Numerical Regularization for Atmospheric Inverse Problems

Adrian Doicu, Thomas Trautmann, and
Franz Schreier

Numerical Regularization for Atmospheric Inverse Problems

Springer

Published in association with
Praxis Publishing
Chichester, UK

PRAXIS

Dr Adrian Doicu
Professor Dr Thomas Trautmann
Dr Franz Schreier
Deutsches Zentrum für Luft- und Raumfahrt
Remote Sensing Technology Institute
Oberpfaffenhofen
Germany

SPRINGER–PRAXIS BOOKS IN ENVIRONMENTAL SCIENCES
SUBJECT *ADVISORY EDITOR*: John Mason, M.B.E., B.Sc., M.Sc., Ph.D.

ISBN 978-3-642-05438-9 e-ISBN 978-3-642-05439-6
DOI 10.1007/978-3-642-05439-6

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010920974

Cover design: Jim Wilkie
Project copy editor: Mike Shardlow
Author-generated LaTex, processed by EDV-Beratung Herweg, Germany

Printed on acid-free paper

Springer is part of Springer Science + Business Media (www.springer.com)

To our families

Table of Contents

Preface

The retrieval problems arising in atmospheric remote sensing belong to the class of the so-called discrete ill-posed problems. These problems are unstable under data perturbations, and can be solved by numerical regularization methods, in which the solution is stabilized by taking additional information into account.

The goal of this research monograph is to present and analyze numerical algorithms for atmospheric retrieval. The book is aimed at physicists and engineers with some background in numerical linear algebra and matrix computations. Although there are many practical details in this book, for a robust and efficient implementation of all numerical algorithms, the reader should consult the literature cited.

The data model adopted in our analysis is semi-stochastic. From a practical point of view, there are no significant differences between a semi-stochastic and a deterministic framework; the differences are relevant from a theoretical point of view, e.g., in the convergence and convergence rates analysis.

After an introductory chapter providing the state of the art in passive atmospheric remote sensing, Chapter 2 introduces the concept of ill-posedness for linear discrete equations. To illustrate the difficulties associated with the solution of discrete ill-posed problems, we consider the temperature retrieval by nadir sounding and analyze the solvability of the discrete equation by using the singular value decomposition of the forward model matrix.

A detailed description of the Tikhonov regularization for linear problems is the subject of Chapter 3. We use this opportunity to introduce a set of mathematical and graphical tools to characterize the regularized solution. These comprise the filter factors, the errors in the state space and the data space, the mean square error matrix, the averaging kernels, and the L-curve. The remaining part of the chapter is devoted to the regularization parameter selection. First, we analyze the parameter choice methods in a semi-stochastic setting by considering a simple synthetic model of a discrete ill-posed problem, and then present the numerical results of an extensive comparison of these methods applied to an ozone retrieval test problem. In addition, we pay attention to multi-parameter regularization, in which the state vector consists of several components with different regularization strengths. When analyzing one- and multi-parameter regularization methods, the focus is on the pragmatic aspects of the selection rules and not on the theoretical aspects

associated with the convergence of the regularized solution as the noise level tends to zero.

At first glance, it may appear that Chapter 4, dealing with statistical inversion theory, is an alien to the main body of the textbook. However, the goal of this chapter is to reveal the similitude between Tikhonov regularization and statistical inversion regarding the regularized solution representation, the error analysis, and the design of regularization parameter choice methods. The marginalizing method, in which the auxiliary parameters of the retrieval are treated as a source of errors, can be regarded as an alternative to the multi-parameter regularization, in which the auxiliary parameters are a part of the retrieval.

Chapter 5 briefly surveys some classical iterative regularization methods such as the Landweber iteration and semi-iterative methods, and then treats the regularizing effect of the conjugate gradient method for normal equations (CGNR). The main emphasis is put on the CGNR and the LSQR implementations with reorthogonalizations. Finally, we analyze stopping rules for the iterative process, and discuss the use of regularization matrices as preconditioners.

The first five chapters set the stage for the remaining chapters dealing with nonlinear ill-posed problems. To illustrate the behavior of the numerical algorithms and tools we introduce four test problems that are used throughout the rest of the book. These deal with the retrieval of O_3 and BrO in the visible spectral region, and of CO and temperature in the infrared spectral domain.

In Chapter 6 we discuss practical aspects of Tikhonov regularization for nonlinear problems. We review step-length and trust-region methods for minimizing the Tikhonov function, and present algorithms for computing the new iterate. These algorithms rely on the singular value decomposition of the standard-form transformed Jacobian matrix, the bidiagonalization of the Jacobian matrix, and on iterative methods with a special class of preconditioners constructed by means of the Lanczos algorithm. After characterizing the solution error, we analyze the numerical performance of Tikhonov regularization with a priori, a posteriori and error-free parameter choice methods.

Chapter 7 presents the relevant iterative regularization methods for nonlinear problems. We first examine an extension of the Landweber iteration to the nonlinear case, and then analyze the efficiency of Newton type methods. The following methods are discussed: the iteratively regularized Gauss–Newton method, the regularizing Levenberg–Marquardt method and the Newton–CG method. These approaches are insensitive to overestimations of the regularization parameter, and depend or do not depend on the a priori information. Finally, we investigate two asymptotic regularization methods: the Runge–Kutta regularization method and the exponential Euler regularization method.

In Chapter 8 we review the truncated and the regularized total least squares method for solving linear ill-posed problems, and put into evidence the likeness with the Tikhonov regularization. These methods are especially attractive when the Jacobian matrix is inexact. We illustrate algorithms for computing the regularized total least squares solution by solving appropriate eigenvalue problems, and present a first attempt to extend the total least squares to nonlinear problems.

Chapter 9 brings the list of nonlinear methods to a close. It describes the Backus–Gilbert method as a representative member of mollifier methods, and finally, it addresses the maximum entropy regularization.

For the sake of completeness and in order to emphasize the mathematical techniques which are used in the classical regularization theory, we present direct and iterative methods for solving linear and nonlinear ill-posed problems in a general framework. The analysis is outlined in the appendices, and is performed in a deterministic and discrete setting. Although discrete problems are not ill-posed in the strict sense, we prefer to argue in this setting because the proofs of convergence rate results are more transparent, and we believe that they are more understandable by physicists and engineers.

Several monographs decisively influenced our research. We learned the mathematical fundamentals of the regularization theory from the books by Engl et al. (2000) and Rieder (2003), the mathematical foundation of iterative regularization methods from the recent book by Kaltenbacher et al. (2008), and the state of the art in numerical regularization from the book by Hansen (1998). Last but not least, the monograph by Vogel (2002) and the book by Kaipio and Somersalo (2005) have provided us with the important topic of regularization parameter selection from a statistical perspective.

This book is the result of the cooperation of more than six years between a mathematically oriented engineer and two atmospheric physicists who are interested in computational methods. Therefore, the focus of our book is on practical aspects of regularization methods in atmospheric remote sensing. Nevertheless, for interested readers some mathematical details are provided in the appendices.

The motivation of our book is based on the need and search for reliable and efficient analysis methods to retrieve atmospheric state parameters, e.g., temperature or constituent concentration, from a variety of atmospheric sounding instruments. In particular, we were, and still are, involved in data processing for the instruments SCIAMACHY and MIPAS on ESA's environmental remote sensing satellite ENVISAT, and more recently for the spectrometer instruments GOME-2 and IASI on EUMETSAT's MetOp operational meteorological satellite. This resulted in the development of the so-called DRACULA (aDvanced Retrieval of the Atmosphere with Constrained and Unconstrained Least squares Algorithms) software package which implements the various methods as discussed in this book. A software package like DRACULA, complemented by appropriate radiative transfer forward models, could not exist without the support we have received from many sides, especially from our colleagues at DLR in Oberpfaffenhofen. To them we wish to address our sincere thanks.

Finally, we would like to point out that a technical book like the present one may still contain some errors we have missed. But we are in the fortunate situation that each author may derive comfort from the thought that any error is due to the other two. In any case, we will be grateful to anyone bringing such errors or typos to our attention.

Oberpfaffenhofen, Germany
March, 2010

Adrian Doicu
Thomas Trautmann
Franz Schreier

1

Remote sensing of the atmosphere

Climate change, stratospheric ozone depletion, tropospheric ozone enhancement, and air pollution have become topics of major concerns and made their way from the scientific community to the general public as well as to policy, finance, and economy (Solomon et al., 2007). In addition to these atmospheric changes related to human activities, natural events such as volcanic eruptions or biomass burning have a significant impact on the atmosphere, while the demands and expections on weather forecasting are steadily increasing (Chahine et al., 2006). Furthermore, the discovery of extrasolar planets with the possibility of hosting life (Des Marais et al., 2002) has brought a new momentum to the subject of planetary atmospheres.

In view of all these developments, atmospheric science comprising various fields of physics, chemistry, mathematics, and engineering has gained new attraction. Modeling and observing the atmosphere are keys for the advancement of our understanding the environment, and remote sensing is one of the superior tools for observation and characterization of the atmospheric state.

In this chapter a brief introduction to atmospheric remote sensing will be given. After a short survey of the state of the atmosphere and some of its threats, the atmospheric sounding using spectroscopic techniques is discussed. A review of the radiative transfer in (Earth's) atmosphere and a general characterization of atmospheric inverse problems will conclude our presentation.

1.1 The atmosphere – facts and problems

The state of planetary atmospheres, i.e., its thermodynamic properties, composition, and radiation field, varies in space and time. For many purposes it is sufficient to concentrate on the vertical coordinate and to ignore its latitude, longitude, and time-dependence. Various altitude regions of the atmosphere are defined according to the temperature structure: troposphere, stratosphere, mesosphere, and thermosphere (Figure 1.1).

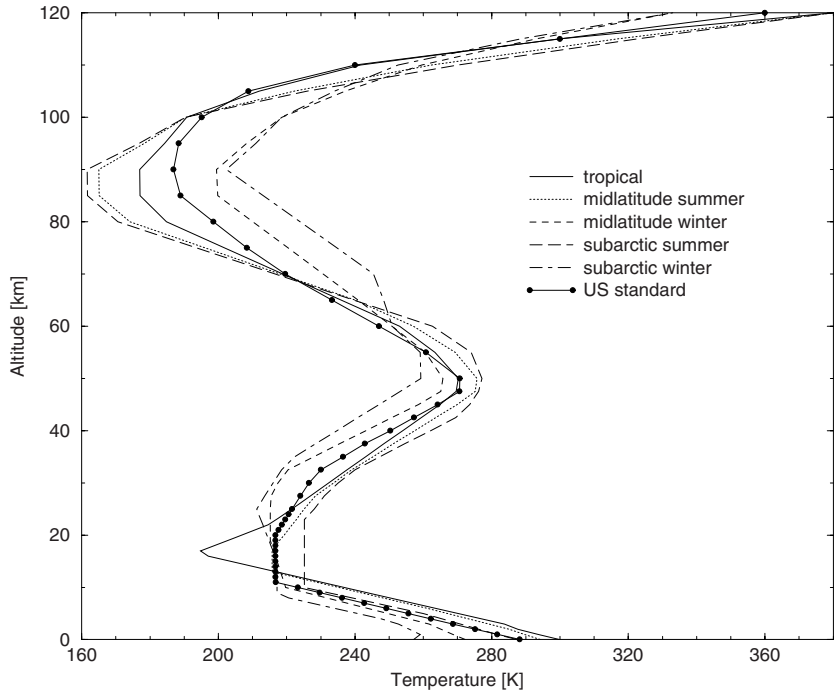

Fig. 1.1. AFGL (Air Force Geophysics Laboratory) reference-atmospheric models: temperatures (Anderson et al., 1986). The circles attached to the US standard profile indicate the altitude levels.

Pressure p decreases monotonically with increasing altitude z; according to the ideal gas law $p = nk_{\mathrm{B}}T$ and the hydrostatic equation $\mathrm{d}p = -g\rho\,\mathrm{d}z$ we have

$$p(z) = p_0 \exp\left(-\int_0^z \frac{\mathrm{d}z}{\bar{H}}\right).$$

Here, n is the number density, g is the gravity acceleration constant, k_{B} is the Boltzmann constant, $\rho = mn$ is mass density, and m is the mean molecular mass ($m \approx 29\,\mathrm{amu} = 4.82 \cdot 10^{-23}\,\mathrm{g}$ for dry air in Earth's lower and mid atmosphere). Ignoring the altitude-dependence of the factors defining the scale height

$$H(z) = \frac{k_{\mathrm{B}}T\,(z)}{mg},$$

yields

$$p\,(z) = p_0 \exp\left(-\frac{z}{\bar{H}}\right), \tag{1.1}$$

where p_0 is the surface pressure ($p_0 = 1\,\mathrm{bar} = 1013.25\,\mathrm{mb}$ for standard STP). Then, assuming a mean atmospheric temperature $T = 250\,\mathrm{K}$, gives the scale height $\bar{H} = 7.3\,\mathrm{km}$.

The terrestrial atmosphere is composed of a large number of gases and various solid and liquid particles (hydrometeors and aerosols), see Figure 1.2. The water- and aerosol-free atmosphere is made up of nitrogen (N_2, 78%) and oxygen (O_2, 21%) with almost constant mixing ratios in the lower and middle atmosphere. Water is present in all three phases

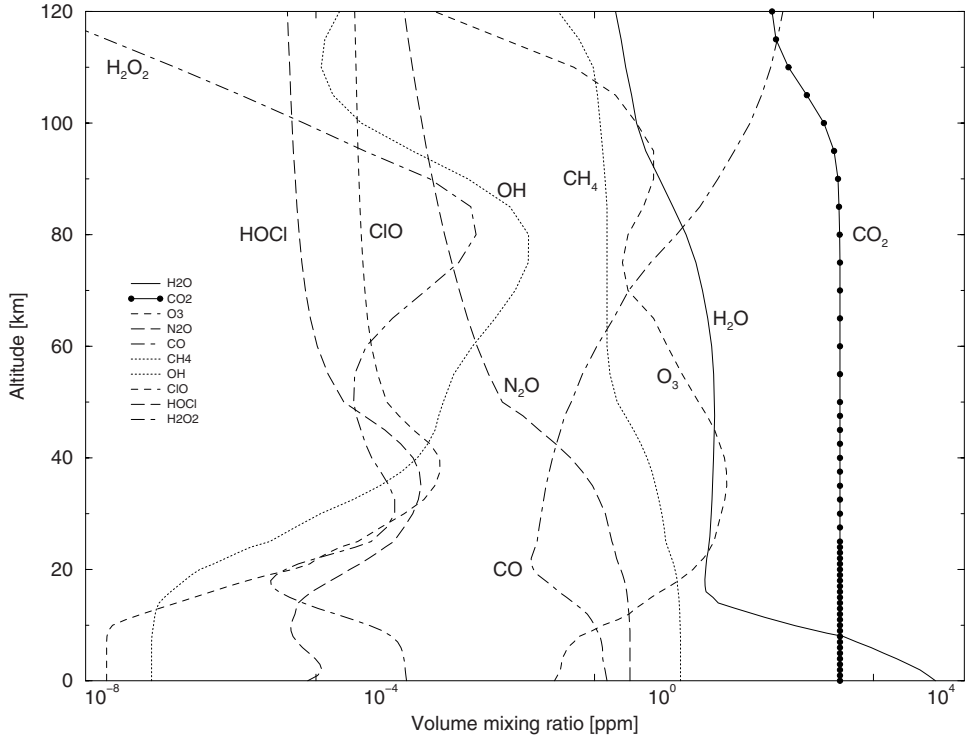

Fig. 1.2. AFGL reference atmospheric models: volume mixing ratios of selected molecules (Anderson et al., 1986).

(vapor, liquid droplets, and ice crystals) and varies significantly in space and time. The remaining 1% of the atmospheric gases are noble gases (0.95%) and trace gases (0.05%). The trace gases, which are mainly carbon dioxide, methane, nitrous oxide and ozone, have a large effect on Earth's climate and the atmospheric chemistry and physics.

Precise knowledge of the distribution and temporal evolution of trace gases and aerosols is important in view of the many challenges of the atmospheric environment.

1.1.1 Greenhouse gases

The greenhouse gases (carbon dioxide CO_2, methane CH_4, tropospheric ozone O_3, chlorofluorocarbons and to a lesser extent water H_2O) are responsible for Earth's natural greenhouse effect which keeps the planet warmer than it would be without an atmosphere. These gases block thermal radiation from leaving the Earth atmosphere and lead to an increase in surface temperature. In the last century, the concentration of greenhouse gases increased substantially: CO_2 from its pre-industrial level of about 280 ppm by more than 30% due to combustion of fossil fuels, and CH_4 by even more than 100%. As a consequence, one expects an average global warming of about $2°$C to $4°$C in the coming century. Hence, substantial changes of the environment can be expected with significant effects for the existing flora and fauna (Solomon et al., 2007).

1.1.2 Air pollution

Pollutants from natural processes and human activities like NO_2 and CO are emitted into the troposphere. In the northern hemisphere, the main source of pollutants is fossil fuel combustion coupled with some biomass burning, while in the southern hemisphere, biomass burning is the primary source. Acid rain produces severe damage to forests and aquatic life, especially in regions with a lack of natural alkalinity. This forms when SO_2 and NO_2 build up in the atmosphere. Sulfur dioxide and nitrogen dioxide are oxidized by reaction with the hydroxyl radical and generate sulfuric acid and nitric acid, respectively. These acids with a pH normally below 5.6 are then removed from the atmosphere in rain, snow, sleet or hail. It should be pointed out that the release of SO_2 into the atmosphere by coal and oil burning is at least two times higher than the sum of all natural emissions.

1.1.3 Tropospheric ozone

Ozone is a toxic and highly oxidizing agent. Photochemical ozone production in the troposphere, also known as summer smog, produces irritation of the respiratory system and reduces the lung function. The majority of tropospheric ozone formation occurs when nitrogen oxides, carbon monoxide and volatile organic compounds react in the atmosphere in the presence of sunlight. High concentrations of ozone arise when the temperature is high, humidity is low, and air is relatively static, and when there are high concentrations of hydrocarbons.

1.1.4 Stratospheric ozone

While ozone behaves like a greenhouse gas in the troposphere, in the stratosphere it helps to filter out the incoming ultraviolet radiation from the Sun, protecting life on Earth from its harmful effects. It is produced from ultraviolet rays reacting with oxygen at altitudes between 20 and 50 km, where it forms the so-called stratospheric ozone layer. In the upper stratosphere, ozone is removed by catalytic cycles involving halogen oxides. In addition, a very substantial depletion of stratospheric ozone over Antarctica and the Arctic has been observed during springtime. The main source of the halogen atoms in the stratosphere is photodissociation of chlorofluorocarbon compounds, commonly called freons, and of bromofluorocarbon compounds known as halons. These compounds are transported into the stratosphere after being emitted at the surface from industrial production. The loss of ozone in the stratosphere is also affected, in a synergistic manner, by the tropospheric emission of greenhouse gases.

1.2 Atmospheric remote sensing

Remote sensing means that measurements are performed at a large distance from the object or the medium to be investigated. The interaction of electromagnetic or acoustic waves with the medium is determined by the state of the medium, and the modification of the waves can be used for the retrieval of the medium's properties. The following discussion

concentrates on measurements of the electromagnetic radiation, but the mathematical tools for the solution of the inverse problem can equally well be applied to acoustic measurements, e.g., SONAR (SOund NAvigation and Ranging) or SODAR (SOund Detection And Ranging).

Remote sensing can be passive or active. Active remote sensing utilizes an artificial radiation source such as a laser emitting light pulses; the laser light is scattered by gas molecules and aerosols and it is partially absorbed by the target gas. A portion of the emitted light is collected by a detector telescope, and the analysis of the recorded laser light reveals information about the composition of the atmosphere. In LIDAR (LIght Detection And Ranging) systems, the transmitter and the detector are usually co-located and the technique is based on backscattering. Radar (radio detection and ranging) systems employ a similar technique using microwave-emitting antennas.

In contrast, passive remote sensing utilizes natural radiation sources. The observation of short-wave solar radiation propagating through the atmosphere, interacting with its constituents and partly being reflected by Earth's surface, and the observation of long-wave thermal emission of both atmosphere and surface are the main approaches. Passive remote sensing can be achieved by analyzing absorption or emission spectra as follows:

(1) Thermal emission. Instruments based upon the emission technique detect the long-wave radiation (infrared or microwave) thermally emitted in the atmosphere along the observer's line-of-sight. The signals from atmospheric constituents can be regarded as thermal 'fingerprints' of the atmosphere, and from the emission line properties, temperature or trace gas concentrations are derived.
(2) Absorption of solar radiation. The upwelling radiation at the top of the atmosphere from the ultraviolet to the near-infrared comprises the solar radiation that has been scattered by air molecules and aerosols, partially absorbed by the target gas and reflected at the Earth's surface. Information on trace gas concentrations is encapsulated in that part of the incoming solar radiation that has been removed by absorption.
(3) Absorption of direct radiation. This category includes occultation instruments that measure solar, lunar, and even stellar radiation directly through the limb of the atmosphere during Sun, Moon and star rise and set events. By measuring the amount of absorption of radiation through the atmosphere, occultation instruments can infer the vertical profiles of trace gas constituents.

A further classification of remote sensing systems is based on the sensor location and the observation geometry (Figure 1.3):

(1) Ground-based systems deployed in laboratory buildings usually observe the atmosphere in an 'uplooking' geometry. Observatories in mountain regions are frequently used with altitudes up to several kilometers, for example, in the Network for Detection of Atmospheric Composition Change (NDACC).
(2) Airborne remote sensing systems work with instruments onboard of aircraft or balloons. Whereas conventional aircraft operate in altitudes more or less confined to the troposphere, some aircraft such as the American ER-2 or the Russian Geophysica can reach altitudes of about 20 km, well in the lower stratosphere. Stratospheric balloons can reach altitudes of almost 40 km, hence permitting observation of the atmosphere in 'limb sounding' geometry.

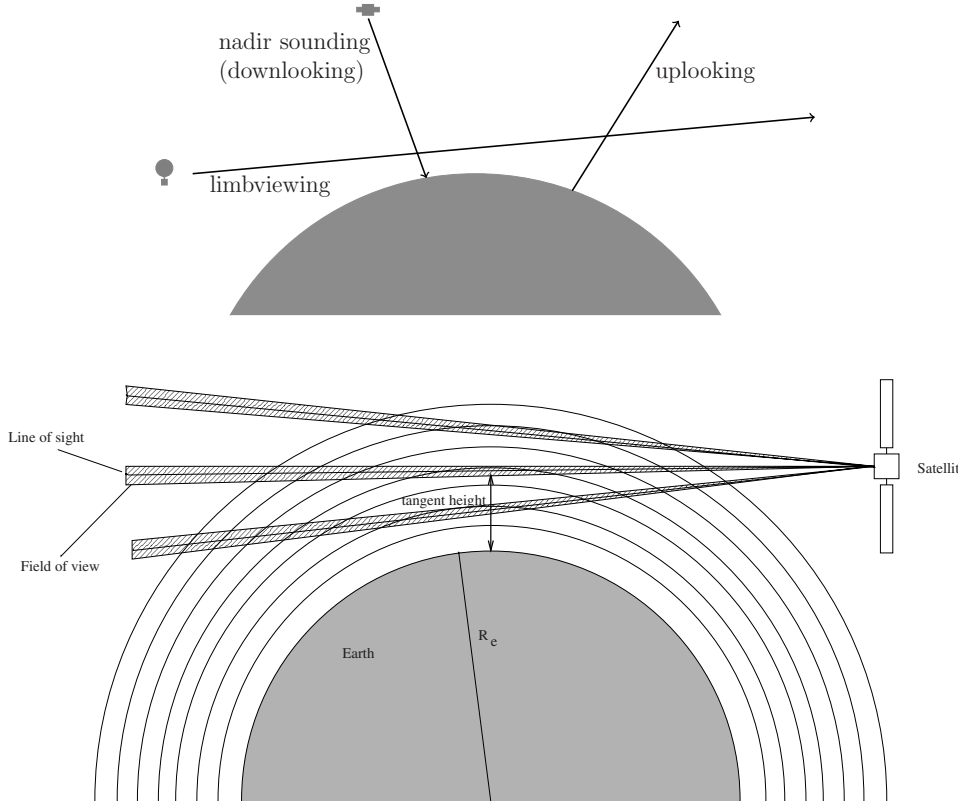

Fig. 1.3. Observation geometries for atmospheric remote sensing.

(3) Spaceborne systems aboard satellites, the Space Shuttle, or the International Space Station (ISS) work in limb viewing or in nadir viewing (downlooking) mode. A large number of sensors for environmental and meteorological studies is mounted on polar orbiting satellites flying at altitudes of about 800 km. Furthermore geostationary satellites with an altitude of about 36 000 km are utilized, especially for meteorological purposes. In contrast, Space Shuttles and the ISS are orbiting at altitudes of about 400 km or less.

Figure 1.4 illustrates the incoming extraterrestrial solar radiation at the top of the atmosphere (TOA) versus wavelength. It is noted that for solar wavelengths beyond 1.4 μm the solar emission curve closely resembles a blackbody radiator having a temperature of about 6000 K. The lower curve depicts a MODTRAN4 (MODerate resolution atmospheric TRANsmission) calculation (Berk et al., 1989) for the downwelling solar flux density reaching the ground. The solar zenith angle has been set to 60°, while for the composition and state of the atmosphere a midlatitude summer case has been adopted. All relevant absorbing atmospheric trace gases, as shown in the figure, were included in the radiative transfer computation which had a moderate spectral resolution of about 20 cm^{-1}. Similarly, in Figure 1.5 we show the infrared spectrum of the Earth atmosphere. The results

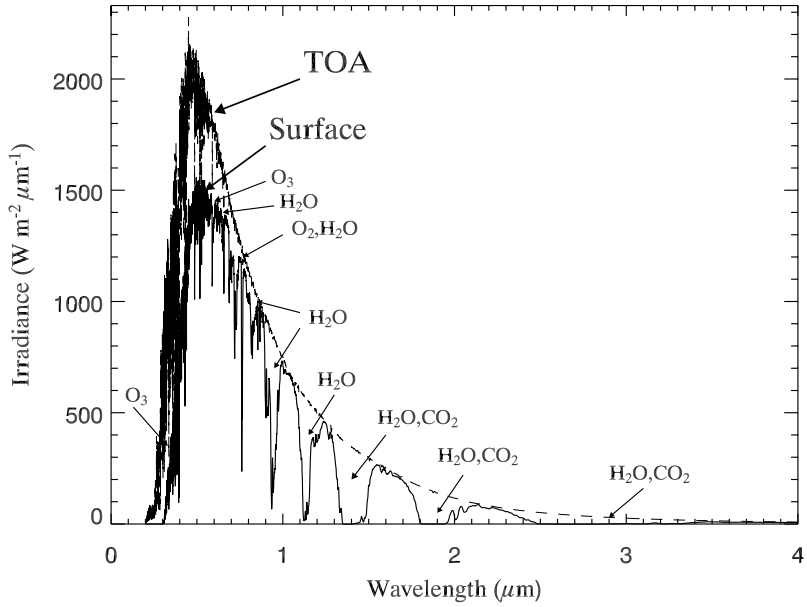

Fig. 1.4. Spectral distribution of the incoming solar flux density at the top of the atmosphere (TOA) and at ground level for a clear sky atmosphere and a nonreflecting ground. The solar zenith angle has been set to 60°. (Adapted from Zdunkowski et al. (2007).)

correspond to a clear sky US standard atmosphere and are also computed with the radiative transfer band model MODTRAN4. Figures 1.4 and 1.5 clearly demonstrate that UV and IR spectra of the terrestrial atmosphere contain a wealth of information about its state, and, in particular, signatures of a large number of molecular absorbers can be identified. Two examples will serve to illustrate the basic principles of atmospheric remote sensing.

In the UV wavelength range 290–330 μm, not only do spaceborne nadir observations of the radiance enable determination of the total column amount of ozone below the subsatellite point but scanning from smaller to larger wavelengths also allows us to 'sound' the atmosphere as a function of increasing distance from the sensor. Ozone molecules absorb solar radiation strongly at short wavelengths, i.e., photons entering the atmosphere are not able to penetrate the ozone layer in the stratosphere (with maximum concentration around 20 or 25 km). On the other hand, photons with higher wavelengths have a better chance to reach a greater depth (lower altitude) before they are absorbed.

Weather forecasting heavily relies on sounding of the atmospheric temperature profile using satellite observations in the infrared or microwave region following the pioneering work of King and Kaplan. King (1956) showed that the vertical temperature profile can be estimated from satellite radiance scan measurements. Kaplan (1959) demonstrated that intensity measurements in the wing of a CO_2 spectral band probe the deeper regions of the atmosphere, whereas observations closer to the band center see the upper part of the atmosphere. Analogously, the complex of O_2 lines in the microwave spectral range can be used. In both cases one utilizes emission from a relatively abundant gas with known and uniform distribution.

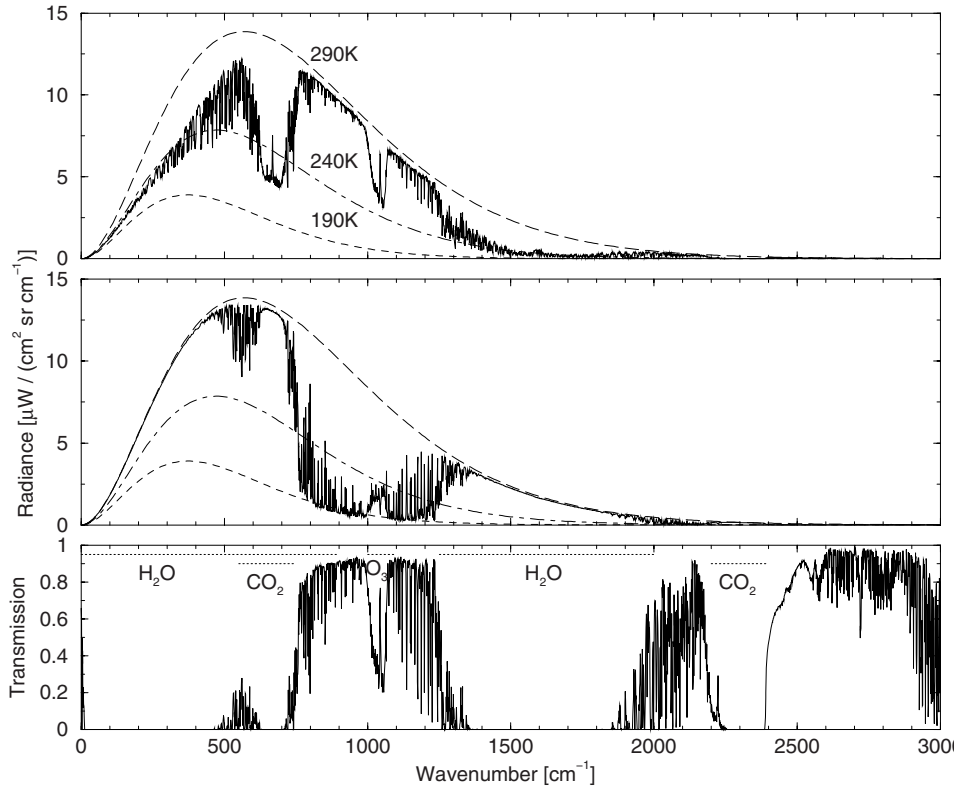

Fig. 1.5. Infrared spectrum of the Earth atmosphere: upwelling radiation seen by an observer above the atmosphere (top), downwelling radiation seen by an observer at sealevel (middle) and atmospheric transmission for a vertical path (bottom). The blackbody radiation according to Planck's function for three representative values and the main absorption bands are indicated too.

In summary, the spectral absorption or emission characteristics combined with monotonically increasing path length allows a mapping between altitude and wavelength, thus providing a direct link between absorber amount or temperature and observed radiation.

1.3 Radiative transfer

In atmospheric remote sensing, the radiation seen by an observer is described by the theory of radiative transfer with an appropriate instrument model. Before discussing radiative transfer models for the UV/vis and IR/mw spectral ranges, we define some quantities of central importance. For a thorough discussion of the material presented in this section we recommend classical textbooks on atmospheric radiation as for example, Goody and Yung (1989), Thomas and Stamnes (1999), Liou (2002), and Zdunkowski et al. (2007).

1.3.1 Definitions

Different variables are used to characterize the 'color' of electromagnetic waves: wavelength λ with units μm, nm, or Å are common in the ultraviolet and visible range, wavenumbers $\nu = 1/\lambda$ in units of cm^{-1} are used in the infrared, and frequencies $\tilde{\nu} = c\nu$ (with c being the speed of light) are employed in the microwave regime. Numerically one has $\nu\,[\mathrm{cm}^{-1}] = 10\,000/\lambda\,[\mu\mathrm{m}] \approx 30\tilde{\nu}\,[\mathrm{GHz}]$.

Monochromatic radiance or intensity is defined as the differential amount of energy $\mathrm{d}E_\lambda$ in a given wavelength interval $(\lambda, \lambda + \mathrm{d}\lambda)$ crossing an area $\mathrm{d}A$ into a solid angle $\mathrm{d}\Omega$, oriented with an angle θ relative to the normal $\mathbf{n}$ of the area, within a time interval $\mathrm{d}t$ (Figure 1.6),

$$I_\lambda = \frac{\mathrm{d}E_\lambda}{\cos\theta\,\mathrm{d}\Omega\,\mathrm{d}t\,\mathrm{d}A\,\mathrm{d}\lambda}. \tag{1.2}$$

The definition of the radiance I_ν is done in a similar manner.

For a beam of radiation traveling in a certain direction, with distances measured by the path variable $s = |\mathbf{r}_1 - \mathbf{r}_2|$, the ratio of the radiances at two different locations defines the transmission

$$\mathcal{T}(\mathbf{r}_1, \mathbf{r}_2) = \frac{I(\mathbf{r}_1)}{I(\mathbf{r}_2)}. \tag{1.3}$$

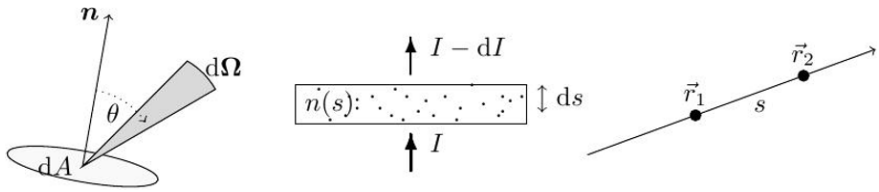

Fig. 1.6. Concepts of radiative transfer. Left: illustration of radiance definition (1.2). Middle: schematics of radiation attenuation $\mathrm{d}I$ traversing a path element $\mathrm{d}s$ with absorber density n. Right: path $s = |\mathbf{r}_1 - \mathbf{r}_2|$ relevant for the definition of optical depth and transmission.

1.3.2 Equation of radiative transfer

A beam of radiation traversing the atmosphere will be attenuated by interactions with the atmospheric constituents, and the extinction (absorption and scattering) is proportional to the amount of incoming radiation, the path distance $\mathrm{d}s$ in the direction $\mathbf{\Omega}$, and the density n of the medium, i.e., $\mathrm{d}I \propto -In\,\mathrm{d}s$ (Figure 1.6). On the other hand, the thermal emission of the medium and the scattering processes will result in an increase of the radiation energy described by a 'source function' $J(\mathbf{r}, \mathbf{\Omega})$. The total change of radiation is given by the equation of radiative transfer

$$\frac{1}{n(\mathbf{r})C_{\mathtt{ext}}(\mathbf{r})}\frac{\mathrm{d}I}{\mathrm{d}s}(\mathbf{r}, \mathbf{\Omega}) = -I(\mathbf{r}, \mathbf{\Omega}) + J(\mathbf{r}, \mathbf{\Omega}). \tag{1.4}$$

The quantity $C_{\mathtt{ext}}$ is called the extinction cross-section, and its product with the number density is the extinction coefficient $\sigma_{\mathtt{ext}} = nC_{\mathtt{ext}}$.

In the absence of any sources, the differential equation can be readily solved and we have (Beer–Lambert–Bouguer law)

$$\mathcal{T}(\mathbf{r}_1, \mathbf{r}_2) = \frac{I(\mathbf{r}_1)}{I(\mathbf{r}_2)} = \exp\left(- \int\limits_{|\mathbf{r}_1 - \mathbf{r}_2|} C_{\mathtt{ext}}(\mathbf{r})\, n(\mathbf{r})\, \mathrm{d}s \right), \tag{1.5}$$

where the integral in the exponent is the so-called (extinction) optical depth between the points $\mathbf{r}_1$ and $\mathbf{r}_2$,

$$\tau_{\mathtt{ext}}(\mathbf{r}_1, \mathbf{r}_2) = \int\limits_{|\mathbf{r}_1 - \mathbf{r}_2|} C_{\mathtt{ext}}(\mathbf{r})\, n(\mathbf{r})\, \mathrm{d}s = \int\limits_{|\mathbf{r}_1 - \mathbf{r}_2|} \sigma_{\mathtt{ext}}(\mathbf{r})\, \mathrm{d}s.$$

Equation (1.4) is a linear first-order differential equation that can be formally integrated giving

$$I(\mathbf{r}_{\mathtt{o}}, \boldsymbol{\Omega}) = I(\mathbf{r}_{\mathtt{s}}, \boldsymbol{\Omega}) \exp\big(-\tau_{\mathtt{ext}}(\mathbf{r}_{\mathtt{o}}, \mathbf{r}_{\mathtt{s}})\big) + \int\limits_{|\mathbf{r}_{\mathtt{o}} - \mathbf{r}_{\mathtt{s}}|} J(\mathbf{r}, \boldsymbol{\Omega}) \exp\big(-\tau_{\mathtt{ext}}(\mathbf{r}_{\mathtt{o}}, \mathbf{r})\big)\, \mathrm{d}s. \tag{1.6}$$

The integral form of the radiative transfer equation (1.6) describes the radiation seen by an observer at $\mathbf{r}_{\mathtt{o}}$; the first term is the source radiation at $\mathbf{r}_{\mathtt{s}}$ (e.g., Earth's surface in case of a downlooking observer) attenuated according to Beer's law (1.5) and the second term represents the radiation due to emission and scattering at intermediate points along the line of sight.

The atmospheric energy budget is essentially determined by solar insolation (roughly in the UV–vis–IR spectral range 0.2–0.35 μm) and emission by the Earth and its atmosphere (in the infrared spectral range 3.5–100 μm). For most practical purposes, these two spectral regions may be treated separately: in the solar spectral range it is justified to neglect the thermal emission of the Earth–atmosphere system, whereas in the infrared the scattering processes are usually important only in the so-called atmospheric window region 8–12.5 μm (Figure 1.5). However, as the clear atmosphere is almost transparent to the infrared radiation in this region, the atmospheric window is of minor importance for remote sensing of trace gases (except for ozone).

1.3.3 Radiative transfer in the UV

The radiation field can be split into two components: the direct radiation, which is never scattered in the atmosphere and reflected by the ground surface, and the diffuse radiation, which is scattered or reflected at least once. Neglecting the thermal emission, the source function J can be decomposed as

$$J(\mathbf{r}, \boldsymbol{\Omega}) = J_{\mathtt{ss}}(\mathbf{r}, \boldsymbol{\Omega}) + J_{\mathtt{ms}}(\mathbf{r}, \boldsymbol{\Omega}), \tag{1.7}$$

where the single and the multiple scattering source functions are given by

$$J_{\mathtt{ss}}(\mathbf{r}, \boldsymbol{\Omega}) = F \frac{\omega(\mathbf{r})}{4\pi} P(\mathbf{r}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_{\mathtt{sun}})\, \mathrm{e}^{-\tau_{\mathtt{ext}}(\mathbf{r}, \mathbf{r}_{\max})},$$

and

$$J_{\mathtt{ms}}(\mathbf{r},\mathbf{\Omega}) = \frac{\omega(\mathbf{r})}{4\pi}\int_{4\pi} P(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}')\, I(\mathbf{r},\mathbf{\Omega}')\ \mathrm{d}\Omega',$$

respectively. In the above relations, $\omega = \sigma_{\mathtt{scat}}/\sigma_{\mathtt{ext}}$ is the single scattering albedo, $\sigma_{\mathtt{scat}}$ is scattering coefficient, F is the incident solar flux, P is the phase function, $\mathbf{\Omega}_{\mathtt{sun}}$ is the unit vector in the sun direction, and $\mathbf{r}_{\max}$ is the point at the top of the atmosphere corresponding to $\mathbf{r}$, that is, $\mathbf{r}_{\max} = \mathbf{r} - |\mathbf{r}_{\max} - \mathbf{r}|\,\mathbf{\Omega}_{\mathtt{sun}}$. It should be pointed out that technically, there is no absolute dividing line between the Earth's atmosphere and space, but for studying the balance of incoming and outgoing energy on the Earth, an altitude at about 100 kilometers above the Earth is usually used as the 'top of the atmosphere'.

An accurate interpretation of the measurements performed by satellite instruments in arbitrary viewing geometries requires the solution of the radiative transfer equation in a three-dimensional inhomogeneous spherical atmosphere. For this type of radiative transfer problems, the Monte Carlo technique (Marchuk et al., 1980) is a possible candidate. In a Monte Carlo simulation the radiance at the top of the atmosphere is determined statistically by simulating a large number of individual photon trajectories through the atmosphere. This method is computationally very expensive in the calculation of the backscattered radiance, because many photons are lost when they leave the atmosphere at other positions and in other directions than the one to the satellite. For atmospheric applications, the so-called backward Monte Carlo method is more efficient. Here, the photons are started from the detector and their path is followed backward to the point where they leave the atmosphere in solar direction. The disadvantages of this method are, however, its poor accuracy for optically thick or weakly absorbing media, and that for each viewing geometry, a new backward calculation has to be performed. Additionally, the required linearization of such Monte Carlo models is a challenging task. Applications of the Monte Carlo method for radiance calculations in a spherical atmosphere can be found in Oikarinen et al. (1999).

Radiative transfer models

In practice, simplified radiative transfer models are used to simulate the radiances at the observer's position and in the direction of the instrument line-of-sight. These can be categorized depending on the assumptions made for the geometry of the model atmosphere.

Plane-parallel radiative transfer calculations have been applied successfully for nadir measurements with solar zenith angles up to 75°. The discrete ordinate method (Stamnes et al., 1988), the doubling-adding method (Hansen, 1971), the finite difference method (Barkstrom, 1975) and the Gauss–Seidel iteration method (Herman and Browning, 1965) have been used to solve the radiative transfer equation in a plane-parallel atmosphere. Further details on the mentioned solution methods can be found in Lenoble (1985).

For nadir viewing geometries with large solar zenith angles and for limb viewing geometries, the so-called *pseudo-spherical approximation* has been developed (Dahlback and Stamnes, 1991). In this approximation, the single scattering radiance is computed in a spherical atmosphere, whereas the multiple scattering radiance is still calculated in a plane-parallel geometry. For limb measurements, the effect of a varying solar zenith angle along the line of sight is accounted for by performing a set of independent pseudo-spherical calculations for different values of the solar zenith angle. This model is equivalent to the *independent pixel approximation* for three-dimensional radiative transfer in clouds, and

can be regarded as a first-order spherical correction to the plane-parallel formulation of the radiative transfer. Solution methods for radiative transfer in a pseudo-spherical atmosphere include the discrete ordinate method (Spurr, 2001, 2002), the finite difference method (Rozanov et al., 2000), and the discrete ordinate method with matrix exponential (Doicu and Trautmann, 2009a).

For a subhorizon Sun as well as for lines of sight with large tangent heights, the independent pixel approximation leads to errors of about 4%. For such problems, the *spherical shell approximation* (Rozanov et al., 2001; Walter et al., 2005; Doicu and Trautmann, 2009e) delivers more accurate results. Here, the atmosphere is approximated by homogeneous spherical shells and no horizontal inhomogeneities in the optical parameters are considered. The radiative transfer equation is solved by means of a Picard iteration with a long or a short characteristic method (Kuo et al., 1996).

Accurate simulations of radiances in ultraviolet and visible spectral regions should take into account that light scattered by the atmosphere is polarized and that approximately 4% of molecular scattering is due to the inelastic rotational Raman component.

Polarization

The radiation and state of polarization of light can be described by the Stokes vector $\mathbf{I} = [I, Q, U, V]^T$, where I is the radiance, Q is a measure for the polarization along the x- and y-axis of the chosen reference frame, U is a measure of the polarization along the $+45^\circ$ and -45° directions, and V describes the circular polarization. The vector radiative transfer equation reads as

$$\frac{\mathrm{d}\mathbf{I}}{\mathrm{d}s}\left(\mathbf{r},\boldsymbol{\Omega}\right) = -\sigma_{\mathtt{ext}}\left(\mathbf{r}\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}\right) + \sigma_{\mathtt{ext}}\left(\mathbf{r}\right)\mathbf{J}\left(\mathbf{r},\boldsymbol{\Omega}\right),$$

where $\mathbf{J}$ is the source term. As in the scalar case, the source function can be split into a single and a multiple scattering component, and we have the representations

$$\mathbf{J}_{\mathtt{ss}}\left(\mathbf{r},\boldsymbol{\Omega}\right) = F\frac{\omega\left(\mathbf{r}\right)}{4\pi}\,\mathrm{e}^{-\tau_{\mathtt{ext}}\left(\mathbf{r},\mathbf{r}_{\max}\right)}\mathbf{Z}\left(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}_{\mathtt{sun}}\right)\begin{bmatrix}1\\0\\0\\0\end{bmatrix},$$

and

$$\mathbf{J}_{\mathtt{ms}}\left(\mathbf{r},\boldsymbol{\Omega}\right) = \frac{\omega\left(\mathbf{r}\right)}{4\pi}\int_{4\pi}\mathbf{Z}\left(\mathbf{r},\boldsymbol{\Omega},\boldsymbol{\Omega}'\right)\mathbf{I}\left(\mathbf{r},\boldsymbol{\Omega}'\right)\,\mathrm{d}\Omega',$$

with $\mathbf{Z}$ being the phase matrix.

The instrumental signal should be simulated with a vector radiative transfer model for two reasons.

First, light reflected from Earth's atmosphere is polarized because of (multiple) scattering of unpolarized light by air molecules and aerosols. Simulations of radiance measurements by a scalar approximation for atmospheric radiative transfer leads to errors of about 10% depending mainly on the viewing scenario (Mishchenko et al., 1994). The scalar radiative transfer errors are small in the spectral regions in which mainly single scattering takes place and significant in the spectral regions in which the amount of multiple scattering

increases because of decreasing gas absorption. For a pseudo-spherical atmosphere, vector radiative transfer models employing the discrete ordinate method (Spurr, 2006, 2008), the successive order of scattering technique (McLinden et al., 2002a) and the discrete ordinate method with matrix exponential (Doicu and Trautmann, 2009b) have been developed. A survey of vector radiative transfer models for a plane-parallel atmosphere can be found in Hansen and Travis (1974).

Second, the different optical devices in the instrument are sensitive to the state of polarization of the incident light. As a result, the radiance that is measured by the detectors, referred to as the polarization-sensitive measurement, is different to the radiance that enters in the instrument. In the calibration process, the instrumental signal is corrected for the polarization sensitivity, whereas the polarization correction factor is determined from broadband on-ground measurements. However, in spectral regions where the state of polarization is varying rapidly with wavelength, the polarization correction is not sufficiently accurate and severely influences the retrieval. To eliminate this drawback, the polarization-sensitive measurement together with the transport of radiation in the atmosphere have been simulated by means of vector radiative transfer models (Hasekamp et al., 2002; McLinden et al., 2002b).

Ring effect

The filling-in of solar Fraunhofer lines in sky spectra and the telluric filling-in of trace gas absorption features in ultraviolet and visible backscatter spectra are known as the Ring effect. Several studies (Kattawar et al., 1981; Joiner et al., 1995) have demonstrated that the main process responsible for the Ring effect is the rotational Raman scattering by molecular O_3 and N_2. In backscatter spectra, the Ring effect shows up as small-amplitude distortion, which follows Fraunhofer and absorption lines. For an inelastically scattering atmosphere, the radiative transfer equation includes an additional source term, the Raman source function, and the single and multiple scattering source terms have to be modified accordingly. Several radiative transfer models have been used to simulate the so-called Ring spectrum defined as the ratio of the inelastic and the elastic scattering radiances. These models include a Monte Carlo approach (Kattawar et al., 1981), a successive order of scattering method (Joiner et al., 1995) and a model which treats rotational Raman scattering as a first-order perturbation (Vountas et al., 1998; Landgraf et al., 2004; Spurr et al., 2008).

As Ring structures appear in the polarization signal, a complete simulation of the polarization-sensitive measurement requires a vector radiative transfer model which simulates Ring structures for all relevant Stokes parameters (Aben et al., 2001; Stam et al., 2002; Landgraf et al., 2004). The calculation of Ring spectra with a vector radiative transfer model is numerically expensive and approximation methods are desirable for large data sets. The numerical analysis performed in Landgraf et al. (2004) reveals that

(1) the polarization Ring spectra of Q and U are much weaker than those of the radiance I due to the low polarization of Raman scattered light;
(2) the combination of both a vector radiative transfer model, simulating the Stokes vector for an elastic scattering atmosphere, and a scalar radiative transfer approach, simulating the Ring spectrum for the radiance is sufficiently accurate for gas profile retrievals but not for applications involving the retrieval of cloud properties.

1.3.4 Radiative transfer in the IR and microwave

Neglecting scattering and assuming local thermodynamical equilibrium, the source function J is given by the Planck function at temperature T,

$$B(\nu, T) = \frac{2hc^2\nu^3}{\exp\left(\dfrac{hc\nu}{k_{\mathtt{B}}T}\right) - 1}. \tag{1.8}$$

The formal solution (1.6), describing the radiance I at wavenumber ν received by an instrument at position $\mathbf{r}_{\mathtt{o}}$, is given by the Schwarzschild equation

$$I(\nu, \mathbf{r}_{\mathtt{o}}) = I(\nu, \mathbf{r}_{\mathtt{s}})\mathcal{T}\left(\nu, \mathbf{r}_{\mathtt{o}}, \mathbf{r}_{\mathtt{s}}\right) + \int_{|\mathbf{r}_{\mathtt{o}} - \mathbf{r}_{\mathtt{s}}|} B\left(\nu, T(\mathbf{r})\right) \frac{\partial \mathcal{T}}{\partial s}\left(\nu, \mathbf{r}_{\mathtt{o}}, \mathbf{r}\right) \mathrm{d}s, \tag{1.9}$$

where $I(\nu, \mathbf{r}_{\mathtt{s}})$ is the background contribution at position $\mathbf{r}_{\mathtt{s}}$. The monochromatic transmission is computed according to Beer's law as

$$\mathcal{T}\left(\nu, \mathbf{r}_{\mathtt{o}}, \mathbf{r}\right) = \exp\left(-\int_{|\mathbf{r}_{\mathtt{o}} - \mathbf{r}|} \sigma_{\mathtt{abs}}(\nu, \mathbf{r}')\, \mathrm{d}s'\right) \tag{1.10}$$

$$= \exp\left(-\int_{|\mathbf{r}_{\mathtt{o}} - \mathbf{r}|} \mathrm{d}s' \sum_m C_{\mathtt{abs}m}\left(\nu, p\left(\mathbf{r}'\right), T\left(\mathbf{r}'\right)\right) n_m\left(\mathbf{r}'\right)\right). \tag{1.11}$$

Here, $\sigma_{\mathtt{abs}}$ is the absorption coefficient, p is the atmospheric pressure, n_m is the number density of molecule m, and $C_{\mathtt{abs}m}$ is its absorption cross-section.

In general, the molecular absorption cross-section is obtained by summing over the contributions from many lines. For an individual line at position $\hat{\nu}$, the cross-section is the product of the temperature-dependent line strength $S(T)$ and a normalized line shape function $g(\nu)$ describing the broadening mechanism(s), that is,

$$C_{\mathtt{abs}m}\left(\nu, p, T\right) = \sum_l S_{ml}\left(T\right) g\big(\nu, \hat{\nu}_{ml}, \gamma_{ml}\left(p, T\right)\big). \tag{1.12}$$

In the atmosphere, the combined effect of pressure broadening, corresponding to a Lorentzian line shape (indices m and l denoting molecule and line will be omitted for simplicity)

$$g_{\mathtt{L}}(\nu, \hat{\nu}, \gamma_{\mathtt{L}}) = \frac{1}{\pi} \frac{\gamma_{\mathtt{L}}}{(\nu - \hat{\nu})^2 + \gamma_{\mathtt{L}}^2}\,, \tag{1.13}$$

and Doppler broadening, corresponding to a Gaussian line shape

$$g_{\mathtt{D}}(\nu, \hat{\nu}, \gamma_{\mathtt{D}}) = \frac{1}{\gamma_{\mathtt{D}}} \left(\frac{\log 2}{\pi}\right)^{\frac{1}{2}} \exp\left(-\log 2 \left(\frac{\nu - \hat{\nu}}{\gamma_{\mathtt{D}}}\right)^2\right), \tag{1.14}$$

can be represented by a convolution, i.e., the Voigt line profile $g_{\mathtt{V}} = g_{\mathtt{L}} \otimes g_{\mathtt{D}}$. Pressure broadening (air-broadening, with self-broadening neglected) and Doppler broadening half-widths are given by

$$\gamma_{\mathtt{L}}(p, T) = \gamma_{\mathtt{L}0} \frac{p}{p_{\mathtt{ref}}} \left(\frac{T_{\mathtt{ref}}}{T}\right)^{\alpha}$$

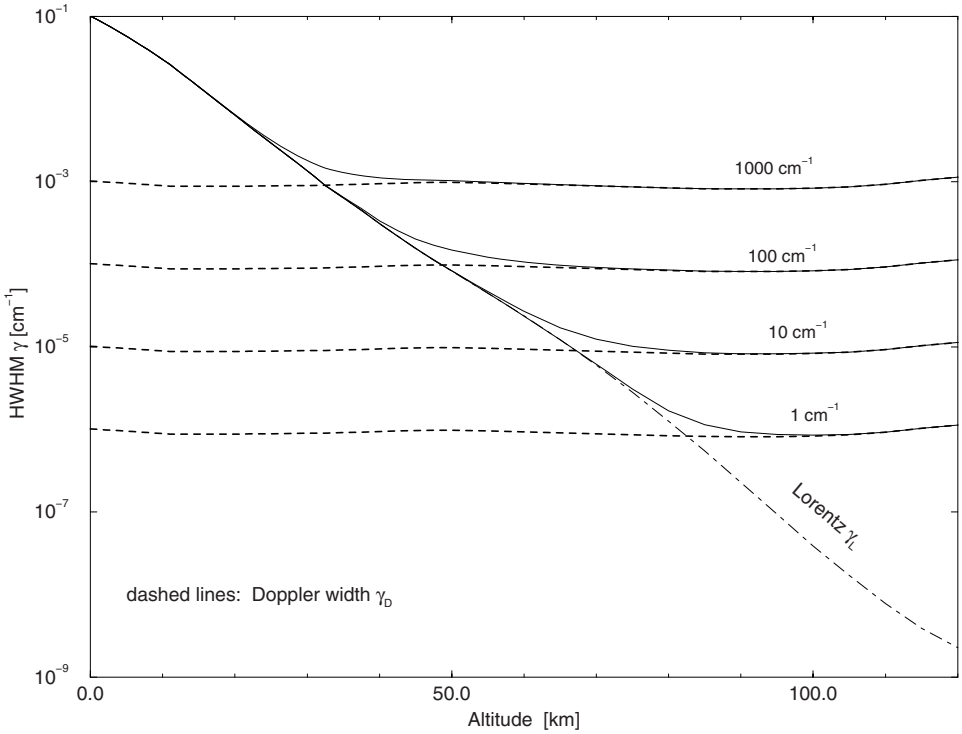

Fig. 1.7. Lorentz, Gauss and Voigt half-widths (HWHM) as a function of altitude in the Earth atmosphere for a variety of line positions $\hat{\nu}$. Pressure and temperature are from US Standard Atmosphere and the molecular mass is 36 amu.

and

$$\gamma_{\mathrm{D}}(T) = \hat{\nu}\sqrt{2\log 2\frac{k_{\mathrm{B}}T}{mc^2}},$$

respectively. Here, $p_{\mathtt{ref}}$ and $T_{\mathtt{ref}}$ are the reference pressure and temperature of line parameters, respectively, m denotes the molecular mass, and α describes the temperature dependence of pressure broadening. Note that pressure broadening dominates in the lower atmosphere; the transition altitude, where Doppler broadening becomes important, moves up from the middle stratosphere to the mesosphere with increasing wavelength (Figure 1.7).

Spectroscopic line parameters required for the calculation of the molecular absorption cross-sections, e.g., the line position $\hat{\nu}$, the line strength S, the temperature exponent α, the air-broadening half-width $\gamma_{\mathrm{L}0}$, and the lower state energy E (required to calculate $S\,(T)$ from the database entry $S\,(T_{\mathtt{ref}})$) have been compiled in various databases such as HITRAN (HIgh-resolution TRANsmission molecular absorption database), GEISA (Gestion et Etude des Informations Spectroscopiques Atmosphériques) and JPL (Jet Propulsion Laboratory) catalog. The latest versions of HITRAN (Rothman et al., 2009) and GEISA (Jacquinet-Husson et al., 2008) list parameters of some million transitions for several dozen molecules from the microwave ($\hat{\nu} = 10^{-6}$ cm^{-1}) to the ultraviolet ($\hat{\nu} \approx 25\,232$

and $\hat{\nu} \approx 35\,877\ \mathrm{cm}^{-1}$, respectively), whereas the JPL catalogue (Pickett et al., 1998) covers millions of rotational transitions in the microwave regime.

At a first glance the forward model appears to be much easier to solve in the infrared than in the ultraviolet as the source function is known. However, for high resolution atmospheric spectroscopy, the line-by-line (lbl) computation of (1.9) and (1.10) remains a challenging task because thousands of spectral lines have to be included in the sum (1.12). Moreover, as the monochromatic wavenumber grid point spacing determined by the half-widths of the spectral lines (cf. Figure 1.7) is very fine, accurate modeling of the spectrum may require thousands or even millions of spectral grid points. Finally, the convolution integral defining the Voigt line profile cannot be solved analytically, and numerical approximations have to be used.

In view of the computational challenges of lbl-modeling, alternative approaches have been used for low to moderate resolution spectra. Band models have been developed since the early days of radiative transfer modeling in meteorology and astrophysics (Goody and Yung, 1989; Liou, 2002; Thomas and Stamnes, 1999; Zdunkowski et al., 2007). More recently, the k-distribution and correlated k methods (Fu and Liou, 1992; Lacis and Oinas, 1991) or exponential sum fitting (Wiscombe and Evans, 1977) have been utilized.

Scattering is usually ignored in lbl models. However, if the analysis of data provided by spaceborne infrared sounders would be confined to clear sky observations only, a large fraction of data would be ignored. For nadir sounding, single scattering can be implemented with moderate effort, but multiple scattering, especially for limb sounding geometries, is still a challenging task. Various attempts have been described by Emde et al. (2004), Höpfner et al. (2002), Höpfner and Emde (2005), and Mendrok et al. (2007).

Intercomparisons of high-quality (laboratory and atmospheric) infrared spectra have revealed discrepancies with accurate model spectra obtained with the lbl approach (1.12). These deviations are commonly attributed to the so-called 'continuum', and a variety of explanations have been given in the literature, e.g., deviations of the far wing line profile from the Lorentzian line shape, contributions from water dimers $(H_2O)_2$ etc. For modeling infrared and microwave spectra, the semi-empirical approach developed by Clough et al. (1989) is widely used (see also Clough et al., 2005), whereas the empirical corrections due to Liebe et al. (1993) are frequently employed in the microwave regime.

When local thermodynamic equilibrium (LTE) is assumed, a local temperature can be assigned everywhere in the atmosphere, and thermal emission can be described by Planck's law of blackbody radiation (1.8). However, because temperature and radiation vary in space and time, the atmosphere is not in thermodynamic equilibrium. Nevertheless, the LTE assumption is justified in the troposphere and stratosphere, where the density of air is sufficiently high so that the mean time between molecular collisions is much smaller than the mean lifetime of an excited state of a radiating molecule. Thus, equilibrium conditions exist between vibrational, rotational and translation energy of the molecule. The breakdown of LTE in the upper atmosphere implies that the source function is no longer given by the Planck function. An adequate description of collisional and radiative processes under non-LTE conditions requires quantum theoretical considerations; see Lopez-Puertas and Taylor (2001) for an in-depth treatment.