S is the strength of the absorption line and is a measure of

how readily a given transition takes place. Line strength depends both on the properties of the single molecule and the populations of molecules in upper and lower states which in turn depend on the temperature of their environment. The line strength S is a product of two distinct factors. The first factor represents the probability that a single molecule in its original lower state absorbs a photon to jump to an upper state. This probability is measured in terms of the *line cross section* σ. The second factor represents the relative populations of the lower and upper states such that

$$S = \sigma(n_\ell - n_u) \tag{3.24}$$

where $n_{\ell,u}$ are the number of molecules in lower and upper states relative to the total number of molecules in all states. It is usually assumed that σ is the same for both the upper and the lower states and that the populations are governed by the Boltzmann distribution as discussed earlier so that the line strength is temperature–dependent. As far as rotational states are concerned, the relative populations are proportional to the factor $g(J) = (2J+1)\exp[-BJ(J+1)/k_BT]$ where B is the rotational constant as defined in relation to (3.9) and k_B is Boltzmann's constant. The shape of this factor is illustrated in Fig. 3.15a and the distribution of line intensities is shown in Fig. 3.15b for the P and R branches of a diatomic molecule. Note that the distribution is asymmetrical because of the different populations of the two branches. This asymmetry is in fact a direct measure of the temperature of the gas. Figure 3.15c presents the actual spectral transmission near 7.78 μm due to absorption by the linear N_2O molecule and demonstrates the features of the hypothetical molecule illustrated in Fig. 3.15b. Mixed amongst the regularly positioned lines of the P and R branches is a second weaker band with a slightly different band center.

3.4.2 Transmission Functions

Absorption of radiation by gases in the Earth's atmosphere is described in terms of *transmission functions*. These functions arise from elementary considerations of radiative transfer which is a topic discussed more extensively in later chapters. For present purposes, we introduce transmission functions with the aid of Fig. 3.16. This diagram provides an illustration of *Lambert's law* which states that the change in intensity along a path (of gas) ds is proportional to the amount of matter (gas) along the path according to

$$dI_\nu = -k_{\nu,v} I_\nu ds \tag{3.25}$$

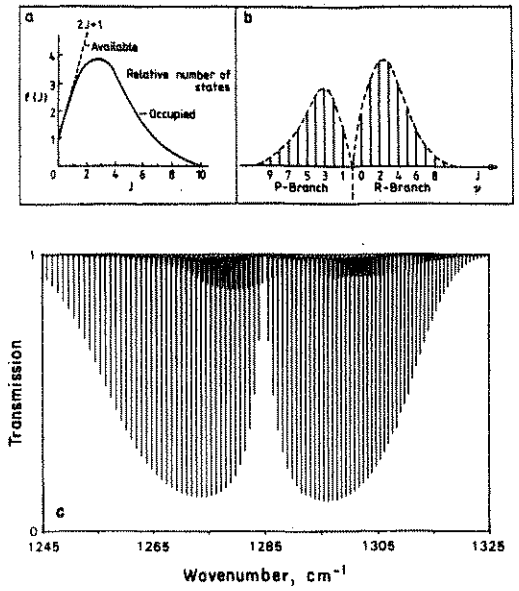

Figure 3.15 (a) The relative occupation of various rotational states. (b) Vibrational–rotational transitions of a diatomic molecule with P and R branches. For a given molecule, the asymmetry of these branches is a measure of its temperature. (c) A synthetic spectrum of N_2O near 7.7μm (after Goody and Yung, 1989).

where $k_{\nu,v}$ is the *volume absorption coefficient*. The dependence of the absorption coefficient on the density of the gas gives rise to a number of possible ways of specifying the absorption coefficient: the *molecular absorption coefficient*, $k_{\nu,n} = k_{\nu,v}/n$, where n is the number density of the absorbing molecules; the *mass absorption coefficient*, $k_{\nu,m} = k_{\nu,v}/\rho_a$, where ρ_a is the density of the absorbing gas; and *the absorption coefficient at s.t.p*, $k_{\nu,s} = k_{\nu,v}n/n_s$, where n_s is Loschmidt's number. The interrelations between these coefficients are discussed in Appendix 2 of Goody and Yung (1989). What is important is the product $k_\nu ds$ which is unitless so that for each type of coefficient there is a corresponding different measure of path length.

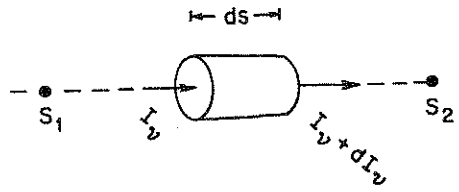

Figure 3.16 A schematic illustration of Lambert's law.

Integrating (3.25) between s_1 and s_2 provides a general solution to Lambert's law in the form

$$I_\nu(s_2) = I_\nu(s_1)\mathcal{T}_\nu(s_1, s_2) \tag{3.26}$$

where $\mathcal{T}_\nu(s_1, s_2)$ is referred to as the *monochromatic transmission function* and is defined as

$$\mathcal{T}_\nu(s_1, s_2) = \exp\left[-\int_{s_1}^{s_2} k_{\nu,v} ds\right] \tag{3.27}$$

Therefore measurement of the radiation flowing from the atmosphere at some level (i.e., at s_2), together with knowledge of the radiation that is incident at some other point along the same direction (i.e., at s_1), provides enough information to obtain the transmission and, with further assumptions, a relatively simple way of inferring the integrated concentration of a particular gas along the path between s_1 and s_2. This is the basic procedure used to determine the path–integrated ozone and water vapor concentrations described in later chapters.

At this point, it is convenient to introduce the quantity

$$\tau(s_1, s_2) = \int_{s_1}^{s_2} k_{\nu,v} ds \tag{3.28}$$

and refer to this as the *optical path*. This quantity is basic to our mathematical description of how radiation interacts with matter. We

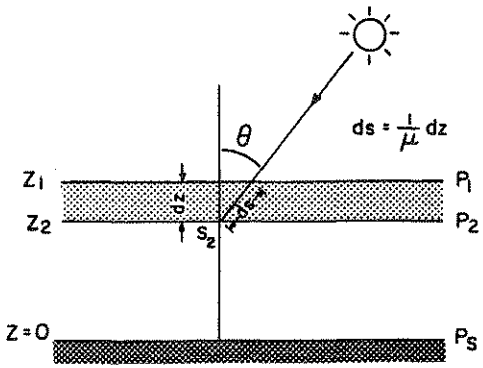

Figure 3.17 A plane parallel atmosphere and the relationship between slant and vertical paths.

shall see a variety of different forms of this quantity throughout this book. It is customary to idealize the atmosphere as a horizontally stratified medium in the manner shown in Fig. 3.17 and to define the path relative to the vertical. Thus transmission along a path tilted from the vertical by an angle θ, the zenith angle, is simply related to the transmission along the vertical path according to

$$\mathcal{T}_\nu(s_1, s_2) = \mathcal{T}_\nu(z_1, z_2, \mu = \cos\theta) = \exp\left[-\tau_\nu(z_1, z_2)/\mu\right] \tag{3.29}$$

where $\tau_\nu(z_1, z_2)$ is now measured along the vertical and is referred to as the *optical depth*.

It is common to use the mass absorption coefficient in the definition of optical depth in describing the transmission along a path through an absorbing gas. Therefore combining (3.28) and (3.29) yields

$$\mathcal{T}_\nu(s_1, s_2) = \exp\left[-\mu^{-1}\int_{z_1}^{z_2} k_{\nu,m}\rho_a dz\right] \tag{3.30}$$

for the slant path transmission function. We define the *optical mass* as

$$u(z_1, z_2) = \int_{z_1}^{z_2} \rho_a dz \tag{3.31}$$

which is often quoted in units of grams per square centimeter. Introduction of the hydrostatic assumption in (3.31), together with the

mixing ratio $r = \rho_a/\rho$, where ρ is the density of air, produces

$$u(p_1, p_2) = \frac{1}{g}\int_{p_2}^{p_1} r\,dp \tag{3.32}$$

where g is the acceleration of gravity, and where p_1 and p_2 are the pressures associated with the altitudes z_1 and z_2, respectively. For uniformly mixed gases, like CO_2 and O_2, r is a constant. If the path through such a gas stretches from a satellite altitude (say $p_2 = 0$) to the surface ($p_1 = p_s$), then

$$u(0, p_s) = \frac{r p_s}{g} \tag{3.33}$$

Thus the absorption path for a uniformly mixed gas is directly proportional to the atmospheric surface pressure p_s; a relationship that has been proposed as a basis for the remote sensing of surface pressure.

For remote sensing, it is important to distinguish between *monochromatic* transmission functions and *band* transmission functions. The former represents the transmission of radiation at one selected wavelength, whereas the latter is the transmission averaged over a range of wavelengths as specified, for example, by the spectral response of a particular instrument. To illustrate this point further, let us suppose that the radiation received at a detector is of the form

$$I_{\Delta\nu} = \int_{\Delta\nu} g(\nu) I_\nu d\nu \tag{3.34}$$

where $g(\nu)$ is the spectral response function of the instrument over its spectral band pass $\Delta\nu$. In terms of the transmission function, the intensity measured at s_2 is

$$I_{\Delta\nu}(s_2) = \int_{\Delta\nu} g(\nu) I_\nu(s_1) \mathcal{T}_\nu(s_1, s_2) d\nu \tag{3.35}$$

where (3.26) has been substituted into (3.34). If the spectral band $\Delta\nu$ is sufficiently narrow that the incident intensity is constant across the band, then the band transmission function becomes

$$\mathcal{T}_{\Delta\nu}(s_1, s_2) = \int_{\Delta\nu} g(\nu) \mathcal{T}_\nu(s_1, s_2) d\nu. \tag{3.36}$$

Therefore instrument properties [in this case the response function $g(\nu)$] directly influence the transmission derived from measurements and must be accounted for in retrieval schemes. Throughout the remainder of this book, the response function will be taken to be understood and is omitted, not because it is unimportant, but merely to simplify matters.

It is also useful to distinguish between transmission functions applicable to the atmosphere where absorption coefficients vary because of varying temperature and pressure along the path (in this case *inhomogeneous* transmission) and transmission functions which are more readily measured in the laboratory under conditions of constant pressure and temperature. A common way of relating these two types of transmission is to assume

$$\exp\left[-\int_{u_1}^{u_2} k_{\nu,m}\,du\right] \approx \exp\left[-k_{\nu,m}(p_o, T_o)\tilde{u}\right] \tag{3.37}$$

where the transmission along an inhomogeneous path is approximated by a transmission function defined along a homogeneous path at a given standard temperature and pressure defined by a scaled optical mass $\tilde{u}$. The scaling typically has the form

$$\tilde{u} = u\left(\frac{p}{p_o}\right)^m \left(\frac{T_o}{T}\right)^{1/2} \tag{3.38}$$

where u is the unscaled path, m is a constant with a value that varies according to the absorbing gas in question and p and T are taken as some representative pressure and temperature of the path, respectively. Other methods which vary in sophistication and accuracy have also been proposed to approximate nonhomogeneous transmission but their details need not concern us here.

3.5 Atmospheric Absorption Spectra

A schematic overview of the atmospheric absorption spectra is given in Fig. 3.18 a, b, and c for the ultraviolet region, the near infrared–far infrared spectral regions, and the microwave region, respectively. The absorption in the ultraviolet beyond about 0.1 μm and in the visible is dominated by transitions in the electronic bands of molecular oxygen and ozone. The absorption spectrum shown in Fig. 3.18a is expressed in terms of the penetration depth of solar radiation. This depth is the altitude at which slightly more than half of the incident solar radiation is absorbed.

particles. As the El Chichon layer decays in time, the size distribution of the aerosol gradually returns to the more typical pre–eruption values of $\alpha \approx 1.3$.

Excursus: Size Distributions from Anomalous Diffraction Theory

The mathematical problem of retrieving aerosol size distributions from multispectral measurements of light extinction reduces to the problem of inverting a Fredholm integral equation

$$\tau_a^*(\lambda) = \pi H \int_0^\infty r^2 Q_{ext}(\mathrm{m}, \lambda, r) n(r) dr \qquad (6.16)$$

for the unknown function $n(r)$. $\tau_a^*(\lambda)$ is obtained from measurements of sunlight using the method described earlier in relation to (6.5) and the function $r^2 Q_{ext}$ is specified in some way. Several approximate analytical solutions to (6.16) exist for the expressions of Q_{ext}, derived from the anomalous diffraction theory formulas introduced in Section 5.5. We start with the ADT expression for a nonabsorbing wavelength written as

$$Q_{ext} = 2 - 4\frac{\sin\rho}{\rho} + 4\frac{1-\cos\rho}{\rho^2} \qquad (6.17)$$

where $\rho = k'r, k' = 4\pi(\mathrm{m}-1)/\lambda$. The inversion of this equation proposed by Fymat (1978) has the form

$$\pi r^2 n(r) = -\frac{1}{2\pi}\int_0^\infty [\cos\rho + \rho\sin\rho](\tau(k') - 2A)dk' \qquad (6.18)$$

where A, the area of the distribution, is the necessary a priori information required to invert (6.16). It is not the intention to dwell on specific methods for evaluating this equation and reference to a number of studies that address this problem are given in the notes at the end of this chapter. All methods, in principle, require extinction measurements for a large number of wavelengths (i.e., for a large number of k's). Klett (1984) introduced a general inversion to (6.16) based on the Laplace transform theory and provides a framework for coping with more practical situations that occur when measurements are made only for a limited number of wavelengths. An example of the Klett method, verified against synthetic

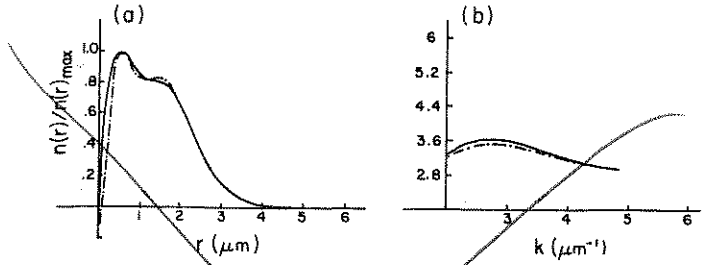

Figure 6.5 (a) Inversion of anomalous diffraction theory extinction data. The input distribution used to simulate particle extinction is plotted as a continuous line and the retrieved distribution is represented by the dashed line. (b) Anomalous diffraction, with a correction factor included (dashed), and Lorenz–Mie extinction (solid) as a function of wavenumber for the bimodal distribution in (a) (after Klett, 1984).

distributions, is given in Fig. 6.5a by the dashed line. Ten extinction values were selected to resolve the $\tau - k'$ spectrum which is also shown in Fig. 6.5b for reference. This is a more complex spectrum than typically predicted by Ångström's turbidity formula (6.9) and is a result of the more complicated bimodal size distribution used to generate the spectrum. The Klett retrieval appears at least capable of reproducing this bimodality.

6.2 More on Extinction–Based Methods

6.2.1 Total Ozone from UV Extinction Measurements

During the 1970s a surge of public concern surfaced about partial destruction of the ozone layer due to anthropogenically produced pollutants. Since then, the substantial depletion of ozone over the Antarctic during the southern hemisphere springtime months was discovered. This has led a significant push into research on ozone chemistry. In recognition of the pressing need for a coordinated international ozone monitoring network, in the 1970s the WMO initiated a special effort to enhance the existing global ozone network. The Dobson ozone spectrometer was selected as the instrument for this worldwide network. Not only do these measurements serve the important purpose of determining hemispheric and global ozone trends,

they also provide a means for "ground–truthing" the satellite ozone observations discussed later in this chapter.

The operational principles of the Dobson spectrometer are discussed in Chapter 3. The spectrometer is used to measure the relative logarithmic attenuation of radiation for pairs of wavelengths in the Hartley–Huggins ultraviolet (UV) ozone bands; one where UV absorption is strong and a second where absorption is weak. These attenuation measurements may be made by observing the direct sun, zenith sky under clear conditions or by observing the zenith sky under cloudy conditions. It is beyond the scope of this book to consider the many different ways of estimating ozone from these spectral measurements. A method based on measurement of direct sunlight is described here, although the basis for the methods associated with zenith sky intensity measurements is developed later in this chapter.

The intensity of the direct sun sensed by an instrument obeys Beer's Law (6.3), which we write in the following way [3]

$$\Delta = \log_{10}(I/I') - \log_{10}(I_o/I'_o) = \\ -\left[\Delta\tau_{d,Ray} + \Delta\tau_{d,O_3} + \Delta\tau_{d,a}\right] m_r \qquad (6.19)$$

obtained from differences of the two adjacent wavelengths (the primes identify quantities at the second wavelength). Of the terms in (6.19), the differential optical depth due to aerosol scattering, namely $\Delta\tau_{d,a}$, is the most uncertain. To overcome this uncertainty two pairs of wavelengths, designated by the superscripts A and D, are used. Then from the subtraction of the logarithm differences of pair D from pair A, we obtain

$$\Delta^{AD} = \Delta^A - \Delta^D$$

and

$$\Delta^{AD} \approx -\left[\Delta\tau^A_{d,O_3} - \Delta\tau^D_{d,O_3}\right] m_r - \left[\Delta\tau^A_{d,Ray} - \Delta\tau^D_{d,Ray}\right] m_r. \qquad (6.20)$$

Because of the more slowly varying wavelength dependence of the aerosol optical thickness [this dependence is predicted, for example, by (6.15)], the difference terms for aerosol are taken to be the same for each wavelength pair so aerosol effects subtract out of (6.20). Various combinations of pairs of wavelengths could be used in this approach but the following were proposed as a standard procedure by WMO: A – 0.3055, 0.3254 μm; D – 0.3176, 0.3398 μm. If we take the decadic optical depth of ozone to be simply the product of the decadic absorption coefficient k_d of ozone and the column integrated ozone amount [4] X_o, then inversion of (6.20) yields

$$X_o = \left[\frac{\Delta^{AD}}{m_r} - \left(\Delta\tau^D_{d,Ray} - \Delta\tau^A_{d,Ray}\right)\right] / \left[\Delta k^A_d - \Delta k^D_d\right] \qquad (6.21)$$

Inserting known values for the absorption coefficient and Rayleigh optical depth for the wavelength pairs A and D leads to the simplification

$$X_o = \frac{\Delta^{AD}}{1.388 m_r} - 0.009 p_o \qquad (6.22)$$

where the dependence of X_o on surface pressure p_o (expressed in units of atmosphere) arises through the dependence of Rayleigh optical depth on pressure. The factor 1.388 is the decadic ozone absorption coefficient difference for the A and D double pair in units of atm–cm^{-1}. The total ozone amount defined by (6.22) has units of milli atm–cm or, alternatively known as a Dobson unit (DU). A DU is the equivalent vertical thickness of atmosphere, in thousandths of a cm, that is occupied by the ozone when concentrated into a uniform layer of pure gas at S.T.P. Values of X_o range from about 200DU to about 400DU.

Total ozone derived from a Dobson spectrometer at the South Pole is shown in Fig. 6.6. These ozone amounts are actually supplemented by ozonesonde observations during the twilight months of March to mid–April and September to mid–October, when it is not possible to make Dobson measurements. Shown are the annual variations of total ozone at the South Pole for 1987, 1986 and portions of the 1978 total ozone record. These results show the marked spring–time (November) increase of ozone and the significant depletion of

[3] In the basic reference to the theory of the Dobson measurement, a different slant path factor is used for each attenuating species. In each case, this factor only differs significantly from $\sec\theta_o$ at very low solar elevations. In our discussion of the Dobson measurements, the factor m_r is used for each species only to simplify matters.

[4] The symbol Ω is most commonly used to represent column ozone and the symbol X is taken to represent the ozone profile in the literature. To avoid confusion with later usage, X_o here refers to total column ozone and $X(p)$ or $X(z)$ refers to the ozone path integrated from the top of the atmosphere to level p or z.

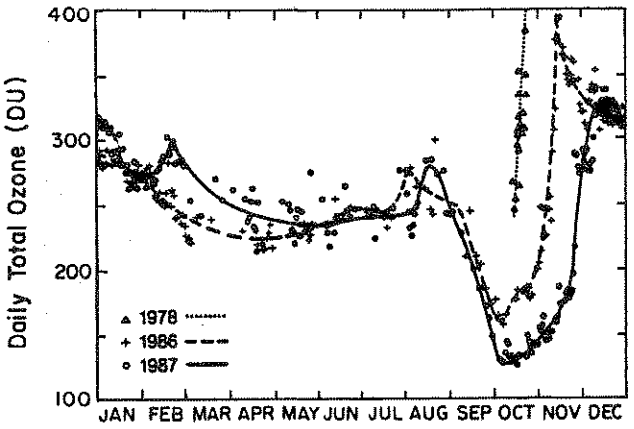

Figure 6.6 Daily total ozone at the south pole in 1986 and 1987 and a portion of the 1978 total ozone record showing how the minimum in October has deepened in the 1980s and how the springtime ozone subsequently increases during (Komhyr et al., 1989b).

total ozone (the so–called ozone hole; Solomon, 1988) observed during the 1980s just prior to the springtime increase of ozone.

6.2.2 Limb Profiling by Extinction—SAGE

One method of obtaining vertical profiles of atmospheric constituents is to measure the radiation emerging from the atmosphere along a horizontal direction. Radiation from the horizon is referred to as *limb* radiation, and the general approach for obtaining profiles of atmospheric constituents derived from limb measurements is referred to as limb sounding. A series of satellite experiments conducted over the past decade have adopted this approach for studying the composition of the stratosphere and limb sounders are now used operationally for stratospheric profiling.

The Stratospheric Aerosol Measurement (SAM) experiment was conducted in the middle to late 1970s and demonstrated the feasibility of solar extinction measurements obtained during satellite sunrise and sunset events observed by a satellite. This experiment served as a prototype for the Stratospheric Aerosol and Gas Experiment (SAGE I, and now SAGE II).

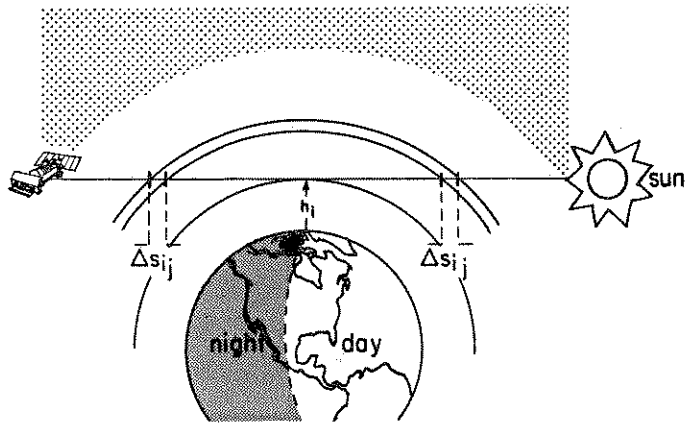

Figure 6.7 The geometry for limb–extinction measurements.

The SAGE I instrument is described in Chapter 3. The SAGE II instrument is similar but has seven channels centered at 1.02, 0.94, 0.6, 0.525, 0.453, 0.448, and 0.385μm. The 0.94μm measurements are used to obtain water vapor amounts, 0.6μm is used for ozone, the difference between 0.452 and 0.448μm used for NO_2 concentrations, and aerosol extinction is deduced from measurements at all wavelengths.

Figure 6.7 illustrates the limb viewing geometry of relevance to this experiment. The instrument on the satellite scans across the solar disc as the sun rises and sets relative to the motion of the satellite. The extinction is taken to be 0 for the highest altitude scan. The ratio of the intensity, I_o, at the center of this scan to the intensity $I(h_i)$ obtained from a scan at a lower altitude h_i provides a measure of the transmittance

$$\mathcal{T}_\lambda = I(h_i)/I_o = \exp(-\tau_\lambda^i) \tag{6.23}$$

where in this instance,

$$\tau_\lambda^i = \int_0^\infty \sigma_{ext,\lambda}(s)ds$$

is the optical thickness along the tangent path defined by h_i at one of the instrument wavelengths. The extinction at any point s along

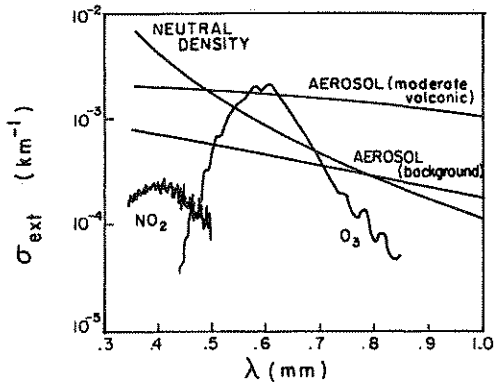

Figure 6.8 The volume extinction at an altitude of 18 km as a function of wavelength (Chu and McCormick, 1979).

this path typically includes several contributions such as

$$\sigma_{ext,\lambda}(s) = \sigma_\lambda^{Ray}(s) + \sigma_\lambda^{O_3}(s) + \sigma_\lambda^{NO_2}(s) + \sigma_\lambda^{a}(s) \qquad (6.24)$$

including contributions from Rayleigh scattering, ozone, NO_2, and aerosol, respectively. An example of the relative contributions to the total extinction by each of these attenuations at an altitude of 18 km is illustrated in Fig. 6.8 for the range of wavelengths relevant to the SAGE II spectrometer.

The basic principle of SAGE is to make measurements at several wavelengths and at many altitudes to arrive at profiles of the attenuators. In the retrieval scheme of Chu and McCormick (1979), the atmosphere is divided into 80 layers. The optical thickness for the kth channel then follows as

$$\tau_k^i = 2 \sum_{j=i}^{N=80} \sigma_{ext,k,j} \Delta s_{ij}, \qquad (6.25)$$

where $\sigma_{ext,k,j}$ is the averaged extinction coefficient in the j^{th} layer, and Δs_{ij} is the path length in the j^{th} layer of the ray of the sun passing through the tangent height at the base of the i^{th} layer. The procedure is to determine τ_k^i from the inversion of (6.23) using the measured intensities. The extinction coefficient for each layer

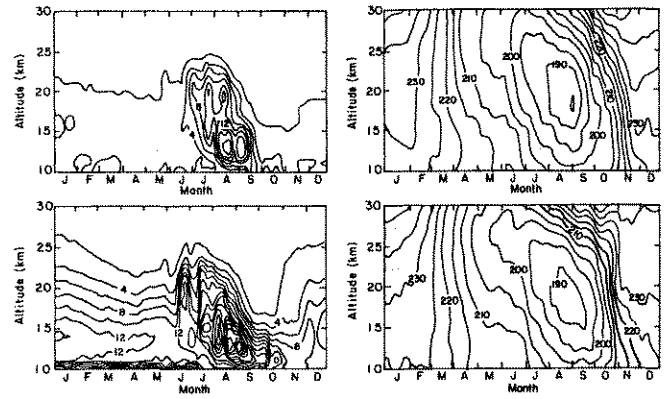

Figure 6.9 Contour plots (left) of the weekly averaged extinction ratio at 1 μm for 1982 (upper) and 1984 (lower) and the corresponding plots of averaged temperature (in K and to the right) for the same years (McCormick and Trepte, 1986).

then follows from the inversion of (6.25) and the extinction by each constituent inferred from either known properties, as in the case of Rayleigh scattering, or from the wavelength dependence of the derived extinction. Details of the inversion algorithms for SAGE I and SAGE II are described by Chu and McCormick (1979) and Chu et al. (1989), respectively. Based on an analysis of the experimental error for typical constituent amounts, Chu and McCormick estimate that the aerosol and ozone profiles can be retrieved to an accuracy of about 10% and the retrieved nitrogen dioxide has an accuracy of about 25% near its peak concentration which occurs between about 25 and 38 km.

SAGE I and II have provided important and unique information about the extinction properties of the atmosphere. For example, we have learned much about the gross properties of polar stratospheric clouds which play a major role in the chemical reactions that produce the Antarctic ozone hole. Some general properties of these clouds

may be inferred from the ratio $\sigma^a_\lambda/\sigma_{ext,\lambda}$ derived from SAGE I data over the South Pole. Figure 6.9 shows the annual variation of weekly averages of this extinction ratio at the South Pole. The aerosol extinction maxima that occur during the winter months indicate the presence of stratospheric clouds between about 10–17 km. These episodes of high extinction coincide with stratospheric temperatures colder than about 190 K, a factor considered important to the formation of polar stratospheric clouds and, in turn, to the chemistry responsible for the destruction of atmospheric ozone.

Another demonstration of the value of these data is given in Fig. 6.10 in the form of a six–year time series of weekly averaged aerosol optical depths integrated from 200 mb upward over the Arctic (dashed) and Antarctic (solid) regions. This time series indicates the regular appearance of polar stratospheric clouds in the Antarctic and the lack of these clouds over the Arctic. Also evident is the general increase of stratospheric aerosol over both poles after the El Chichon volcanic eruption.

Not only can SAGE detect volcanic clouds, but observations from SAGE also allow us to map their movement as they are dispersed by the stratospheric winds. Plate 1 (refer to the front of the book) shows a map of the optical depth derived by integrating the 1.02 μm SAGE extinction profiles from 2 km above the tropopause to the highest level of observation for the period in 1991 from mid–June to late July. This map highlights regions where optical depths exceed 0.1 and shows how the Pinatubo volcanic cloud has spread around the equatorial region into both hemispheres (McCormick and Viega, 1992). The line profiles below the SAGE map are the vertical profiles of lidar backscatter (a topic discussed more in Chapter 8) which were measured at Aspendale, Australia, during the same time as SAGE observations. These shows accentuated levels of backscatter between 20 and 24 km associated with the Pinatubo ash cloud.

6.3 Scattering as a Source of Radiation

So far we have only considered scattering as a process that removes photons from a particular direction such as along an instrument's line–of–sight. The radiative transfer equation used in extinction–based retrievals described earlier assumes only single scattering which is reasonable for tenuous scattering media. In dense clouds or under highly turbid skies, photons from the sun can actually reappear again along the direction of the sun when scattered

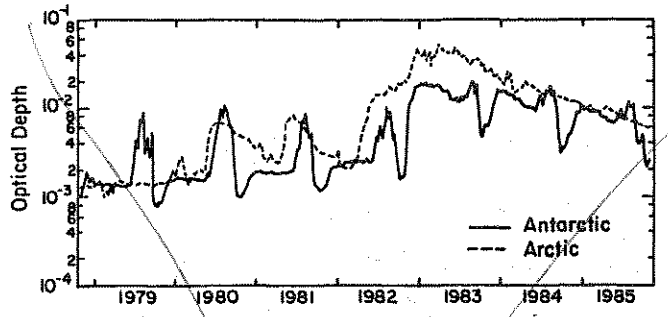

Figure 6.10 Weekly averaged optical depths integrated upward from 200 mb obtained from SAM II observations. The solid line represents Antarctic measurements; the dashed represents Arctic measurements (McCormick and Trepte, 1987).

a multiple of times. In fact, many scattering problems of interest to the atmospheric sciences have to deal with multiple scattering. Multiple scattering of sunlight, for instance, gives rise to many observable phenomena that cannot be explained from single scattering arguments alone. Single scattering predicts a sky that is of uniform brightness and color contrary to what we observe. The whiteness and brightness of clouds is also a result of multiple scattering. Reflection of visible and microwave radiation from various surfaces is largely a product of multiple scattering. Multiple scattering is thus relevant to many topics of remote sensing, and especially to methods based on reflected sunlight.

To account for multiple scattering, we need to introduce a mathematical expression for the reappearance of photons along a specified direction. Unfortunately, to do so requires that we pay more attention to what we mean by direction so that we can talk about changes in direction that occur after a scattering event. We do this using a frame of reference as a basis and describe all scattering processes relative to this frame of reference. Discussion of a particular type of frame of reference used here, and the general definition of a direction vector set on this framework, is relegated to Appendix 1. A few exercises to familiarize the reader with these geometric concepts are

$$I(0;\mu) = \mathcal{B}^* e^{-\tau^*/\mu} + \mathcal{B}^* \left(1 - e^{-\tau^*/\mu}\right)$$

$$(\mathcal{B}^* - \mathcal{B}_o)\left[1 - \frac{\mu}{\tau^*}\left(1 - e^{\tau^*/\mu}\right)\right] \qquad (7.10)$$

which is no longer isotropic but decreases as the instrument scans toward the horizon (*limb darkening*).

In general, we learn from these simple examples that the amount of limb darkening or brightening depends upon the vertical distribution of temperature (that is on both $\mathcal{B}^*$ and $\mathcal{B}_o$) and on τ^*. There is also generally less angular variation in the upwelling intensity than for the downwelling field, particularly at wavelengths where τ^* is small, because the atmosphere transmits some of the radiation incident upon it from the underlying surface. As the optical depth increases, the angular variation of the downward intensity decreases because most of the radiation originates nearer the detector. The upward intensity generally changes relatively slowly with angle, except in the limb.

7.2 The Remote Sensing of Sea Surface Temperature

Except for a few selected examples, most topics in this book deal with the remote sensing of atmospheric properties. There are many surface properties that influence the atmosphere in significant ways. It is also important for studies of the atmosphere to monitor these properties. Sea and land surface temperatures, for example, fundamentally impact upon both the weather and the climate of Earth. In fact, global monitoring of sea surface temperature (SST), in particular, looms as one of the most important exercises in the study of the Earth's climate and the potential for climate change. Some appreciation of the need for remote measurements of SST is provided in Fig. 7.4 which exemplifies the current distribution of in situ SST observations from ships and buoys. The density of observations is greatest in the oceans of the northern hemisphere, and data coverage is sparse over the southern hemisphere oceans. This situation is obviously inadequate for climate studies of SSTs and is clearly inadequate for monitoring global temperature.

The remote sensing of surface temperature is commonly done using measurements of the Earth–atmosphere emission at various wavelengths. The general idea is to measure the emission at wavelengths where the atmosphere is most transparent. Unfortunately,

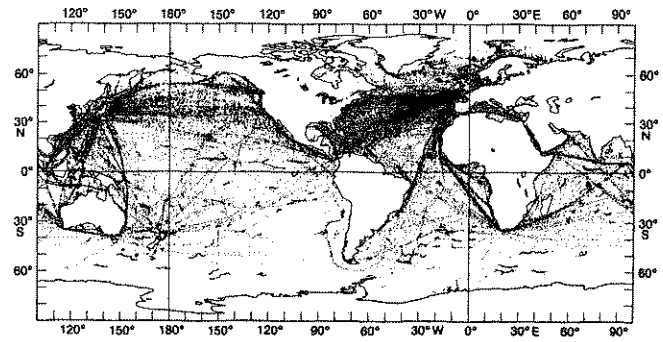

Figure 7.4 Distribution of surface marine in situ (ship and buoy) observations for October 1986. Drifting buoys are distinguished as nearly continuous wiggly lines (after Reynolds, 1988).

the atmosphere is never completely transparent in the main "windows" even for clear skies; therefore, some method is needed to correct for the atmospheric emission. A technique that attempts to provide such a correction is the "split–window" approach which uses measurements at multiple wavelengths. This multichannel approach (referred to as MCSST) originates from the spectral transfer equation, (7.6a), written in the form

$$I(\tau \approx 0, +\mu) = \mathcal{B}(T_s)\mathcal{T}(\tau^*, \mu) + \mathcal{B}(T_a)[1 - \mathcal{T}(\tau^*, \mu)] \qquad (7.11)$$

where the emission from the surface is approximated by the blackbody function $\mathcal{B}(T_s)$ at the surface temperature T_s. In (7.11), $\mathcal{T}(\tau^*, \mu) = e^{-\tau^*/\mu}$ is the transmittance of the atmosphere along the slant path defined by μ and extending from the surface to the satellite. We also express the emission by the atmosphere in terms of an effective blackbody temperature T_a defined from the relationship

$$\mathcal{B}(T_a) = [1 - \mathcal{T}(\tau^*, \mu)]^{-1} \int_0^{\tau^*} \mathcal{B}(t) e^{-t/\mu} \frac{dt}{\mu} \qquad (7.12)$$

The split–window technique uses observations at two channels in an attempt to eliminate the term containing T_a from (7.11) and solve for T_s. To explore this approach, suppose we measure the two

intensities at adjacent wavelengths λ_1 and λ_2. We represent these measurements by I_1 and I_2 and it follows from (7.11) that

$$I_1 = \mathcal{B}_1(T_s)\mathcal{T}_1(\tau_1^*, \mu) + \mathcal{B}_1(T_a)[1 - \mathcal{T}_1(\tau_1^*, \mu)] \tag{7.13a}$$

$$I_2 = \mathcal{B}_2(T_s)\mathcal{T}_2(\tau_2^*, \mu) + \mathcal{B}_2(T_a)[1 - \mathcal{T}_2(\tau_2^*, \mu)] \tag{7.13b}$$

Prabhakara et al. (1974) claim that the value of T_a varies by less than 1 K across the 10.5–12.5 μm window so it is also reasonable to assume a common value for T_a. It is also reasonable to expect that the surface emissivities are the same at adjacent wavelengths (here we conveniently assume this emissivity to be unity).

To solve for T_s, we make use of Taylor's theorem to arrive at the following

$$\mathcal{B}_\lambda(T) \approx \mathcal{B}_\lambda(T_a) + \frac{\partial \mathcal{B}_\lambda}{\partial T}(T - T_a)$$

where the partial derivative is evaluated at $T = T_a$. Application of this equation to both wavelengths and further elimination of the $T - T_a$ factor yields

$$\mathcal{B}_2(T) \approx \mathcal{B}_2(T_a) + \frac{\partial \mathcal{B}_2/\partial T}{\partial \mathcal{B}_1/\partial T}[\mathcal{B}_1(T) - \mathcal{B}_1(T_a)] \tag{7.15}$$

We use this expression twice, once to approximate our observation I_2, which we write in terms of a brightness temperature $T_{b,2}$ according to $I_2 = \mathcal{B}_2(T_{b,2})$, and a second time to approximate $\mathcal{B}_2(T_s)$ to obtain

$$\mathcal{B}_1(T_{b,2}) = \mathcal{B}_1(T_s)\mathcal{T}_2 + \mathcal{B}_1(T_a)[1 - \mathcal{T}_2] \tag{7.16}$$

Eliminating $\mathcal{B}_1(T_a)$ from (7.13a) and (7.16) yields the split–window equation

$$\mathcal{B}_1(T_s) = I_1 + \eta[I_1 - \mathcal{B}_1(T_{b,2})] \tag{7.17}$$

where

$$\eta = \frac{1 - \mathcal{T}_1}{\mathcal{T}_1 - \mathcal{T}_2} \tag{7.18}$$

This relationship is further approximated and linearized in terms of the brightness temperatures,

$$T_s \approx T_{b,1} + \eta[T_{b,1} - T_{b,2}] \tag{7.19}$$

In practice, the actual split–window technique is rarely used in the form given by either (7.17) or (7.19) but these serve merely as a justification for regressing the surface temperature as a linear function

of the measured brightness temperatures. The more typical form of this regression is

$$SST = aT_{11} + b(T_{11} - T_{12}) - c \tag{7.20}$$

where the coefficients a, b, and c are empirically derived from in situ observations such as obtained from drifting buoys. In this expression, $T_{11,12}$ are the brightness temperatures of the $11\mu m$ and $12\mu m$ channels of the AVHRR instrument. When comparing the satellite observations to conventional observations, it is important to note that the satellite temperatures correspond to the temperature of a surface layer just a few millimeters thick (this is referred to as the "skin" temperature). In situ measurements, on the other hand, are bulk measurements of the temperature of a layer of water perhaps a few meters deep. The regression approach of the MCSST is an attempt, in part, to tune the satellite skin temperatures to the in situ bulk temperature. Part of this tuning also accounts for the fact that the atmosphere is not completely transparent at the window wavelengths, especially over the moist tropics (discussion of Table 3.1 and Problem 7.6 offer more quantitative perspectives of these atmospheric effects).

The multichannel approach described here is presently used operationally by NOAA; the precise details of the operational algorithm are described by McClain et al. (1985). The approach actually uses both the $11\mu m$ and $12\mu m$ channels of the AVHRR and the $3.7\mu m$ window channel and has a form similar to that of (7.20). An important part of the algorithm, and one of relevance to many retrieval problems involving infrared emission measurements, is the need to establish whether or not clouds appear in the field of view. Undetected clouds are a source of bias when temperatures associated with the emission from clouds are mistakenly mixed with the SSTs.

Another problem associated with the MCSST approach is that the retrieval of SST is affected by other changes in the atmospheric infrared opacity that occur, for example, with increased concentrations of stratospheric aerosol from volcanic eruptions. Figure 7.5 shows a monthly time series of in situ and satellite observations for the 1982–1986 period for two eastern Pacific regions: one in the tropics (labeled Nino–3) and one in the northern midlatitudes (Namais). The satellite–derived SSTs are lower overall than are the in situ values by about 0.5 degrees; however, the bias is approximately 2 degrees during the period of the El Chichon eruption.

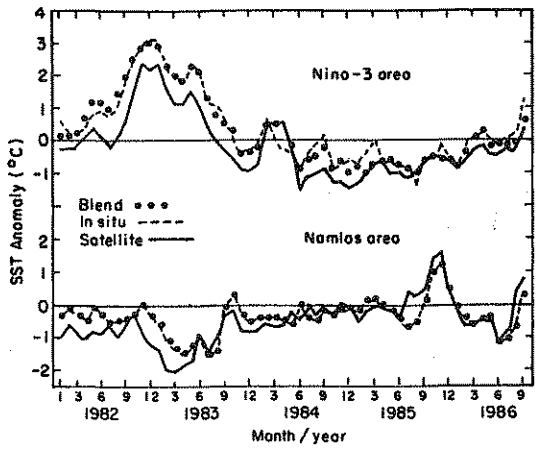

Figure 7.5 Time series of in situ (dashed) and satellite (solid line) SST anomalies for the period shown. The blended data are given as dotted curves. The tropical Pacific region is referred to as Nino–3 and the midlatitude region is referred to as Namais (after Reynolds, 1988).

Reynolds (1988) describes a method which attempts to blend the SST observations from ships and buoys with the satellite observations in an attempt to remove biases of this type. The philosophy is to use the in situ observations as a kind of "benchmark" and the satellite observations are used to define the shape of fields in regions where there is little in situ data. This approach attempts to overcome these biases and is used to develop a climatological SST data set. An example of the comparison of the blended product to the in situ and satellite data is also shown in Fig. 7.5 for the two east Pacific regions.

7.3 Examples of Path–Integrated Quantities

The microwave radiation emitted from the surface of the Earth and sensed by an instrument on a satellite is modulated by processes of emission, absorption, and scattering in the atmosphere; thus, it depends upon the properties of the Earth's surface, atmospheric constituents, and hydrometeors (water droplets, rainfall and ice crystals). The microwave emission from the surface of the Earth depends both on the physical temperature and the emissivity of the surface. Over oceans, values of emissivity typically vary between 0.4 to 0.5, and the emissivity depends on the surface wind speed (refer to Section 4.5). Thus the emission from the ocean is relatively constant and provides a low–level background signal for observing atmospheric emission. Over land, however, both temperature and emissivity are highly variable. The temperature of the land varies with solar insolation, both diurnally and in response to clouds. Land temperature is also a function of the surface albedo, evaporation and evapotranspiration, wind speed, and other factors. The land–surface emissivity is dependent on the thickness, type, and water content of the vegetation canopy and the moisture content and type of the soil. This makes the microwave background signal over land highly variable and measurements subsequently more difficult to interpret. Because of this difficulty, the methods now described are applied only over oceans or over large water bodies on land.

A method that provides vertically integrated water vapor (we refer to this as the *precipitable water*) and the vertically integrated cloud liquid water amount is now described. This method makes use of measurements of the microwave emission by the atmosphere and Fig. 7.6 offers a physical perspective on the approach. Shown are theoretically derived brightness temperature spectra from 6.6 to 37 GHz calculated by a radiative transfer model assuming conditions relevant to a tropical atmosphere. The solid, almost straight, line is the brightness temperature that would be measured over the ocean in the absence of any atmospheric absorption. A calm ocean surface therefore appears cold and the overlying atmosphere appears warm against this cold background due to the increased emission by water vapor and liquid water. The amount of warming is related to the amount of increased emission by the atmosphere associated with the amount of water vapor and liquid water distributed along the path. The solid curve indicates the absorption line centered at 22.235 GHz and is derived for a water vapor overburden of 34.2 kgm^{-2}. The dashed curve is the spectrum produced when a cloud layer containing 0.5 kgm^{-2} of liquid water droplets (a value perhaps typical of low–level stratiform clouds) is added to the atmosphere. Cloud droplets at these wavelengths are assumed to be Rayleigh particles (Section 5.3) and produce a systematic increase in emission at all frequencies over and above that due to water vapor. The dif-

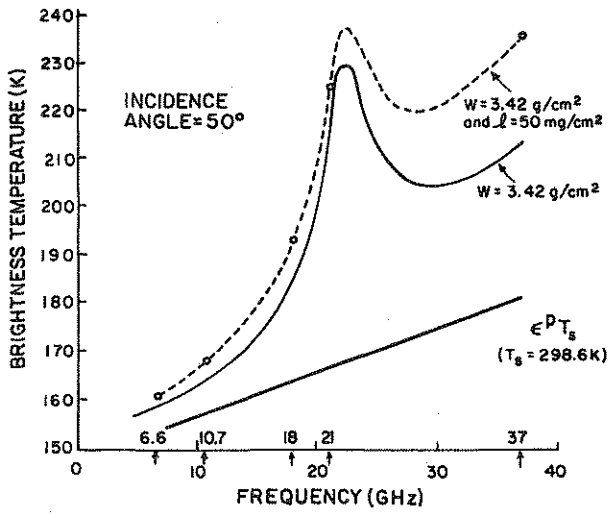

Figure 7.6 Brightness temperature spectra between 6 and 37 GHz for no atmosphere (solid line), a tropical model atmosphere with 34.2 kgm^{-2} of precipitable water (solid curve), and the addition of 0.5 kgm^{-2} of cloud liquid water (from Prabhakara et al., 1982).

ference in brightness temperature associated with the emission by cloud droplets and water vapor increases as the spectral frequency increases. Also shown for reference are the spectral positions of the channels of two satellite microwave radiometers, the Scanning Multichannel Microwave Radiometer (SMMR) flown on the experimental Nimbus 7 satellite (upward pointing arrows) and the Special Sensor Microwave Imager (SSM/I – downward pointing arrows) which has been used operationally since July 1987 as part of the Defense Military Satellite Program.

7.3.1 Microwave Radiative Transfer

It is important to realize that a significant portion of the microwave intensity emitted from the atmosphere toward the ocean surface is reflected back to the atmosphere. Furthermore, the emissivity is polarized by an amount dependent on the viewing direction (Section

4.5). Equation (7.6a) can be readily modified to account for these factors in the following way,

$$T_B = \epsilon^p T_s e^{\tau^*/\mu} + \int_0^{\tau^*} T(t) e^{-(t-\tau)/\mu} d(\frac{t}{\mu})$$

$$+\mathcal{R}^p e^{-\tau^*/\mu} \int_{\tau^*}^{0} T(t) e^{-(\tau^*-t)/\mu} d(\frac{t}{\mu}) \qquad (7.21)$$

where T_B is the brightness temperature measured at the satellite altitude,[3] T_s is the sea surface temperature, ϵ^p is the emissivity of the ocean surface with the given polarization state p, and $\mathcal{R}^p = (1 - \epsilon^p)$ is the surface reflectivity. The first term of the right-hand side of (7.21) is the surface emission term, the second defines the integrated atmosphere emission, and the third corresponds to the downwelling radiation emitted by the atmosphere, reflected at the surface and then transmitted to the satellite sensor. A simplification to (7.21) can be made if it is assumed that the absorption by water vapor is confined to the boundary layer. Thus

$$\int_0^{\tau^*} T e^{-(t-\tau)/\mu} d(\frac{t}{\mu}) \approx T_s \int_0^{\tau^*} e^{-(t-\tau)/\mu} d(\frac{t}{\mu}) \approx T_s \left(1 - e^{-\tau^*/\mu}\right) \qquad (7.22)$$

where we have specifically assumed that the emitting temperature of the water vapor is the same as the sea surface temperature. Refinements to this approximation to include vertical variations of temperature are relatively simple to make but lead only to relatively small corrections to what is presented below so these are omitted here. On substitution of (7.22) into (7.21) we obtain

$$T_B \approx T_s \left[1 - \mathcal{T}^2(\mu)(1 - \epsilon^p)\right] \qquad (7.23)$$

where $\mathcal{T}(\mu) = e^{-\tau^*/\mu}$ is the transmissivity along the direction defined by μ.

[3] We can invoke the Rayleigh–Jeans distribution for $\mathcal{B}$ for many of the microwave frequencies of interest in this book and thus replace this function simply by the temperature T (refer to the discussion in Section 2.5 for more details).

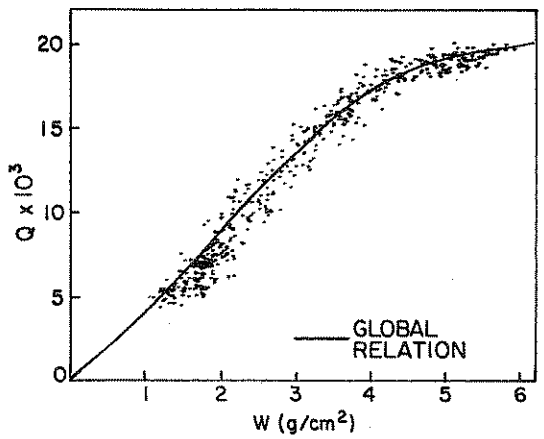

Figure 7.7 Scatter plot of sea–level water vapor mixing ratio and precipitable water for July data in the northern hemisphere and for January data in the southern hemisphere (Liu, 1986).

Excursus: Surface–Level Humidity and Heat Fluxes Over the Ocean

Atmospheric scientists have long been interested in the relationship between vertically integrated path quantities, like precipitable water, and surface observations of properties such as the dew–point temperature, mixing ratio, and vapor pressure. Over the past few years, the association between precipitable water derived from microwave emission measurements and surface level mixing ratio has been studied. A relationship between these quantities is particularly important over the ocean regions where observations are lacking and where this information is crucial for studies of air–sea interaction.

Using 17 years of sounding data from 49 ocean stations, Liu (1986) proposes an empirical relationship between mixing ratio near the surface and precipitable water, an example of which is shown in Fig. 7.7. This relationship provides us with a way of estimating the flux of water vapor E via the bulk parameterization

$$E = \rho_{air} C_E v (q_s - q)$$

where ρ_{air} is the density of surface air, C_E is a specified transfer

coefficient, q_s is the saturation humidity at the surface, v is the near–surface wind speed and q is the humidity of air at some level just above the surface (usually within the surface layer such that v and q are observed at the same height). This value of q is determined from satellite observations of precipitable water using the a priori relationship between q and w. The latent heat flux then follows as the product of E and the latent heat of vaporization L. In principle, the wind speed at the surface can also be estimated from spaceborne sensors using methods like those described earlier. It is therefore possible to estimate E and LE from satellite observations provided we know C_E and ρ_{air}. Liu and Niiler (1984) propose this approach for estimating the monthly mean latent heat fluxes and claim an accuracy of approximately 20 Wm^{-2} for these fluxes.

7.3.2 A Microwave Method for Integrated Water Vapor and Cloud Water

Both the vertically integrated water vapor and cloud liquid water contents are important quantities in the study of the Earth's climate system and especially relevant to topical problems of global change. Figure 7.6 also illustrates how the emission of microwave radiation arises from emission by vapor and by cloud liquid water droplets. This emission is unpolarized whereas the microwave radiation from the ocean surface is polarized to some degree (as described in Section 4.5). The stronger the absorption and thus the stronger the emission by the atmosphere, the more obscured is this polarization. A scheme that exploits this property of emission to derive vertically integrated water vapor and cloud liquid water contents will now be described. The method begins with the approximation (7.23) to analyze the 19.35 GHz and 37 GHz brightness temperatures for the two measured polarization states. This equation is differenced relative to these two polarizations to produce

$$\Delta T_B = T_s \left(\mathcal{R}^V - \mathcal{R}^H\right) \mathcal{T}^2 \tag{7.24}$$

at each frequency where $\mathcal{R}^{V,H} = 1 - \epsilon^{V,H}$ represents the surface reflectivity in the horizontal (H) and vertical (V) polarization states. These surface reflectivities are functions of wind speed as we have previously described. The square of the atmospheric transmission is

$$\mathcal{T}^2 = \mathcal{T}_w^2 \mathcal{T}_W^2 \mathcal{T}_{ox}^2$$

which includes the transmission factor for O_2 and factors for cloud liquid water $\mathcal{T}_W = \exp(-k_\ell W/\mu)$ and water vapor $\mathcal{T}_w =$

$\exp(-k_w w/\mu)$ where k_ℓ is the absorption coefficient of liquid water (the form of this coefficient is described in Section 5.3) and k_w is the absorption coefficient for vapor. With some rearrangement of (7.24), it follows that

$$k_\ell W + k_w w = -\frac{\mu}{2} \ln \frac{\Delta T_B}{T_s(\mathcal{R}^V - \mathcal{R}^H)\mathcal{T}_{ox}^2} \tag{7.25}$$

This equation, applied to both the 19 GHz and 37 GHz measurements, constitutes a set of linear simultaneous equations which can be solved for w and W given values of $\mathcal{T}_{ox,19}$, $\mathcal{T}_{ox,37}$, Δk_w, and Δk_W and modeled values of $\mathcal{R}^V$ and $\mathcal{R}^H$. These surface reflectivities are functions of wind speed as previously noted in Chapter 4.

Knowing just what values to use for the absorption coefficients is perhaps the greatest source of uncertainty not only to the retrieval described here but also to most retrievals of water vapor based on measurements of atmospheric emission. Values of k_ℓ can be taken from a number of sources and the values of k_W in principle follow from the particle scattering theories discussed in Chapter 5. The temperature dependence of k_W arising from the temperature dependence of the refractive index of water at these microwave frequencies is yet another source of uncertainty.

Examples of the precipitable water w derived from this retrieval method are given in Figs. 7.8a and b as four–year seasonal averages of w for December–January–February and June–July–August, respectively. The uncertainty in w is estimated to be about 3 kgm^{-2} based on comparisons with near coincident radiosonde data. The distribution of monthly mean w broadly follows the distribution of SST (Stephens, 1990) with the largest amount over the warmest waters of the equatorial western Pacific Ocean.

One significant factor that has limited the wide use of microwave liquid water data is the general lack of independent data to verify retrievals. Greenwald et al. (1992) attempt such a verification using a limited amount of independent ground–based microwave measurements of W as well as an estimate derived from AVHRR measurements of reflected sunlight. Comparisons between near coincident SSM/I satellite values of W with those of both AVHRR and the surface microwave measurements are presented in the form of a scatter plot in Figs. 7.9a and b, respectively, for stratocumulus clouds off the west coast of California. The error bars are meant to signify the extent of spatial variability associated with the measurements and represent one standard deviation above and below the mean value.

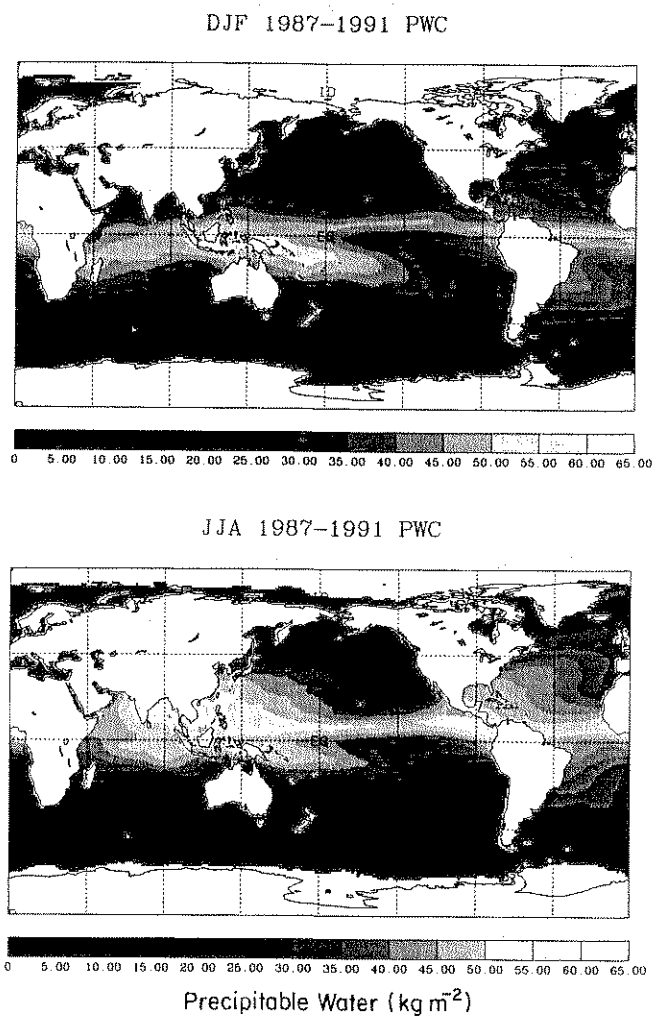

Figure 7.8 (a) The DJF averaged precipitable water derived as an average of DJFs from four years of data. (b) Same as (a) but for the JJA season (Jackson and Stephens, 1993).

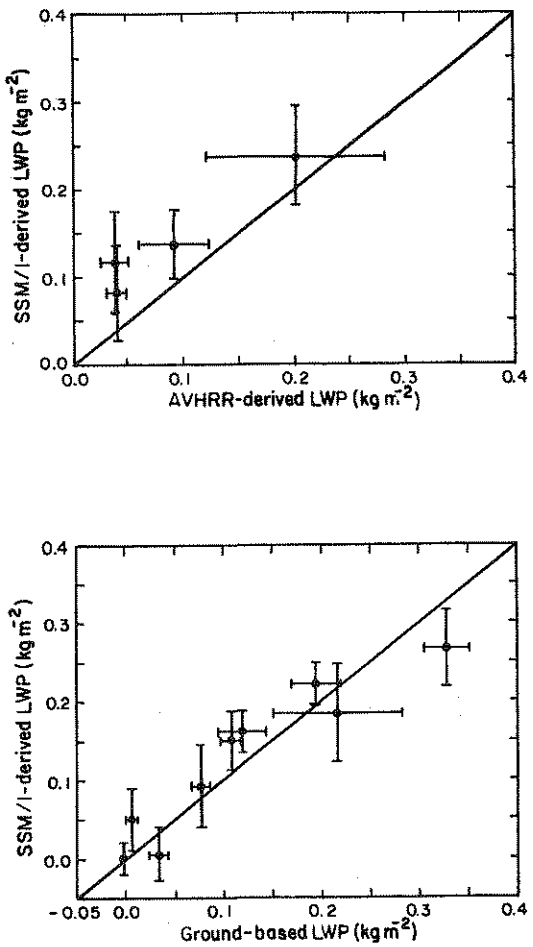

Figure 7.9 Scatter plot of W derived from the microwave retrieval method based on SSM/I measurements with those derived from AVHRR (a) and surface microwave measurements (b). The horizontal and vertical lines about each point are meant to convey the extent of variability in the measurements (Greenwald et al., 1993).

It is yet to be shown if this level of agreement is reasonable and these comparisons do not constitute a rigorous test of the approach, especially when applied to provide global distributions of W. With this in mind, the distribution of W derived from emission measurements over oceans are shown in Fig. 7.10 for August 1987 and February 1988. The main characteristics of global cloudiness are reproduced in these maps including the predominant regions of cloudiness associated with the Inter–Tropical Convergence Zone (ITCZ), the South Pacific Convergence Zone (SPCZ) and the cloudiness associated with midlatitude storminess.

7.4 Passive Sensing of Rainfall

Precipitation, in the form of rainfall, is a key component of the Earth's hydrological cycle. The need to obtain better observations of precipitation globally as well as the need to incorporate more realistic cloud processes in models of the large scale atmospheric motion has prompted more interest in the possibility of space–borne observations of precipitation.

A number of methods exist for the remote sensing of rainfall. Two of these, based on radiation emitted by clouds and rainfall, will now be described and Fig. 7.11 provides a convenient framework for contrasting these two techniques. The diagram presents measurements of radiation at a number of frequencies as a function of elapsed time along a flight path of an instrumented aircraft. The emissions at 18, 37, 92, and 183 GHz are presented in terms of brightness temperature as is the emission at 11 μm. The bottom shaded portion of each panel is the rainfall rate deduced by radar (methods for estimating rainfall from radar measurements are described in Chapter 8). As the aircraft overflies the cloud system, the 11μm brightness temperature decreases from about 245 to 205 K. This decrease occurs at the edge of the deepening outflow anvil from the thunderstorm complex. The microwave temperatures do not show a corresponding drop indicating that the cloud is transparent at these frequencies and presumably composed of ice crystals much smaller than the wavelength of the radiation in question. Passage of the aircraft across the coastline from ocean to land is indicated by an increase in the microwave brightness temperatures due to the sharp rise in surface emissivity at these frequencies. Over the area of deep convection and heavy rain, small decreases (10 to 15 K) in the 11μm brightness temperatures associated with overshooting cells are ob-