



Institut für Geophysik, Astrophysik und Meteorologie  
Karl-Franzens-Universität Graz



**Wissenschaftlicher Bericht No. 11**

Institut für Geophysik, Astrophysik und Meteorologie (IGAM), Karl-Franzens-Universität Graz  
Universitätsplatz 5, A-8010 Graz, Austria

**Temperature Profiling by the  
Infrared Atmospheric  
Sounding Interferometer (IASI):  
Advanced Retrieval Algorithm and  
Performance Analysis**

by  
Elisabeth Weisz

Graz, May 2001

This report contains the Ph.D. thesis of Dr. Elisabeth Weisz, which she worked out during the years 1999 - 2001 at the IGAM/UG.



## **ACKNOWLEDGMENTS**

I would like to thank my supervisor Prof. Gottfried Kirchengast for his overall support and attendance during the last years. The work was financially supported by the START research award of G. Kirchengast funded by the Bundesministerium für Bildung, Wissenschaft und Kultur (BMBWK, Vienna, Austria) and managed under Programme No. Y130-CHE of the Fonds zur Förderung der wissenschaftlichen Forschung (FWF, Austria).

Thanks to all members at the Institute for Geophysics, Astrophysics, and Meteorology (IGAM), University of Graz, Austria. Special thanks to my colleagues Dr. Jeffrey Lerner, Dr. Andrea Steiner, and Dr. Ulrich Foelsche for the good teamwork.

I thank M. Matricardi (ECMWF, Reading, U.K.) for providing an early version of the RTIASI model and for related valuable advice. Further thanks are given to the ISSWG gathered by EUMETSAT, especially to P. Schluessel (EUMETSAT, Darmstadt, Germany), and to J. Eyre (The Met. Office, Bracknell, U.K.) for fruitful comments on various topics; P. Schluessel is also specifically thanked for providing a consolidated version of the RTIASI model (which is a property of EUMETSAT) and the recent IASI level 1c noise model. Thanks to the ECMWF for providing the atmospheric analysis fields, which have been kindly pre-processed for my applications by A. Steiner and U. Foelsche.

My sincere thanks go to my parents, my brothers and sisters, and my friends for all the support given to me.



# TABLE OF CONTENTS

	Introduction	1
<b>1</b>	<b>Atmospheric Radiative Transfer</b>	<b>5</b>
1.1	Introduction	5
1.2	Basic Radiation Quantities	9
1.3	Fundamental Laws	11
1.4	The Radiative Transfer Equation	14
1.4.1	General RT Equation	14
1.4.2	RT Equation in the Plane-parallel Atmosphere	17
1.5	Absorption Spectra of Atmospheric Gases	23
1.5.1	Absorption in the Infrared	24
1.5.2	Spectral Line Shape and Line Strength	28
1.5.3	Computation of Transmittance	32
<b>2</b>	<b>Meteorological Satellites and Instruments</b>	<b>35</b>
2.1	Satellite Orbits	35
2.1.1	Newton's and Kepler's Laws	35
2.1.2	Classical Orbital Elements	36
2.1.3	Meteorological Satellite Orbits	40
2.2	Meteorological Satellite Instruments	43
2.2.1	Types of Instruments	43
2.2.2	Data Transmission	44
2.2.3	Operational and Future Instruments	45
2.3	Operational Meteorological Satellites	49
2.3.1	Satellites in Geostationary Orbit	49
2.3.2	Satellites in Polar Orbits	53
<b>3</b>	<b>The IASI Instrument</b>	<b>55</b>
3.1	IASI Objectives and Main Specifications	56

3.2	Introduction to Fourier Transform Spectroscopy	60
3.2.1	Fourier Transform Analysis and Interferometry	61
3.2.2	Apodization and Instrument Response Function	65
3.3	IASI Functional Components	68
3.4	IASI Data and Products	71
<b>4</b>	<b>Retrieval of Physical Parameters from Measurements</b>	<b>73</b>
4.1	General Considerations on Retrieving Atmospheric Parameters	73
4.2	The Forward Model	75
4.2.1	Theoretical Introduction	75
4.2.2	The Fast Transmittance Model RTIASI	78
4.3	The Inverse Model	84
4.3.1	Optimal Linear Inversion	84
4.3.2	Optimal Non-linear Inversion	92
<b>5</b>	<b>IASI Temperature Retrieval and Performance Analysis</b>	<b>99</b>
5.1	Retrieval Inputs	99
5.1.1	Simulated Measurements and Measurement Error	99
5.1.2	<i>A priori</i> Profiles and <i>a priori</i> Error	101
5.2	Channel Selection	106
5.2.1	Information Content Theory	107
5.2.2	Maximum Sensitivity Approach	109
5.2.3	Channel Clustering	110
5.3	Standard Temperature Profile Retrieval	112
5.4	The "Hurricane Floyd" Case Study: Performance Analysis	115
5.4.1	Temperature Field Retrieval	116
5.4.2	Temperature Profile Retrieval Characteristics	124
5.4.3	Empirical Error Analysis	135
5.5	Retrievals for a Globally Distributed Ensemble	141
<b>6</b>	<b>Summary and Conclusions</b>	<b>149</b>
	Appendix	153
	List of Symbols	159
	List of Acronyms	165
	References	169

## INTRODUCTION

In spaceborne remote sensing sensors onboard of satellites acquire information about the Earth's surface (land and ocean) and atmosphere by means of detection, recording and analysis of electromagnetic (EM) radiation. The Earth absorbs the incoming solar radiation, which has a maximum at visible wavelengths and emits radiation mainly in the infrared range of the EM spectrum. Thus the amount of absorbed energy is balanced by the amount of radiation emitted back to space to maintain thermal equilibrium as observed over long time-scales (a year or longer). The characteristics of the modification of radiation when passing through the atmosphere depend on the amount and properties of the atmospheric constituents. Conversely, information on the Earth-atmosphere system and its constituents may be retrieved from the measured radiation. Earth observation by remote sensing is commonly restricted to the ultraviolet, visible, infrared (IR) and microwave region of the EM spectrum. In terms of wavelength the regions primarily exploited occupy the range from about 0.2  $\mu\text{m}$  ( $1 \mu\text{m} = 10^{-6}$  meter) to about 1 meter. In this work we primarily concentrate on the thermal IR portion, which is considered to range from about 3  $\mu\text{m}$  to about 20  $\mu\text{m}$ .

The emitted energy is recorded over finite wavelength bands by radiometers mounted on satellites. The main components of the instrument are detectors, which measure the received radiation and convert them to electric signals, and data processing components, which translate detected radiation into digital data. The first experimental meteorological satellite was launched in 1960 by the United States. Since then many satellites followed and the radiative data measured from spaceborne instruments became indispensable in research and operational meteorology. Today's global system of operational meteorological satellites includes geostationary types, which fly at an altitude of about 36000 km, and polar-orbiting satellites circling the Earth in low earth orbits (at about 800 km above surface). The last decade of development and technical improvements of satellites and instruments supports spaceborne remote sensing in an unprecedented way, providing imagery and data for a multitude of applications.

Temperature and gaseous profiles may be retrieved from the radiation measured by the sensors by detecting the upwelling radiance in discrete wavelength bands. Radiance measured in the center of

an absorption band arises from the upper atmospheric layers, while measurements at the wings will sense deeper into the atmosphere. Temperature profiles are obtained in the thermal IR by observations in the absorption band of carbon dioxide ( $\text{CO}_2$ ), which is a relatively abundant gas of known and uniform distribution. Other constituents absorbing in the thermal IR are water vapor ( $\text{H}_2\text{O}$ ), ozone ( $\text{O}_3$ ), nitrous oxide ( $\text{N}_2\text{O}$ ), methane ( $\text{CH}_4$ ), and carbon monoxide ( $\text{CO}$ ). Atmospheric windows, where attenuation is minimal, are used to obtain surface and cloud properties. By observing at very high spectral resolution (high number of narrow bands) at several wavelengths, different layers of the atmosphere can be sensed. The physical relation between the measured radiance at the top of the atmosphere (TOA) and temperature is described by the radiative transfer equation, which gives the radiance at TOA as a sum of an emission term from the surface and of weighted Planck intensities emitted from every atmospheric layer. The weighting functions are related to the transmittances, which depend on the absorption coefficient and the density of the relevant absorbing gas. The Planck intensity describes the spectral radiance emitted by a blackbody. In order to comprehensively resolve the vertical structure of temperature profiles, sensors furnishing high spectral resolution, i.e., measuring in a large number of appropriate narrow channels, are necessary.

New instruments such as the Infrared Atmospheric Sounding Interferometer (IASI) have the potential to provide substantial improvement in accuracy and vertical resolution of retrieved temperature profiles over current instruments supporting weather forecasting and meteorological research. IASI will fly onboard the polar-orbiting METOP satellite series operated by EUMETSAT and ESA, with the first launch planned for late 2005. The instrument measures infrared radiation in more than 8000 channels, which will support retrievals of high resolution profiles but which also poses new problems on data reduction. Efficient retrieval algorithms need to be developed to fully exploit the potential of the instrument. Such retrieval algorithms rely on two processes. The first is referred to as the forward model, which calculates the radiance emitted by the atmosphere corresponding to a certain state of the atmosphere. In this work we will employ the fast transmittance model RTIASI (Matricardi and Saunders, 1999) to simulate IASI measurements. The inversion of the forward model constitutes the second process, termed the inverse model, which estimates the temperature from measured radiance. A physical-statistical approach, incorporating the underlying radiative transfer physics and models of the uncertainties in the measurements and of the prior knowledge on the state, has been utilized to solve the inverse problem.

This report starts with an introduction on radiative transfer (Chapt. 1), followed by an overview on atmospheric sounding in current operational meteorology (Chapt. 2). A comprehensive description of the future instrument IASI (Chapt. 3), on which we focused the work, and a detailed overview on the optimal estimation retrieval methodology employed (Chapt. 4), complete the more general part. In the specific part (Chapt. 5), where most of our own original work is contained, we carefully investigate the potential of IASI to provide accurate temperature profiles. We show and discuss the results of a range of computational experiments regarding channel selection, theoretical retrieval accuracy and resolution, retrieval scheme characteristics, as well as empirical error estimates. A summary and a collection of main conclusions (Chapt. 6) close the report.



# 1 ATMOSPHERIC RADIATIVE TRANSFER

## 1.1 Introduction

The Earth and its atmosphere exchange energy with space by means of radiative transfer to maintain thermal equilibrium. Radiation from the sun and the Earth is modified by various physical processes (absorption, emission, scattering) as it propagates through the atmosphere. A quantitative description of the radiation field resulting from the interaction between matter and radiant energy is known as the theory of radiative transfer.

The major atmospheric constituents are molecular nitrogen  $N_2$  (78% by volume) and molecular oxygen  $O_2$  (21% by volume). The remaining ~1% includes water vapor ( $H_2O$ ), carbon dioxide ( $CO_2$ ), ozone ( $O_3$ ) and other trace gases (Tab. A3). Constituents may be classified in permanent and variable gases. Aerosols, clouds and precipitation vary highly in space and time. Although of small amount, variable gases such as water vapor are extremely important for radiative transfer.

The atmosphere can be categorized by its thermal structure, which is determined in large parts by radiative transfer, as well as by vertical air motions. Fig. 1.1 illustrates a global mean vertical temperature profile versus height (in km) and pressure (in kPa, kilo Pascal). Usually we use millibar (mbar) units, with  $1 \text{ mbar} = 10^2 \text{ Pa} = 1 \text{ hPa}$ . Pressure  $p$  decreases with height  $z$  according to the hydrostatic balance  $dp/dz = -\rho g$ , depending on the acceleration of gravity  $g$  and the air density  $\rho$ , described by  $\rho = p/RT$  (temperature  $T$ , specific gas constant  $R$ ).

The atmosphere is heated by solar radiation, but primarily from the surface by absorption of terrestrial radiation and by turbulent heat transfer. Temperature decreases with height at an average lapse rate of  $6.5^\circ\text{C}$  per kilometer in the troposphere, which is dominated by convective motions. The stratosphere is prevailed by radiative processes. Absorption of ultraviolet radiation by ozone leads to a temperature increase. The temperature value at the stratopause may exceed  $0^\circ\text{C}$  (273 Kelvin). Convective motion and radiative processes are determinative in the mesosphere, where the temperature decreases again. In the thermosphere temperature rises due to the absorption of extreme ultraviolet radiation by  $O_2$  and atomic oxygen ( $O$ ). From surface to 100 km turbulent air

motions yield homogenous mixing ratios (i.e., 0.78 for  $N_2$  and 0.21 for  $O_2$ ). Therefore the mean molecular mass (Tab. A2) of the mixture stays constant up to 100 km. This well-mixed region is called the homosphere. Above 100 km diffusion stratifies constituents according to their molecular mass. Thus, heavier species (e.g.,  $O_2$ ) decrease more rapidly with altitude than light species (e.g., He). In this region, which is termed the heterosphere, the mean molecular mass decreases monotonically with height. Above about 200 km atomic oxygen (O) becomes the predominant gas.

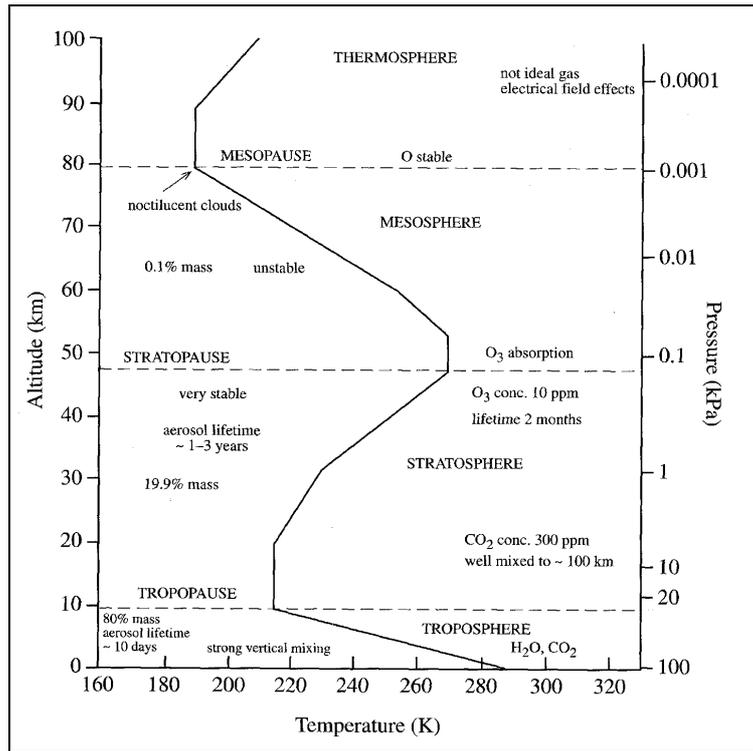


Figure 1.1: Typical vertical temperature profile (from Wells, 1997).

The incident solar flux (i.e., radiant energy per unit time and unit area) at the top of the atmosphere (TOA) is about  $1372 \text{ Wm}^{-2}$ , known as the solar constant and commonly denoted by  $S$ . The global-mean solar flux incident on TOA is  $S / 4$ , about  $343 \text{ Wm}^{-2}$ . The distribution of the solar flux averaged over a certain period is termed solar insolation and is a function of latitude as well as of the orbital characteristics of the Earth around the sun. Over long time scales the Earth-atmosphere system is in thermal equilibrium. Consequently, absorbed solar fluxes are balanced by thermal infrared fluxes emitted by the Earth-atmosphere system. The globally averaged energy budget is illustrated in Fig. 1.2 where the global-mean incident solar flux of  $343 \text{ Wm}^{-2}$  is represented by 100 units. The energy balance can be expressed by

$$(1 - A)S\pi r^2 = 4\pi r^2 \sigma T_e^4, \quad (1.1.1)$$

where  $4\pi r^2$  represents the surface area of a spherical Earth with radius  $r$ . The left hand side of Eq. 1.1.1 describes the incident solar flux of cross-sectional area  $\pi r^2$ . The albedo  $A$  defines the portion of incoming solar flux reflected back to space. The right hand side represents the flux emitted by the Earth-atmosphere system based on the Stefan-Boltzmann law, which describes the flux  $F = \pi B$  emitted by a blackbody as  $\pi B = \sigma T^4$ , where  $B$  is the Planck constant and  $\sigma$  the Boltzmann constant (Tab. A1). Inserting  $S = 1372 \text{ Wm}^{-2}$  and  $A = 0.3$  in Eq. 1.1.1 gives a surface Earth temperature of about 255 Kelvin (K). However, the standard profile in Fig. 1.1 depicts a surface temperature of about 288 K. This discrepancy can be explained by the greenhouse effect, which is controlled by strong absorption of infrared radiation by atmospheric constituents such as water vapor ( $\text{H}_2\text{O}$ ) and clouds, carbon dioxide ( $\text{CO}_2$ ), ozone ( $\text{O}_3$ ), methane ( $\text{CH}_4$ ), nitrous oxide ( $\text{N}_2\text{O}$ ), aerosols and chlorofluorocarbons (CFCs).

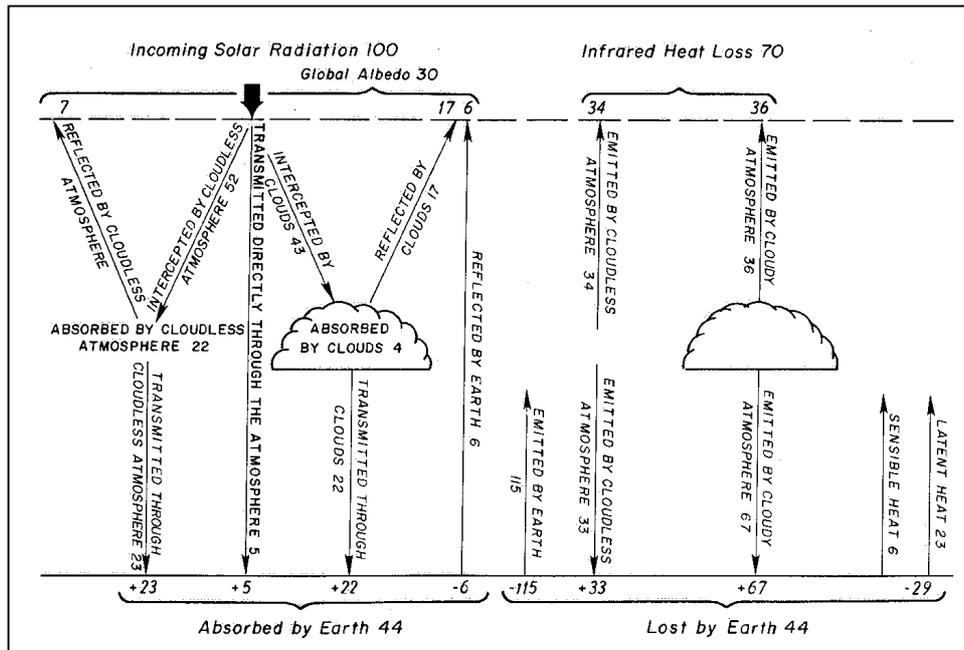
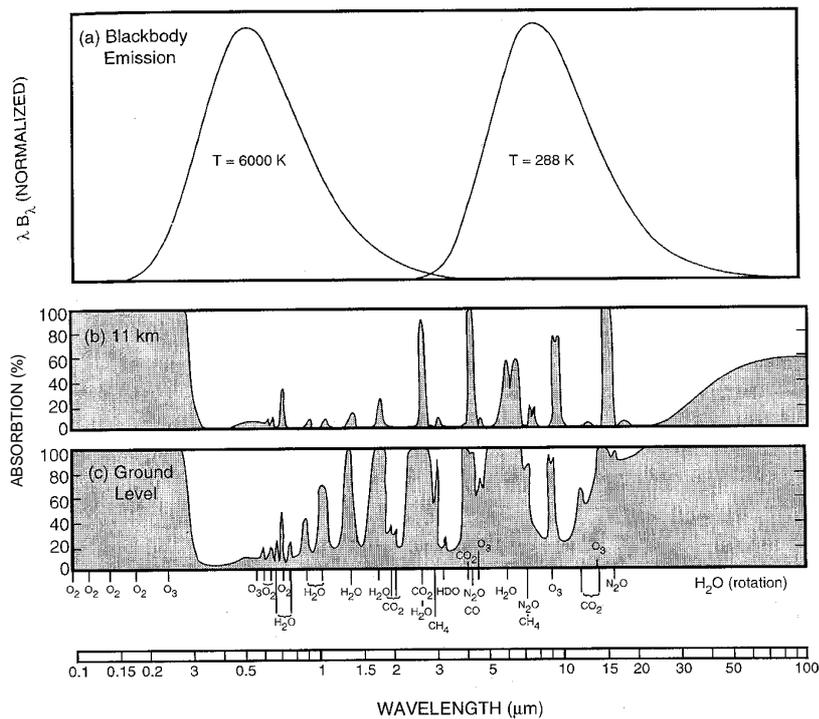


Figure 1.2: Energy balance of the Earth-atmosphere system, where global-mean incoming solar energy is regarded as 100 units (from Liou, 1992).

Due to their distinct temperature, the sun and the Earth emit electromagnetic radiation in different spectral ranges. Both virtually behave as blackbodies. A blackbody is a perfect emitter and complete absorber of radiation at all wavelengths. Fig. 1.3a shows the blackbody emission curves for 6000 K (sun) and 288 K (Earth). The radiation from the sun consists roughly of 10% ultraviolet ( $0.01\text{--}0.4 \mu\text{m}$ ), 45% visible ( $0.4\text{--}0.8\mu\text{m}$ ) and 45% infrared radiation (most in  $0.8\text{--}100\mu\text{m}$ ), with a maximum at  $0.48 \mu\text{m}$  and is referred to as shortwave (SW) radiation. The Earth emits wavelengths with a peak intensity at about  $10 \mu\text{m}$  in the infrared range from about 3 to  $100 \mu\text{m}$ , known as

longwave (LW) radiation. The blackbody spectra are modified by various mechanisms brought into play by atmospheric constituents as a function of wavelength (Fig. 1.3b, Fig. 1.3c). SW radiation is attenuated primarily by reflection by clouds and aerosols, scattering by air molecules smaller than the wavelength (Rayleigh scattering) and by absorption. Wavelengths shorter than  $0.1 \mu\text{m}$  are absorbed by  $\text{N}_2$ ,  $\text{O}_2$  and atomic nitrogen (N) and oxygen (O) at altitudes above 100 km and wavelengths shorter than  $0.24 \mu\text{m}$  are absorbed by  $\text{O}_2$ , which is then dissociated into two oxygen atoms above about 50 km. Atomic oxygen recombines with  $\text{O}_2$  to form ozone. Ozone ( $\text{O}_3$ ) absorbs radiation between  $0.2$  and  $0.3 \mu\text{m}$  at high levels (above 50 km) leading to dissociation of the  $\text{O}_3$  molecule into molecular and atomic oxygen. Ozone also absorbs in the visible range above  $0.4 \mu\text{m}$ , which becomes important at altitudes around 25 km. Absorption by  $\text{O}_3$  constitutes the primary source of temperature increase in the stratosphere. Water vapor ( $\text{H}_2\text{O}$ ) absorbs in the near IR ( $\sim 0.6\text{--}3 \mu\text{m}$ ) in several narrow bands from about  $0.9$  to  $2.1 \mu\text{m}$ . Note that the atmosphere is quite transparent to electromagnetic radiation in the region of peak emission from the sun, i.e., around  $0.4 \mu\text{m}$ .



**Figure 1.3:** (a) Normalized blackbody curves for sun (6000 K) and Earth (288 K), (b) atmospheric absorption as a function of wavelength at 11 km, and (c) atmospheric absorption at ground level. (From Salby, 1996.)

In the infrared (IR) extending from  $0.76 \mu\text{m}$  to  $100 \mu\text{m}$  most attenuation is by absorption, primarily due to the effects of  $\text{H}_2\text{O}$ ,  $\text{CO}_2$  and  $\text{O}_3$ . In addition, methane ( $\text{CH}_4$ ), nitrous oxide ( $\text{N}_2\text{O}$ ) and to a lesser extent carbon monoxide ( $\text{CO}$ ), sulfur dioxide ( $\text{SO}_2$ ), ammonia ( $\text{NH}_3$ ) and CFCs have narrow absorption bands in the infrared. Except for  $\text{O}_3$  virtually all the absorption occurs in the

troposphere, where absolute concentration is large. As mentioned above, gases absorbing in the infrared are referred to as greenhouse gases, since LW radiation is trapped in the lower atmosphere by repeated absorption and re-emission. Only in the atmospheric window at wavelengths of 8–12  $\mu\text{m}$  absorption is weak enough such that infrared radiation passes freely through the atmosphere.

## 1.2 Basic Radiation Quantities

Electromagnetic (EM) radiation consists of varying electric and magnetic fields, which oscillate perpendicular to each other with the same amplitude. The EM wave propagates perpendicular to both fields with the speed of light  $c$ . It can be described by its wavelength  $\lambda$  (distance between crests) or by the frequency  $\nu$  (rate at which the fields oscillate observed at one point). Wavelength is measured in meters (m) or in fractions of meters such as  $\mu\text{m}$  ( $10^{-6}$  m), whereas frequency is specified in Hertz (Hz or 1/s). These quantities are related by

$$c = \nu\lambda . \quad (1.2.1)$$

Another important quantity especially when dealing with the infrared portion of the spectrum is the wavenumber  $\nu = 1/\lambda$  in units of  $\text{m}^{-1}$  (or often  $\text{cm}^{-1}$  for convenience).

As will be seen later every wavelength can be associated with a photon of energy  $h\nu$ , where  $h$  is the Planck constant (Tab. A1). Photon energy is calculated in units of Joule (J or  $\text{kgm}^2\text{s}^{-2}$ ) and converted to electron volts (eV) with  $1\text{eV} = 1.60219 \times 10^{-19}$  J.

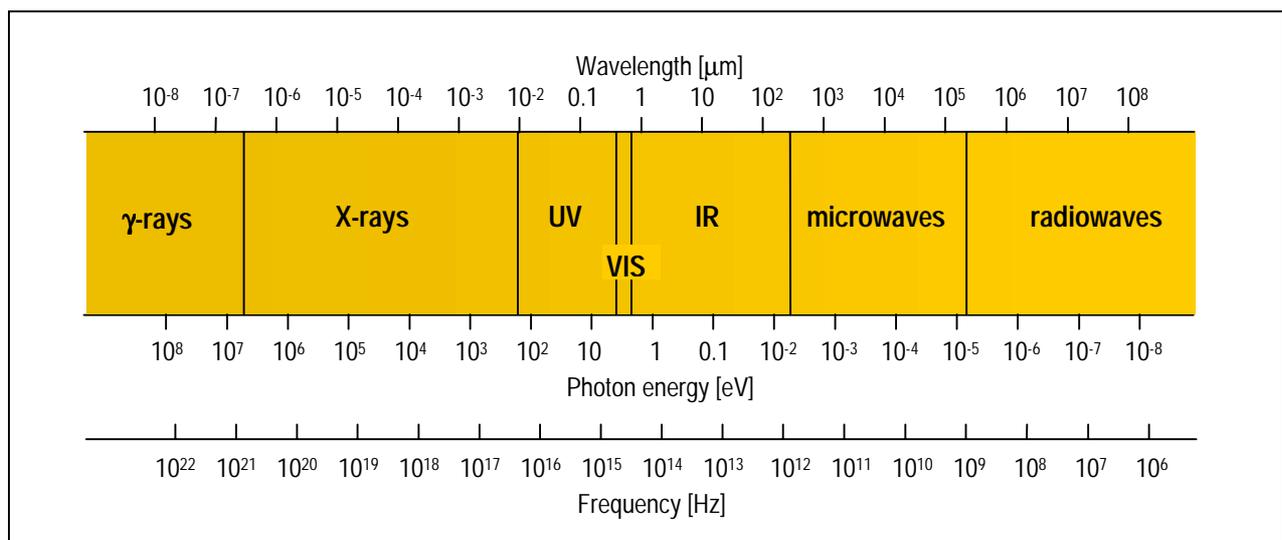


Figure 1.4: The electromagnetic spectrum composed of gamma-rays, x-rays, ultraviolet (UV), visible (VIS), infrared (IR), microwaves and radiowaves.

Electromagnetic radiation comprising photon energy levels from high (high frequency, short wavelengths) to low (low frequency, long wavelengths) gives the electromagnetic spectrum illustrated in Fig. 1.4. The spectral ranges in Fig. 1.4 can be subdivided in smaller regions. Frequency, wavelength and wavenumber ranges for the infrared (IR) subranges is given in Tab. 1.1. Note that the classification of the EM spectrum may differ slightly in various literature. The IR region can also be divided in two categories based on their radiation properties. The reflected IR covers wavelengths from the visible end to approximately 3  $\mu\text{m}$  while the thermal IR represents the radiation emitted by the Earth's surface in form of heat covering the remaining wavelengths up to 100  $\mu\text{m}$ .

Name of band	Frequency [Hz]	Wavelength [ $\mu\text{m}$ ]	Wavenumber [ $\text{cm}^{-1}$ ]
Visible (violet, blue, green, yellow, orange, red)	$7.89 \times 10^{14} - 3.95 \times 10^{14}$	0.38 – 0.76	26316 – 13158
Near Infrared (NIR)	$3.95 \times 10^{14} - 2 \times 10^{14}$	0.76 – 3	13158 – 3333
Thermal Infrared (TIR)	$2 \times 10^{14} - 1 \times 10^{14}$	3 – 15	3333 – 667
Far Infrared (FIR)	$1 \times 10^{14} - 3 \times 10^{12}$	15 – 100	667 – 100

Table 1.1: Frequency, wavelength and wavenumber ranges for the visible and infrared spectral regions.

Radiation depends on direction. This aspect requires the introduction of the solid angle  $\Omega$ , which is defined as the angle subtended by the center of a sphere by an area on its surface. It is measured in steradian (sr) and can be regarded as the conical slice of space. The solid angle of an object, which completely surrounds a point is  $4\pi$  [sr]. Mathematically, the differential element of the solid angle is

$$d\Omega = \sin \theta \, d\theta \, d\phi = d\mu \, d\phi, \quad (1.2.2)$$

where  $\theta$  is the zenith angle and  $\phi$  represents the azimuth angle (Fig. 1.5) and  $\mu = \cos \theta$ .

A radiation beam occupies the solid angle element  $d\Omega$  and can be attenuated proportional to the density and absorbing characteristics of the prevailing medium. Energy passing through this beam decreases due to absorption and scattering out of the beam, while energy increases if energy is emitted or scattered into  $d\Omega$ . We consider radiation incident in direction  $\hat{\mathbf{k}}$  on an area element with unit vector  $\hat{\mathbf{n}}$  (see Fig. 1.4).

The radiant flux is defined as the radiant energy per unit time, whereas radiant flux density is the flux crossing a unit area. Radiant quantities associated with a specific wavenumber (wavelength or frequency) are termed monochromatic or spectral. The main quantity is the monochromatic radiance (or intensity)  $I_\nu$ , defined as the energy per unit time per unit wavenumber (or wavelength)

per unit solid angle crossing a unit area perpendicular to the radiation beam (Fig. 1.5). If the wavenumber is specified by  $\text{cm}^{-1}$  the unit of radiance is  $\text{Wm}^{-2}\text{cm sr}^{-1}$ . If intensity can be considered independent of direction such as  $I_\nu = I_\nu(\hat{\mathbf{k}})$  in all directions  $\hat{\mathbf{k}}$ , then the flux is  $F_\nu = \pi I_\nu$ . A summary of commonly used radiant quantities is given in Tab. 1.2.

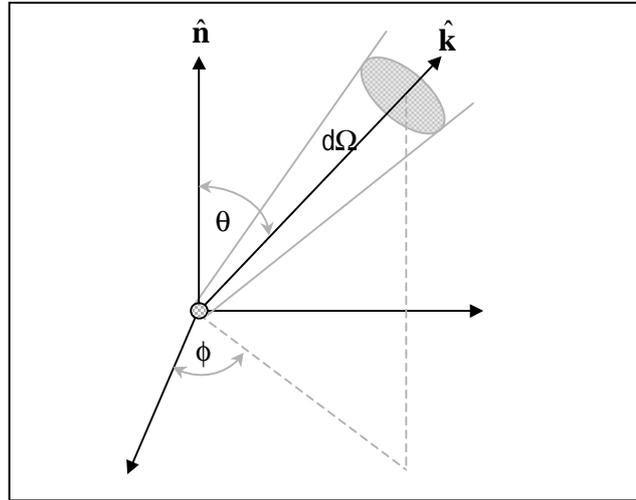


Figure 1.5: Radiation beam occupying the increment of solid angle  $d\Omega$ .

Quantity	Symbol	Explanation	Unit (SI)
Radiant energy	Q		J (Joule)
Radiant flux	F	Power (i.e. radiant energy per unit time)	W (Watt), $\text{J s}^{-1}$
Radiant flux density	M/E	Radiant flux (Power) per unit area	$\text{W m}^{-2}$
Radiant exitance	M	Radiant flux density emerging an area	$\text{W m}^{-2}$
Irradiance	E	Radiant flux density incident on an area	$\text{W m}^{-2}$
Radiance	I or R	Radiant flux density per unit solid angle	$\text{W m}^{-2} \text{sr}^{-1}$

Table 1.2: Basic radiation quantities.

### 1.3 Fundamental Laws

All bodies above absolute zero temperature emit radiation. A body which absorbs all incoming radiation is known as a blackbody. To explain observed blackbody spectra Max Planck (1858–1947) postulated that the radiation of frequency  $\nu$  assumes energy amounts only in discrete quanta described by

$$E = h\nu, \quad (1.3.1)$$

where  $h$  is the Planck constant. Individual molecules absorb or emit radiation in discrete pieces, in photons, carrying energy multiples of  $h\nu$ . *Planck's law* describes the spectral intensity  $B_\nu(T)$ , often specified in  $\text{Wm}^{-2}\text{cm sr}^{-1}$ , emitted by the blackbody, given its temperature (in Kelvin),

$$B_\nu(T) = \frac{2h\nu^3 c^2}{\exp[hc\nu/kT] - 1} \quad (1.3.2)$$

at wavenumber  $\nu$  (in  $\text{cm}^{-1}$ ), where  $k$  is the Boltzmann constant and  $c$  is the speed of light (Tab. A1). The Planck function (Eq. 1.3.2) can be more compactly written as

$$B_\nu(T) = \frac{c_1 \nu^3}{\exp[c_2 \nu/T] - 1}, \quad (1.3.3)$$

where the first and the second radiation constant are given by  $c_1 = 2hc^2$  and  $c_2 = hc/k$ , respectively. For the exact values see the appendix, Tab. A1. The Planck function can also be expressed in terms of wavelength

$$B_\lambda(T) = \frac{c_1}{\lambda^5 (\exp[c_2/\lambda T] - 1)}. \quad (1.3.4)$$

Fig. 1.6 shows the Planck function versus wavelength for several temperatures. For long wavelengths (low frequency) the Planck radiation may be approximated by the *Rayleigh-Jeans law*

$$B_\lambda(T) = \frac{c_1 T}{c_2 \lambda^4}, \quad (1.3.5)$$

which constitutes the result of classical physics to describe emitted blackbody radiation. For a wavelength of order of centimeters and millimeters, Eq. 1.3.5 simply states that the Planck intensity is proportional to temperature of the blackbody.

Integrating  $B_\nu(T)$  over all wavenumbers gives the total flux emitted by a blackbody described by the *Stefan-Boltzmann Law*

$$F = \pi B(T) = \pi \int_0^\infty B_\nu(T) d\nu = \sigma T^4, \quad (1.3.6)$$

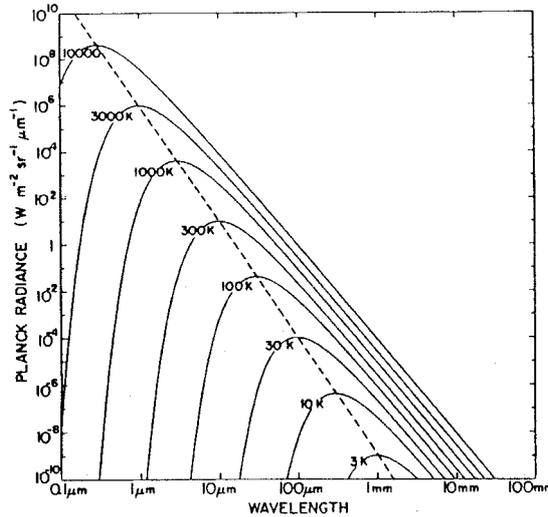
where  $\sigma$  is the Stefan-Boltzmann constant (Tab. A1). For a radiance  $I_\nu$  emitted by an arbitrary object there exists an associated temperature referred to as the *brightness temperature*  $T_B$  at which a blackbody would emit the same radiation. Hence we may write

$$T_B \equiv B_\nu^{-1}(I_\nu) = (c_2 \nu) / \ln[1 + (c_1 \nu^3 / I_\nu)]. \quad (1.3.7)$$

For any temperature  $T$  the Planck function has a single maximum at a certain wavelength which follows from the *Wien's displacement law*

$$\lambda_{\max} = \frac{2897}{T} \quad (1.3.8)$$

with  $T$  in units of K and  $\lambda_{\max}$  in units of  $\mu\text{m}$ .



**Figure 1.6:** Planck function versus wavelength for indicated temperatures (from Kidder and Vander Haar, 1995).

The wavelength of maximum intensity  $\lambda_{\max}$  decreases with increasing temperature (Fig. 1.6). For instance, solar temperature of 6000 K yields radiation concentrated around 0.5  $\mu\text{m}$ , whereas the 288 K mean surface temperature of the Earth corresponds to a wavelength of about 10  $\mu\text{m}$ .

Most solids and liquids behave like blackbodies and the preceding laws described by Eqs. 1.3.2 to 1.3.7 can be applied. But gases do not behave like blackbodies and their emission is less than for an equivalent blackbody. *Kirchhoff's law* applies for the so-called gray bodies stating that

$$\alpha_{\nu} = \varepsilon_{\nu} \quad (1.3.9)$$

where the absorptance  $\alpha_{\nu}$  and the emittance  $\varepsilon_{\nu}$  of the body are given by

$$\alpha_{\nu} = \frac{\text{absorbed radiation at } \nu}{\text{incident radiation at } \nu} \quad (1.3.10)$$

$$\varepsilon_{\nu} = \frac{\text{emitted radiation at } \nu}{B_{\nu}(T)} \quad (1.3.11)$$

Therefore, for a gray body the emitted flux is  $\varepsilon_{\nu}\sigma T^4$  and the absorbed flux is  $\alpha_{\nu}\sigma T^4$ . Eq. 1.3.9 states that the body emits and absorbs radiation of a certain wavenumber with equal efficiency.

Note that Kirchhoff's law is valid only if the material is in local thermodynamic equilibrium. This is a good assumption for the atmospheric regions below 80 km. Incident radiation can be modified not only by absorption but it can as well be reflected and scattered away at various angles or just transmitted through the medium. Thus, we define the reflectance  $\rho_\nu$  and the transmittance  $T_\nu$  as

$$\rho_\nu = \frac{\text{reflected radiation at } \nu}{\text{incident radiation at } \nu}, \quad (1.3.12)$$

$$T_\nu = \frac{\text{transmitted radiation at } \nu}{\text{incident radiation at } \nu}. \quad (1.3.13)$$

Since energy is conserved, we have  $\alpha_\nu + \rho_\nu + T_\nu = 1$ . Note that the quantities described by Eqs. 1.3.10 to 1.3.13 are unitless, since they constitute ratios (to irradiance).

## 1.4 The Radiative Transfer Equation

The description of this section is mostly based on Salby (1996) and Hanel et al. (1992).

### 1.4.1 General RT Equation

The RT equation gives the change in radiance as the radiation passes through the atmosphere. If we consider a beam of radiation, there are four processes that can modify the radiation: absorption (A), emission (B), scattering out of the beam (C), and scattering into the beam (D) from other directions, respectively. We may write

$$\frac{dI_\nu}{ds} = -A + B - C + D, \quad (1.4.1)$$

giving the rate of change of radiance of wavenumber  $\nu$  for an incremental distance  $ds$  (known as slant path). The negative and positive signs in Eq. 1.4.1 reflect losses or gains of the energy passing through the radiation beam. Interactions A and C represent portions lost from the radiation field by absorption and scattering, respectively. Terms B and D represent portions gained by the radiation field by emission and scattering into the beam.

To describe the (scattering) processes one considers radiation incident on a volume element  $dV$  (containing a number of particles) in direction  $(\mu, \phi)$  and scattered by  $dV$  through the scattering angle  $\psi$  into the direction  $(\mu', \phi')$ . The directions are specified by  $\mu$  (the cosine of the zenith angle  $\theta$ ) and by the azimuth angle  $\phi$ . The scattering geometry is illustrated in Fig. 1.7.

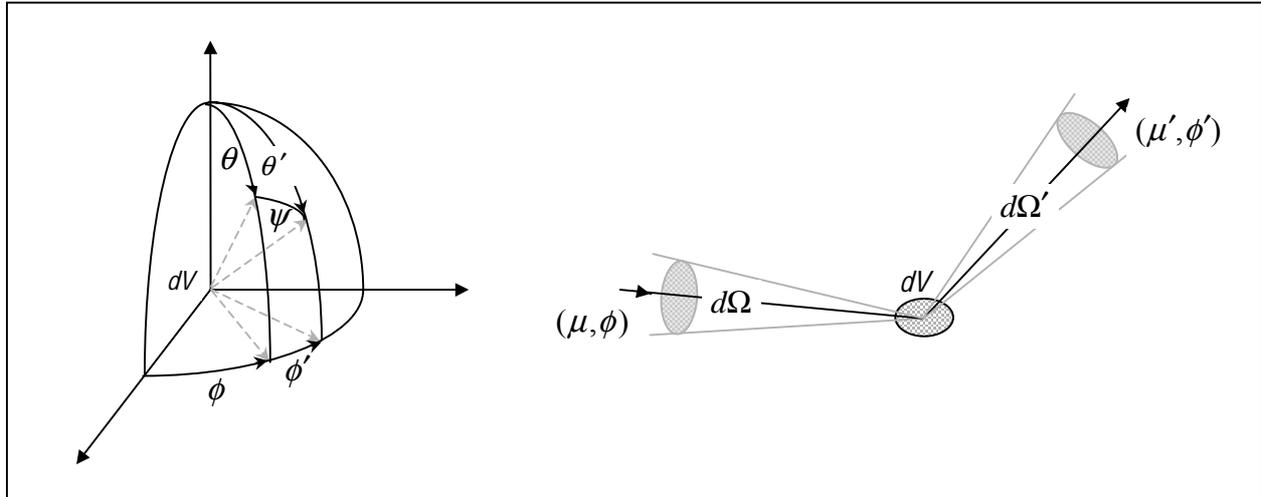


Figure 1.7: Scattering geometry. Radiation incident in direction  $(\mu, \phi)$  and scattered through angle  $\psi$  in direction  $(\mu', \phi')$  where  $\mu = \cos \theta$  and  $\mu' = \cos \theta'$ .

### ABSORPTION

Lambert's (or Beer's, or Lambert-Beer's, or Beer-Lambert's) Law describes the attenuation of radiance of intensity  $I_\nu$  passing through a medium with thickness  $ds$  according to

$$dI_\nu = -I_\nu \rho \sigma_{av} ds. \quad (1.4.2)$$

$\rho$  is the density of the absorber and  $\sigma_{av}$  is the mass absorption coefficient (or specific absorption cross section), which describes the characteristic absorption spectrum of each atmospheric species in SI units of  $\text{m}^2\text{kg}^{-1}$ . It can be regarded as an illuminated area lost by absorption. Another quantity often referred to as the length absorption coefficient is defined as  $\rho \sigma_{av}$  and measured in  $\text{m}^{-1}$ . Term A in Eq. 1.4.1 is then  $\rho \sigma_{av} I_\nu$  representing the portion lost from the incoming direction  $(\mu, \phi)$  by absorption in  $dV$ .

Integrating Eq. 1.4.2 along a finite path gives

$$I_\nu(s) = I_\nu(0) \exp \left[ - \int_0^s \rho(s') \sigma_{av}(s') ds' \right], \quad (1.4.3)$$

where  $I_\nu(s)$  and  $I_\nu(0)$  are the radiation intensities entering and leaving the layer. If no scattering and emission is considered, the transmittance  $T_\nu$  is

$$T_\nu = \frac{I_\nu(s)}{I_\nu(0)} = \exp \left[ - \int_0^s \rho(s') \sigma_{av}(s') ds' \right]. \quad (1.4.4)$$

EMISSION

Particles in  $dV$  emit radiation into the solid angle element  $d\Omega$ . The basis of thermal emission is blackbody radiation described by the fundamental laws in Sect. 1.3. Kirchhoff's law states that a body emits the same amount as it absorbs. Hence term B in Eq. 1.4.1 can be described by  $\rho\sigma_{av}B_v(T)$ .

SCATTERING OUT

This process may be considered as a scattering of a certain fraction of the incident radiation into  $d\Omega'$ . The scattering cross section  $\sigma_{sv}$  symbolizes the fractional area removed through scattering from the area illuminated by a beam. Lambert's Law (Eq. 1.4.2) applies here as well and term C in Eq. 1.4.1 has the form  $\rho\sigma_{sv}I_v$ , where  $\rho$  is the density of the absorber.

SCATTERING IN

This process may be considered as a scattering of a certain fraction of the incident radiation into  $d\Omega$ . Radiation is scattered into the beam from different directions. The directionality is described in the scattering phase function  $p_v(\psi)$ , which gives the portion of the radiation from the direction  $(\mu, \phi)$  into the direction  $(\mu', \phi')$ , where  $\psi$  is the angle between the incidence direction  $(\mu, \phi)$  and the scattering direction  $(\mu', \phi')$ . In other words, the phase function describes the angular distribution of radiation scattered through the scattering angle  $\psi$ . The phase function  $p_v(\psi) = p_v(\mu, \phi; \mu', \phi')$  is normalized according to

$$\frac{1}{4\pi} \int_0^{4\pi} p_v(\psi) d\Omega' = 1, \quad (1.4.5)$$

with  $d\Omega' = \sin\theta' d\theta' d\phi'$ . Now, the term  $\rho\sigma_{sv} \frac{1}{4\pi} p_v(\psi) d\Omega'$  represents the part of radiation lost through scattering from the beam in direction  $(\mu, \phi)$  scattered into the beam in direction  $(\mu', \phi')$ . Thus, term D can be regarded as the directionally averaged incoming radiation weighted by the phase function

$$D = \frac{\rho\sigma_{sv}}{4\pi} \int_0^{4\pi} I_v(\mu', \phi') p_v(\psi) d\Omega'. \quad (1.4.6)$$

Combining terms A to D of Eq. 1.4.1 yields

$$\frac{dI_v}{ds} = \rho\sigma_{av}(-I_v + B_v(T)) - \rho\sigma_{sv}\left(I_v + \frac{1}{4\pi} \int_0^{4\pi} I_v(\mu', \phi') p_v(\psi) d\Omega'\right). \quad (1.4.7)$$

Effects of scattering and absorption can be gathered into the *extinction cross section*

$$k_v = \sigma_{av} + \sigma_{sv}. \quad (1.4.8)$$

The *single scattering albedo*  $\omega_v$  is defined to be the ratio of the radiant portion scattered in all directions (i.e., scattering cross section) to the portion extinguished by absorption and scattering (i.e., extinction cross section). Thus we have

$$\omega_v = \frac{\sigma_{sv}}{k_v}, \quad (1.4.9)$$

describing the fraction of radiation lost through scattering out of the beam. Then,  $1 - \omega_v = \sigma_{av} / k_v$  represents the portion lost by absorption, also known as the absorption number.

Utilizing these quantities the radiative transfer equation becomes

$$\frac{dI_v}{ds} = \rho k_v \left[ -I_v + (1 - \omega_v) B_v(T) - \frac{\omega_v}{4\pi} \int_0^{4\pi} I_v(\mu', \phi') p_v(\psi) d\Omega' \right]. \quad (1.4.10)$$

It is convenient to use the following shorter version

$$\boxed{\frac{1}{\rho k_v} \frac{dI_v}{ds} = -I_v + J_v}, \quad (1.4.11)$$

with the total so-called source function  $J_v$  combining emission and scattering given by

$$\boxed{J_v = (1 - \omega_v) B_v(T) - \frac{\omega_v}{4\pi} \int_0^{4\pi} I_v(\mu', \phi') p_v(\psi) d\Omega'}. \quad (1.4.12)$$

## 1.4.2 RT Equation in the Plane-parallel Atmosphere

### GENERAL CONSIDERATIONS

Within the framework of a plane-parallel atmosphere we assume that the curvature of the Earth is ignored, the absorbing medium is homogeneous and the radiation field is horizontally isotropic.

This implies that variations in intensity and atmospheric parameters only occur in the vertical direction.

The slant path  $ds$  can be converted to a vertical incremental depth  $dz$  using

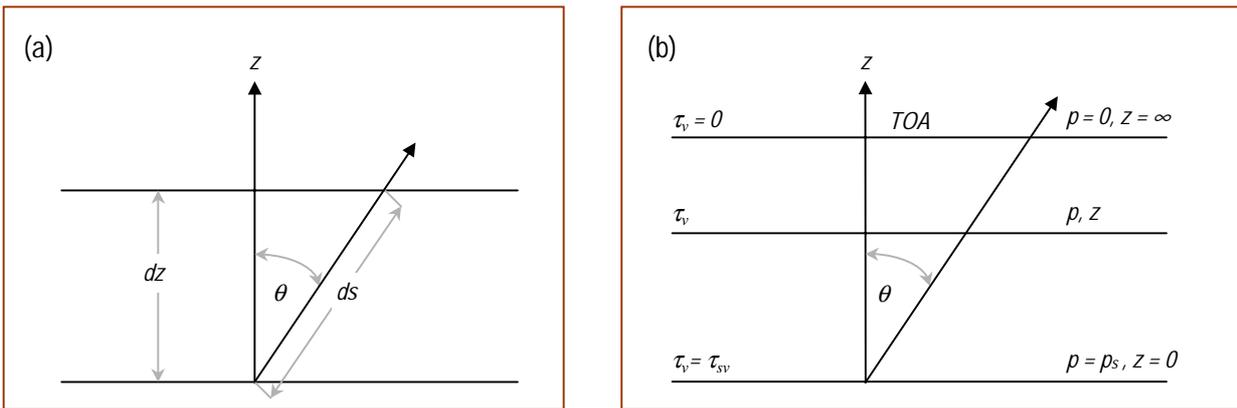
$$ds = dz / \mu, \quad (1.4.13)$$

where  $\mu = \cos \theta$ , i.e., the cosine of the zenith angle (Fig. 1.8). Therefore intensity becomes a function of the vertical path  $z$  and the zenith angle  $\theta$ .

We introduce the optical depth  $\tau_v$  defined as

$$\tau_v = \int_z^{\infty} \rho(z') k_v(z') dz', \quad (1.4.14)$$

which is measured downward from the top of the atmosphere ( $\tau_v = 0$ ) to the surface value  $\tau_{sv}$ .



**Figure 1.8:** Plane parallel atmosphere. (a) Relation of depth to slant path. (b) Coordinate system in optical depth  $\tau_v(z)$ , pressure  $p$  and height  $z$ . Subscript  $s$  denotes surface and TOA stands for top of the atmosphere.

Optical depth along the density-weighted path length  $u$  is defined as  $\tau_v = \int_0^u k_v(u') du'$ . Note that at the level of unit optical depth ( $\tau_v = 1$ ) the absorption of radiation is strongest, and if  $k_v$  is large, then the level of unit optical depth will be high in the atmosphere.

Fig. 1.8b shows the coordinate system in optical depth, pressure and height  $z$ . Since the incremental optical depth is  $d\tau = -k_v(z)\rho(z)dz$  from Eq. 1.4.14 the RT equation can be written in optical depth coordinates as

$$\mu \frac{dI_v}{d\tau_v} = I_v - J_v, \quad (1.4.15)$$

and the source function is given by

$$J_v = (1 - \omega_v)B_v(\tau_v) + \frac{\omega_v}{4\pi} \int_0^1 \int_{-1}^1 I_v(\tau_v, \mu', \phi') p_v(\mu, \phi; \mu', \phi') d\mu' d\phi', \quad (1.4.16)$$

with  $I_v = I_v(\tau_v, \mu, \phi)$ ,  $J_v = J_v(\tau_v, \mu, \phi)$  and  $B_v(T) = B_v(T(\tau_v)) = B_v(\tau_v)$ .

In practice, this monochromatic field consists of a diffuse component originating from thermal emission of the atmosphere and the surface as well as a component (diffuse and direct) from the sun (Hanel et al., 1992). Diffuse components arise from radiation undergoing multiple scattering events. The component from the atmosphere dominates in the thermal and far infrared, whereas solar emission contributes at visible and near infrared wavelengths. The source function separated in diffuse and direct components can be written as

$$J_v = (1 - \omega_v)B_v(\tau_v) + \frac{\omega_v}{4\pi} \int_0^1 \int_{-1}^1 I_v(\tau_v, \mu', \phi') p_v(\mu, \phi; \mu', \phi') d\mu' d\phi' + \frac{\omega_v F_0}{4} \exp[-\tau_v / \mu] p_v(\mu, \phi; -\mu_0, \phi_0). \quad (1.4.17)$$

The additional term accounts for the direct component associated with the exponential attenuation of unscattered solar flux. The position of the sun is denoted by  $(\mu_0, \phi_0)$ , where  $\mu_0$  is the cosine of the solar zenith angle  $\theta_0$  and  $\pi F_0$  is the direct solar flux.

We now concentrate on the upwelling (upward) intensity, i.e., from the surface ( $\tau_v = \tau_{s,v}$ ) to the TOA ( $\tau_v = 0$ ). In this case we have  $0 < \mu \leq 1$ . We multiply both sides of Eq. 1.4.15 by the factor  $\exp[-\tau_v / \mu]$  to get an exact differential on the left hand side,

$$\mu \frac{d(\exp[-\tau_v / \mu] I_v)}{d\tau_v} = \exp[-\tau_v / \mu] J_v, \quad (1.4.18)$$

with the source function (Eq. 1.4.17). We integrate from  $\tau_v$  at height  $z$  to  $\tau_{s,v}$  at the surface and represent the source function by its individual components

$$\begin{aligned}
I_v(\tau_v, \mu, \phi) \exp[-\tau_v / \mu] = & \\
& I_v(\tau_{s,v}, \mu, \phi) \exp[-\tau_{s,v} / \mu] \\
& + \frac{1}{\mu} \int_{\tau_v}^{\tau_{s,v}} (1 - \omega_v) \exp[-\tau'_v / \mu] B_v(\tau'_v) d\tau'_v \\
& + \frac{\omega_v}{4\pi\mu} \int_{\tau_v}^{\tau_{s,v}} \exp[-\tau'_v / \mu] \int_0^1 \int_{-1}^1 I_v(\tau'_v, \mu', \phi') p_v(\mu, \phi; \mu', \phi') d\mu' d\phi' d\tau'_v \\
& + \frac{\omega_v F_0}{4\mu} \int_{\tau_v}^{\tau_{s,v}} \exp\left[-\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)\tau'_v\right] p_v(\mu, \phi; -\mu_0, \phi_0) d\tau'_v.
\end{aligned} \tag{1.4.19}$$

Solving for the upward intensity  $I_v(\tau_v, \mu, \phi)$  we find (Hanel et al., 1992)

$$\begin{aligned}
I_v(\tau_v, \mu, \phi) = & I_v(\tau_{s,v}, \mu, \phi) \exp[-(\tau_{s,v} - \tau_v) / \mu] \\
& + \frac{1}{\mu} \int_{\tau_v}^{\tau_{s,v}} (1 - \omega_v) \exp[-(\tau'_v - \tau_v) / \mu] B_v(\tau'_v) d\tau'_v \\
& + \frac{\omega_v}{4\pi\mu} \int_{\tau_v}^{\tau_{s,v}} \int_0^1 \int_{-1}^1 \exp[-(\tau'_v - \tau_v) / \mu] \exp[-(\tau_{s,v} - \tau'_v) / \mu'] I_v(\tau_{s,v}, \mu', \phi') p_v(\mu, \phi; \mu', \phi') d\mu' d\phi' d\tau'_v \\
& + \frac{\omega_v F_0}{4\mu} \int_{\tau_v}^{\tau_{s,v}} \exp[-\tau'_v / \mu_0] \exp[-(\tau'_v - \tau_v) / \mu] p_v(\mu, \phi; -\mu_0, \phi_0) d\tau'_v \\
& + \frac{\omega_v}{4\pi\mu} \int_{\tau_v}^{\tau_{s,v}} \int_0^1 \int_{-1}^1 \exp[-(\tau'_v - \tau_v) / \mu] I_v(\tau'_v, \mu', \phi') p_v(\mu, \phi; \mu', \phi') d\mu' d\phi' d\tau'_v.
\end{aligned} \tag{1.4.20}$$

The first term on the right hand side of Eq. 1.4.20 describes radiation arriving from the surface and attenuated by the atmosphere between  $\tau_{s,v}$  and  $\tau_v$ . The second term represents the radiation thermally emitted at by the atmosphere between  $\tau_{s,v}$  and  $\tau_v$ . The third term accounts for radiation from the surface in the direction  $(\mu', \phi')$  that is scattered at  $\tau'_v$  in the direction  $(\mu, \phi)$ . The fourth term represents radiation from the sun at level  $\tau'_v$  before undergoing a scattering event. Finally, the last term describes radiation that has undergone one or more scattering processes before scattered at  $\tau'_v$  into the direction  $(\mu, \phi)$ .

NO SCATTERING

Many situations allow the assumption that scattering is negligible, especially when we consider infrared radiation in the absence of clouds. In a non-scattering atmosphere the RT equation takes the form

$$\boxed{\mu \frac{dI_v}{d\tau_v} = I_v - B_v(\tau_v)}, \quad (1.4.21)$$

which is known as Schwarzschild's equation. The upward intensity at the top of the atmosphere (TOA) can be found from Eq. 1.4.20

$$I_v(0, \mu, \phi) = I_v(\tau_{s,v}, \mu, \phi) \exp[-\tau_{s,v} / \mu] + \int_0^{\tau_{s,v}} \exp[-\tau'_v / \mu] B_v(\tau'_v) \frac{d\tau'_v}{\mu}, \quad (1.4.22)$$

recalling that  $\tau_v = 0$  at TOA.

For many practical applications the atmosphere can be regarded as divided into thin layers. The emitted intensity of the  $j$ -th layer can be described by the second term on the right hand side of Eq. 1.4.20,

$$I_v(\tau_{j-1,v}, \mu) = B_v(\tau_{j,v}) [1 - \exp[-(\tau_{j,v} - \tau_{j-1,v})]], \quad (1.4.23)$$

within the layer the Planck function can be considered constant. If we repeat Eq. 1.4.23 for every layer, the sum of these contributions will represent the upwelling intensity at TOA

$$I_v(0, \mu) = \sum_{j=1}^n \exp[-\tau_{j-1,v} / \mu] B_v(\tau_{j,v}) [1 - \exp[-(\tau_{j,v} - \tau_{j-1,v})]], \quad (1.4.24)$$

where  $\tau_{0,v} = 0$ ,  $\tau_{n-1,v} = \tau_{s,v}$  and  $B_v(\tau_{j,v})$  is the mean Planck function between  $\tau_{j,v}$  and  $\tau_{j-1,v}$  (layer above layer  $j$ ). Note that the optical depth  $\tau_{s,v}$  at the surface can be regarded as the total optical depth.

The net flux at level  $\tau$  is given by

$$F_v(\tau) = 2\pi \int_{-1}^1 I(\tau, \mu) \mu d\mu. \quad (1.4.25)$$

If we consider overhead viewing (i.e.,  $\mu = 1$ ), then we may write, instead of Eq. 1.4.22,

$$I_v(0) = I_v(\tau_{s,v}) \exp[-\tau_{s,v}] + \int_0^{\tau_{s,v}} \exp[-\tau'_v] B_v(\tau'_v) d\tau'_v. \quad (1.4.26)$$

We introduce the monochromatic transmittance (or transmission function)  $T_v(u)$  for an inhomogeneous path length  $u$  as

$$T_v(u) = \exp[-\tau_v] = \exp\left[-\int_0^u k_v(u') du'\right], \quad (1.4.27)$$

where  $k_v = \sigma_{av}$  is the absorption coefficient in  $\text{m}^2\text{kg}^{-1}$  in the absence of scattering, and the path length  $u$  for an absorber with density  $\rho$  is given by

$$u = \int_0^z \rho(z') dz'. \quad (1.4.28)$$

If a homogenous path length is assumed, we have  $T_v(u) = \exp[-k_v u]$ .

In terms of transmittances the solution of the RT equation without scattering is expressed by

$$I_v = I_v(\tau_{s,v})T_v(\tau_{s,v}) + \int_{\tau_{sv}}^0 B_v(\tau'_v) \frac{dT_v(\tau'_v)}{d\tau'_v} d\tau'_v. \quad (1.4.29)$$

Note that  $\frac{dT_v(\tau_v)}{d\tau_v} = -\exp[-\tau_v]$ , and  $T_v(\tau_{s,v})$  represents the surface transmittance to TOA and

$T_v(\tau_v)$  the atmospheric transmittance from level  $\tau_v$  to TOA. The Earth's surface may be considered as a blackbody and it is convenient to substitute  $I_v(\tau_{s,v}, \mu, \phi) = B(T(\tau_{s,v}))$ . To be strict, the infrared radiation from the surface is slightly less than that of a blackbody and an emissivity coefficient  $\varepsilon$  with values between 0.90 and 0.95 should be introduced such that  $I_v(\tau_{s,v}, \mu, \phi) = \varepsilon_{s,v} B(T(\tau_{s,v}))$ .

In height coordinates Eq. 1.4.29 becomes

$$I_v = B_v(0)T_v(0) + \int_0^\infty B_v(T(z)) \frac{dT_v(z)}{dz} dz, \quad (1.4.30)$$

or in pressure coordinates,

$$I_v = B_v(p_s)T_v(p_s) + \int_{p_s}^0 B_v(T(p)) \frac{dT_v(p)}{dp} dp, \quad (1.4.31)$$

where  $p_s$  represents the surface pressure. The weighting function  $W_v(p)$  may be defined in terms of the logarithm of pressure

$$W_\nu(p) = -\frac{dT_\nu(p)}{d \ln p}, \quad (1.4.32)$$

representing the weight of the Planck intensity contributions to the upwelling radiance. Thus,  $W_\nu(p)$  determines those height regions which are sensed from space at wavenumber  $\nu$ .

## 1.5 Absorption Spectra of Atmospheric Gases

General absorption spectra have been shown in Fig. 1.3. We will concentrate now on absorption of infrared radiation by atmospheric gases such as water vapor, carbon dioxide and ozone. Fig. 1.9 depicts absorption spectra of several atmospheric constituents in the infrared range. It seems that the molecules produce continuous absorption bands. For instance, water vapor absorbs strongly in a band centered around  $6.3 \mu\text{m}$ . At higher resolution we can find that the absorption features in Fig. 1.9 consist of a complex array of lines. If we consider a large number of photons interacting with a volume element containing a large number of molecules, it is sufficient to use classical physics to discuss radiative transfer (Hanel et al., 1992).

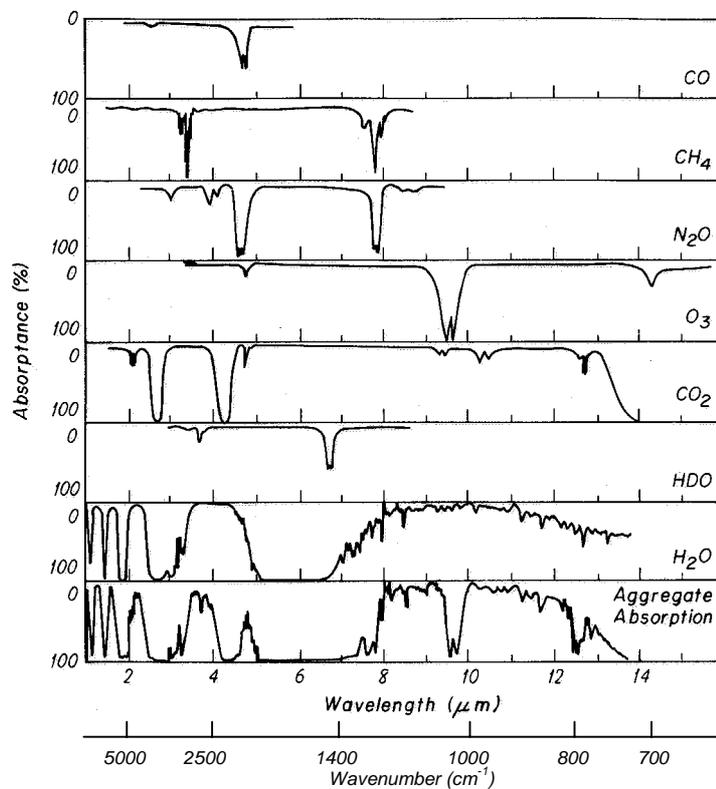


Figure 1.9: Absorption of thermal infrared radiation by atmospheric gases (after Liou, 1992).

Since single photons interact with individual atoms and molecules, we need to utilize quantum mechanics, where the atoms and molecules change their internal energy by absorption and

emission of energy in discrete amounts, in order to get improved insight in the problem. A molecule consists of several atoms, which are bound together by sharing valence electrons. Individual atoms vibrate with respect to one another, while the molecule as a whole rotates around any spatial axis. The internal energy of a molecule comprises electronic ( $E_{el}$ ), vibrational ( $E_{vib}$ ) and rotational ( $E_{rot}$ ) energies. The energy in each form assumes only discrete values and is therefore called quantized. The kinetic energy of molecular translation denoted by  $E_{trans}$ , which assumes values in a continuous range, completes the total energy of an isolated molecule, thus

$$E = E_{trans} + E_{el} + E_{vib} + E_{rot} . \quad (1.5.1)$$

The magnitudes of these forms of energies are of different orders such that electronic transition spectra cover the ultraviolet and visible part of the spectra, the vibrational spectra occupy the near infrared and rotational spectra the far infrared and microwave regions. Molecules can alter their states by absorbing or emitting discrete quantities from the radiation field. In absorption the molecule captures a photon and moves to a higher level of internal energy. This process modifies the radiation field and produces spectral absorption lines. In emission the molecule releases a photon and transits to a lower energy level. This process produces spectral emission lines. In general all internal energy forms vary simultaneously during absorption and emission, particularly pure electronic and pure vibrational spectra do not exist. As mentioned above electronic transitions, where orbital electrons jump between discrete energy levels, occur mostly in the UV and in the visible portion, concerning molecules like  $O_2$  and  $O_3$ . Vibrational transitions, where molecule change their vibrational energy states, occur mostly in the infrared portion. The energy states are characterized by the vibrational quantum number  $v$  obeying the selection rule  $\Delta v = \pm 1, \pm 2, \pm 3, \dots$ . For each level there exist rotational levels characterized by the rotational quantum number  $J$  with  $\Delta J = \pm 1$  for diatomic molecules and  $\Delta J = -1, 0, +1$  for polyatomic molecules. Therefore vibrational transitions are coupled with rotational transitions. In the atmosphere most molecules are in the ground vibrational states implying that observed transitions are fundamental. Fundamental refers to transition from the ground state to the first excited state ( $\Delta v = \pm 1$ ). Pure rotational transitions ( $\Delta v = 0$ ) occur in the far-infrared and microwave regions.

### 1.5.1 Absorption in the Infrared

We shall consider absorption spectra of the principal absorbing gases mainly in the thermal infrared region. The absorption bands important in thermal infrared radiation are summarized in Tab. 1.3. The second column gives the central wavenumber  $\nu$  of the particular band. The information given below is mostly based on Liou (1992).

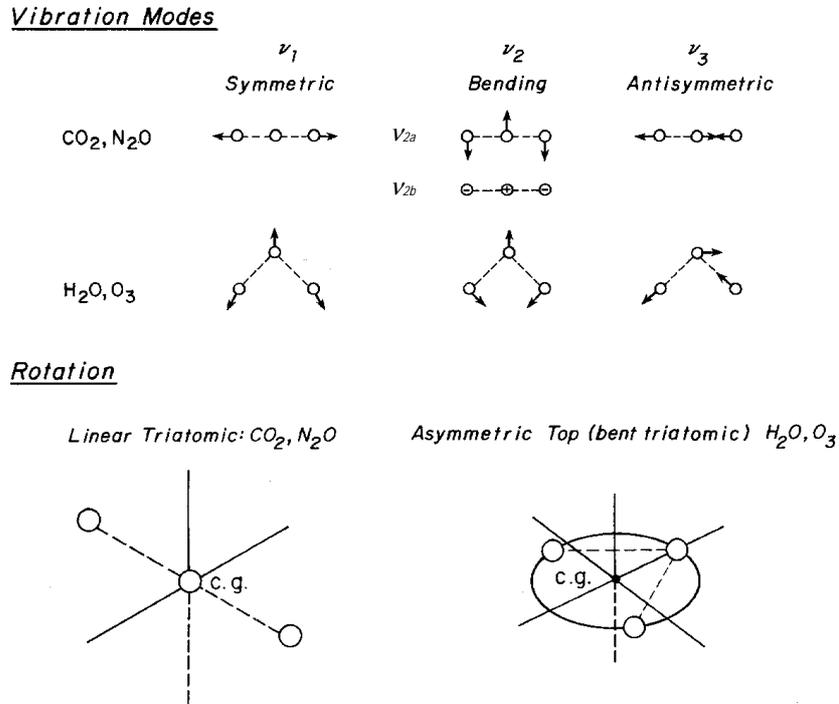
### CARBON DIOXIDE

The linearly symmetric CO<sub>2</sub> molecule with the carbon atom in the middle has three vibrational modes illustrated schematically in Fig. 1.10. The CO<sub>2</sub> molecule has several isotopes which are present in the atmosphere. The most significant are C<sup>12</sup>O<sub>2</sub><sup>16</sup>, C<sup>13</sup>O<sub>2</sub><sup>16</sup> and C<sup>13</sup>O<sup>16</sup>O<sup>18</sup> with relative abundances of about 98.8, 1.1 and 0.4 %, respectively. The fundamental symmetrical stretch mode  $\nu_1$  may be found near 7.5  $\mu\text{m}$ . This mode does not have an electric dipole moment and is therefore radiatively inactive (i.e., it does not interact with incident radiation). CO<sub>2</sub> has no pure rotational transitions since only molecules with a permanent electric dipole moment produce pure rotational spectra. The 15  $\mu\text{m}$  band represents the bending mode  $\nu_2$  which is twofold degenerate and consists of  $\nu_{2a}$  and  $\nu_{2b}$  (Fig. 1.10). The  $\nu_2$  fundamental transition is coupled with rotational transitions corresponding to  $\Delta J = -1, 0, +1$ . Line series originating from these changes are referred to as P, Q, and R branches. P and R branches spread out from the band center, whereas the lines from the Q branch are clustered around the band center. The  $\nu_3$  fundamental vibration-rotation band is responsible for high absorption at wavenumber 4.3  $\mu\text{m}$ . Since this mode is parallel (Fig. 1.10), it does not possess a Q branch. In addition to the strong absorption bands at 4.3  $\mu\text{m}$  and 15  $\mu\text{m}$ , in which absorption is primarily due to the fundamental  $\nu_2$  and  $\nu_3$  transitions, CO<sub>2</sub> has bands centered near 10.6, 9.4, 5.2, 4.8, 2.7, 2.0, 1.6, and 1.4  $\mu\text{m}$  (Zuev, 1974). The CO<sub>2</sub> molecule has also several combination (simultaneous transitions in two vibrational modes) and overtone (vibrational transitions between nonadjacent levels) bands (see Tab. 1.3).

### WATER VAPOR

The H<sub>2</sub>O molecule has an asymmetric configuration (bent triatomic) with the oxygen atom at the vertex. The most important isotopes of H<sub>2</sub>O are H<sub>2</sub><sup>16</sup>O, H<sub>2</sub><sup>18</sup>O and HD<sup>16</sup>O which are present in the atmosphere with 99.9, 0.2 and 0.02 %, respectively. The molecule has three vibrational modes (Fig. 1.10), where  $\nu_2$  has the lowest wavenumber. The fundamental  $\nu_2$  band centered around 6.3  $\mu\text{m}$  is the most intensive absorption band in the vibration-rotation spectrum. The centers of the band  $\nu_3$  and  $\nu_2$  are located at about 2.66  $\mu\text{m}$  and 2.74  $\mu\text{m}$ , respectively. Other vibration-rotation bands of H<sub>2</sub>O can be identified between 1.8 and 0.7  $\mu\text{m}$ . The fine structure of the vibration-rotation spectrum is very complex and consists of thousands of individual lines. The pure rotational absorption band occurs in the wavenumber range  $\sim 0\text{--}1000\text{ cm}^{-1}$  that corresponds to wavelengths of several centimeters to about 10  $\mu\text{m}$ . Inspecting Fig. 1.9 we notice an (almost transparent) atmospheric window from 8–12  $\mu\text{m}$  ( $1250\text{--}830\text{ cm}^{-1}$ ) intervened by the ozone absorption band at

9.6  $\mu\text{m}$ . But also  $\text{H}_2\text{O}$  absorbs continuously in small amounts in this region, caused by collision broadening between  $\text{H}_2\text{O}$  molecules and between  $\text{H}_2\text{O}$  and non-absorbing molecules such as  $\text{N}_2$ .



**Figure 1.10:** Vibration modes of triatomic atmospheric molecules and rotation axes for linear and asymmetric top molecules (after Liou, 1992).

## OZONE

The  $\text{O}_3$  molecule has an asymmetric top configuration (Fig. 1.10). Electronic transitions occur in the ultraviolet region (Hartley and Huggins band) and in the visible region (Chappius band), see, e.g., Salby (1996). The atmosphere mostly contains three ozone isotopes, these are  $\text{O}_3^{16}$ ,  $\text{O}^{16}\text{O}^{18}\text{O}^{16}$  and  $\text{O}^{16}\text{O}^{16}\text{O}^{18}$ . The  $\nu_1$  and  $\nu_3$  fundamental vibration modes are situated at 9.1 and 9.6  $\mu\text{m}$  and constitute the 9.6  $\mu\text{m}$  band located in the atmospheric window 8–12  $\mu\text{m}$  (1250–830  $\text{cm}^{-1}$ ). This band contains a large number of fine-structure lines. The  $\nu_2$  fundamental band is centered at 14.27  $\mu\text{m}$ . Overtone and combination frequencies of  $\text{O}_3$  vibrations produce several bands between 2.7 and 5.7  $\mu\text{m}$  of which the 4.75  $\mu\text{m}$  band is the most intensive one.

### METHANE

The CH<sub>4</sub> molecule exhibits a spherical top configuration. Since it has no permanent electric dipole moment, CH<sub>4</sub> has no pure rotational spectrum. The  $\nu_1$  band is fully symmetric and  $\nu_2$  is twofold degenerate. Both are inactive and have centers near 3.4 and 6.6  $\mu\text{m}$ . The active  $\nu_3$  and  $\nu_4$  bands are threefold degenerate and are centered at 3.3 and 7.7  $\mu\text{m}$ , respectively. In addition, methane possesses a rich spectrum of overtone and combination frequencies primarily located in the solar spectrum.

### NITROUS OXIDE

The N<sub>2</sub>O molecule is a linear asymmetric molecule with an electric dipole moment. Strong electronic bands are located in the far ultraviolet region. The three fundamental modes  $\nu_1$ ,  $\nu_2$ ,  $\nu_3$  are active in the IR and have centers at 7.8, 17.0 and 4.6  $\mu\text{m}$ , respectively. The  $\nu_4$  fundamental band overlaps the  $\nu_4$  fundamental band of methane. N<sub>2</sub>O has many bands of overtones and combination frequencies. For instance, the region between 1.3 and 2.5  $\mu\text{m}$  contains 18 bands.

### CARBON MONOXIDE

The fundamental rotational band lies near 4.67  $\mu\text{m}$ , while overtone and combination frequencies lie between 1.2 and 2.3  $\mu\text{m}$ . The pure rotational spectrum can be found in the FIR and microwave region.

### CHLOROFLUOROCARBONS

Methylchloride (CH<sub>3</sub>Cl), carbon tetrachloride (CCl<sub>4</sub> or CFC-10), trichlorofluoromethane (CFCl<sub>3</sub> or CFC-11), methylchloroform (CH<sub>3</sub>CCl<sub>3</sub>), and dichlorodifluoromethane (CF<sub>2</sub>Cl<sub>2</sub> or CFC-12) have several fundamental transitions primarily located in the atmospheric window at 8–12  $\mu\text{m}$ . These gases may significantly contribute to the greenhouse effect if their atmospheric abundance is increased.

Band	$\nu$ [cm <sup>-1</sup> ]	Transition	Band interval [ $\mu\text{m}$ ]	Band interval [cm <sup>-1</sup> ]
CO <sub>2</sub> 15 $\mu\text{m}$	667	$\nu_2$ ; P,Q,R, combination	18.2 – 12.5	550 – 800
CO <sub>2</sub> 10.6 $\mu\text{m}$	961 1063.8	Overtone, combination	11.8 – 9.1	850 – 1100
CO <sub>2</sub> 4.3 $\mu\text{m}$	2349	$\nu_2$ ; P,Q,R, combination	4.8 – 4.2	2100 – 2400
H <sub>2</sub> O rotation		P, R	> 1 cm – 10 $\mu\text{m}$	~ 0 – 1000
H <sub>2</sub> O 6.3 $\mu\text{m}$	1594.8	$\nu_2$ ; P,R	15.6 – 3.6	640 – 2800
H <sub>2</sub> O continuum			~ 50 – 8.3	~ 200 – 1200
O <sub>3</sub> 9.6 $\mu\text{m}$	1110 1043	$\nu_1$ ; P,R $\nu_3$ ; P,R	10.5 – 8.3 16.7 – 12.5	950 – 1200 600 – 800
O <sub>3</sub> 14.2 $\mu\text{m}$	705	$\nu_2$ ; P,R	16.7 – 12.5	600 – 800
O <sub>3</sub> 4.8 $\mu\text{m}$	2105	Overtone, combination		
CH <sub>4</sub> 7.6 $\mu\text{m}$	1306.2	$\nu_4$	10.5 – 6.1	950 – 1650
N <sub>2</sub> O 7.9 $\mu\text{m}$	1285.6	$\nu_1$	8.3 – 7.4	1200 – 1350
N <sub>2</sub> O 17.0 $\mu\text{m}$	588.8	$\nu_2$	19.2 – 15.2	520 – 660
N <sub>2</sub> O 4.5 $\mu\text{m}$	2223.5	$\nu_3$	4.7 – 4.4	2120 – 2270
CFC's			14.3 – 7.7	~700 – 1300

Table 1.3: Absorption bands of several molecules in the thermal infrared region (after Liou, 1992).

## 1.5.2 Spectral Line Shape and Line Strength

When examining discrete transitions between molecular energy levels one assumes that the spectra consist of individual monochromatic lines. But actual lines occupy a finite band of wavelengths and their shape depends on the atmospheric conditions under which the absorption and emission occurs. We can distinguish three types of the spectral line shape occurring in the Earth's atmosphere, namely the Lorentz profile, the Doppler profile, and the (hybrid) Voigt profile.

The spectral width of the absorption line is described by the shape factor  $f$ . The absorption cross section  $k_\nu = \sigma_{a\nu}$  at wavenumber  $\nu$  is expressed by

$$k_\nu = S f(\nu - \nu_0), \quad (1.5.2)$$

where  $\nu_0$  is the line center and  $S$  is the line strength given by

$$S = \int_0^\infty k_\nu d\nu. \quad (1.5.3)$$

LORENTZ PROFILE

In the lower atmosphere radiative transitions are disturbed by molecule collisions resulting in a broadening of spectral lines. This type of broadening is also referred to as pressure or collisional broadening and is essential for the IR bands of CO<sub>2</sub> and H<sub>2</sub>O at altitudes below 30 km. The spectral width can be described by the Lorentz line shape

$$f_L(\nu - \nu_0) = \frac{\alpha_L}{\pi(\nu - \nu_0)^2 + \alpha_L^2}, \quad (1.5.4)$$

where  $\alpha_L$  is the half-width given to a good approximation by

$$\alpha_L(p, T) = \alpha_0 \frac{p}{p_0} \left( \frac{T_0}{T} \right)^{\frac{1}{2}}, \quad (1.5.5)$$

which is inversely proportional to the mean time between collisions. The reference half-width  $\alpha_0$  is approximately 0.1 cm<sup>-1</sup> evaluated at standard temperature  $T_0$  (273 K) and standard pressure  $p_0$  (1013.25 mbar). The value for  $\alpha_L$  for collisions of H<sub>2</sub>O molecules with N<sub>2</sub> lies in the range 0.03–0.1 cm<sup>-1</sup> at  $p = 1013.25$  mbar (Zuev, 1974). The Lorentz line shape is illustrated in Fig. 1.11.

Natural broadening is another fundamental source of line width although it is not so significant for vibrational and rotational transitions in the IR. Natural broadening results from finite lifetime of excited states and the introduced line width can also be modeled by the Lorentz shape (Eq. 1.5.4) but the half-width is described by

$$\alpha = (2\pi\bar{t})^{-1}, \quad (1.5.6)$$

where  $\bar{t}$  represents the mean lifetime of the excited states. Natural broadening occurs in the visible and ultraviolet bands such as in the Schumann–Runge bands of O<sub>2</sub>; at longer wavelength it is negligibly small under all natural (i.e., non-laboratory) conditions.

DOPPLER PROFILE

Higher in the atmosphere the pressure becomes small and above about 40 km the temperature-generated thermal velocities of the molecules become significant. The so-called Doppler broadening results from the thermal random molecular motion and the associated frequency shifts

$$\nu = \nu_0 \left( 1 \pm \frac{v}{c} \right), \quad (1.5.7)$$

where  $c$  is the velocity of light. The Boltzmann distribution gives the probability of the velocity ( $v$ ) components according to

$$p(v) = \sqrt{\frac{m}{2\pi kT}} \exp\left[-\frac{mv^2}{2kT}\right], \quad (1.5.8)$$

where  $m$  denotes the mass of the molecule,  $k$  is the Boltzmann constant and  $T$  is the temperature. Based on this, the Doppler line shape factor can be evaluated as

$$f_D(v - v_0) = \frac{1}{\alpha_D \pi} \exp\left[-\frac{(v - v_0)^2}{\alpha_D^2}\right], \quad (1.5.9)$$

and the Doppler half-width is

$$\alpha_D = \frac{v_0}{c} \sqrt{2RT}, \quad (1.5.10)$$

where  $R$  is the individual gas constant. For the line of the rotational  $H_2O$  band about  $200 \text{ cm}^{-1}$  the Doppler half-width is  $3.5 \times 10^{-4} \text{ cm}^{-1}$  (Zuev, 1974). Inspecting Fig. 1.11 reveals that the absorption coefficient of a Doppler line is more intense around the line center but weaker at the wings compared to the absorption coefficient of a Lorentz line.

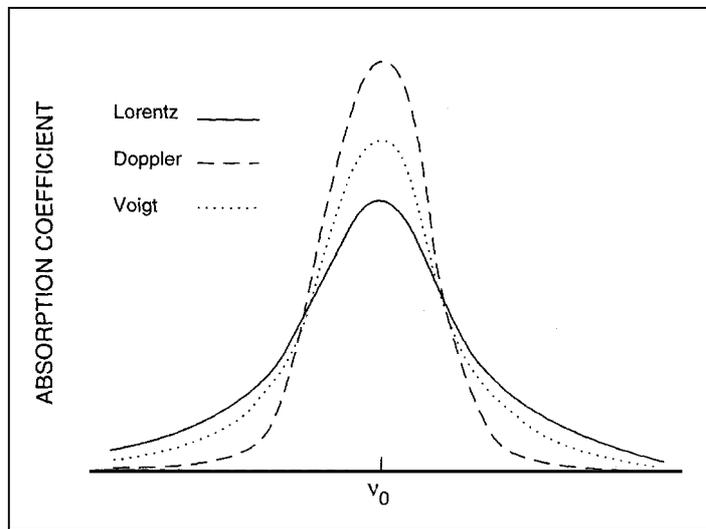


Figure 1.11: Lorentz, Doppler and Voigt line shape (from Salby, 1996)

### VOIGT PROFILE

In atmospheric regions from about 20 to 50 km the line shapes are determined by both collisional- and Doppler-broadening processes, which have been expressed individually by  $f_L(v - v_0)$  and

$f(v - v_0)$ , respectively. The Voigt line shape, in fact a hybrid Doppler-Lorentz line shape, can be written as

$$\begin{aligned} f_v(v - v_0) &= \int_{-\infty}^{\infty} f_L(v' - v_0) f_D(v - v') dv' \\ &= \frac{1}{\pi^{\frac{1}{2}}} \frac{\alpha_L}{\alpha_D} \int_{-\infty}^{\infty} \frac{1}{(v' - v_0)^2 + \alpha_L^2} \exp\left[-\frac{(v - v')^2}{\alpha_D^2}\right] dv', \end{aligned} \quad (1.5.11)$$

(Liou, 1992). The Voigt profile gives a line shape between the Lorentz and the Doppler line as illustrated in Fig. 1.11 and can thus be used at heights where both line shapes play a role.

The detailed form of the line strength  $S = \int_0^{\infty} k_\nu dv$ , Eq. 1.5.3, is provided by quantum theory and involves the probability of a single molecular transition from state  $j$  to  $i$

$$S(T) = \frac{g_i I_a}{Q(T)} \frac{8\pi^3 v_{ij}}{3hc} (1 - \exp[-hc v_{ij} / kT]) \exp[-hc E_i / kT] |R_{ij}|^2 \times 10^{-36} [\text{cm}^{-1}/(\text{mol cm}^{-2})], \quad (1.5.12)$$

from Liou (1992). The quantity  $g_i$  is the nuclear spin degeneracy,  $E_i$  is the energy of the lower state  $i$ ,  $v_{ij}$  is the line-center wavenumber,  $I_a$  is the natural isotopic abundance and  $Q(T)$  is the total internal partition sum and  $h$ ,  $k$ , and  $c$  are the Planck constant, Boltzmann constant and the speed of light, respectively. The transition probability  $R_{ij}$  is given by

$$R_{ij} = \int \psi_i^* \mathbf{M} \psi_j dV, \quad (1.5.13)$$

where  $\mathbf{M}$  is the matrix form of the relevant dipole moment (electric or magnetic) and  $\psi$  is the wave function of the molecular state obtained by solving the Schrödinger equation (\* denotes conjugate complex). The internal partition function depends on the internal energy and the temperature. Under atmospheric conditions  $Q(T) \cong 1$  holds due to vibration, and for rotational transitions we may approximate the internal partition function by  $Q(T) \cong T$  for linear molecules, such as  $\text{CO}_2$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$ , and by  $Q(T) \cong T^{3/2}$  for non-linear molecules, such as  $\text{H}_2\text{O}$ ,  $\text{O}_3$ ,  $\text{CH}_4$ , respectively. The expression Eq. 1.5.12 may hence be approximated for a reference temperature  $T_r$  in the form

$$S(T) \cong S(T_r) \left(\frac{T_r}{T}\right)^n \exp\left[-\frac{hc E_i}{K} \left(\frac{1}{T} - \frac{1}{T_r}\right)\right], \quad (1.5.14)$$

where  $n = 1$  for linear molecules and  $n = \frac{3}{2}$  for non-linear molecules.

### 1.5.3 Computation of Transmittance

Evaluation of the upwelling radiance according to Eq. 1.4.29,

$$I_v = B_v(\tau_{s,v})T_v(\tau_{sv}) + \int_{\tau_{s,v}}^0 B_v(\tau'_v) \frac{dT_v(\tau'_v)}{d\tau'_v} d\tau'_v, \quad (1.5.15)$$

requires accurate calculation of the spectral transmittance. The monochromatic transmittance is given by

$$T_v(u) = \exp[-\tau_v] = \exp\left[-\int_0^u k_v(u') du'\right]. \quad (1.5.16)$$

The absorption coefficient  $k_v$  is a function of pressure and temperature and is given (see Eq. 1.5.2) by

$$k_v = S f(\nu - \nu_0), \quad (1.5.17)$$

where  $f(\nu - \nu_0)$  is given by the shape profile appropriate for the application described, i.e., the Lorentz profile (Eq. 1.5.4), Doppler profile (Eq. 1.5.9), or Voigt profile (Eq. 1.5.11), respectively.

The absorption lines for many molecules such as H<sub>2</sub>O, CO<sub>2</sub>, O<sub>3</sub>, N<sub>2</sub>O, and CH<sub>4</sub> have been intensively examined and associated parameters including line position in cm<sup>-1</sup>, line strength in cm<sup>-1</sup>/(mol cm<sup>-2</sup>), line half-width in cm<sup>-1</sup>atm<sup>-1</sup> and lower energy state in cm<sup>-1</sup> are documented and listed in databases such as the HITRAN database.

Transmittance computation accounting for all absorption lines in a given spectral interval is referred to as line-by-line calculation. The monochromatic transmittance at wavenumber  $\nu$  containing  $N$  absorption lines can be described by a product of the transmittances for each individual absorption line according to

$$T_\nu = \prod_{j=1}^N T_{\nu_j}, \quad (1.5.18)$$

with the transmittance for absorption line  $j$

$$T_{\nu_j} = \exp\left[-\int_0^u k_{\nu_j}(u') du'\right], \quad (1.5.19)$$

where  $j = 1, 2, \dots, N$ . Thus Eq. 1.5.18 can be rewritten as

$$T_v = \exp \left[ - \int_0^u \sum_{j=1}^N k_{vj}(u') du' \right], \quad (1.5.20)$$

and the absorption coefficient may be expressed in terms of line strength  $S$  and line shape  $f$

$$k_v(p, T) = \sum_{j=1}^N S_j f_{vj}(p, T). \quad (1.5.21)$$

The spectral intervals for which  $k_v$  is calculated have to be smaller than the half-width in order to resolve individual lines. CO<sub>2</sub> and O<sub>3</sub> absorb in the upper stratosphere, where Doppler broadening is significant. The Doppler half-width in the 15 μm band of CO<sub>2</sub> and in the 9.6 μm band of O<sub>3</sub> is approximately 0.0005–0.0001 cm<sup>-1</sup>. Since the spectral ranges of these bands are about 300 cm<sup>-1</sup> (see Tab. 1.3) more than half a million points need to be calculated to resolve individual lines. In the troposphere absorption is primarily accomplished by water vapor covering a range of about 15000 cm<sup>-1</sup>. The collision half-width is > 0.01 cm<sup>-1</sup>. This requires a computation of about one million points. Based on numerical experiments spectral resolutions of 0.01, 0.002 and 0.005 cm<sup>-1</sup> have been found adequate for H<sub>2</sub>O, O<sub>3</sub>, and CO<sub>2</sub>, respectively (Liou, 1992).

Over a spectral interval including the wavenumber  $\nu_i$  we have from Eq. 1.5.20

$$\begin{aligned} T_{\bar{\nu}} &= \sum_i \exp \left[ - \int_u \sum_j k_{ij}(u) du \right] \frac{\Delta \nu_i}{\Delta \nu} \\ &= \sum_i \exp \left[ - \sum_n \sum_j k_{ij}(p_n, T_n) \Delta u_n \right] \frac{\Delta \nu_i}{\Delta \nu}, \end{aligned} \quad (1.5.22)$$

where  $j$  is the line index,  $n$  is the index of the atmospheric levels with pressure  $p_n$  and temperature  $T_n$ , and  $i$  is the wavenumber index sampling over  $\bar{\nu}$ . Transmittance computation according Eq. 1.5.22 involves summation over the absorption lines in the spectral interval and the inhomogeneous path considered. In addition, the absorption coefficient depends on the line strength  $S = S(T)$  and the line shape  $f = f(p, T)$ . These line-by-line calculations are computationally very expensive and may be impractical for many applications. Thus great effort is made to find simplified approaches to compute spectral transmittance.

Many RT models rely on band models, which present the gross characteristics of absorption in a spectral range. Band models assume a homogeneous atmosphere, where the absorption coefficient is independent of pressure and temperature. This assumption allows analytic forms for transmittance in terms of simple exponential functions. The assumption of homogenous path may not be appropriate for applications to realistic atmospheres. There are several approximations of

the inhomogeneous path, which account for the pressure and temperature dependence of  $k_\nu$ . More details on band models may be found in Liou (1992). Chapt. 4 presents an example of a fast transmittance model suitable for instruments with high spectral resolution. A description of fast transmittance models is also given in Hannon et al. (1996).

## 2 METEOROLOGICAL SATELLITES AND INSTRUMENTS

Since the first meteorological satellite (TIROS-1) was launched in 1960, space-borne observations became indispensable in studies of the Earth and its atmosphere. Today an international network of geostationary satellites, stationed above the equator, and polar orbiting satellites, circling the Earth at lower altitudes, provide data to be processed for meteorology, climate, ocean, land or ecology applications. A broad introduction to the various applications and uses of present meteorological satellites is given by EUMETSAT and CGMS (Coordination Group for Meteorological Satellites) at <http://www.wmo.ch/hinsman/>.

### 2.1 Satellite Orbits

The main sources of information given in this section are Sellers (1994), and Kidder and Vonder Haar (1995).

#### 2.1.1 Newton's and Kepler's Laws

The basic mathematical laws governing the orbits of heavenly bodies are known since the work of Johannes Kepler (1571–1630) and Isaac Newton (1642–1727). Newton's law of universal gravitation

$$F = \frac{Gm_1m_2}{r^2}, \quad (2.1.1)$$

derived from his second law of motion  $F = ma$  (mass  $m$ , acceleration  $a$ ), describes the force  $F$  of attraction between two point masses  $m_1$  and  $m_2$  at a distance  $r$ , and where  $G$  is the universal gravitation constant (Tab. A1). To keep a satellite (small mass  $m$ ) in a circular orbit around the Earth (big mass  $m_e$ ) the force due to gravitational attraction Eq. 2.1.1 needs to be compensated by the centripetal force acting on the satellite due to the orbital velocity ( $v$ ) of the satellite according to

$$\frac{Gm_e m}{r^2} = \frac{mv^2}{r}. \quad (2.1.2)$$

The centripetal force can also be written in terms of the angular velocity  $\omega = v/r$ . The orbit of a satellite is independent of the mass of the satellite, since  $m$  can be eliminated from Eq. 2.1.2. The period  $T$  of the satellite to travel once around the orbit is  $T = 2\pi r / v = 2\pi / \omega$ , thus

$$T = 2\pi \sqrt{\frac{r^3}{\mu}}, \quad (2.1.3)$$

since  $v$  is derived from Eq. 2.1.2 as  $v = \sqrt{\mu/r}$  yielding  $\omega = \sqrt{\mu/r^3}$ . The quantity  $\mu = Gm_e$  is the gravitational parameter of the Earth (Tab. A2). For example, the period of a polar-orbiting satellite at a distance of 850 km above the Earth surface (thus  $r = 7220$  km with  $r_e \sim 6370$  km) is approximately 102 minutes.

Newton's laws of motions and formula Eq. 2.1.1 gave a mathematical explanation of the three basic laws of planetary orbits discovered by Johannes Kepler a few decades earlier. Kepler's laws can be summarized as follows:

1. All planets move on elliptical paths around the sun as one focal point.
2. The radius vector from the sun to the planet covers equal areas in equal times.
3. The ratio of the square of the period of revolution to the cube of the semi-major axis of the elliptical orbit is the same for all planets moving around the sun.

### 2.1.2 Classical Orbital Elements

These three laws utilize six orbital parameters, known as classical orbital elements, which completely determine and define the orbit of a planet or a satellite in space. In the following we consider a satellite moving around the Earth according to the geometry of the orbit shown in Fig. 2.1. First of all, the size of the elliptical orbit is determined by the semi-major axis  $a$  (i.e., the half length of the ellipse). In a gravity field the total mechanical energy  $E$  of the satellite (of mass  $m$ ), given by a sum of its kinetic and potential energy, is

$$E = \frac{1}{2}mv^2 - \frac{\mu m}{r}. \quad (2.1.4)$$

Correspondingly, the specific energy  $e$  is

$$e = \frac{v^2}{2} - \frac{\mu}{r} = -\frac{\mu}{2a}, \quad (2.1.5)$$

which is conserved. The same is true for the angular momentum  $L = mr^2\omega$ . To obtain conservation of energy and momentum the satellite needs to move faster when it is closer to the Earth near perigee and slower when it is farther from the Earth near apogee, as stated in Kepler's second law. As indicated in Eq. 2.1.5 the *semi-major axis*  $a$  as our first classical orbital element can be related to the specific mechanical energy of the satellite. The exact derivation of that relationship can be found in Sellers (1994). The shape of the orbit is specified by the *eccentricity*  $\varepsilon$ , which is the ratio of the distance between the foci to the length of the ellipse. In order to describe the orientation of the orbit in space we introduce a geocentric equatorial coordinate system, where the fundamental plane is the Earth's equatorial plane (see Fig. 2.1). The z-axis is aligned with the spin axis of the Earth, the x-axis points in the direction of vernal equinox (i.e., where the sun is crossing the equatorial plane from the southern to the northern hemisphere) and the y-axis completes the right-hand rule. The orientation of the orbit plane in space is described by two orbital parameters. The *inclination*  $i$  is the angle between the equatorial plane and the satellite orbit plane. If  $i$  equals  $0^\circ$  or  $180^\circ$ , the satellite moves on an equatorial orbit, whereas exact polar orbits are orbits with  $i = 90^\circ$ . If  $i < 90^\circ$  or  $i > 90^\circ$ , the orbit is called prograde (moving in direction of the Earth's rotation) or retrograde (moving against the Earth's rotation), respectively. The ascending node is the point, where the satellite crosses the equator from to the southern hemisphere to the northern hemisphere as indicated in Fig. 2.1.

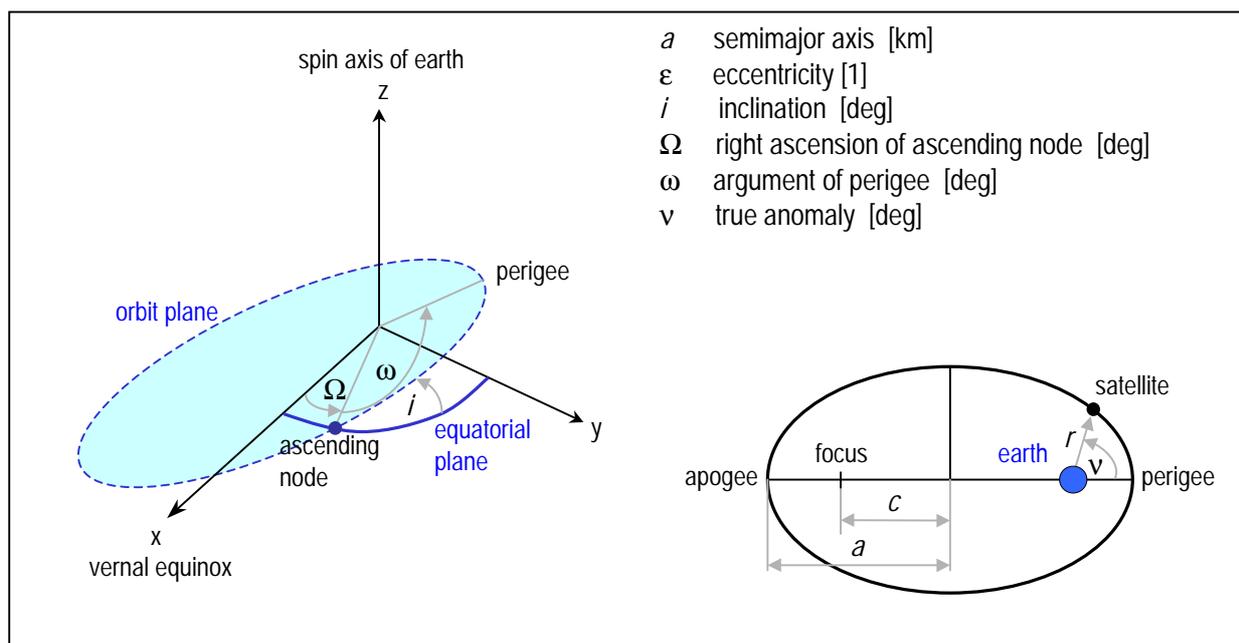


Figure 2.1: Geometrical description of classical orbital parameters.

The angle measured in the equatorial plane from the x-axis direction to the ascending node is termed *right ascension of ascending node*  $\Omega$  (or longitude of ascending node).

Note, that  $\Omega$  is not defined, if the inclination  $i$  equals  $0^\circ$  or  $180^\circ$  (i.e., equatorial orbit). The orientation of the orbit itself within the plane is specified by the angle measured in the orbit plane from the ascending node to the perigee and is called *argument of perigee*  $\omega$ . The argument of perigee is undefined if the considered orbit is equatorial or if it is circular (i.e.,  $\varepsilon = 0$ ). Finally, the *true anomaly*  $v$ , as the angle between the radius vector  $r$  and the perigee, determines the satellite's position in the orbit plane. It is not defined for circular orbits ( $\varepsilon = 0$ ). Some space agencies specify the *mean anomaly*  $M$  instead of the true anomaly  $v$ . The mean anomaly  $M$  is an angle defined as

$$M = n(t - t_p), \quad (2.1.6)$$

where  $n$  is the mean motion (i.e., the average angular rate) of the satellite given by

$$n = \sqrt{\frac{\mu}{a^3}}, \quad (2.1.7)$$

and  $t$  is the actual time whereas  $t_p$  is the time of perigee passage. As mentioned in Sect. 2.1.1, the period is given by  $T = 2\pi / n$ . According to Eq. 2.1.6,  $M = 0$  when the satellite is at perigee and the rate of change of  $M$  is given by  $n$ . Note, that  $M$  (in contrast to  $v$ ) is an angle without a physical meaning, since a time interval is considered.

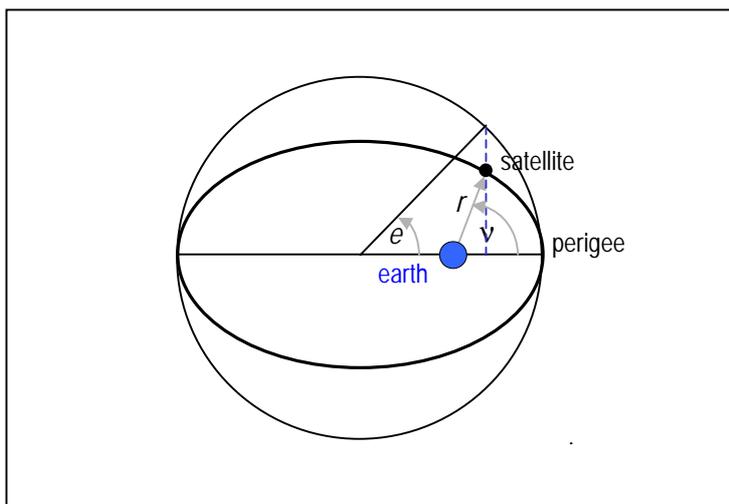


Figure 2.2: Geometric description of eccentric anomaly  $e$ .

We can relate the mean anomaly  $M$  to the true anomaly  $\nu$  by using Kepler's equation

$$M = e - \varepsilon \sin e, \quad (2.1.8)$$

where  $\varepsilon$  is eccentricity. The angle  $e$  is termed the eccentric anomaly and relates motion on an ellipse to motion on a circle. Therefore  $e$  can be found by circumscribing the elliptical orbit with a circle as shown in Fig. 2.2. Then  $e$  can be connected to the true anomaly  $\nu$  via

$$\cos e = \frac{\varepsilon + \cos \nu}{1 + e \cos \nu}. \quad (2.1.9)$$

With Eq. 2.1.6 and Eq. 2.1.8 the time of flight between two known positions can be calculated, as well as future orbital positions, if the time of flight is given. In addition to the orbital elements, the *epoch time*, which is the specific time at which the position and the remaining elements are observed, is specified.

If the classical orbital elements stay constant (except  $\nu$  or  $M$ ) the orbit is said to be Keplerian. Usually satellite orbits are only nearly Keplerian orbits. The most significant cause of the deviation from a true Keplerian orbit is the non-spherical gravitational field, since the Earth is non-spherical and inhomogeneous. The equatorial radius exceeds the polar radius by approximately 21 km. The gravitational potential  $U$  of the Earth can be approximated (to first order) by

$$U = -\frac{\mu}{r} \left[ 1 + \frac{1}{2} J_2 \left( \frac{r_{ee}}{r} \right)^2 (1 - 3 \sin^2 \delta) + \dots \right], \quad (2.1.10)$$

(Kidder and Vonder Haar, 1995), where  $\mu = Gm_e$  is the gravitational parameter,  $J_2$  is the quadrupole gravitational coefficient of the Earth (Tab. A2),  $r_{ee}$  is the equatorial Earth radius and  $\delta$  is the declination angle (angle measured from the equatorial plane to the point in space). Due to the coefficient  $J_2$  being non-zero, the rate of change of  $M$  needs to be modified according to

$$\frac{dM}{dt} = \tilde{n} = n \left[ 1 + \frac{3}{2} J_2 \left( \frac{r_{ee}}{a} \right)^2 (1 - \varepsilon)^{-\frac{3}{2}} \left( 1 - \frac{3}{2} \sin^2 i \right) \right], \quad (2.1.11)$$

where  $n$  is the mean motion, given by Eq. 2.1.7,  $a$  is the semi-major axis and  $i$  is inclination angle. Depending on the inclination  $i$  the satellite moves faster or slower than it would do in the unperturbed orbit. Besides  $M$ , the second impact of this force due to the non-spherical gravitational field is a precession of the orbit about the z-axis. The rate of change of the right ascension of ascending node  $\Omega$  can be written as

$$\frac{d\Omega}{dt} = -\tilde{n} \left[ \frac{3}{2} J_2 \left( \frac{r_{ee}}{a} \right)^2 (1 - \varepsilon^2)^{-2} \cos i \right], \quad (2.1.12)$$

where  $\tilde{n}$  is the anomalistic mean motion, given by Eq. 2.1.11. The argument of perigee  $\omega$  is also precessing according to

$$\frac{d\omega}{dt} = \tilde{n} \left[ \frac{3}{2} J_2 \left( \frac{r_{ee}}{a} \right)^2 (1 - \varepsilon^2)^{-2} \left( 2 - \frac{5}{2} \sin^2 i \right) \right]. \quad (2.1.13)$$

The period for a perturbed orbit is  $T = 2\pi / \tilde{n}$ , termed as the anomalistic period. In practice, the nodal period, which is the time from one ascending node to the next ascending node, is used. It can be well approximated by  $T \sim 2\pi / (\tilde{n} + d\omega/dt)$  (Kidder and Vonder Haar, 1995).

### 2.1.3 Meteorological Satellite Orbits

The classical orbit elements describe various orbits used for current applications. In satellite meteorology and remote sensing two specific orbit types have been found particularly useful, geostationary and near-polar sun-synchronous ones. Typical values of the classical orbital elements of a geostationary and a sun-synchronous orbit by means of two exemplary satellites GOES-10 and NOAA-15 are given in Tab. 2.1. GOES stands for Geostationary Operational Environmental Satellite, whereas the acronym NOAA stands for National Oceanic and Atmospheric Administration.

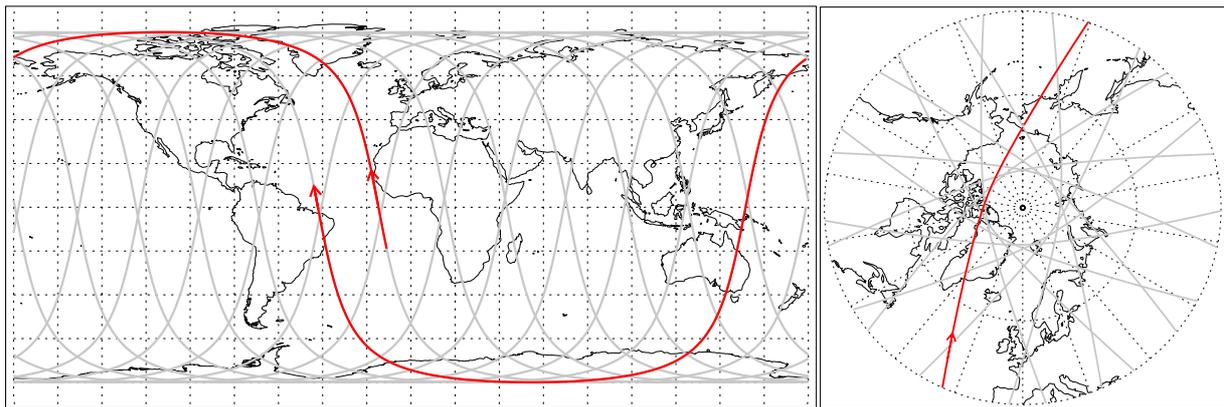
name	orbit type	$a$ [km]	$\varepsilon$ [deg]	$i$ [deg]	$\Omega$ [deg]	$\omega$ [deg]	$M$ [deg]
GOES-10	geostationary	42166.20	0.00027	0.12686	101.778	256.009	92.7534
NOAA-15	sun-synchronous	7187.00	0.00183	98.6165	60.1285	11.9681	116.673

Table 2.1: Orbital elements for the geostationary GOES-10 (epoch: 05/02/01 6:00) and the sun-synchronous NOAA-15 (epoch: 29/01/01 20:00) satellites (from: <http://psbsgi1.nesdis.noaa.gov:8080/noaasis.html>).

According to the table above the sun-synchronous satellites fly in a near-polar orbit at a height of about 850 km with a period of approximately 100 minutes. The geostationary satellites, at a height of about 36000 km, circle the Earth with a period of 24 hours and therefore remain almost fixed over one point above the equator. Due to their different orbit geometry, flight height and orbital period these two major types of meteorological satellites have different but complementary objectives. Both are thus necessary for providing a complete weather monitoring system.

### THE SUN-SYNCHRONOUS ORBIT

As mentioned before the oblateness of the Earth (and some minor effects such as the gravitational attraction of other heavenly bodies) causes the satellite orbit to deviate from a Keplerian orbit, in which the classical elements (except for  $M$ ) can be considered constant throughout the flight. The non-spherical gravitational field of the Earth yield a small change of the mean motion and a slow, linear change in  $\Omega$  and  $\omega$  according Eq. 2.1.11 to Eq. 2.1.13. The associated precession of the orbit has a useful application in creating an orbit, which can be synchronized with the sun by choosing  $d\Omega/dt$  (Eq. 2.1.12) to equal the rotation rate of the Earth around the sun. Since it takes the Earth one tropical year (being the time interval between two crosses of the sun through the vernal equinox, see Tab. A2) to rotate around the sun the rotation rate can be calculated to be approximately  $1.99 \times 10^{-7}$  rad/sec. If we insert  $a = 7228$  km (altitude  $\sim 850$  km) and eccentricity  $\varepsilon = 0$  in Eq. 2.1.12, an inclination  $i$  of  $98.8^\circ$  makes the orbit sun-synchronous. This means that the orbit plane appears fixed with respect to the sun. The typical orbital period for typical orbit heights is found to be 98 to 102 minutes, yielding about 14 to 15 orbits per day. Due to Earth's rotation, successive orbits appear to shift to the west from an Earth-based viewpoint and satellites on a sun-synchronous orbit view the entire Earth twice a day. Fig. 2.3 shows the ground-track for a sun-synchronous satellite due to the orbital plane and the rotation of the Earth beneath the satellite yielding global daily coverage. Thus satellites in sun-synchronous orbits are used for global weather-studies, long-term forecasting and for environmental monitoring.



**Figure 2.3:** Successive ground tracks of a satellite in a sun-synchronous near-polar orbit with one orbit high-lighted utilizing a cylindrical equidistant projection (right plot) and a polar stereographic projection (left plot).

Sun-synchronous orbits are classified by their equatorial crossing times (ECT), defined as the local time when the satellite crosses the equator. For example, satellites in an “early morning” orbit

ascend (crossing the equator northward) around 04:00 to 06:00 local time and descend (crossing the equator southward) over the same location (longitude) around 16:00 to 18:00 local time. For instance, the AVHRR (Advanced Very High Resolution Radiometer), to be described below, mounted on polar-orbiting satellites provides complete strips of roughly 3000 km width from pole to pole of every point of the Earth twice in an 24-hours period.

Since the inclination  $i$  is about  $98^\circ$  the sun-synchronous orbits are also referred to as polar orbits, although the orbits do not actually go over the poles (Fig. 2.3). To distinguish a sun-synchronous polar orbit from a geostationary one, where satellites fly at much higher altitudes, it is also termed a low earth orbit (LEO).

### THE GEOSTATIONARY ORBIT

An orbit is called geo-synchronous if the satellite moves with the same angular velocity as the Earth. If, in addition, the eccentricity  $\varepsilon = 0$  and the inclination  $i = 0$ , yielding an equatorial and circular orbit, the term geostationary is used. With an orbital period of roughly 24 hours the satellite appears stationary to an Earth-based observer. To be geo-synchronous, the orbit needs to be at about 6.6 Earth radii from the Earth's center. The exact semi-major axis  $a$  can be calculated to be 42168 km if the oblateness of the Earth is taken into account. Those orbits are typically used for communication and meteorological applications such as weather forecasting and extensive data analysis, where a continuous view of the mid-latitude and tropical areas of a particular geographical region can be exploited. Since satellites in geostationary orbits provide continuous day and night observations, they are very useful for monitoring severe weather events (e.g., hurricanes). In order to get global coverage of the equatorial and mid-latitudinal area, a network of at least 5 satellites is needed as we will see in the next section on operational satellites in geostationary orbits.

## 2.2 Meteorological Satellite Instruments

This section should serve as an introduction of satellite instruments for meteorological applications. The interested reader is referred to more detailed literature, e.g., Kidder and Vonder Haar (1995) or to <http://www.wmo.ch/hinsman/>; sources on which the description below has been largely based.

### 2.2.1 Types of Instruments

The instruments of the meteorological payload of a satellite collect electromagnetic energy reflected or emitted by the Earth and its atmosphere. Traditionally any instrument designed to quantitatively measure radiation in some interval of the electromagnetic spectrum is called *radiometer*. The *photometer* measures the narrow spectral band including the visible wavelengths and the *polarimeter* records the polarization of the light. If the sensor contains components (such as a prism or diffraction grating) to break radiation into discrete wavelengths and separates (or disperses) them at different angles, the term *spectrometer* is used. *Spectroradiometers* measure dispersed radiation in bands rather than in small wavelength steps. The *sounder* collects radiation data to be used to create vertical profiles of atmospheric temperature, moisture, ozone or other atmospheric constituents.

Sensors can be classified into passive and active instruments. *Passive* sensors investigate radiation, which comes from natural sources, such as the Earth and its atmosphere or the sun. *Active* refers to measurements of radiation, which is generated, transmitted and received by the sensor system. For instance, radar systems transmit high-frequency radio waves and receive them, after they have been reflected by an object.

The main part of all radiometric devices are *detectors*, which convert the incoming radiant energy to electrical signals to be amplified and recorded. A main underlying principle of detector technology is the photoelectric effect, where electromagnetic radiation incident on metal releases electrons. The resulting electric current can be recorded as a measure of the intensity of the incoming radiation. The effect is dependent on the wavelength of the incident radiation and on the material. Detectors of different materials are designed to receive radiation in different channels, where each channel analyses different wavelengths of the electromagnetic spectrum. The most common detectors are photon detectors, classified as photoconductive and photovoltaic (photodiode) devices, usually made of intrinsic semiconductive material.

Non-imaging systems report the measurements as electrical signals of different strength, whereas in *imaging systems* the electrical charge (created by the detector) is recorded on a film to generate images of the radiation field. Besides detectors *optical components* (including telescopes) to focus radiation on the detectors, elements to limit and record the signal and the scan assembly make up the sensor module of the instrument.

Imaging systems (except cameras) view the scene by scanning. Thus the sensors gather the data while the satellite is moving by sensing the scene below point by point along successive lines. The range of angles scanned of the system is called field of view (FOV) or swath width (i.e., the length of one scan line). The area viewed at a given instant is known as the instantaneous field of view (IFOV). The IFOV can be square or circular and is usually specified at nadir (sub-satellite point) in degrees or kilometers. The scanning in moving direction can be done continuous or in a so-called step and dwell mode, where the distance between two consecutive scan lines is termed scan separation. The scanning of the path covering an area to the side (specifying the swath width) can be done in across-track or in along-track mode. The *across-track* mode, also referred to as whiskbroom scanning, sweeps the scene perpendicular to the moving direction in adjacent lines. Usually this is accomplished by an oscillating mirror looking off at various angles. Each scan line consists of contiguous spatial elements emanating radiation, which is characterized by the associated picture element (pixels) in the sensor. Instruments operating in an *along-track* mode (pushbroom scanning) contain an array of numerous detectors to receive simultaneously radiation coming from the spatial cells below. In this mode the pixels in the sensor correspond to these individual detectors. While the sensor moves forward the detectors are exposed to the radiation coming from the next line.

### 2.2.2 Data Transmission

All meteorological data are continually broadcast to Earth by transmitting devices to be processed by receiving stations. This data stream of high rate is called HRPT (High Resolution Picture Transmission). Transmission of data in an analogue (television-like) signal is of a coarser resolution, since only single channels are used and some scan lines are neglected; this is called APT (Automatic Picture Transmission). Most polar-orbiting satellites transmit APT format at a VHF (Very High Frequency, in orders of  $10^8$  Hz) frequency band, usually at 137 MHz. APT transmits 120 picture lines in a minute. Digital tape recorders onboard the satellite are used to store the data before transmitting them to Earth. Data from a full orbit with a 4 km resolution is called GAC (Global Area Coverage) data, whereas data collected in 10 minutes of full resolution is

termed LAC (Local Area Coverage) data. Another method to transmit satellite images (visible, infrared, and water vapor) is the international WEFAX (Weather Facsimile) usually used by geostationary satellites. The satellite transmits the full-disk image of the Earth to a ground station, normally a CDA (Command and Data Acquisition) station, on an S-band (1.55–4.2 GHz, wavelength range 75–150 mm) microwave frequency, usually at 1691 MHz. Images are currently transmitted every half hour with about 4 minutes transmission time per image. The image is reformatted with political boundaries attached before it is sent back to the satellite on a similar S-band frequency. The satellite relays the data to several field service stations. This analogue WEFAX format has been the transmission standard on geostationary satellites for decades and will be replaced by LRIT (Low Rate Information Transmission) on satellites launched after 2000.

### 2.2.3 Operational and Future Instruments

In the following listing (in alphabetical order) we concentrate on main meteorological instruments, although many other instruments, such as for monitoring, tracking or data storage and transmission are on board the satellite.

#### AIRS

The Advanced Infrared Radiation Sounder (AIRS) on polar-orbiting satellites is a high spectral resolution spectrometer covering about 2400 bands in visible and infrared regions. It accurately measures temperature, humidity, surface temperature and cloud parameters for weather prediction and climate studies. The instrument consists of a grating spectrometer to measure the infrared portion and a photometer to cover the visible and the near-infrared range. The instrument produces 90 pixels per scan line across-track. Together with AMSU and HSB (Humidity Sounder Brazil), AIRS is scheduled to fly first on the EOS (Earth Observation System) AQUA satellite operated by NASA (National Aeronautics and Space Administration). Launch of the AQUA satellite is scheduled in fall 2001.

#### AMSU-A, AMSU-B

Data from the AMSU (Advanced Microwave Sounding Unit), which is a multi-channel microwave radiometer on polar-orbiting satellites, can be processed to obtain temperature profiles even under cloudy conditions. AMSU-A measures in 15 microwave channels (from about 23 to 90 GHz) in a stepping mode with an IFOV at nadir of about 45 km, a swath-width of about 2100 km and a scan separation of about 53 km. The number of pixels produced per scan is 30. AMSU-B has 5

microwave channels (90 to 184 GHz) to provide in total 20 channels in conjunction with AMSU-A for atmospheric sounding. AMSU-B scans across-track with an IFOV of 16 km and the same swath width as AMSU-A.

### ASCAT

The ASCAT (Advanced Scatterometer) is an active-type instrument to be flown on the polar-orbiting METOP satellites and measures wind speed and direction over the ocean. ASCAT is a C-band (4.2–5.75 GHz, wavelength range 37.5–75 mm) radar with a center frequency at 5.25 GHz. The instrument developed by ESA (European Space Agency) will also monitor snow and ice cover over land and sea, but the primary product is the radar backscattering coefficient provided over a swath width of 2×500 km.

### AVHRR

The AVHRR (Advanced Very High Resolution Radiometer) instrument is a six-channel (one visible, two near-infrared, three infrared) spectral imager, which succeeded the VHRR (Very High Resolution Radiometer) device on polar-orbiting satellites. The main objectives of the latest version AVHRR/3 are global cloud imagery, mapping of surface temperature, sea-ice, snow-cover, and vegetation. AVHRR consists of a rotating scan mirror, a telescope, internal optics and electronics, as well as detectors. As the mirror rotates (at 360 rpm) the telescope scans the field of view across-track with a resolution of about 1.1 km (square IFOV at nadir) with a swath-width of about 3000 km. Per scan approximately 2000 Earth views (pixels) are provided. Data are transmitted within the high resolution HRPT as well in the APT data stream, where two channels and every third scan lines are selected.

### GOME

The near-future GOME (Global Ozone Monitoring Experiment) instrument is GOME-2 to be flown on the polar-orbiting METOP satellites (first launch 2005). It is an across-track scanning spectrometer measuring the radiance backscattered from the Earth and the atmosphere in the ultraviolet and visible range to create profiles of atmospheric ozone and other trace gases such as the nitrogen compounds NO, NO<sub>2</sub> or NO<sub>3</sub>, the halogen compounds ClO, OClO and BrO, as well as HCHO and SO<sub>2</sub>.

### GRAS

The new sounding instrument GRAS (GNSS Receiver for Atmospheric Sounding) is scheduled to also fly onboard the polar-orbiting METOP series. The instrument looks tangentially through the Earth's atmosphere to receive radio signals from navigation satellites belonging to the Global Navigation Satellite System (GNSS), including the U.S. GPS (Global Positioning System) and the Russian GLONASS (Global Navigation Satellite System). The received signal has been affected due to refraction by gradients of atmospheric temperature and humidity. Therefore, information on these parameters can be retrieved. The instrument can make about 1000 soundings a day with a vertical resolution of  $< 1$  km.

### HIRS

The current instrument HIRS/3 of the HIRS (High Resolution Infrared Radiation Sounder) series flown on polar-orbiting satellites senses data in 20 channels (one visible, 19 infrared) to provide temperature profiles in cloud-free conditions of high resolution. The instrument utilizes a stepping mirror to accomplish across-track scanning with a scan separation of 42 km. The field of view is about 20 km and the swath is roughly 2100 km wide.

### MHS

The new instrument MHS (Microwave Humidity Sounder) measures in five microwave channels (90-190 GHz) and produces profiles of atmospheric humidity. In addition, it can be used to measure liquid water content in clouds and to estimate precipitation rates. The circular IFOV at nadir has a diameter of about 16 km, the swath-width is about 2100 km and 90 pixels are produced per scan. The instrument is an advanced version of AMSU-B with a very similar measurement concept.

### MSU

On polar-orbiting satellites the MSU (Microwave Sounding Unit) is a microwave radiometer to be used for temperature sounding in the presence of clouds. As the instrument scans, it measures microwave radiation in 4 channels (within the 50-60 GHz oxygen band) at eleven beam positions over a swath about 2000 km wide.

### SEM

The SEM (Space Environment Monitor) instrument is a multi-channel spectrometer, which senses the flux of charged particles at satellite altitude to monitor the local space environment. Since very energetic particles of the solar wind can change the energy input to the magnetosphere, ionosphere and higher atmosphere, it is essential to understand the space environment. The current SEM-2 is onboard the polar-orbiting NOAA satellites and will be onboard the European METOP satellites.

### TOVS, ATOVS

The TOVS (TIROS Operational Vertical Sounder) instrument suite contains HIRS/2, MSU and the SSU (Stratospheric Sounding Unit). The advanced TOVS package ATOVS includes the instruments AMSU-A and AMSU-B (instead of MSU) and the new HIRS/3. TIROS stands for Television and Infrared Observational Satellite.

### VISSR

The VISSR (Visible and Infrared Spin Scan Radiometer) consists of a telescope, mirror, reflectors, lenses and detectors for visible and infrared wavelengths. While the spacecraft is spinning (at 100 rpm) the instrument scans one line (west-to-east) of the Earth's surface to generate images. The telescope moves one step in the north-south direction to scan the next line. Onboard a geostationary satellite a full-disc image requires 2500 scans and is generated every 30 minutes.

## 2.3 Operational Meteorological Satellites

This section on current meteorological satellites should only serve as an overview and the interested reader is referred to more detailed literature and internet-sites such as the sites owned by the space agencies operating the particular satellites. Selected addresses will be mentioned at relevant locations in the text.

### 2.3.1 Satellites in Geostationary Orbit

Current operational geostationary satellites are stationed above the equator at different longitudes to cover 5 sectors as depicted in Fig. 2.4. Information on the major geostationary satellites, retrieved from the internet-based information of the World Meteorological Organization (WMO), last updated in December 2000, is given in Tab. 2.2. The inclination of these satellite orbits lies between 0.0 and 1.0 degrees, the typical altitude is 35800 km and the orbital period is approximately 1435 minutes.

satellite	operator	launch date	location* [deg]
GOES-10	USA	04/1997	-135
GOES-8	USA	04/1994	-75
METEOSAT-7	Europe	02/1997	0
METEOSAT-5	Europe	03/1991	63
GOMS-1	Russia	11/1994	76
INSAT-1D	India	06/1990	74
INSAT-2E	India	04/1999	83
FY-2B	China	06/2000	105
GMS-5	Japan	03/1995	140

Table 2.2: Current operational geostationary satellites (status December 2000). [ \* current longitude location (not at launch date)]

Beside the satellites mentioned in Tab. 2.2 many other meteorological satellites are currently in orbit, e.g., serving as back-up satellites for satellites on failure or in a pre-operational mode such as experimental satellites to test on-board instrumentation.

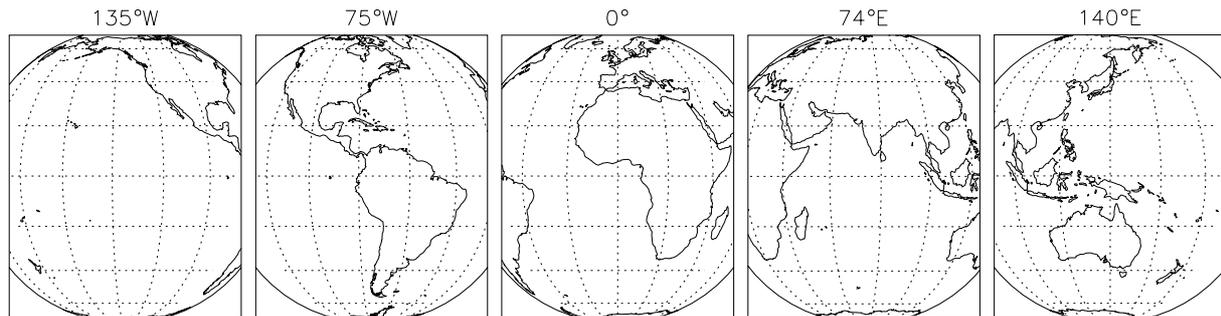


Figure 2.4: Global coverage of the international network of geostationary satellites.

### GOES (USA)

The U.S. satellite program under NOAA (National Oceanic and Atmospheric Administration) usually maintains two geo-stationary satellites of fully operational status belonging to the GOES (Geostationary Operational Environmental Satellites) series. Spacecraft GOES-E, stationed at 75°W, is viewing the eastern part of America and the western Atlantic Ocean, whereas GOES-W, stationed at 135°W, covers the western part of the American continent and the eastern Pacific Ocean. Following two SMS (Synchronous Meteorological Satellites) the first GOES-1 was launched in 1975, succeeded by GOES-2 to GOES-7 in 1994. These early satellites were spin-stabilized and viewed the Earth only about 10% of the time. The current satellites of the GOES I–M series (becoming GOES 8–12 after launch), also referred to as GOES-NEXT, are of an advanced design, where GOES-8 and GOES-10 are fully operational as GOES-E and GOES-W, respectively. GOES-9 and GOES-11 are back-up satellites (in storage) coming into operation if one of the current satellites fails. GOES-12 is scheduled to be launched in 2001. The existing series will be expanded by the GOES-N to GOES-Q missions with the first launch scheduled for late 2002. The satellites of the new series I–M weigh about 2000 kg and are stabilized by three axes to view the Earth all the time. Thus the main body of the satellite containing the imaging and sounding instruments always points towards the Earth. The sounder measures in 19 channels (one visible, 18 infrared). The field of view, sampled every 10 km, is 8 km. The imaging radiometer, consisting mainly of a mirror scanning system, a telescope and detectors, has five channels (one visible, four infrared). It scans west to east in one line, steps in north-south direction (~ 8 km) and begins the next line from east to west. The scanning rate for a full-disc image is approximately 25 minutes. Detailed information on operational status, instruments, products and other technical features of the GOES-satellites are available, e.g., under <http://www.goes.noaa.gov/>.

### METEOSAT (EUROPE)

The first Meteosat satellite was launched in 1977 by the European Space Agency (ESA). The current satellite, maintained by EUMETSAT (European Organization of Meteorological Satellites) is Meteosat-7 located (as its predecessors) at 0° longitude over the equator. The old Meteosat-5 satellite (launched in 1991) was positioned at 63° longitude in 1998 to support the INDOEX (International Indian Ocean Experiment) mission and the coverage of the Indian ocean by Meteosat-5 will be extended for a few more years after the end of this experiment. Meteosat-6 serves as a back-up spacecraft of Meteosat-7. While the spacecraft is spinning (at 100rpm) the radiometer onboard, that is MVIRI (Meteosat Visible and Infrared Imager) scans one line of Earth in three channels (one visible, one infrared, one water vapor) to provide images every 30 minutes. The next generation MSG (Meteosat Second Generation) of improved performance is under development as a cooperation between ESA and EUMETSAT. It will measure in 12 channels and create images of improved resolution every 15 minutes. The interested reader is referred to <http://www.eumetsat.de/>.

### INSAT (INDIA)

The INSAT satellites are operated by ISRO (Indian Space Research Organization). INSAT-1D was launched in 1990 and moved to longitude 74°E in 1999. It carries a VHRR (Very High Resolution Radiometer) with two spectral channels (one visible, one infrared). INSAT-2E is located at 83°E and the meteorological payload has an additional water vapor channel. The new generation INSAT 3A to 3E will be built and launched within the next 10 years. See <http://www.isro.org/> for further details.

### GMS (JAPAN)

The Japan Meteorological Agency (JMA) launched its first GMS (Geostationary Meteorological Satellite) satellite in 1977. The current satellite is GMS-5 located at longitude 140°E and launched in 1995. The spinning section of the satellite contains the VISSR instrument to image the Earth, clouds and several weather phenomena. Image data are disseminated by S-VISSR (stretched VISSR data) and WEFAX. The mission of GMS-5 will be continued by its successor MTSAT (Multipurpose Transportation Satellite) to be launched in 2003. The WebPage of Japan's meteorological agency is located at <http://www.kishou.go.jp/english/>.

### FENG-YUN (CHINA)

The current Feng-Yun-2 satellite, termed FY-2B, is operated by CMA (China Meteorological Institute) and NMSC (National Meteorological Satellite Center) and stationed at 105°E since 2000. The main payload is a VISSR (Visible and Infrared Spin Scan Radiometer) instrument producing hourly full-disk images of the Earth, a SEM as well as data transmission systems. FY-2B will be followed by FY-2C/D/E within the next ten years. Detailed information on China's geostationary satellites may be found on <http://www.nmc.gov.cn/>.

### GOMS (RUSSIA)

The Russian Planeta-C Meteorological Space System includes Russia's first geostationary weather satellite GOMS-1 (also termed ELEKTRO) operating at 77°E. The spacecraft is three-axes stabilized and weighs about 2400 kg. Images are produced by the STR (Scanning Television Radiometer) instrument with two channels (one visible, one infrared). The spacecraft also carries a space radiation and monitoring system as well as communications facilities to transmit images and data at 7.5 GHz to receiver stations. The GOMS-2 will carry an imager with an additional infrared water vapor channel and is scheduled to be launched in 2003. Further reading can be done at <http://sputnik.infospace.ru/>.

### 2.3.2 Satellites in Polar Orbits

Current operational meteorological satellites in (near-) polar low earth orbits are summarized in Tab. 2.3.

satellite	operator	orbit type*	launch date	inclination [deg]	altitude** [km]	period [min]
NOAA-15	USA	s-s (m)	05/1998	98.6	807	101
NOAA-16	USA	s-s (a)	09/2000	98.8	853	102
DMSP-F15	USA	s-s (m)	12/1999	98.9	837	102
DMSP-F13	USA	s-s (em)	03/1997	98.9	842	102
FY-1C	China	s-s (m)	05/1999	98.7	869	102
METEOR 2-21	Russia	non s-s	08/1993	82.5	934	104
METEOR 3-5	Russia	non s-s	08/1991	82.6	1184	109

Table 2.3: Current polar-orbiting satellites (status: December 2000). [ \* s-s...sun-synchronous, a...afternoon, m...morning, em...early morning, \*\* altitude at perigee ]

#### NOAA (USA)

In 1970 the first satellite of the NOAA-series was launched, following the original TIROS (Television and Infrared Observational Satellite) series. The latter, being the first satellites dedicated to meteorology, started with TIROS-1, which returned the first image of the Earth in 1960. In 1978 the first satellite of the next generation TIROS-N was launched to support the POES (Polar-orbiting Operational Environmental Satellite) program as a cooperation of NOAA, NASA, the United Kingdom and France. NOAA maintains a configuration of two POES satellites (complementing two GOES spacecrafts), one in morning and one in afternoon orbit. The current satellites are NOAA-15 and NOAA-16 representing the first satellites of the advanced NOAA-K/L/M/N series with improved imaging and sounding capabilities. The primary instruments onboard are AVHRR, the ATOVS suite and SEM. Data are transmitted to ground via HRPT and APT systems. The interested reader is referred to <http://www.osd.noaa.gov/>.

NOAA and EUMETSAT are currently establishing the IJPS (Initial Joint Polar System) to be established by NOAA and EUMETSAT to ensure operational coverage from polar orbit in the upcoming decades. The European component is the EPS (EUMETSAT Polar System) based on the METOP satellites to be described within the next chapter.

### DMSP (USA)

The DMSP (Defense Meteorological Satellite Program) operated by the U.S. air force includes two satellites in low earth orbit. The current satellites DMSP F-13 and DMSP F-15 carry SSM/I (Special Sensor Microwave Imager), SSM/T-1 (Special Sensor Microwave Temperature Sounder) and SSM/T-2 (Atmospheric Water Vapor Profiler). The data are available to civilian users through NOAA. More information may be found under <http://www.ngdc.noaa.gov/dmsp/>.

### METEOR (RUSSIA)

Satellites of the METEOR series are in a near-polar non-synchronous orbit with an inclination of about 82° at a height of about 1000 km. Meteor-3M will be the advanced series of polar-orbiting satellites. Russia also supports the OKEAN satellites for oceanographic applications and the RESURS mission for land-use/land-cover mapping. Details may be found under <http://sputnik.infospace.ru/>.

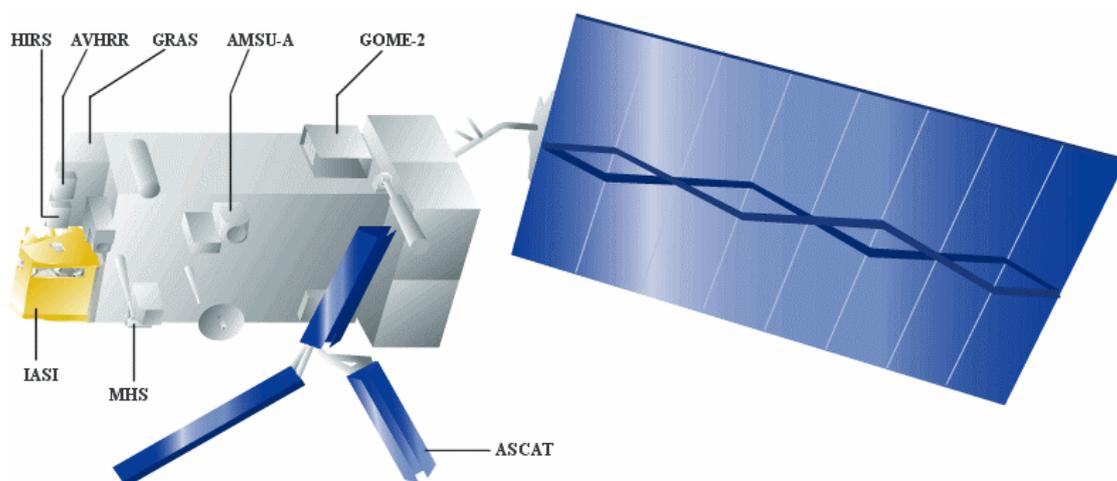
### FENG-YUN (CHINA)

The FY-1C satellite as the successor of FY-1A and FY-1B carries the MVISR (Multispectral Visible and Infrared Scan Radiometer) which measures in ten channels (four visible, three near infrared, three infrared). The resolution is about 1.1 km at nadir. Data are transmitted by the HRPT and CDPT (Chinese Delayed Picture Transmission). Details can be found at the WebPage of the Chinese National Meteorological Satellite Center at <http://www.cma.gov.cn/>.

### 3 THE IASI INSTRUMENT

The Infrared Atmospheric Sounding Interferometer IASI is being developed by CNES (Centre National d'Etudes Spatial) for EUMETSAT (European Organization for the Exploitation of Meteorological Satellites) as a core payload of the METOP satellites developed by ESA (European Space Agency) to be operated by EUMETSAT. The launch of the first of three METOP satellites planned for the EUMETSAT Polar System (EPS) is scheduled for late 2005. The orbit is sun-synchronous with a 9:30 equator crossing local time. This morning orbit will complement the afternoon orbit of the NOAA satellites in the Initial Joint Polar system (IJPS), which constitutes a coordinated European/U.S.A. plan to ensure operational data continuity. The first 6 months in orbit will be dedicated to commissioning and validation, followed for IASI by a pre-processing assessment phase of about 18 months. The operational phase will then be for the rest of the 5-years lifetime of IASI.

Besides the instrument onboard the satellite, the IASI system includes the data processing software in the EPS ground segment. CNES is in charge of the technical management and the data processing software while EUMETSAT is responsible for operating IASI and operational data processing/distributing.



**Figure 3.1:** METOP-1 satellite and meteorological payload instruments

(from: <http://sads.cnes.fr:8060/IASI/>).

The METOP-satellites are dedicated to operational meteorology as well as climate monitoring. The companion instruments of IASI on the METOP polar platform are AMSU-A, MHS, HIRS, AVHRR, GOME-2, GRAS and ASCAT (Fig. 3.1). For descriptions of these instruments see Section 2.2.3. This suite of instruments provides high resolution temperature and humidity soundings, global images of clouds and weather systems, and information on land/sea surfaces and ocean surface winds. In addition METOP will carry a data collection system to gather information from ground-based systems and supports Search and Rescue services as well. Image data will be transmitted continuously via METOP HRPT (High Rate Picture Transmission) and LRPT (Low Rate Picture Transmission).

The following sections may be regarded as a summary of information and various reports provided by CNES and EUMETSAT available at <http://sads.cnes.fr:8060/IASI/> such as the IASI Science Plan (Camy-Peyret and Eyre, 1999). We note that an introductory description to IASI and its objectives has also been given in Hiebler (1998).

### 3.1 IASI Objectives and Main Specifications

#### OBJECTIVES

The IASI instrument mounted on the METOP satellite platform and its main characteristics are illustrated in Fig. 3.2.

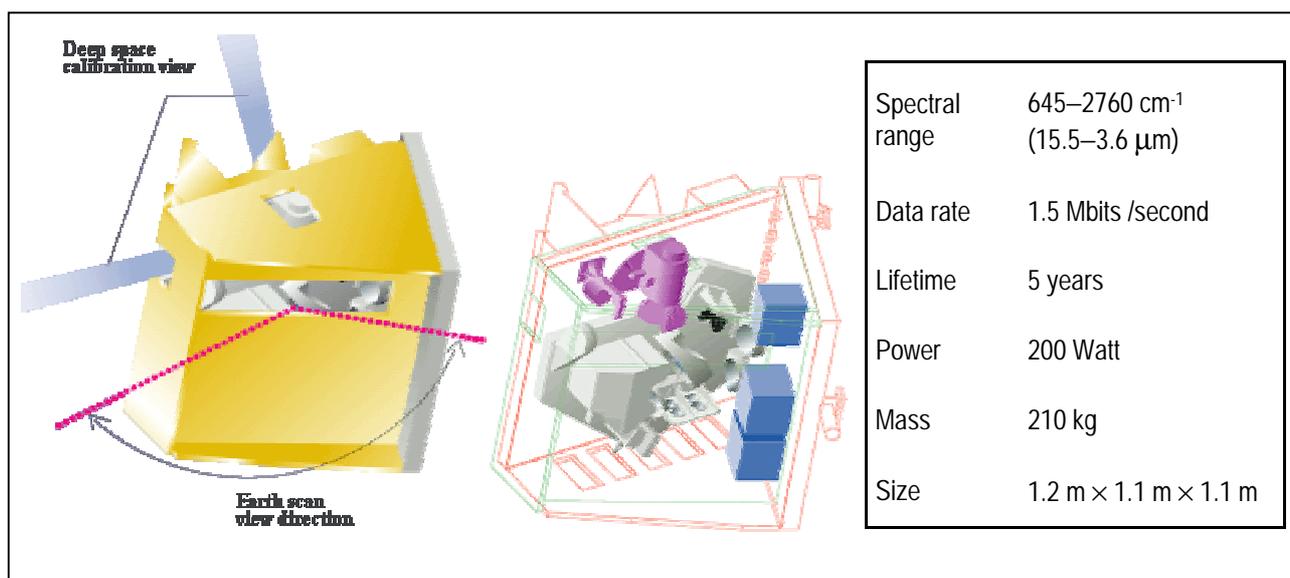


Figure 3.2: IASI Sensor (from: <http://sads.cnes.fr:8060/IASI/>) and main characteristics table.

IASI data will have potential for use in various operational and research applications. To exploit this potential, intensive and efficient research and development is required to prepare successful interpretation and application of IASI data. Current scientific activity, coordinated by the IASI Sounding Science Working Group (ISSWG), established by CNES and EUMETSAT in 1995, involves

- Radiative Transfer modeling to compare simulated radiances with IASI radiances
- Simulation of IASI instrument processes, as well as on-board and on-ground data ingest processes
- Development of data pre-processing methods including combination of IASI data with data from other METOP instruments
- Inverse modeling to retrieve temperature, humidity, minor constituents and many other geophysical variables
- Monitoring, quality control and validation of IASI data and products
- Preparatory studies of appropriate interfaces between IASI data and applications

The primary objective of IASI is the measurement of atmospheric profiles of temperature and humidity with improved resolution and accuracy as compared to current infrared instruments. The most important targets are to retrieve

- *temperature* profiles with an average accuracy of 1 K and 1 km vertical resolution
- *humidity* profiles with an accuracy of 10% and 1 to 2 km vertical resolution

Further products, which can be retrieved from IASI measurements include *cloud parameters* (cover, cloud top temperature, liquid water amount), column amounts of  $O_3$ ,  $N_2O$ ,  $CH_4$  and  $CO$ , *sea- and land-surface temperature* and *outgoing longwave radiation*.

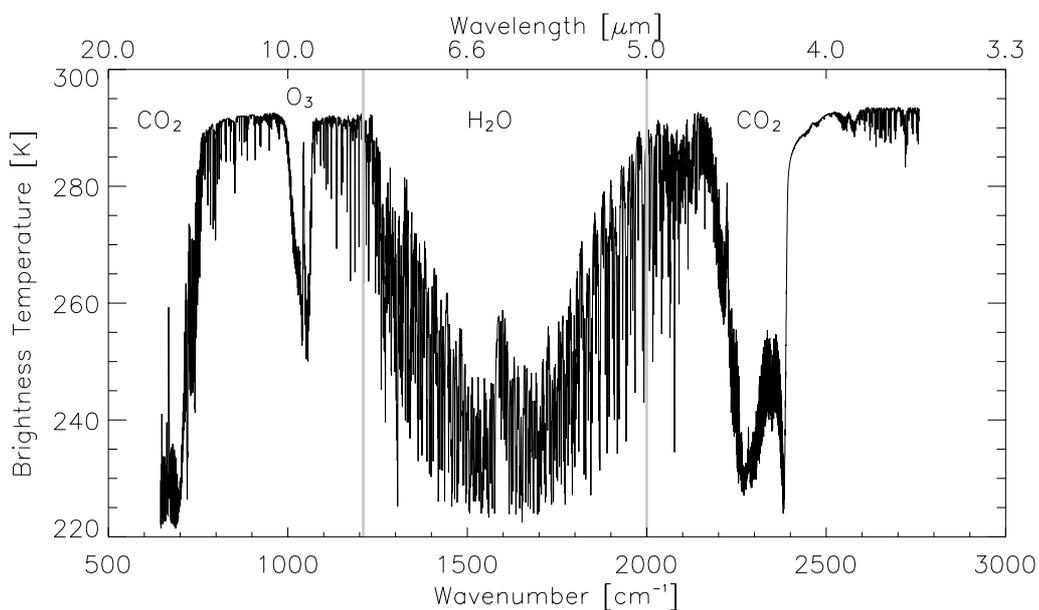
The mission objectives assigned to IASI and its companion meteorological instruments include

- *Operational meteorology*: comprises many activities concerning operational meteorology coordinated by the World Weather Watch (WWW) of the World Meteorological Organization (WMO) like forecasting beyond 12 hours and operational *Numerical Weather Prediction* (NWP) based on assimilation techniques and computing facilities. The high spectral resolution of IASI conjugated with the performance of the remaining METOP-instruments will provide temperature and humidity profiles of improved accuracy and vertical resolution, which will increase the quality of current NWP.

- *Climate monitoring and global change*: to understand global changes of the climate systematic observations of parameters such as sources and sinks of greenhouse gases, clouds which influence radiative transfer, oceans, land surfaces and the hydrological circle are needed.
- *Atmospheric chemistry*: increased concentrations of trace gases will effect the climate and the chemical equilibrium. Since IASI and its companion instruments will observe atmospheric constituents such as CH<sub>4</sub>, O<sub>3</sub>, and SO<sub>2</sub>, valuable information about chemical processes and their impact on atmospheric composition and climate processes can be obtained.

### SPECTRAL SPECIFICATIONS

The IASI sounder covers the spectral wavenumber range from 645 cm<sup>-1</sup> to 2760 cm<sup>-1</sup> (corresponding wavelength range is 15.5 μm to 3.62 μm) with a constant sampling interval of 0.25 cm<sup>-1</sup>. The nominal spectral sampling was determined by the linespacing in the CO<sub>2</sub> absorption bands, the final choice of 0.25 cm<sup>-1</sup> was found adequate to provide sufficient spectral resolution for all defined objectives. The IASI spectral range (depicted in Fig. 3.3) includes, in particular, the large water vapor absorption band at 1250–2000 cm<sup>-1</sup> (8–5 μm, center 6.3 μm) and two strong CO<sub>2</sub> bands, one near 645 cm<sup>-1</sup> (15.5 μm), the other near 2325 cm<sup>-1</sup> (4.3 μm). Finite spectral signatures of interest derive from O<sub>3</sub>, CH<sub>4</sub>, N<sub>2</sub>O, CO, and SO<sub>2</sub> (see Tab. 1.3, Sect. 1.5.1 for details).



**Figure 3.3:** Brightness temperature spectrum covering the IASI spectral range from 645 cm<sup>-1</sup> to 2760 cm<sup>-1</sup> (15.5 μm – 3.62 μm). Grey lines divide the full range into three spectral bands.

In order to specify the spectral resolution of the instrument we first need to introduce a few terms, which will be discussed in Sect. 3.2.2. in more detail. The spectrum measured by the instrument

equals the ideal spectrum convolved by an appropriate instrument function, termed as the Instrumental Spectral Response Function (ISRF). The spectral resolution  $\delta\nu$  is then defined as the full width at half maximum of the ISRF and is  $0.35 \text{ cm}^{-1}$ . Apodization refers to the removal of false sidelobes of the instrumental spectrum, introduced by finite optical path displacements in the interferometer. Resolution slightly degrades due to apodization, leading to  $0.5 \text{ cm}^{-1}$  resolution for the apodized spectrum. A further relevant instrumental parameter is the shape error index  $\epsilon$ , which specifies the error of the knowledge of the ISRF.

As indicated in Fig. 3.3 the full spectral range is subdivided into three spectral bands. Tab. 3.1 gives the boundaries of the three bands, their unapodized resolution  $\delta\nu$ , the shape error index  $\epsilon$  and the parameters, which may be retrieved in the particular band. Note that the apodized resolution will be slightly lower.

	range [ $\text{cm}^{-1}$ ]	range [ $\mu\text{m}$ ]	$\delta\nu$ [ $\text{cm}^{-1}$ ]	$\epsilon$	retrieved parameter
Band 1	645 – 1210	15.5 – 8.26	$\leq 0.35$	$\leq 0.046$	temperature, $\text{O}_3$ , surface and cloud properties
Band 2	1210 – 2000	8.26 – 5.0	$\leq 0.39$	$\leq 0.056$	humidity, $\text{N}_2\text{O}$ , $\text{CH}_4$ , $\text{SO}_2$
Band 2	2000 – 2760	5.0 – 3.62	$\leq 0.50$	$\leq 0.092$	temperature, $\text{CO}$ , $\text{N}_2\text{O}$ , $\text{CH}_4$ , surface and cloud properties

Table 3.1: IASI spectral bands, resolution  $\delta\nu$ , shape error index  $\epsilon$ , and retrievable parameters.

### RADIOMETRIC SPECIFICATIONS

The IASI *sounder* measures radiance from a blackbody within the temperature range 4 to 315 Kelvin. The radiometric noise is specified in terms of noise equivalent to temperature difference (NEAT) for a given reference temperature of 280 K. This includes all noise contributions such as noise induced by detectors, amplifiers, analogue/digital converter, and digital data processing and errors due to field-of view motion, fluctuations of wavelength and radiometric calibration or the knowledge of ISRF. Tab. 3.2 gives the radiometric noise for a reference temperature of 280 K for selected wavenumbers.

wavenumber [ $\text{cm}^{-1}$ ]	NEAT [K]	wavenumber [ $\text{cm}^{-1}$ ]	NEAT [K]
700	0.157	1700	0.16
1000	0.165	2000	0.13
1200	0.095	2500	0.7
1500	0.111	2700	1.6

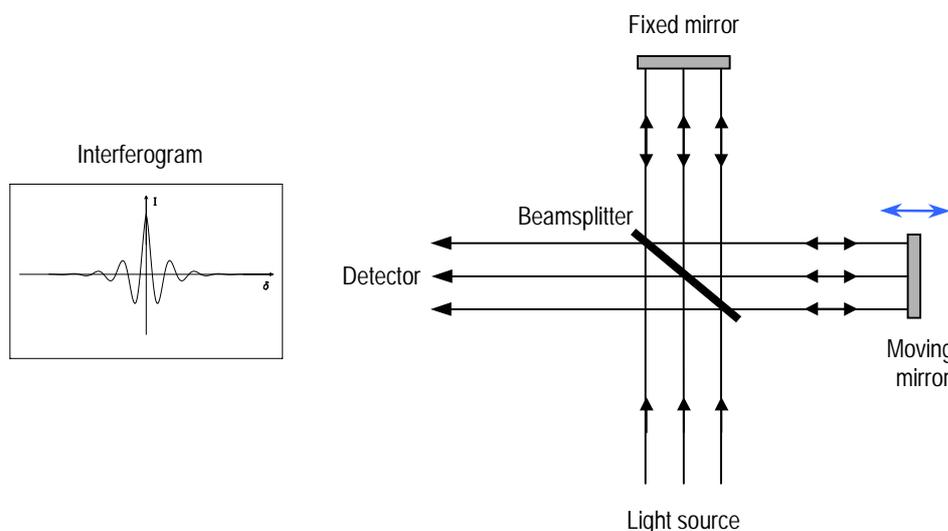
Table 3.2: IASI radiometric noise in terms of NEAT for selected wavenumbers (see also Tab. A5)

IASI calibration relies on measurements of cold and hot targets once every scan line. The absolute calibration accuracy is better than 0.5 K at 280 K. The calibration homogeneity depends on parameters such as spectral position, geometry, orbital and lifetime repeatability, each nominally introducing an error less than 0.15 K.

The IASI *imager* (mainly for supporting cloud detection) is a broad band radiometer with one infrared channel and measures a blackbody radiance within 4–315 K. NE $\Delta$ T will be less than 0.5 K at a reference temperature of 280 K. The calibration accuracy is better than 1 K at 280 K.

## 3.2 Introduction to Fourier Transform Spectroscopy

The IASI instrument concept is based on the Michelson interferometer principle illustrated in Fig. 3.4. The incoming radiation beam is amplitude-divided by the beamsplitter. One part goes to the fixed mirror, where it is reflected back to the beamsplitter, where it is reflected towards the detector. The other part is reflected by the beamsplitter, goes to a moving mirror, where it is reflected. It is then transmitted through the beamsplitter towards the detector.



**Figure 3.4:** Optical diagram of the basic Michelson interferometer

The moving mirror can be translated smoothly within a distance (arm displacement)  $\pm L$  to obtain an optical path difference (OPD) denoted by  $\delta$  within  $2L$  of the two beams at the detector. The variation of the OPD produces an interferogram when the two phase-shifted beams interfere with each other. Thus the interferogram corresponds to the detection of the signal versus OPD.

Each wavenumber produces its own characteristic flux pattern of the combined beams. For a polychromatic source the interferogram is the sum of the fluxes of each wavelength. The spectrum

(radiance versus wavenumber) is obtained by Fourier Transform (FT) analysis of the interferogram.

Albert A. Michelson (1852–1931) constructed his two-beam interferometer more than a century ago to measure the motion of earth relative to an ‘ether’ in the famous Michelson-Morley experiment. But only half a century later the interferometer has been recognized as a powerful spectrometer due to its advantages over conventional instruments such as very large resolving power, high wavenumber accuracy, fast scan times and large wavenumber range per scan. This advantages arise basically from two concepts (Bell, 1972). The first is termed the throughput or *étendue advantage* expressed by Jaquinot, which states that the flux throughput and brightness can be considered constant in a lossless optical system from the source to the detector. The size of the throughput is proportional to the power (energy at high resolution) and it was found that more power can be put through an interferometer than through the best grating spectrometers. For the latter the power is limited by the area of the entrance slit. The second advantage is the *multiplex advantage* first stated by Fellgett: within a given range the interferometer measures all spectral intervals simultaneously. The conventional grating spectrometer, on the other hand, receives information only in a narrow band (determined by the exit slit) at a given time.

### 3.2.1 Fourier Transform Analysis and Interferometry

As mentioned above the desired radiance spectrum can be obtained by Fourier Transform Analysis of the interferogram. The Fourier Transform  $F$  is mathematically defined as

$$F(y) = \int_{-\infty}^{\infty} f(x) \exp[-i2\pi yx] dx \equiv F\{f(x)\}, \quad (3.2.1)$$

and the inverse FT  $F^{-1}$  as

$$f(x) = \int_{-\infty}^{\infty} F(y) \exp[i2\pi xy] dy \equiv F^{-1}\{F(y)\}. \quad (3.2.2)$$

These two formulas constitute the Fourier transform pair, used to move from the spatial or temporal ( $x$ -)domain into a frequency ( $y$ -)domain. Eq. 3.2.1 can be rewritten as

$$F(y) = F_c(y) - iF_s(y), \quad (3.2.3)$$

where  $F_c(y)$  is the Fourier cosine transform given by

$$F_c(y) = \int_0^{\infty} f(x) \cos(2\pi yx) dx \equiv F_c \{f(x)\}, \quad (3.2.4)$$

and  $F_s(y)$  is the Fourier sine transform given by

$$F_s(y) = \int_0^{\infty} f(x) \sin(2\pi yx) dx \equiv F_s \{f(x)\}. \quad (3.2.5)$$

Note that if  $f(x)$  is an even function,  $F_c(y)$  can be used instead of  $F(y)$  given by Eq. 3.2.1.

The following derivation is mainly based on Bell (1972). A plane monochromatic wave at point  $z$  and time  $t = 0$  can be described by

$$y(z) = a(v) \cos(2\pi v z) dv, \quad (3.2.6)$$

where  $v$  is the wavenumber and  $a(v)$  is the (spectral) amplitude, whereas a wavetrain composed of many waves of different wavenumber and amplitude has an amplitude given by

$$y(z) = \frac{1}{v} \int_0^{\infty} a(v) \cos(2\pi v z) dv, \quad (3.2.7)$$

and  $a(v)/\bar{v}$  can be regarded as an amplitude density. To apply FT analysis we want the amplitude limits to be  $\pm \infty$ . Since

$$\int_0^{\infty} f(v) \cos(2\pi v z) dv = \frac{1}{2} \int_{-\infty}^{\infty} f(v) \exp[i2\pi v z] dv \quad (3.2.8)$$

for an arbitrary real and symmetric function  $f(v)$ , we can rewrite Eq. 3.2.7 as

$$y(z) = \frac{1}{2} \int_{-\infty}^{\infty} b(v) \exp[i2\pi v z] dv = \int_{-\infty}^{\infty} E(v) \exp[i2\pi v z] dv, \quad (3.2.9)$$

where we defined  $\frac{1}{2} b(v) \equiv E(v)$  as the electric field amplitude of wavenumber  $v$ . As stated in Eq. 3.2.9 the electric field as a function of wavenumber  $E(v)$  is the FT of the electric field as a function of position  $y(z)$ . Two waves from each arm of the interferometer with the optical path difference  $\delta$  after recombination can be expressed by

$$y_1(z) = \int_{-\infty}^{\infty} E(v) \exp[i2\pi v z] dv \quad (3.2.10)$$

$$y_2(z) = \int_{-\infty}^{\infty} E(v) \exp[i2\pi v (z - \delta)] dv. \quad (3.2.11)$$

The intensity  $B(\nu)$  (irradiance or flux density) of an electric field is proportional to the square of its amplitude, that is,

$$B(\nu) = \frac{1}{2} c \epsilon_0 E^2(\nu), \quad (3.2.12)$$

where  $c$  is the speed of light and  $\epsilon_0$  is the electric permittivity of free space (Tab. A1). Applying the law of superposition  $y(z) = y_1(z) + y_2(z)$  we get

$$\begin{aligned} y(z) &= \int_{-\infty}^{\infty} E(\nu)(1 + \exp[-i2\pi\nu\delta]) \exp[i2\pi\nu z] d\nu \\ &= \int_{-\infty}^{\infty} E_R(\delta, \nu) \exp[i2\pi\nu z] d\nu, \end{aligned} \quad (3.2.13)$$

where the resultant electric field as a function of wavenumber and OPD  $\delta$  is

$$E_R(\delta, \nu) = E(\nu)(1 + \exp[-i2\pi\nu\delta]). \quad (3.2.14)$$

The intensity of the resultant field is then

$$B_R(\delta, \nu) = \frac{1}{2} c \epsilon_0 E_R^*(\delta, \nu) E_R(\delta, \nu), \quad (3.2.15)$$

where  $*$  denotes complex conjugate.

Substituting Eq. 3.2.14 and utilizing  $\cos \alpha = \frac{1}{2}(\exp[i\alpha] + \exp[-i\alpha])$  we get

$$\begin{aligned} B(\delta, \nu) &= c \epsilon_0 E^2(\nu)[1 + \cos(2\pi\nu\delta)] \\ &= 2B(\nu)[1 + \cos(2\pi\nu\delta)]. \end{aligned} \quad (3.2.16)$$

As mentioned previously, the total flux is the sum (integral) of all individual flux densities. Thus

$$\begin{aligned} I_R(\delta) &= \int_0^{\infty} B(\delta, \nu) d\nu \\ &= c \epsilon_0 \left\{ \int_0^{\infty} E^2(\nu) d\nu + \int_0^{\infty} E^2(\nu) \cos(2\pi\nu\delta) d\nu \right\} \\ &= 2 \left\{ \int_0^{\infty} B(\nu) d\nu + \int_0^{\infty} B(\nu) \cos(2\pi\nu\delta) d\nu \right\}. \end{aligned} \quad (3.2.17)$$

Usually  $[I_R(\delta) - \frac{1}{2}I_R(0)]$  is referred to as the interferogram, that is, the oscillation of  $I_R(\delta)$  about  $\frac{1}{2}I_R(0)$ . The value at zero path difference (ZPD), i.e.,  $\delta = 0$ , is obtained via Eq. 3.2.17 as

$$\begin{aligned}
I_R(0) &= 2c\epsilon_0 \int_0^{\infty} E^2(\nu) d\nu \\
&= 4 \int_0^{\infty} B(\nu) d\nu.
\end{aligned} \tag{3.2.18}$$

Thus the interferogram can be evaluated as

$$\begin{aligned}
[I_R(\delta) - \frac{1}{2}I_R(0)] &= c\epsilon_0 \int_0^{\infty} E^2(\nu) \cos(2\pi\nu\delta) d\nu \\
&= 2 \int_0^{\infty} B(\nu) \cos(2\pi\nu\delta) d\nu.
\end{aligned} \tag{3.2.19}$$

Finally, the Fourier cosine transform of Eq. 3.2.19 yields the basic equation of Fourier transform spectroscopy, which reads

$$\boxed{B(\nu) = 2 \int_0^{\infty} [I_R(\delta) - \frac{1}{2}I_R(0)] \cos(2\pi\nu\delta) d\delta,} \tag{3.2.20}$$

stating that  $B(\nu) = F_C \{ [I_R(\delta) - \frac{1}{2}I_R(0)] \}$ , i.e.,  $B(\nu)$ , the intensity at wavenumber  $\nu$  is the Fourier cosine transform of the interferogram, the flux versus the optical path difference  $\delta$ . First one needs to determine  $I_R(\delta)$ . In the case of IASI a mechanical drive moves the mirror (corner cube) by  $\pm 1$  cm, thus the OPD is explored within 2 cm. Then  $I_R(0)$  is determined at ZPD and  $B(\nu)$  is calculated according Eq. 3.2.20 in order to obtain the full spectrum.

Note that Eq. 3.2.17 is valid only for symmetric interferometers, since we assumed that the electric field is a real quantity. If the mirrors in the two arms of the interferometer reflect the incident beams unequally, the complex Fourier transform for the computation of  $B(\nu)$  should be used. For the general case, considering symmetric or asymmetric interferometers, we have

$$\boxed{B(\nu) = \int_{-\infty}^{\infty} [I_R(\delta) - \frac{1}{2}I_R(0)] \exp[-i2\pi\nu\delta] d\delta,} \tag{3.2.21}$$

expressing that  $B(\nu) = F \{ [I_R(\delta) - \frac{1}{2}I_R(0)] \}$ . Note that constant terms have been omitted. For details on asymmetric interferometers the reader is referred to Bell (1972).

Another fact is worth mentioning: the interferogram can be viewed as the autocovariance function of the electric field amplitude. In general, the Wiener-Khinchine Theorem states, that the inverse Fourier transform of the autocovariance function of the electric field amplitude is the spectrum. In

other words, the interferogram is the inverse FT of the spectrum as can be directly seen in Eq. 3.2.19 and implicitly in the more general Eq. 3.2.21.

### 3.2.2 Apodization and Instrument Response Function

The basic Fourier transform integrals Eq. 3.2.20 and Eq. 3.2.21 have infinite limits for the optical path difference, but in practice we can only measure the interferogram and calculate the spectrum out to some finite OPD. This produces sidelobes or the phenomena called ‘ringing’. The necessary modification to rectify these sidelobes is termed apodization.

The spectrum calculated for finite OPD is called *instrumental spectrum*  $B_I(\nu)$  given by

$$B_I(\nu) = \int_{-L}^L [I_R(\delta) - \frac{1}{2}I_R(0)] \exp[-i2\pi\nu\delta] d\delta, \quad (3.2.22)$$

where L is the maximum arm displacement. This spectrum would introduce sidelobes due to finite OPD. If we multiply the integrand by an *apodization function*  $A(\delta)$  the resulting spectrum is

$$B_I(\nu) = \int_{-L}^L [I_R(\delta) - \frac{1}{2}I_R(0)] A(\delta) \exp[-i2\pi\nu\delta] d\delta, \quad (3.2.23)$$

expressing that  $[I_R(\delta) - \frac{1}{2}I_R(0)] A(\delta) = F^{-1}\{B_I(\nu)\}$ . The spectrum Eq. 3.2.23 will be closer to the true spectrum than the unapodized spectrum Eq. 3.2.22 because sidelobes will be removed by a certain factor. The expense will be a widening of the spectrum. Let us, for reference, first choose a rectangular function for  $A(\delta)$  defined by

$$A(\delta) = \begin{cases} 1 & |\delta| < L \\ 0 & |\delta| > L \end{cases}. \quad (3.2.24)$$

We may introduce a function  $F(\nu)$  defined as

$$F\{A(\delta)\} \equiv F(\nu). \quad (3.2.25)$$

Since we have  $[I_R(\delta) - \frac{1}{2}I_R(0)] = F^{-1}\{B(\nu)\}$  for the ideal (true) spectrum, we may write

$$F^{-1}\{B_I(\nu)\} = F^{-1}\{B(\nu)\} F^{-1}\{F(\nu)\}. \quad (3.2.26)$$

The convolution theorem (the FT of the convolution of two functions equals the product of the FTs of the functions) then gives

$$F^{-1}\{B_I(\nu)\} = F^{-1}\{B(\nu) \otimes F(\nu)\}, \quad (3.2.27)$$

where  $\otimes$  denotes convolution. The FT of a rectangular function gives a sinc function defined as  $\text{sinc}(x) = \sin(x)/x$ . Therefore, the FT of the rectangular apodizing function  $A(\delta)$  given in Eq. 3.2.24 can be evaluated as

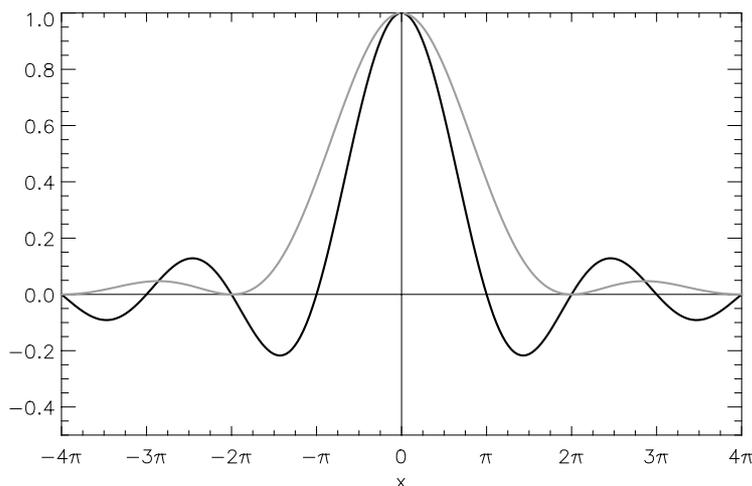
$$F(\nu) = 2L\text{sinc}(2\pi\nu L). \quad (3.2.28)$$

In general,  $F(\nu)$ , representing the FT of the apodizing function, is termed the *Instrument Spectral Response Function* (ISRF) or just instrument function. Finally, we may write Eq. 3.2.27 in the form

$$\boxed{B_I(\nu) = B(\nu) \otimes \text{ISRF}}, \quad (3.2.29)$$

highlighting that the instrumental spectrum equals the convolution of the true spectrum with the instrumental spectral response function.

The impact of apodization may be shown by means of a monochromatic source. The spectrum of a monochromatic source with wavenumber  $\nu_1$  can be represented as a positive Dirac delta function  $\frac{1}{2}\{\delta_D(\nu - \nu_1) + \delta_D(\nu + \nu_1)\}$ , where  $\delta_D$  denotes the Dirac delta symbol to distinguish it from the OPD  $\delta$ . In this case the interferogram (i.e., the inverse FT of the spectrum) is simply a cosine wave given by  $\cos(2\pi\nu_1\delta)$ . Now, the (unapodized) spectrum of a monochromatic source calculated with finite OPD is  $B_I(\nu_1) = 2L\text{sinc}(2\pi\nu_1 L)$ , which can be identified as Eq. 3.2.28 with  $\nu = \nu_1$ . The function  $\text{sinc}(x)$  with  $x = 2\pi\nu_1 L$  is indicated as the black line in Fig. 3.5.



**Figure 3.5:** Black line:  $\text{sinc}(x)$  versus  $x$ , gray line:  $\text{sinc}^2(x/2)$  versus  $x$ .

We notice quite significant (and deep) sidelobes. A triangular apodization function  $A(\delta) = 1 - |\delta|/L$  yields after FT an instrument function described by  $L\text{sinc}^2(\pi\nu L)$ , represented by  $\text{sinc}^2(x/2)$  as the gray line in Fig. 3.5. We see that this apodization reduces the depth of the sidelobes but results in a broadening of spectral features.

The *spectral resolution*  $\delta\nu$  describes the line width of a very narrow line introduced by the instrument. For a perfectly monochromatic source the (unapodized) spectrum is  $B_I(\nu_1) = 2L\text{sinc}(2\pi\nu_1 L)$ . Investigating Fig. 3.5 we find that the FWHM (full width at half maximum) of that function is  $1.21\pi$  and the spectral resolution in terms of wavenumber is

$$\delta\nu = 1.21/2L \quad (3.2.30)$$

where  $L$  is the maximum arm displacement of the interferometer (i.e., the unapodized spectral resolution of IASI is  $\sim 0.3 \text{ cm}^{-1}$ ). Note that the resolution for the (triangular) apodized spectra gives  $\delta\nu = 1.77/2L$  according to Fig. 3.4, which is slightly higher than in Eq. 3.2.30 due to increased FWHM. For given  $L$  the sinc-function as the FT of the rectangular apodization function yields the highest resolution, but introduces quite strong sidelobes. In Tab. 3.3 a few commonly used apodization functions, their instrument functions and FWHM's are summarized.

Name	Apodizing function	Instrument function	FWHM
Bartlett (triangular)	$1 - \frac{ \delta }{L}$	$L\text{sinc}^2(\pi\nu L)$	$1.77\pi$
Cosine	$\cos\left(\frac{\pi\delta}{2L}\right)$	$\frac{4L \cos(2\pi\nu L)}{\pi(1 - 16L^2\nu^2)}$	$1.64\pi$
Gaussian*	$\exp\left(-\frac{\delta^2}{2\sigma^2}\right)$	$2 \int_0^L \cos(2\pi\nu\delta) \exp\left(-\frac{\delta^2}{2\sigma^2}\right) d\delta$	—
Hamming	$0.54 + 0.46 \cos\left(\frac{\pi\delta}{L}\right)$	$\frac{L(1.08 - 0.64L^2\nu^2)\text{sinc}(2\pi\nu L)}{1 - 4L^2\nu^2}$	$1.82\pi$
Hanning	$\cos^2\left(\frac{\pi\delta}{2L}\right)$	$\frac{L\text{sinc}(2\pi\nu L)}{1 - 4L^2\nu^2}$	$2.00\pi$
Uniform (rectangular)	1	$2L\text{sinc}(2\pi\nu L)$	$1.21\pi$

**Table 3.3:** Selected apodization functions, their corresponding instrument functions and the FWHM (full width at half maximum) of the instrument function. [\*  $\sigma^2$  is the Gaussian variance, FWHM depends on  $\sigma^2$  ]

### 3.3 IASI Functional Components

The path of atmospheric radiation through the instrument is functionally described in Fig. 3.6. The IASI sounder can be separated into three modules. These are the optical system, the analogue electronic chain and the digital signal processing. The single components and the IASI Integrated Imager System are described below.

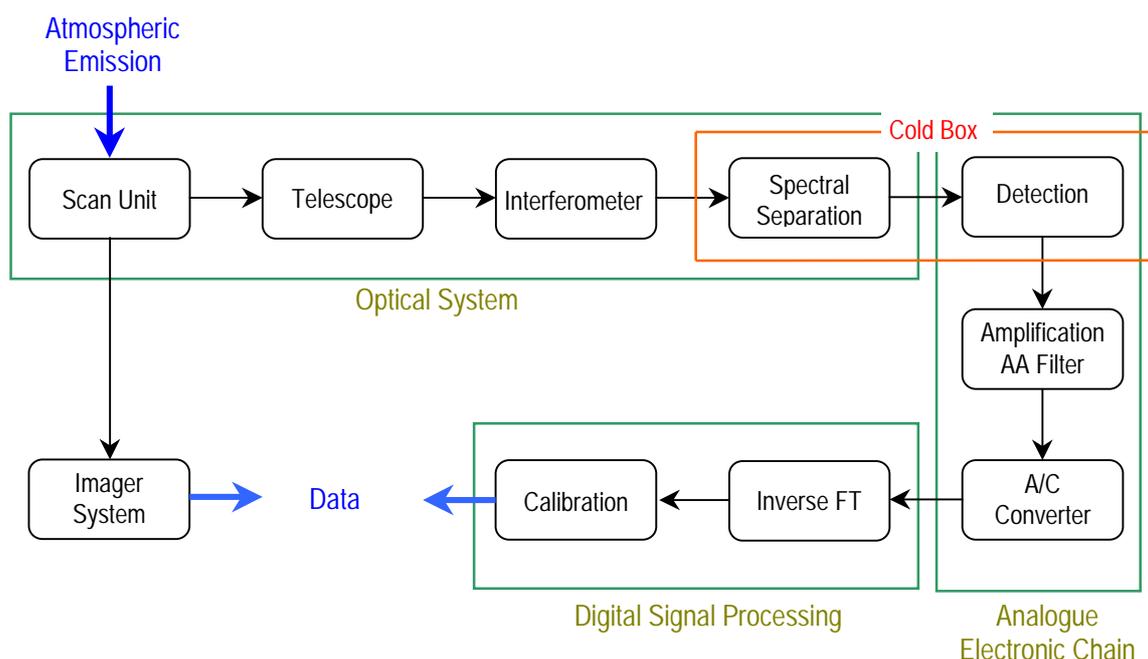


Figure 3.6: Functional description of the IASI instrument

#### THE SCAN UNIT AND THE TELESCOPE

The scan unit (rotating mirror) is shared by sounder and imager and provides scanning according to the geometric pattern depicted in Fig. 3.7. IASI scans across-track in a ‘step and dwell’ mode with a field of view of  $48.3^\circ$  to each side, which gives a swath-width of approximately  $\pm 1030$  km.

The instrument makes 30 Earth views (pixels) per scan, one every 216 ms (step time), with an IFOV size of  $3.33^\circ \times 3.33^\circ$ . The dwell time for each pixel is 150 ms. The size of the IFOV at nadir is about  $48 \times 48$  km<sup>2</sup>. In each pixel 4 sub-pixels with a diameter of  $0.84^\circ$  or 12 km are analyzed simultaneously by 4 detectors, arranged in form of a  $2 \times 2$  array. The total scan period, including 30 Earth views, calibration measurements and returning to the starting position, lasts 8 seconds.

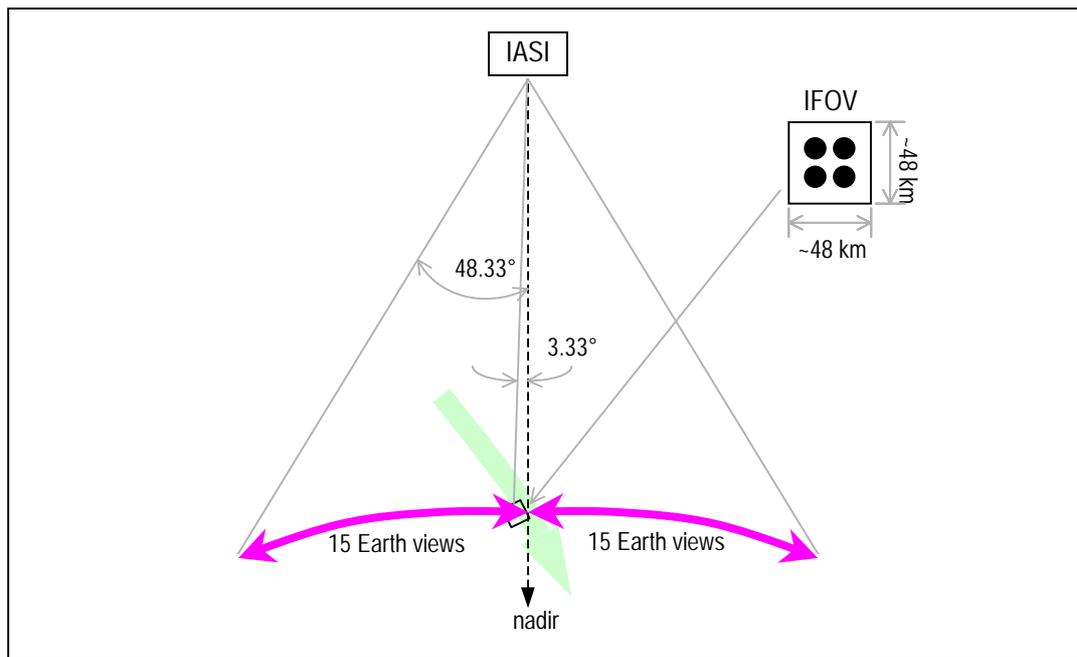


Figure 3.7: IASI scanning geometry.

The calibration is done by viewing a blackbody of known temperature (270–300K) mounted inside the instrument, as well as viewing a cold blackbody (deep space). Then a scan separation step of about 53 km in platform-moving direction is accomplished.

An afocal telescope transfers each 150 ms aperture stop of the scanning mirror into the interferometer entrance mirror (the entry of signal into the interferometric assembly is sketched in Fig. 3.8 as mirror C).

### THE INTERFEROMETER

The heart of the instrument and the main part of the optical configuration of IASI, shown in Fig. 3.8, is a Michelson Interferometer. The incident radiation (mirror C) is divided into two beams by a beamsplitter. One part of the radiation beam goes through a fixed mirror, which in case of IASI is a corner cube reflector, which reflects the beam back to the beamsplitter. There it is reflected to the folder and the focusing mirror. The other beam is reflected from the beamsplitter and goes to a movable corner cube reflector, which moves linearly by  $\pm 1$  cm corresponding to an optical path difference (OPD) of 2 cm. The reflected beam then transmits through the beamsplitter and finally reaches the folder and focusing mirror. The focusing mirror recombines the two beams at the detector. As mentioned above, an interferometer does not measure the incident radiative spectrum (radiance versus wavenumber) but the interferogram, which is the recording of the detected signal versus the optical path difference of the two beams (see, e.g., Eq. 3.2.19).

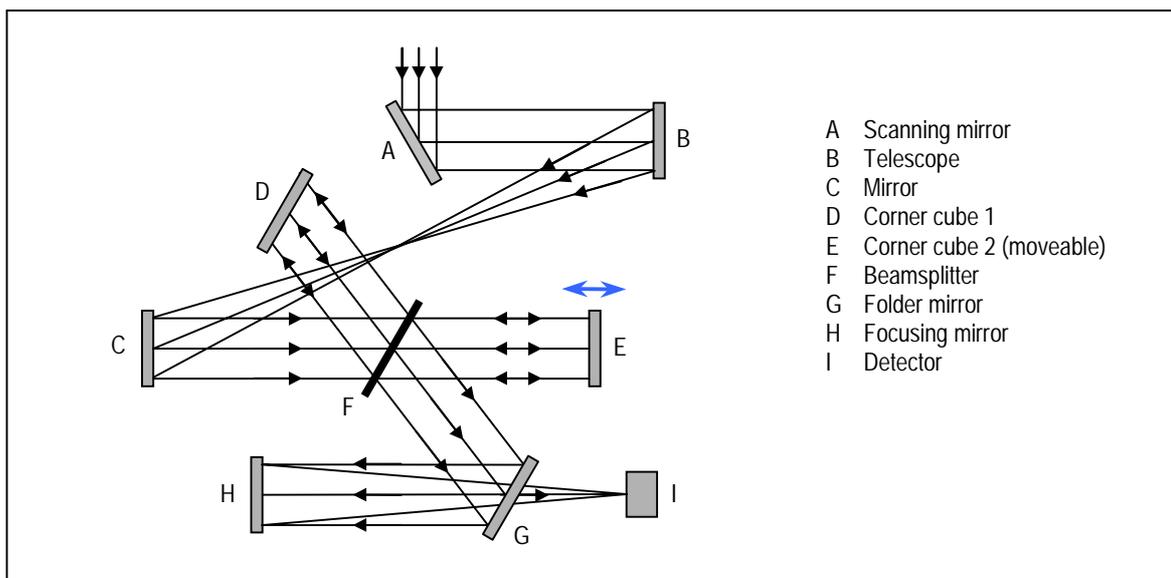


Figure 3.8: Optical configuration of the IASI Michelson interferometer and associated components.

### SPECTRAL SEPARATION

The radiometric performance to be secured during spectral separation and detection requires a cooling of the cold box hosting these functions to below 100 K. The output of the interferometer, the relative amplitude versus optical path difference summed over all frequencies, is separated into three bands,  $645\text{--}1210\text{ cm}^{-1}$ ,  $1210\text{--}2000\text{ cm}^{-1}$ , and  $2000\text{--}2760\text{ cm}^{-1}$ , which are optically separated by two dichroic plates.

### DETECTION, AMPLIFICATION AND ANTI-ALIASING FILTER

For each pixel the detection is performed by three conductors, one for each of the three wavenumber bands. Band 1 and Band 2 use a photoconductive detector made of mercury cadmium telluride (HgCdTe) and a photovoltaic detector made of the same semiconducting material, respectively. Band 3 uses a photovoltaic detector made of indium antimonide (InSb).

Each band is pre-amplified to obtain three identical circuits, which are then amplified and subject to a Butterworth anti-aliasing filter.

### ANALOGUE / DIGITAL CONVERTER

This process highly depends on constant time increment sampling at very precise and stable known increments of the optical path difference (OPD). An auxiliary monochromatic laser beam is used to measure the OPD and triggers the sampling of the interferogram. The sampling is triggered and the full signal is digitized, converting the optical interferogram into a digital interferogram.

### INVERSE FOURIER TRANSFORM AND CALIBRATION

All digitized interferograms are processed on-board for detection and correction of spurious effects (basically non-linearities). Then the interferogram is Fourier transformed to obtain a digital spectrum. Calibration with the two calibration targets (cold space and internal black body) is performed to produce calibrated spectra. These spectra are, after some proper formatting, transmitted to the EPS-ground station as Level 0 data.

### THE IMAGER SYSTEM

The IASI imager system consists of the imager optics, detectors, amplifiers and the A/D converter. The imager optics share the scan mirror with the sounder system. It images the IFOV on a detector matrix, composed of 128×128 photovoltaic detectors, representing the physical field stops for the imager pixels. The physical aperture stop is conjugated with the scanning mirror. The 128×128 imager pixels are compressed to 64×64, the signal is amplified, digitized and transmitted to ground station for further processing.

## **3.4 IASI Data and Products**

IASI data and products are currently classified in Level 0 to 3, where all data except Level 0 are processed on ground.

Level 0: Raw telemetry data, including housekeeping, on-board calibration, verification, elimination of redundant information.

Level 1a: Data after decoding, spectral calibration, radiometric post-calibration, IASI/AVHRR co-registration, locating and dating.

Level 1b: Resampled radiance spectra and imager data, corrected from all geometrical and instrumental effects based on calibration.

Level 1c: Level 1b after apodization.

Level 2a: Standard geophysical products derived from IASI 1c data: profiles of temperature, humidity, ozone, trace gas (such as CO, N<sub>2</sub>O, CH<sub>4</sub>) distributions, surface temperature, outgoing longwave radiance and cloud characteristics, etc.

Level 2b: Geophysical products derived after co-processing of IASI and companion meteorological instruments, similar to Level 2a data, but with more accuracy and/or resolution, as well as being representative of a wider range of cloudy conditions.

Level 3: Geophysical products, gridded and time-averaged, possibly in combination with information from other sources than EPS.

## 4 RETRIEVAL OF PHYSICAL PARAMETERS FROM MEASUREMENTS

### 4.1 General Considerations on Retrieving Atmospheric Parameters

Atmospheric constituents absorb electromagnetic radiation at specific wavelengths and emit radiation at the same wavelengths. The intensity of the radiation depends on the temperature of the emitting gas according to the Planck intensity at wavenumber  $\nu$ ,

$$B_\nu(T) = \frac{2h\nu^3 c^2}{\exp(hc\nu/kT) - 1} \quad (4.1.1)$$

(see Sect. 1.3). The radiometer onboard a meteorological satellite measures the incoming radiation. The measured radiance  $I_\nu$  constituting the solution of the forward radiative transfer equation at wavenumber  $\nu$  in pressure coordinates is

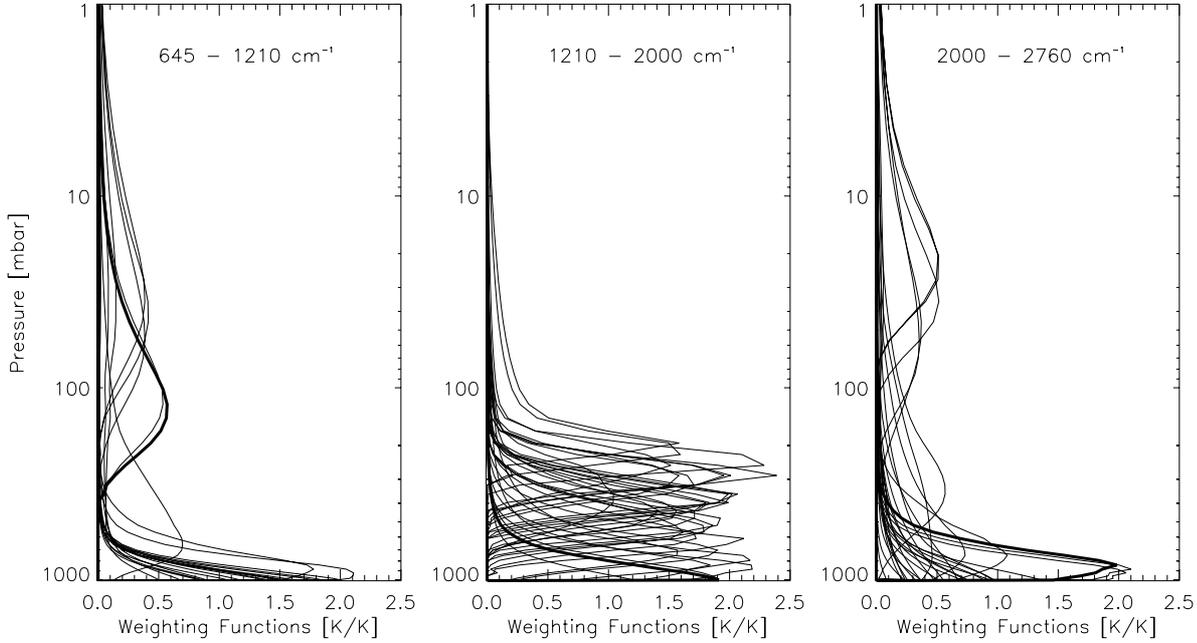
$$I_\nu = \int_0^\infty B_\nu(T(p)) W_\nu(p) dp, \quad (4.1.2)$$

where  $B_\nu(T(p))$  is the Planck function at temperature  $T$  and pressure level  $p$ . The measurement can also be specified in terms of brightness temperature  $T_B$ , which is the blackbody temperature corresponding to the radiance  $I_\nu$ , thus  $T_B = B_\nu^{-1}(I_\nu)$  (see Eq. 1.3.7). The quantity  $W_\nu(p)$  is given by

$$W_\nu(p) = \frac{\partial T_\nu(p)}{\partial \ln p}, \quad (4.1.3)$$

where  $T_\nu(p)$  is the transmittance from level  $p$  to the measuring instrument. In Eq. 4.1.2 clouds and other absorbers have been ignored, as well as the contribution from the surface. The latter assumption is justified if we assume that the transmittance from surface to the instrument is zero. Note that in Eq. 4.1.2 we integrate from TOA to the surface, where the latter is regarded as an infinitely distant isothermal emitter. The radiance equals the Planck function weighted by  $W_\nu(p)$ ; the latter is therefore termed weighting function. The radiance contains contributions from a range

of atmospheric pressure levels and  $W_v(p)$  gives the relative contribution from each level to the radiation measured by a space-borne instrument. Fig. 4.1 shows for the sake of illustration weighting functions according to Eq. 4.1.3 for every 100<sup>th</sup> channel within the IASI spectral range from 645 cm<sup>-1</sup> to 2760 cm<sup>-1</sup>.



**Figure 4.1:** Weighting functions  $[dT_v(p)/d \ln p]$  for every 100<sup>th</sup> channel in the IASI main spectral bands corresponding to the U.S. standard mid-latitude summer profile.

Investigating Eq. 4.1.2 and Eq. 4.1.3 we notice that if the mixing ratio (intrinsic in the transmittance) of one atmospheric constituent is constant and well-known (e.g., CO<sub>2</sub>), temperature as the only unknown can be determined from the measured radiance. If we make a set of measurements at certain wavenumbers whose weighting function cover the whole atmosphere sufficiently, then temperature profiles can be estimated. Eq. 4.1.2 may be rewritten for a set of closely spaced wavenumbers

$$I_{\bar{\nu}_i} = \int_0^{\infty} B_{\bar{\nu}_i}(T(p)) W_{\nu_i}(p) dp, \quad (4.1.4)$$

where  $i=1, \dots, m$  and  $\bar{\nu}$  is some representative wavenumber, since the Planck function does not vary much in small spectral intervals. Measurements near the center of an absorption band contain information of high levels, while measurements in the wings of the bands (where absorption is weak) yield information from deeper levels. Alternatively, if the temperature profile is known, total mass of the absorbing gas can be determined.

The atmospheric parameters are coupled to the measured radiation through radiative transfer. The information that can be extracted is limited by the complex nature of the RT process and by factors concerning the measurement process including spectral range and resolution, observational geometry and instrumental noise.

The computational methodology to retrieve atmospheric parameters can be divided into two general processes:

- the forward problem
- the inverse problem

The so-called forward problem contains a model of the radiative transfer process to calculate radiances according to Eq. 4.1.4. In other words, we simulate the radiance spectrum for a given state of the atmosphere for various atmospheric conditions. The calculated radiance is then compared with the radiance recorded by the instrument. The parameters of the model are adjusted until a reasonable fit to the data is obtained. The calculation of the radiance can be done by line-by-line transmittance models or by fast transmittance models. The latter have the ability to keep pace with the multitude of observations and are essential for the new generation of atmospheric sounders, which measure in thousands of spectral channels (such as IASI and AIRS).

Only if the forward problem is modeled appropriately, we can solve the associated inverse problem, the latter being the estimation of the atmospheric parameters from the measured radiance. The general inversion problem is to determine a profile of the Planck function from known weighting functions and measured radiance. Then  $T(p)$  can be found by inverting the Planck function algebraically. Eq. 4.1.4 cannot be easily inverted, since it is underconstrained and the unknown is a continuous function. Several assumptions and simplifications have to be made to find estimates of the true parameters based on available noisy measurements within computational constraints.

## 4.2 The Forward Model

### 4.2.1 Theoretical Introduction

Given a physical system we need to develop a mathematical model, which represents the actual behavior of the system for investigating its structure and modes.

The relation between the measurement vector  $\mathbf{y}$  and the actual state  $\mathbf{x}$  is given by a forward model (function)  $\mathbf{f}$  such that

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad (4.2.1)$$

where  $\boldsymbol{\varepsilon}$  is the vector of measurement noise, which will be discussed in detail in Sect. 4.3. Note that  $\mathbf{f}$  is only a model of the underlying physics of the measurement. The real physics may be too complex and in some aspects too uncertain to be entirely known and described. Thus the forward model should be constructed in such way that the detailed physics is processed to an adequate accuracy (e.g., Maybeck, 1979). The quantity  $\mathbf{f}$  is a linear or nonlinear function of the state vector. In the case where  $\mathbf{y}$  is assumed to be linearly related to  $\mathbf{x}$  we may write

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon}. \quad (4.2.2)$$

$\mathbf{K}$  is called the weighting (function) matrix or observation matrix, since it constitutes the observability condition, which defines our ability to determine the actual state from the measurements (Gelb, 1974). If the dimension of  $\mathbf{y}$  is  $m$  and the dimension of  $\mathbf{x}$  is  $n$  then the matrix  $\mathbf{K}$  is of dimension  $m \times n$ . For a particular measurement we have

$$y_i = \sum_{j=1}^n (\mathbf{k}_i)_j x_j + \varepsilon_i, \quad (4.2.3)$$

where the vector  $\mathbf{k}_i$  (of dimension  $n$ ) constitutes the  $i^{\text{th}}$  row of  $\mathbf{K}$  corresponding to the  $i^{\text{th}}$  measurement and is termed *weighting function* (not to be confused with  $W_v(p)$ , though, see notes a few paragraphs below). Thus every single measurement in  $\mathbf{y}$  consists of the complete state vector  $\mathbf{x}$ , where each element  $x_i$  is weighted by the corresponding element  $K_{ij}$ . We can say that the weighting functions describe the mapping of the state vector into the measurement (see, e.g., Rodgers, 2000).

In many situations the considered observation system is governed by a nonlinear relation between the vector to be measured and the one to be estimated. In order to obtain practical estimation algorithms we expand  $\mathbf{f}$  in a Taylor series about an initial guess profile  $\mathbf{x}_0$ . Thus we have

$$\mathbf{y} = \mathbf{f}(\mathbf{x}_0) + \left. \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) + \dots + \boldsymbol{\varepsilon}. \quad (4.2.4)$$

The matrix of partial derivatives (or *Jacobian matrix*) of the forward model with respect to  $\mathbf{x}$ , evaluated at  $\mathbf{x} = \mathbf{x}_0$ , is denoted by  $\mathbf{K}_0$  and usually termed weighting function matrix (i.e., as in the linear case).

Note that this weighting function is different from that defined in Eq. 4.1.3. Here we consider the derivative of the forward model with respect to the state, whereas in Sect. 4.1 the derivative of the true transmittance with respect to the logarithm of the pressure was considered, which may be

referred to as the atmospheric weighting functions. As will be seen later, their shape can differ much although both represent weights of the unknown quantity.

If we include terms of higher order in Eq. 4.2.4 better approximations of the exact relationship between state and measurements may be achieved. If we ignore higher-order terms, which is appropriate for nearly linear or moderately non-linear problems, we rewrite Eq. 4.2.4 as

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{K}_0 (\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\varepsilon} . \quad (4.2.5)$$

Now we could construct optimal estimation techniques to deduce an estimate of the state with minimum error, assuming that we know the measurement dynamics and their uncertainty for the particular observation situation.

As described in the preceding chapters we measure the radiance, emitted by surface and atmosphere, which reaches the instrument onboard a meteorological satellite. Our measurement vector  $\mathbf{y}$  comprises all measurements taken at a certain moment. The dimension of  $\mathbf{y}$  (denoted by  $m$ ) is the number of channels, which the instrument can record. The corresponding atmospheric state, presented in our case by temperature, at the time of that measurement is described by  $\mathbf{x}$ . The dimension of  $\mathbf{x}$  (denoted by  $n$ ) is the number of pressure levels.

It should be mentioned that we restrict ourselves to a situation described by Eq. 4.2.5 at a certain time. Measurement system descriptions and analysis are often done in the time domain involving measurement dynamics and time-varying characteristics of the parameters to be estimated. For instance, a very common filtering technique for estimating the state of a system is the Kalman Filter. For details see, e.g., Brown and Hwang (1997) and Weisz (1998).

The connection between the measurement vector  $\mathbf{y}$  and the state vector  $\mathbf{x}$  in our context is the radiative transfer equation as described in Chapt. 1. Since IASI is not operating yet, we need a radiative transfer algorithm, which computes simulated radiances in a fast and accurate way. The basic task of such a forward model is the calculation of transmittances (or optical depths) under different assumptions to be substituted in the radiative transfer equation and to derive the radiance for a certain wavenumber and temperature. In the last years a number of fast transmittance models have been developed and investigated for various satellite instruments (e.g., Eyre, 1991; Susskind et al., 1983; Hannon et al. 1996; Sherlock, 2000a). In this work we employ the fast radiative transfer model RTIASI (Matricardi and Saunders, 1999) to simulate IASI measurements. We describe this model in more detail below.

## 4.2.2 The Fast Transmittance Model RTIASI

The model has been developed for studies and exploitation of IASI radiances before the actual launch of the satellite as well as for developing retrieval and data assimilation algorithms. The following description of the model in this section is largely based on Eyre (1991), Matricardi and Saunders (1999), and Matricardi (1999).

RTIASI provides fast transmittance (FT) coefficients, which have been computed for a set of atmospheric profiles representing the range of variations in temperature and absorber amount found in the real atmosphere. The model performs on 43 pressure levels (from 0.1 mbar, which corresponds approximately a height of 65 km, to surface). Except water vapor and ozone all gases (such as CO, CO<sub>2</sub>, N<sub>2</sub>O, N<sub>2</sub>, O<sub>2</sub>, CH<sub>4</sub>, CFC11 and CFC12) are assumed to stay constant in a spatial and temporal sense and are therefore called fixed gases. The FT coefficients are then used to compute transmittances (or optical depths) for any desired input profile. By solving the radiative transfer equation the radiances and brightness temperatures for the IASI wavenumber range are evaluated. In addition, RTIASI calculates the Jacobian matrix  $\mathbf{K}_o$  (Eq. 4.2.5) as well.

### 4.2.2.1 The Fast Transmittance Coefficients

To derive the FT coefficients for the fixed gases and for water vapor a set of 42 profiles has been selected from the TIGR (TOVS Initial Guess Retrieval) dataset. For ozone 33 profiles selected from 383 NESDIS profiles have been used. These profiles contain data for temperature and absorber amount of the variable gases (water vapor and ozone) at the 43 standard pressure levels (Tab. A4). Line-by-line transmittances, computed for these profiles and 6 scan angles with the GENLN2 line-by-line transmittance and radiance model, become the regression data for the FT coefficients.

The atmosphere is divided into layers and the optical depth is calculated within each layer for each wavenumber. The line parameters for each spectral line at a certain wavenumber  $\nu$  were obtained from the HITRAN (High Resolution Transmission) database. The HITRAN molecular database contains the line parameters such as half-width, line position, line strength and lower energy state of about one million spectral lines for 35 molecules. The line strength and width are then modified according the temperature and pressure in the considered layer. Doppler broadening and the important continuum absorption bands have been considered in the calculations. For the profile set described above, line-by-line transmissions were computed over the IASI spectral range with a resolution of 0.001 cm<sup>-1</sup>. This was done along the path from each pressure level to space for 6 scan angles, namely those for which the secants equals 1.0, 1.25, 1.5, 1.75, 2.0 and 2.25. Since IASI

measures a spectrum whose natural resolution is much higher (“infinite resolution”) than the instrument resolution, the radiances (and transmittances) have to be convolved by the ISRF (Instrument Spectral Response Function) of the IASI instrument. RTIASI computes the radiances using convolved transmittances instead of convolving monochromatic radiances (polychromatic approximation). This approximation introduces an error, which however, is lower than the radiometric noise and is therefore acceptable for most purposes.

With these convolved level-to-space line-by-line transmittances three sets (for fixed gases, water vapor and ozone) of FT coefficients, denoted by  $a$ , have been evaluated according to the following regression scheme

$$-\ln\left(\frac{T_{j,v}}{T_{j-1,v}}\right) = \sum_{k=1}^M a_{j,v,k} Q_{j,k} , \quad (4.2.6)$$

where  $T_{j,v}$  is the convolved transmittance at wavenumber  $v$  from level  $j$  to space (level  $j-1$  is the level above) and  $M$  is the number of profile dependent predictors  $Q$ . The ratio of transmittances for adjacent levels on the left side of Eq. 4.2.6 is sometimes referred to as the effective layer transmittance  $T_{eff,j,v}$ . Since

$$T_{j,v} = \exp[\tau_{j,v}] , \quad (4.2.7)$$

we can rewrite Eq. 4.2.6 for the convolved level-to-space optical depth  $\tau_{j,v}$ ,

$$\tau_{j,v} - \tau_{j-1,v} = \sum_{k=1}^M a_{j,v,k} Q_{j,k} . \quad (4.2.8)$$

The predictors  $Q$  are functions of viewing angle, pressure, temperature and absorber amount (for more details see Hannon et al., 1996) and have been calculated for each profile and all viewing angles. RTIASI uses 10 predictors for the fixed gases and 14 predictors for water vapor.

The regression made on the optical depths (Eq. 4.2.8) rather than on the transmittances (Eq. 4.2.6) was found to yield better accuracy. Thus the FT coefficients are generated for each profile according to Eq. 4.2.8 for the fixed gases, water vapor, and ozone separately.

#### 4.2.2.2 Transmittances

The given input profile to RTIASI should contain data of temperature, water vapor and ozone volume mixing ratio and surface variables like emissivity, pressure, temperature and skin temperature.

Optical depths are evaluated according to Eq. 4.2.8 with predictor values calculated from the input profile variables, and then converted to transmittances. The total (convolved) transmittance for all the gases is given by

$$T_{j,v} = T_{j,v}^F \cdot T_{j,v}^W \cdot T_{j,v}^O, \quad (4.2.9)$$

where the upper indices  $F$ ,  $W$  and  $O$  stand for fixed gases, water vapor and ozone, respectively. It should be mentioned that for real non-monochromatic channels the convolution of all the gases is different from the product given in Eq. 4.2.9. For details see Matricardi and Saunders (1999).

A set of total transmittances for the three IASI main bands is shown in Fig. 4.2 based on the U.S. standard mid-latitude summer input profiles of temperature, water vapor, and ozone..

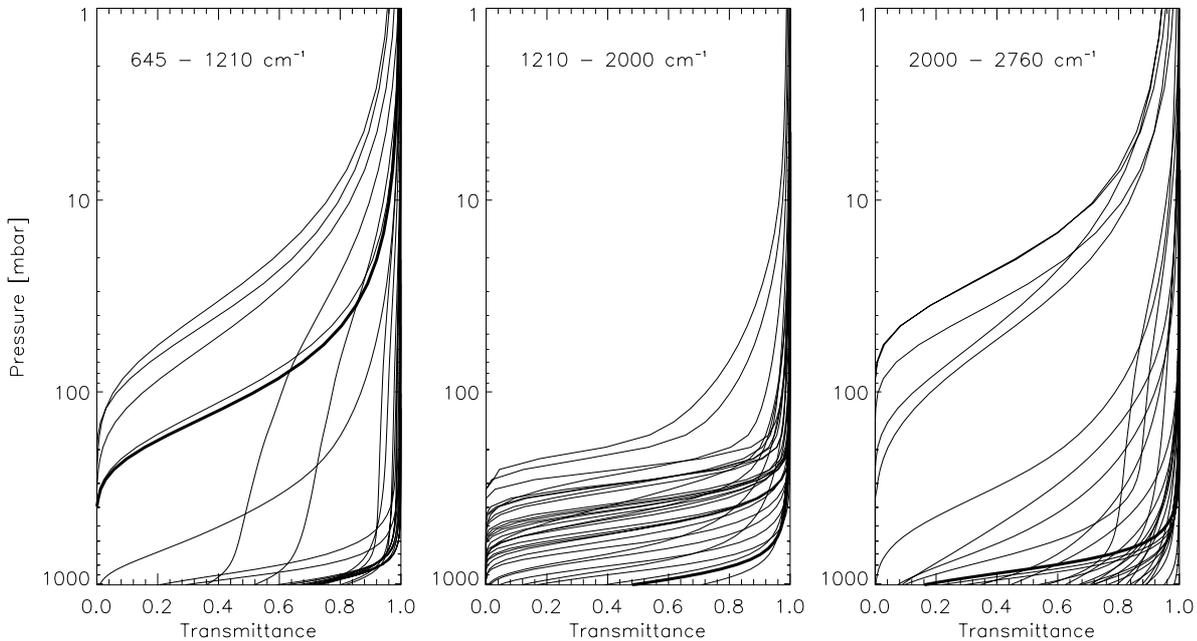


Figure 4.2: Total transmittances of every 100<sup>th</sup> IASI channel for the U.S. standard mid-latitude summer profiles for the three IASI main spectral bands calculated by RTIASI.

#### 4.2.2.3 Radiative Transfer Equation

The radiative transfer scheme of RTIASI used to calculate simulated IASI level 1c radiances assumes a plane-parallel atmosphere in local thermodynamic equilibrium without any scattering. The upwelling radiance in channel  $i$  (at wavenumber  $\nu_i$ ) at the top of the atmosphere (TOA) for a certain zenith angle is given by

$$R_i = (1 - N_i)R_i^c + N_iR_i^o. \quad (4.2.10)$$

The clear-column and the overcast radiance are denoted with  $R_i^c$  and  $R_i^o$ , respectively. and  $N_i$  is the effective fractional cloud cover (assumed to be in a single layer). The clear-column radiance is

$$R_i^c = R_{s,i} + R_{a,i} , \quad (4.2.11)$$

where  $R_{s,i}$  is the contribution from the surface and  $R_{a,i}$  is the contribution from the atmosphere.

The former term is given by

$$R_{s,i} = \varepsilon_{s,i} B_i(T_s) T_{s,i} , \quad (4.2.12)$$

where  $\varepsilon_{s,i}$  is the surface emissivity,  $B_i(T_s)$  is the Planck function for the surface skin temperature  $T_s$  and  $T_{s,i}$  is the transmittance from surface to space. The atmospheric contribution includes upward ( $\uparrow$ ) and downward ( $\downarrow$ ) (i.e., reflected from the surface) emission according to

$$R_{a,i} = R_i^\uparrow + R_i^\downarrow = \sum_{j=1}^N R_{i,j}^\uparrow + \sum_{j=1}^N R_{i,j}^\downarrow , \quad (4.2.13)$$

with

$$R_{i,j}^\uparrow = \frac{1}{2} (B_i(T_i) + B_i(T_{j-1})) \cdot (T_{i,j-1} - T_{i,j}) \quad (4.2.14)$$

and

$$R_{i,j}^\downarrow = R_{i,j}^\uparrow (1 - \varepsilon_{s,i}) (\tau_{s,i})^2 / (T_{i,j} T_{i,j-1}) . \quad (4.2.15)$$

$T_{i,j}$  is the transmittance from level  $j$  to space and  $T_j$  is the temperature at level  $j$ . Since the temperature varies with height and we consider different path (layer) lengths,  $T_i$  is defined as the Curtis-Godson weighted layer mean temperature (for details see, e.g., Liou, 1992). Recall that the levels are numbered from the top down, such that  $j = 0$  indicates the TOA level with  $p_j = 0$  and transmittance  $T_{i,j} = 1$ , whereas  $j = N$  constitutes the surface with  $p_j = p_s$  and  $T_{i,j} = T_{s,i}$ . In addition to the two terms at the right hand side, Eq. 4.2.13 also includes a small atmospheric contribution from the surface to the first layer above surface.

The overcast radiance in Eq. 4.2.10 at the level  $J$  for a cloud-top (subscript  $ct$ ) is defined as

$$R_{i,J}^o = \varepsilon_{i,ct} B_i(T_{ct}) \tau_{i,ct} + \sum_{j=1}^J R_{i,j}^\uparrow , \quad (4.2.16)$$

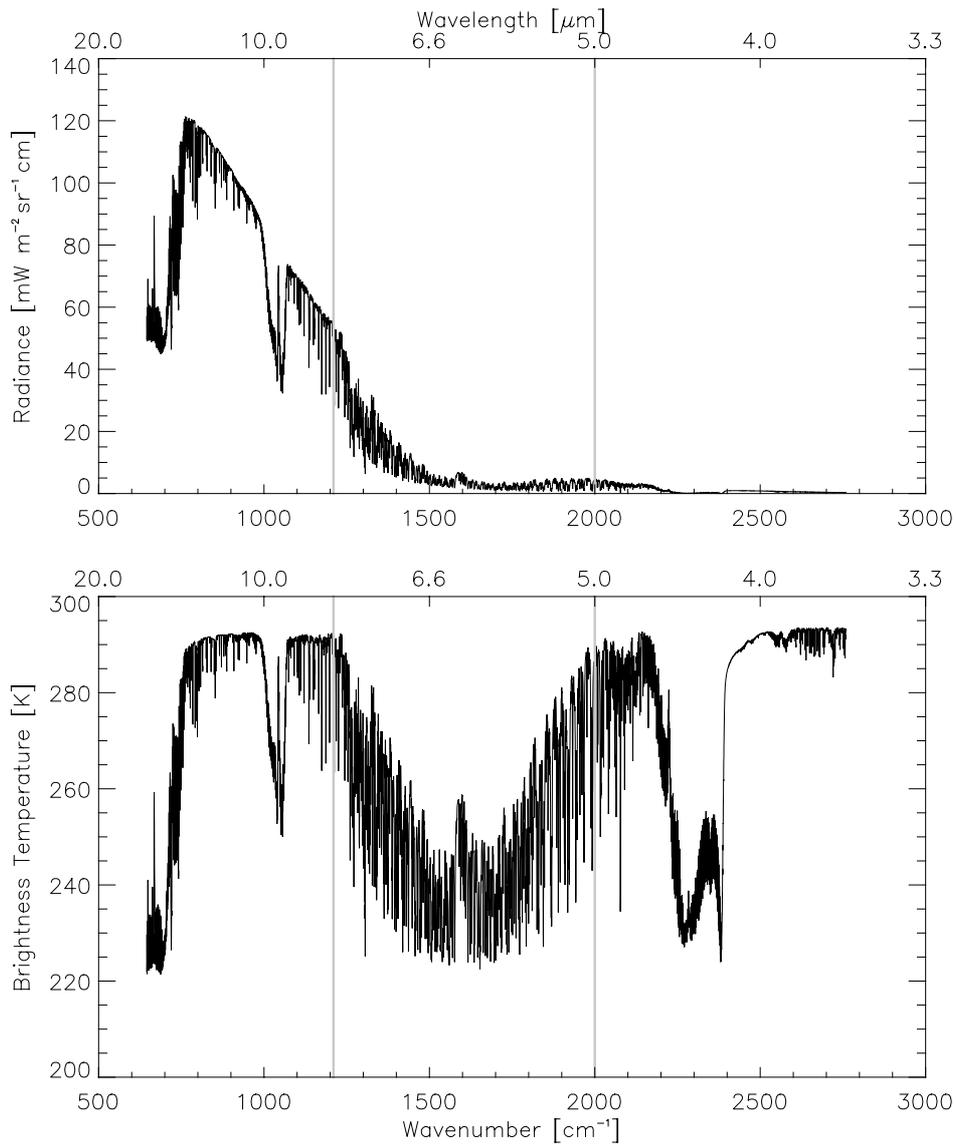
where  $\varepsilon_{i,ct} = 1$  is assumed. Thus the overcast radiance at the cloud-top pressure  $p_{ct}$  lying between level  $J$  and  $J-1$  can be written as

$$R_i^o = (1 - f_{ct})R_{i,J}^o + f_{ct}R_{i,J-1}^o \quad (4.2.17)$$

with the factor  $f_{ct}$  given by

$$f_{ct} = \frac{(p_J - p_{ct})}{(p_J - p_{J-1})} \quad (4.2.18)$$

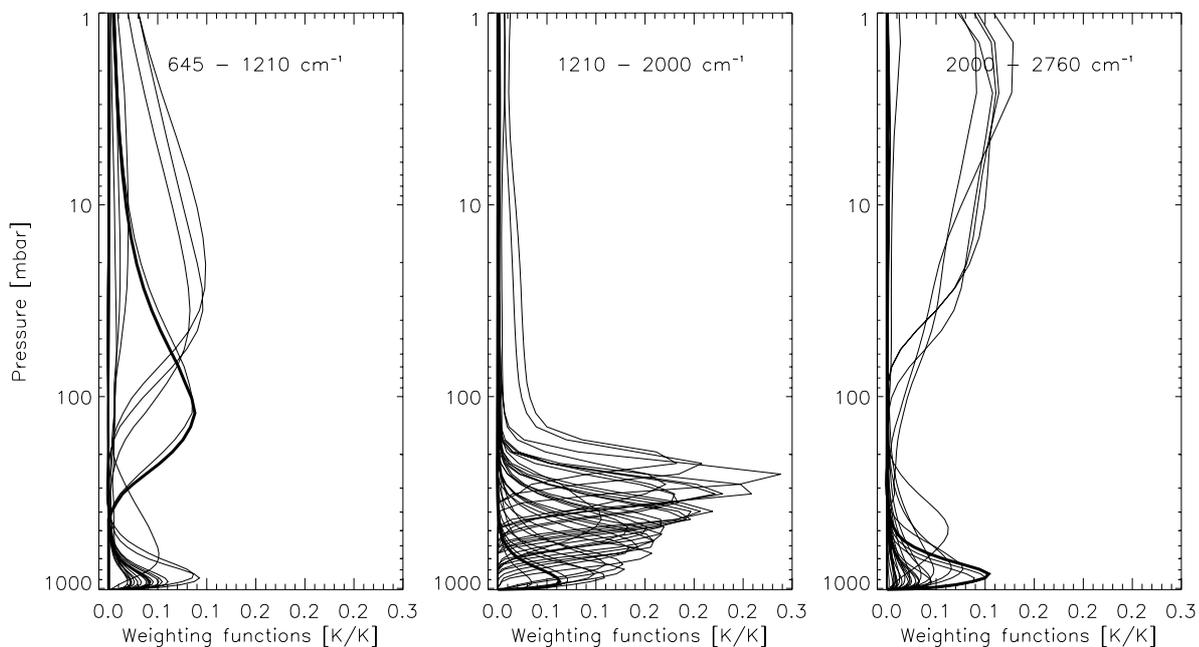
After this short overview on the RTIASI model algorithm we show radiances (calculated by RTIASI) in Fig. 4.3 (top) for the U.S. standard mid-latitude summer atmosphere. After converting these radiances to brightness temperatures we obtain the spectrum shown in Fig. 4.3 (bottom).



**Figure 4.3:** Radiances (top) and brightness temperatures (bottom) versus wavenumber and wavelength calculated by RTIASI for the U.S. standard mid-latitude summer atmosphere. Gray lines indicate limits of the IASI main spectral bands.

#### 4.2.2.4 Weighting Functions

The RTIASI model also comprises a program to calculate the Jacobian matrix  $\mathbf{K}$  of partial derivatives of the brightness temperatures with respect to the input profile variables. The weighting functions (i.e., the rows of  $\mathbf{K}$ ) reflect the sensitivity of the forward model to incremental changes in temperature (or humidity). The gradient matrix code is based on the adjoint version of the model described above. Details on the adjoint technique can be found in, e.g., Thepaut and Moll (1990) or Talagrand (1989). The model outputs the exact partial derivatives, i.e., the increment in brightness temperature if one element of the input profile is perturbed by 1 K or 1 ppmv. The resulting weighting functions corresponding to the same channels as in Fig. 4.1. and Fig. 4.2 are depicted in Fig. 4.4.



**Figure 4.4:** Weighting functions (rows of  $\mathbf{K}$ ) of every 100<sup>th</sup> IASI channel for the U.S. standard mid-latitude summer atmosphere for the three IASI main spectral bands calculated by RTIASI.

Inspecting Fig. 4.4 we find that no channels peak at the surface, whereas there are quite some channels peaking at the surface in the atmospheric weighting functions, i.e., in  $W_v(p) = dT_v / (p) d \ln p$  (see Fig. 4.1). This property, directly rooting in the form of the RT equation, suggests that the forward model involving Jacobians  $\mathbf{K}$  may not be the most proper item to model window channels, which allow surface parameter retrievals. Nevertheless, if we do not focus on the surface and the boundary layer, the forward model will be adequate for application in atmospheric retrievals. As mentioned earlier, the weighting functions give the relative contribution from a given atmospheric layer to the spectral radiance at TOA. The maximum contribution occurs

at the layer of unit optical depth. The width of the weighting functions determines the vertical resolution of retrieval soundings. In the case illustrated in Fig. 4.4 the weighting functions peaking in the troposphere (middle panel) have narrow peaks compared to those associated with wavenumbers associated with from the stratosphere (e.g., third panel). The latter region can thus be expected to pose more difficulties for retrievals than the troposphere, due to reduced sensitivity of  $\mathbf{K}$ .

## 4.3 The Inverse Model

The description below is mostly based on Rodgers (2000), Maybeck (1979; 1982), and Gelb (1974).

### 4.3.1 Optimal Linear Inversion

Based upon a set of measurements  $\mathbf{y}$  we need to find an accurate inversion algorithm to extract the actual state  $\mathbf{x}$ . The estimation of the true state vector, denoted by  $\hat{\mathbf{x}}$  and referred as the *a posteriori* estimate, should be guided by a suitable criterion of optimality. For instance, we could choose the estimate with minimum variance of all possible estimates. Alternatively, the optimal estimate could be regarded as the one, which maximizes an appropriate likelihood function. Basic information and details on various estimation methods can be found in any book on statistical analysis, e.g., Mardia et al. (1974), and Anderson (1984).

A more general way to deal with the inverse problem – and the best way to understand the underlying problem – is the *Bayesian approach*, where the propagation of conditional probability densities is investigated. Once the solution, which is the probability density  $p(\mathbf{x}|\mathbf{y})$  of  $\mathbf{x}$  conditioned on the measurements  $\mathbf{y}$ , is found, we specify the optimality criterion as mentioned above. Usually the mean of the resulting probability density is defined to be the optimal estimate. Applying the same assumptions the various estimation methods yield identical results.

We consider the linear measurement system given by Eq. 4.2.2, namely

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon}, \quad (4.3.1)$$

where  $\mathbf{y}$  is the measurement vector (dimension  $m$ ),  $\mathbf{x}$  is the state vector (dimension  $n$ ),  $\mathbf{K}$  is the weighting function matrix (dimension  $m \times n$ ) and  $\boldsymbol{\varepsilon}$  is the measurement error vector (dimension  $m$ ). The measurement errors are assumed to be Gaussian (i.e., normally distributed) noise with zero mean and associated with a positive-definite covariance matrix  $\mathbf{S}_{\boldsymbol{\varepsilon}}$  given by

$$E[\mathbf{y} - \mathbf{Kx}] = E[\boldsymbol{\varepsilon}] = 0, \quad (4.3.2)$$

$$\mathbf{S}_\varepsilon = E[(\mathbf{y} - \mathbf{Kx})(\mathbf{y} - \mathbf{Kx})^T] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T], \quad (4.3.3)$$

where  $E$  is the expectation value operator. Random variables with zero mean are said to be unbiased. Gaussian statistics is a good approximation, since many processes as well as errors in real measurements are well described by a Gaussian distribution. The measurement vector  $\mathbf{y}$ , obtained from a given  $\mathbf{x}$  according Eq. 4.3.1 is also Gaussian, since it is the sum of a known vector  $\mathbf{Kx}$  and a Gaussian vector  $\boldsymbol{\varepsilon}$ , with mean and covariance given by

$$E[\mathbf{y}] = E[\mathbf{Kx} + \boldsymbol{\varepsilon}] = \mathbf{Kx} + E[\boldsymbol{\varepsilon}] = \mathbf{Kx}, \quad (4.3.4)$$

$$E[(\mathbf{y} - \mathbf{Kx})(\mathbf{y} - \mathbf{Kx})^T] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \mathbf{S}_\varepsilon. \quad (4.3.5)$$

Thus  $\mathbf{S}_\varepsilon$  is not only the covariance of the measurement error  $\boldsymbol{\varepsilon}$  but also of the measurements. The probability density function (*pdf*) of  $\mathbf{y}$  conditioned on  $\mathbf{x}$  (i.e., describing the forward model) is given by

$$p(\mathbf{y} | \mathbf{x}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{S}_\varepsilon|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{Kx})\right]. \quad (4.3.6)$$

Thus  $p(\mathbf{y} | \mathbf{x})$  is normally distributed with mean  $\mathbf{Kx}$  and covariance  $\mathbf{S}_\varepsilon$  and the quantity  $p(\mathbf{y} | \mathbf{x}) d\mathbf{y}$  is the probability that, for a given  $\mathbf{x}$ ,  $\mathbf{y}$  lies in the volume element  $(\mathbf{y}, \mathbf{y}+d\mathbf{y})$ .

Now we like to combine the measurement data taken from the actual system with our information of the underlying physics (incorporated in  $\mathbf{K}$ ) and the statistical description of the uncertainty of the system, Eqs. 4.3.2 and 4.3.3, to find the best (optimal) estimate of the true state.

If our optimality criterion demands an estimate, which maximizes Eq. 4.3.6, we need to minimize the exponent in the square brackets. In least-square estimation this is equivalent to minimizing a cost function

$$J = (\mathbf{y} - \mathbf{Kx})^T \mathbf{S}_\varepsilon (\mathbf{y} - \mathbf{Kx}) \quad (4.3.7)$$

by solving  $\partial J / \partial \mathbf{x} = 0$ . We find the minimum as

$$\hat{\mathbf{x}} = (\mathbf{K}^T \mathbf{S}_\varepsilon \mathbf{K})^{-1} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{y}. \quad (4.3.8)$$

Returning to the Bayesian point of view we can exploit Bayes' rule to find a posterior *pdf*  $p(\mathbf{x} | \mathbf{y})$  of the event  $\mathbf{x}$  provided that  $\mathbf{y}$  occurred according to

$$\boxed{p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}}. \quad (4.3.9)$$

Thus we need to know statistical models for both  $\mathbf{y}$  and  $\mathbf{x}$ , since  $p(\mathbf{x})$  and  $p(\mathbf{y})$  are the prior probability density functions of the state and the measurements, respectively, implying our knowledge on  $\mathbf{x}$  and  $\mathbf{y}$  before the actual measurement is made. It should be mentioned that  $p(\mathbf{y})$  just serves as a normalizing factor in Bayes' rule and can be omitted in the minimization problem.

Let  $\mathbf{x}_{ap}$  denote an estimate of the normally distributed  $\mathbf{x}$  (remember  $\mathbf{y}$  and  $\boldsymbol{\varepsilon}$  are assumed to be Gaussian) known before the measurements are taken. The mean and the *a priori* estimation error covariance matrix ( $n \times n$ ) shall be defined as

$$E[\mathbf{x}] = \mathbf{x}_{ap}, \quad (4.3.10)$$

$$E[(\mathbf{x} - \mathbf{x}_{ap})(\mathbf{x} - \mathbf{x}_{ap})^T] = \mathbf{S}_{ap}. \quad (4.3.11)$$

Thus the prior knowledge of  $\mathbf{x}$  is described by the Gaussian *pdf*

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{S}_{ap}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1} (\mathbf{x} - \mathbf{x}_{ap})\right]. \quad (4.3.12)$$

We find that the error  $\mathbf{x} - \mathbf{x}_{ap}$  committed by using  $\mathbf{x}_{ap}$  as an estimate of  $\mathbf{x}$  before the measurements are processed is by definition unbiased, since

$$E[\mathbf{x} - \mathbf{x}_{ap}] = E[\mathbf{x}] - \mathbf{x}_{ap} = \mathbf{x}_{ap} - \mathbf{x}_{ap} = 0 \quad (4.3.13)$$

and the covariance of  $\mathbf{x} - \mathbf{x}_{ap}$  equals the state covariance  $\mathbf{S}_{ap}$  given in Eq. 4.3.11.

The quantity  $\mathbf{K}\mathbf{x}_{ap}$  can be regarded as a measurement prediction before the measurements are taken. We have

$$E[\mathbf{y}] = E[\mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon}] = \mathbf{K}E[\mathbf{x}] + E[\boldsymbol{\varepsilon}] = \mathbf{K}\mathbf{x}_{ap}, \quad (4.3.14)$$

and the associated prior measurement error covariance matrix can be evaluated according to

$$\begin{aligned} \mathbf{S}_{\varepsilon,ap} &= E[(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})^T] \\ &= E[(\mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon} - \mathbf{K}\mathbf{x}_{ap})(\mathbf{K}\mathbf{x} + \boldsymbol{\varepsilon} - \mathbf{K}\mathbf{x}_{ap})^T] \\ &= \mathbf{K}^T E[(\mathbf{x} - \mathbf{x}_{ap})(\mathbf{x} - \mathbf{x}_{ap})^T] \mathbf{K} + E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] \\ &= \mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_{\varepsilon}. \end{aligned} \quad (4.3.15)$$

Therefore the associated *pdf* of the predicted measurements can be written as

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{m}{2}} |\mathbf{S}_{\varepsilon, ap}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})^T \mathbf{S}_{\varepsilon, ap}^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})\right]. \quad (4.3.16)$$

Again the error  $\mathbf{y} - \mathbf{K}\mathbf{x}_{ap}$  committed by the estimate  $\mathbf{K}\mathbf{x}_{ap}$  of  $\mathbf{y}$  prior to the measurements has zero mean,

$$E[\mathbf{y} - \mathbf{K}\mathbf{x}_{ap}] = E[\mathbf{y}] - \mathbf{K}\mathbf{x}_{ap} = \mathbf{K}\mathbf{x}_{ap} - \mathbf{K}\mathbf{x}_{ap} = 0. \quad (4.3.17)$$

Hence, the error  $\mathbf{y} - \mathbf{K}\mathbf{x}_{ap}$  and the measurement vector  $\mathbf{y}$  share the same covariance  $\mathbf{S}_{\varepsilon, ap}$  described by Eq. 4.3.15.

Substituting Eqs. 4.3.6, 4.3.12 and 4.3.16 for  $p(\mathbf{y} | \mathbf{x})$ ,  $p(\mathbf{x})$  and  $p(\mathbf{y})$  into Bayes' rule, Eq. 4.3.9, we obtain for the posterior *pdf*, which corresponds to the situation, where the measurements have been taken,

$$p(\mathbf{x} | \mathbf{y}) = A \exp\left[-\frac{1}{2}\{(\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_{\varepsilon}^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}) + (\mathbf{x} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1} (\mathbf{x} - \mathbf{x}_{ap}) - (\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})^T \mathbf{S}_{\varepsilon, ap}^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})\}\right] \quad (4.3.18)$$

with

$$A = \frac{|\mathbf{S}_{\varepsilon, ap}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}} |\mathbf{S}_{\varepsilon}|^{\frac{1}{2}} |\mathbf{S}_{ap}|^{\frac{1}{2}}} = \frac{|\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_{\varepsilon}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}} |\mathbf{S}_{\varepsilon}|^{\frac{1}{2}} |\mathbf{S}_{ap}|^{\frac{1}{2}}}. \quad (4.3.19)$$

If  $p(\mathbf{x} | \mathbf{y})$  is a Gaussian distribution, it should be possible to rewrite Eq. 4.3.18 as

$$p(\mathbf{x} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\hat{\mathbf{S}}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \hat{\mathbf{S}}^{-1} (\mathbf{x} - \hat{\mathbf{x}})\right]. \quad (4.3.20)$$

Then  $p(\mathbf{x} | \mathbf{y})$  is completely specified by its mean  $E[\mathbf{x}] = \hat{\mathbf{x}}$ , representing our optimal estimate, and by the associated estimation (or retrieval) error covariance matrix  $\hat{\mathbf{S}}$  defined by

$$\hat{\mathbf{S}} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T]. \quad (4.3.21)$$

Equating term  $A$  (Eq. 4.3.19) with the fraction preceding the exponential function in Eq. 4.3.20 we find that

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{S}_{\varepsilon} \mathbf{S}_{ap} (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_{\varepsilon})^{-1} \\ &= (\mathbf{K}^T \mathbf{S}_{\varepsilon}^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1}. \end{aligned} \quad (4.3.22)$$

Inserting this result into Eq. 4.3.20 and equating with the exponential term of Eq. 4.3.18 gives the following expression for  $\hat{\mathbf{x}}$  identities after regrouping and use of matrix identities,

$$\begin{aligned}\hat{\mathbf{x}} &= (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1} (\mathbf{K} \mathbf{S}_\varepsilon^{-1} \mathbf{y} + \mathbf{S}_{ap}^{-1} \mathbf{x}_{ap}) \\ &= \mathbf{x}_{ap} + (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{K} \mathbf{x}_{ap}) \\ &= \mathbf{x}_{ap} + \hat{\mathbf{S}} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{K} \mathbf{x}_{ap}) .\end{aligned}\quad (4.3.23)$$

In order to obtain the second line of Eq. 4.3.23 the identity  $\mathbf{S}_{ap}^{-1} = \hat{\mathbf{S}}^{-1} - \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}$  has been inserted. Hence, the probability density function Eq. 4.3.18 is indeed a Gaussian density with mean  $\hat{\mathbf{x}}$  and covariance  $\hat{\mathbf{S}}$ , given by Eq. 4.3.23 and Eq. 4.3.22, respectively. Since the Gaussian density is symmetric the optimal estimate is not only the conditional mean (i.e., the expected value averaged over  $p(\mathbf{x} | \mathbf{y})$ ), but also the conditional mode, which maximizes the posterior *pdf* (Eq. 4.3.18) and is therefore the most likely state. The difference  $\mathbf{y} - \mathbf{K} \mathbf{x}_{ap}$  between the measurement and a measurement estimate, is sometimes termed residual (vector).

The quantity  $\hat{\boldsymbol{\varepsilon}} = (\mathbf{x} - \hat{\mathbf{x}})$  is the error committed by using  $\hat{\mathbf{x}}$  as an estimate of the true state  $\mathbf{x}$ . The error  $\hat{\boldsymbol{\varepsilon}}$  again is normally distributed with zero mean (i.e.,  $\hat{\boldsymbol{\varepsilon}}$  is unbiased) according to

$$E[\hat{\boldsymbol{\varepsilon}}] = E[\mathbf{x} - \hat{\mathbf{x}}] = E[\mathbf{x}] - \hat{\mathbf{x}} = \hat{\mathbf{x}} - \hat{\mathbf{x}} = 0 . \quad (4.3.24)$$

The covariance of  $\hat{\boldsymbol{\varepsilon}}$  equals the covariance of the state  $\mathbf{x}$  itself, i.e., the covariance is  $\hat{\mathbf{S}}$  given by Eq. 4.3.21 and Eq. 4.3.22. Geometrically this means that the shape of the *pdf* of  $\hat{\boldsymbol{\varepsilon}}$  equals the one of  $\mathbf{x}$ , shifted by the mean of the *pdf* of  $\mathbf{x}$  (Eq. 4.3.20). The formula for  $\hat{\mathbf{S}}$  (Eq. 4.3.22) can be verified by inserting the second line of Eq. 4.3.23 in  $\hat{\mathbf{S}} = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T]$  and evaluating the indicated expectation.

Note that if we do not consider any prior information, Eq. 4.3.23 reduces to the standard weighted least-squares estimate given by Eq. 4.3.8.

Minimization of a cost function  $J$  as mentioned above can also be done in a Bayesian manner using the following cost function with an arbitrary positive-semidefinite matrix  $\mathbf{S}$  according to

$$J = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{S} (\mathbf{x} - \hat{\mathbf{x}}) p(\mathbf{x} | \mathbf{y}) dx_1 \dots dx_n , \quad (4.3.25)$$

which constitutes the expectation of  $(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{S} (\mathbf{x} - \hat{\mathbf{x}})$ . Solving  $\partial J / \partial \mathbf{x} = 0$  gives

$$\hat{\mathbf{x}} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{x} | \mathbf{y}) dx_1 \dots dx_n , \quad (4.3.26)$$

i.e., the expectation value or the mean of  $\mathbf{x}$ . Evaluating Eq. 4.3.26 we obtain the same result as Eq. 4.3.23 (Gelb, 1974). Thus  $\hat{\mathbf{x}}$  is the conditional mean minimizing any cost function of the general quadratic form  $(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{S}(\mathbf{x} - \hat{\mathbf{x}})$ . Since  $\mathbf{S}$  is arbitrary the identity matrix  $\mathbf{I}$  can be chosen instead and we can write  $J = E[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T] = \text{trace}[\hat{\mathbf{S}}]$ . Thus we could minimize the trace of  $\hat{\mathbf{S}}$  to find that the minimum variance estimate is just the conditional mean as described above.

Another common method to find the optimal estimate is to maximize an appropriate likelihood function  $L$ , which is often defined as the natural logarithm of the Gaussian *pdf*  $p(\mathbf{x} | \mathbf{y})$  described by (4.3.18). Thus we may write

$$\begin{aligned} L &= \ln p(\mathbf{x} | \mathbf{y}) \\ &= \ln(A) - \frac{1}{2}[(\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x})] - \frac{1}{2}[(\mathbf{x} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1}(\mathbf{x} - \mathbf{x}_{ap})] \\ &\quad + \frac{1}{2}[(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})^T (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})] \end{aligned} \quad (4.3.27)$$

with  $A$  given by Eq. 4.3.19. Performing differentiation on  $L$  gives

$$\frac{\partial L}{\partial \mathbf{x}} = -\mathbf{K}^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}) + \mathbf{S}_{ap}^{-1}(\mathbf{x} - \mathbf{x}_{ap}) \quad (4.3.28)$$

Equating to zero and solving for  $\mathbf{x}$  yields again the same result as described by Eq. 4.3.23.

Hence, we found that the (unbiased) optimal estimate  $\hat{\mathbf{x}}$  described by Eq. 4.3.23 is the conditional mean of the *pdf* (Eq. 4.3.20), i.e., the minimum variance (or mean square error) estimate and the mode of the *pdf*, i.e., the maximum likelihood estimate.

In the literature two different forms for estimating  $\hat{\mathbf{x}}$  are known depending upon the size of the matrix being inverted. The  $n$ -form is the one given by Eq. 4.3.23. We recapitulate

$$\boxed{\hat{\mathbf{x}} = \mathbf{x}_{ap} + \hat{\mathbf{S}} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})} \quad (4.3.29)$$

with the estimation error covariance matrix

$$\boxed{\hat{\mathbf{S}} = (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1}} \quad (4.3.30)$$

This formula needs an inversion of the  $n \times n$  matrix  $\hat{\mathbf{S}}^{-1} = (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})$ . If we substitute for  $\hat{\mathbf{S}}$  the form given by the first line of Eq. 4.3.22, that is,  $\hat{\mathbf{S}} = \mathbf{S}_\varepsilon \mathbf{S}_{ap} (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1}$ , into Eq. 4.3.29, we find the  $m$ -form according to

$$\boxed{\hat{\mathbf{x}} = \mathbf{x}_{ap} + \mathbf{S}_{ap} \mathbf{K}^T (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})} \quad (4.3.31)$$

Here the  $m \times m$  matrix  $\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon$  has to be inverted. If not otherwise stated we will use the  $n$ -form defined by Eq. 4.3.29 and Eq. 4.3.30 throughout the remaining chapters, as in our context the dimension  $n$  of the state vector is always smaller than the dimension  $m$  of the measurement vector.

The matrix of partial derivatives of the inverse model (i.e., the retrieval) with respect to the measurements is termed *gain function matrix*  $\mathbf{G}$  (or contribution function matrix) of dimension  $n \times m$  and defined as

$$\begin{aligned} \mathbf{G} &= \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{y}} \\ &= (\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \\ &= \hat{\mathbf{S}} \mathbf{K}^T \mathbf{S}_\varepsilon^{-1}. \end{aligned} \quad (4.3.32)$$

The columns of  $\mathbf{G}$  (gain functions) reflect how each measurement contributes to the retrieved state. Note that the second and third line of Eq. 4.3.32 represent the  $n$ -form of the gain function matrix corresponding to the  $n$ -form of  $\hat{\mathbf{x}}$  given by Eq. 4.3.29. The  $m$ -form of  $\mathbf{G}$  is

$$\mathbf{G} = \mathbf{S}_{ap} \mathbf{K}^T (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1}, \quad (4.3.33)$$

which can be easily verified by differentiation of the  $m$ -form of  $\hat{\mathbf{x}}$  (Eq. 4.3.31) with respect to  $\mathbf{y}$  according to the first line of Eq. 4.3.32. It should be mentioned that  $\hat{\mathbf{S}}$  can be written as

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{S}_\varepsilon \mathbf{S}_{ap} (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1} \\ &= \mathbf{S}_{ap} - \mathbf{S}_{ap} \mathbf{K}^T (\mathbf{K}^T \mathbf{S}_{ap} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1} \mathbf{K} \mathbf{S}_{ap} \\ &= \mathbf{S}_{ap} - \mathbf{G} \mathbf{K} \mathbf{S}_{ap} \end{aligned} \quad (4.3.34)$$

with the  $m$ -form of  $\mathbf{G}$  (Eq. 4.3.33). Note that, when using the  $m$ -formulation,  $\mathbf{G}$  needs to be computed first to calculate  $\hat{\mathbf{S}}$  according to Eq. 4.3.34.

We can rewrite Eq. 4.3.29 and Eq. 4.3.31 into the form

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{x}_{ap} + \mathbf{G}(\mathbf{y} - \mathbf{K} \mathbf{x}_{ap}) \\ &= (\mathbf{I} - \mathbf{G} \mathbf{K}) \mathbf{x}_{ap} + \mathbf{G} \mathbf{y}. \end{aligned} \quad (4.3.35)$$

Thus  $\mathbf{G}$  is always the term premultiplied with the residual  $\mathbf{y} - \mathbf{K} \mathbf{x}_{ap}$  in the estimation equation whatever form we use for  $\hat{\mathbf{x}}$ . The first line of Eq. 4.3.35 indicates that the estimate is the sum of a weighted difference between the actual measurements  $\mathbf{y}$  and a measurement prediction  $\mathbf{K} \mathbf{x}_{ap}$  and the *a priori* state estimate  $\mathbf{x}_{ap}$ .

Investigating expression Eq. 4.3.32 or equivalently Eq. 4.3.33, we see that for very precise measurements (i.e., small  $\mathbf{S}_\varepsilon$ ) the gain will be large and according to Eq. 4.3.35 we may conclude that the measurement data contribute more to the estimate  $\hat{\mathbf{x}}$  than the prior estimate  $\mathbf{x}_{ap}$ . On the other hand, if the uncertainty of  $\mathbf{x}_{ap}$  is small (i.e., small  $\mathbf{S}_{ap}$ ), reflecting an accurate initial estimate, while the measurement is uncertain,  $\mathbf{G}$  will be small. Therefore, in the latter case, the retrieval will mostly depend on its prior information and will be less influenced by the measurements.

Another useful quantity is the *averaging kernel matrix*  $\mathbf{A}$  of dimension  $n \times n$ , defined as the matrix of partial derivatives of the retrieval with respect to the true state vector. With  $\hat{\mathbf{x}}$  described by Eq. 4.3.35 and  $\mathbf{y} = \mathbf{K}\mathbf{x} + \varepsilon$  we find

$$\mathbf{A} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} = \mathbf{G}\mathbf{K} \quad (4.3.36)$$

and instead of Eq. 4.3.35 we may formulate

$$\begin{aligned} \hat{\mathbf{x}} &= (\mathbf{I} - \mathbf{A})\mathbf{x}_{ap} + \mathbf{A}\mathbf{x} + \mathbf{G}\varepsilon \\ &= \mathbf{x}_{ap} + \mathbf{A}(\mathbf{x} - \mathbf{x}_{ap}) + \mathbf{G}\varepsilon . \end{aligned} \quad (4.3.37)$$

The matrix  $\mathbf{A}$  can be evaluated according to Eq. 4.3.36 in  $n$ - or  $m$ -form depending on the form used for  $\mathbf{G}$ . The rows of  $\mathbf{A}$  are called the averaging kernels reflecting how a true state element is reproduced in the retrieved state. In the ideal inverse case a certain element of  $\hat{\mathbf{x}}$  originates entirely from the corresponding element of the true state, i.e.,  $\mathbf{A}$  would be the identity matrix. In reality the averaging kernels have peaks at the appropriate state vector element and certain widths. The full width at half maximum (FWHM) can be regarded as a measure of the spatial resolution of the observing system.

So far we found an algorithm, which provides a state estimate  $\hat{\mathbf{x}}$  by combining new measurement data  $\mathbf{y}$  with an initial state estimate  $\mathbf{x}_{ap}$ . We may summarize the main results of the ( $n$ -form) linear optimal estimation problem as follows

Optimal state estimate	$\hat{\mathbf{x}} = \mathbf{x}_{ap} + \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{K}\mathbf{x}_{ap})$	(4.3.38)
Estimation error covariance matrix	$\hat{\mathbf{S}} = (\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1}$	(4.3.39)
Gain function matrix	$\mathbf{G} = \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}$	(4.3.40)
Averaging kernel matrix	$\mathbf{A} = \mathbf{G}\mathbf{K}$	(4.3.41)

### 4.3.2 Optimal Non-linear Inversion

In many situations the measurement  $\mathbf{y}$  is a non-linear function of the state  $\mathbf{x}$  plus an unbiased Gaussian noise  $\boldsymbol{\varepsilon}$  according to

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad (4.3.42)$$

where  $\mathbf{f}$  is the forward model function introduced in Sect. 4.1. This problem is considerably more difficult to solve than the linear version. The problem is generally not Gaussian and distinct optimality criteria lead to distinct estimates (Gelb, 1974). Conceptually, we need to evaluate  $p(\mathbf{x} | \mathbf{y})$  given by Bayes' rule, Eq. 4.3.9, and Eq. 4.3.20 as described in the previous section, but now we replace  $\mathbf{K}\mathbf{x}$  by  $\mathbf{f}(\mathbf{x})$ . The computation involves functional integral difference equations. Therefore one seeks to approximate the function  $\mathbf{f}$  in a way which is suitable for the considered situation. Usually Taylor series of  $\mathbf{f}$  expanded about an appropriate state estimate and truncated to some low order give way to practically feasible algorithms.

We present the most common solutions of the non-linear measurement system, Eq. 4.3.42, of various accuracy, practicability and computational expense.

#### 4.3.2.1 Linearization

Here we seek to incorporate non-linearities while still using a linear estimation algorithm. Thus we require that the estimate is constrained to be a linear function of the measurements. We introduce an arbitrary reference state  $\mathbf{x}_l$ , which may be determined by some means and called linearization state vector in this context. The measurement vector  $\mathbf{y}$  can now be approximated by a Taylor series truncated to first order according to

$$\begin{aligned} \mathbf{y} &= \mathbf{f}(\mathbf{x}_l) + \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_l} (\mathbf{x} - \mathbf{x}_l) + \boldsymbol{\varepsilon} \\ &= \mathbf{y}_l + \mathbf{K}_l (\mathbf{x} - \mathbf{x}_l) + \boldsymbol{\varepsilon}. \end{aligned} \quad (4.3.43)$$

The  $m \times n$  matrix  $\mathbf{K}_l$  comprises the partial derivatives of  $\mathbf{f}$  with respect to  $\mathbf{x}$  evaluated at the reference state  $\mathbf{x}_l$ . If we follow the steps as in the linear case (Sect. 4.3.2) as described by Eqs. 4.3.1 to 4.3.23, we can replace the initial measurement estimate  $E[\mathbf{y}] = \mathbf{K}\mathbf{x}_{ap}$ , appearing in the residual of Eq. 4.3.23, by

$$\begin{aligned} E[\mathbf{y}] &= \mathbf{y}_l + \mathbf{K}_l (E[\mathbf{x}] - \mathbf{x}_l) + E[\boldsymbol{\varepsilon}] \\ &= \mathbf{y}_l + \mathbf{K}_l (\mathbf{x}_{ap} - \mathbf{x}_l) \end{aligned} \quad (4.3.44)$$

recalling that  $E[\mathbf{x}] = \mathbf{x}_{ap}$  before the measurement are taken and  $E[\boldsymbol{\varepsilon}] = 0$ . Thus we may write

$$\hat{\mathbf{x}} = \mathbf{x}_{ap} + \hat{\mathbf{S}}\mathbf{K}_l^T\mathbf{S}_\varepsilon^{-1}[(\mathbf{y} - \mathbf{y}_l) - \mathbf{K}_l(\mathbf{x}_{ap} - \mathbf{x}_l)] \quad (4.3.45)$$

with the retrieval error covariance

$$\hat{\mathbf{S}} = (\mathbf{K}_l^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}_l + \mathbf{S}_{ap}^{-1})^{-1}. \quad (4.3.46)$$

The estimate Eq. 4.3.45 is an approximation of the conditional mean of the associated *pdf* with unbiased retrieval error  $\hat{\boldsymbol{\varepsilon}} = (\mathbf{x} - \hat{\mathbf{x}})$ . It should be mentioned that low accuracy or big errors can occur, if the true and the linearization vector  $\mathbf{x}_l$  differ significantly. Therefore an iteration procedure as will be described in the succeeding section is desirable.

In the literature one can find estimation algorithms, where the linearization is done about the prior state estimate  $\mathbf{x}_{ap}$ . In this context  $\mathbf{x}_{ap}$  is an estimate at a previous time-step processing previous measurements, i.e., it is assumed to be already a good estimate of the true state. Following Eqs. 4.3.43 to 4.3.46 replacing  $\mathbf{x}_l$  and  $\mathbf{y}_l$  by  $\mathbf{x}_{ap}$  and  $\mathbf{y}_{ap}$  yields

$$\hat{\mathbf{x}} = \mathbf{x}_{ap} + \hat{\mathbf{S}}\mathbf{K}_{ap}^T\mathbf{S}_\varepsilon^{-1}[\mathbf{y} - \mathbf{y}_{ap}] \quad (4.3.47)$$

with  $\mathbf{y}_{ap} = \mathbf{f}(\mathbf{x}_{ap})$  and  $\mathbf{K}_{ap} = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{ap}}$ . In Kalman Filtering, where measurements are processed

sequentially, this result is called extended Kalman Filter algorithm.

### 4.3.2.2 Iterated Linearization

In certain applications the solution given by Eq. 4.3.45 with Eq. 4.3.46 can be improved by subsequent relinearization about the recent estimate. Instead of Eq. 4.3.45 we find, with the iteration index  $i$ ,

$$\mathbf{x}_{i+1} = \mathbf{x}_{ap} + \mathbf{S}_i\mathbf{K}_i^T\mathbf{S}_\varepsilon^{-1}[\mathbf{y} - \mathbf{y}_i - \mathbf{K}_i(\mathbf{x}_{ap} - \mathbf{x}_i)] \quad (4.3.48)$$

and

$$\mathbf{S}_i = (\mathbf{K}_i^T\mathbf{S}_\varepsilon^{-1}\mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1}, \quad (4.3.49)$$

where  $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$  and  $\mathbf{K}_i = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_i}$ .

The starting point  $\mathbf{x}_i = \mathbf{x}_0$  is a first guess profile and usually  $\mathbf{x}_0 = \mathbf{x}_{ap}$  is implemented. Therefore the first estimate  $\mathbf{x}_1$  may be generated using Eq. 4.3.47 and will serve as a better estimate than any linearization profile to update  $\mathbf{K}$  (and hence  $\mathbf{S}$ ) and to evaluate a new estimate  $\mathbf{x}_2$ . We note that

with  $\mathbf{x}_i \rightarrow \hat{\mathbf{x}}$  and  $\mathbf{S}_i \rightarrow \hat{\mathbf{S}}$  as  $i \rightarrow \infty$ , Eq. 4.3.48 is practically repeated until a convergence criterion is reached or no further improvements of  $\mathbf{x}_{i+1}$  over  $\mathbf{x}_i$  is realized by further iterations.

It should be considered that any additional iteration will increase the computational time significantly, since the weighting function matrix  $\mathbf{K}_i$  has to be computed at each iteration step.

The same iteration equation can be found if we repeat the steps as in the linear case (Eq. 4.3.1 to Eq. 4.3.20) and obtain the *pdf* of  $\mathbf{x}$  conditioned on  $\mathbf{y}$  according to

$$p(\mathbf{x} | \mathbf{y}) = A \exp\left[-\frac{1}{2}\{(\mathbf{y} - \mathbf{f}(\mathbf{x}))^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1}(\mathbf{x} - \mathbf{x}_{ap})\}\right], \quad (4.3.50)$$

where

$$A = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{S}_\varepsilon|^{\frac{1}{2}} |\mathbf{S}_{ap}|^{\frac{1}{2}}}, \quad (4.3.51)$$

omitting the normalizing factor  $p(\mathbf{y})$  in Bayes' rule given by Eq. 4.3.9. As mentioned in Sect. 4.3.1 one method to find the estimate  $\hat{\mathbf{x}}$  is to maximize the likelihood function  $L$ , which is usually chosen to be the natural logarithm of  $p(\mathbf{x} | \mathbf{y})$  described by Eq. 4.3.50, thus

$$\begin{aligned} L &= \ln p(\mathbf{x} | \mathbf{y}) \\ &= -\frac{1}{2}[(\mathbf{y} - \mathbf{f}(\mathbf{x}))^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}))] - \frac{1}{2}[(\mathbf{x} - \mathbf{x}_{ap})^T \mathbf{S}_{ap}^{-1}(\mathbf{x} - \mathbf{x}_{ap})] + \ln(A). \end{aligned} \quad (4.3.52)$$

The derivative of that function is

$$\mathbf{g}(\mathbf{x}) = \frac{\partial L(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{K}^T \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})) + \mathbf{S}_{ap}^{-1}(\mathbf{x} - \mathbf{x}_{ap}), \quad (4.3.53)$$

where  $\mathbf{K} = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}$ . Unfortunately the equation  $\mathbf{g}(\mathbf{x}) = 0$  can be difficult to solve if the problem considered is highly non-linear. But we can apply the Newtonian iteration method to find an iterative solution of this problem (Rodgers, 2000). The solution of the general optimization problem  $\mathbf{g}(\mathbf{x}) = 0$  is given by

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left[ \frac{\partial \mathbf{g}(\mathbf{x}_i)}{\partial \mathbf{x}} \right]^{-1} \mathbf{g}(\mathbf{x}_i). \quad (4.3.54)$$

The derivative of  $\mathbf{g}$  can be evaluated as

$$\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} = -\frac{\partial \mathbf{K}^T(\mathbf{x})}{\partial \mathbf{x}} \mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x})) + \mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K} + \mathbf{S}_{ap}^{-1} \quad (4.3.55)$$

being the Hessian matrix of  $L$ , i.e., the matrix of second derivatives of  $L$  with respect to  $\mathbf{x}$  evaluated at  $\mathbf{x}_i$ . We see that Eq. 4.3.55 involves the first derivative of the forward model  $\mathbf{f}$  (i.e.,  $\mathbf{K}$ ) and the derivative of  $\mathbf{K}$  (i.e., the Hessian of  $\mathbf{f}$ ). The latter may be difficult to evaluate. But in many situations where the product of the noise (modeled by  $\mathbf{S}_\varepsilon$ ) and the non-linearity, described by  $\mathbf{y} - \mathbf{f}(\mathbf{x})$ , is small the first term on the right hand side of Eq. 4.3.55, involving the Hessian of  $\mathbf{f}$ , can be ignored. Employing that assumption and solving Eq. 4.3.54 we find for the iterative solution of the (moderately) non-linear estimation problem

$$\mathbf{x}_{i+1} = \mathbf{x}_i + (\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1} [\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{f}(\mathbf{x}_i)) - \mathbf{S}_{ap}^{-1} (\mathbf{x}_i - \mathbf{x}_{ap})], \quad (4.3.56)$$

and after some rearrangements we obtain

$$\mathbf{x}_{i+1} = \mathbf{x}_{ap} + (\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1} \mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y} - \mathbf{f}(\mathbf{x}_i) - \mathbf{K}_i (\mathbf{x}_{ap} - \mathbf{x}_i)]. \quad (4.3.57)$$

With  $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$  and  $\mathbf{S}_i = (\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1}$  this is found to be identical to Eq. 4.3.48. Situations which are regarded grossly non-linear, i.e., where the entire right hand side of Eq. 4.3.55 needs to be considered, will be discussed in the next section.

### 4.3.2.3 Second Order Solution

If the situation considered can be regarded only slightly non-linear, we can apply the linearization and iterated linearization methods described by Sect. 4.3.2.2. The resulting estimates are approximations of the optimal minimum variance (conditional mean) or maximum probability (conditional mode) estimates. If non-linearities are sufficiently important, improved performance can be achieved by including second order terms (or even higher order terms) in the expansion of  $\mathbf{f}(\mathbf{x})$  about the linearization point  $\mathbf{x}_l$ . Thus  $\mathbf{f}(\mathbf{x})$  may be approximated by

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_l) + \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_l} (\mathbf{x} - \mathbf{x}_l) + \frac{1}{2} \left. \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^2} \right|_{\mathbf{x}=\mathbf{x}_l} (\mathbf{x} - \mathbf{x}_l)(\mathbf{x} - \mathbf{x}_l)^T. \quad (4.3.58)$$

For simplification we first consider the case where we linearize about a prior state vector instead of a linearization vector. Motivated by the (linear) estimation Eq. 4.3.48 we require for the estimate

$$\hat{\mathbf{x}} = \mathbf{x}_{ap} + \mathbf{G}(\mathbf{y} - E[\mathbf{y}]) \quad (4.3.59)$$

with the matrix  $\mathbf{G}$  to be determined. Furthermore, we require that the retrieval error  $\mathbf{x} - \hat{\mathbf{x}}$  is unbiased and the desired estimate is an approximation to the conditional mean. Then  $\mathbf{y}$  can be approximated by the expectation of Eq. 4.3.58, with  $\mathbf{x}_l$  replaced by  $\mathbf{x}_{ap}$ , that is,

$$\begin{aligned}
E[\mathbf{y}] &= \mathbf{f}(\mathbf{x}_{ap}) + \mathbf{K}_{ap} (E[\mathbf{x}] - \mathbf{x}_{ap}) + \frac{1}{2} \left. \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^2} \right|_{\mathbf{x}=\mathbf{x}_{ap}} E[(\mathbf{x} - \mathbf{x}_{ap})(\mathbf{x} - \mathbf{x}_{ap})^T] + E[\boldsymbol{\varepsilon}] \\
&= \mathbf{f}(\mathbf{x}_{ap}) + \mathbf{K}_{ap} (\mathbf{x}_{ap} - \mathbf{x}_{ap}) + \frac{1}{2} \partial^2 (\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap}) \\
&= \mathbf{f}(\mathbf{x}_{ap}) + \frac{1}{2} \partial^2 (\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap}),
\end{aligned} \tag{4.3.60}$$

where  $\mathbf{K}_{ap} = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{ap}}$ . As indicated, the second-order term in Eq. 4.3.60 can be summarized by the operator  $\partial^2 (\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})$ . The resulting quantity is a vector of length  $m$  whose  $j$ -th element is defined by

$$\partial_j^2 (\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap}) = \text{trace} \left[ \left. \frac{\partial^2 f_j(\mathbf{x}_{ap})}{\partial \mathbf{x}^2} \right|_{\mathbf{x}=\mathbf{x}_{ap}} \mathbf{S}_{ap} \right]. \tag{4.3.61}$$

Since the residual of Eq. 4.3.59 is determined by Eq. 4.3.60, we need to find the matrix  $\mathbf{G}$ . One method is to minimize  $J = E[(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}})] = \text{trace}[\hat{\mathbf{S}}]$ , where we estimate  $\mathbf{x} - \hat{\mathbf{x}}$  by its expectation, substitute Eq. 4.3.59 for  $\hat{\mathbf{x}}$  and solve for  $\mathbf{G}$  as shown in Gelb (1974). Another approach is to introduce  $p(\mathbf{y} | \mathbf{x})p(\mathbf{x}) = p(\mathbf{y}, \mathbf{x})$  in Bayes' rule (Eq. 4.3.9), since the *pdf*'s on the right side are all conditioned on the situation before the actual measurements are taken. Thus we have for the joint *pdf* of  $\mathbf{x}$  and  $\mathbf{y}$

$$p(\mathbf{y}, \mathbf{x}) = p(\mathbf{x} | \mathbf{y})p(\mathbf{y}). \tag{4.3.62}$$

Using this result and recalling that  $E[\mathbf{x}] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbf{x} p(\mathbf{y}, \mathbf{x}) dx_1 \dots dx_n$ , it can be shown (e.g., Maybeck, 1982) that for two functions  $g(\mathbf{x})$  and  $h(\mathbf{y})$  we have

$$E[h(\mathbf{x})g(\mathbf{y})] = E[E[h(\mathbf{x}) | \mathbf{y}]g(\mathbf{y})], \tag{4.3.63}$$

where  $E[h(\mathbf{x}) | \mathbf{y}]$  is the conditional expectation of  $h(\mathbf{x})$  given a certain value of  $\mathbf{y}$ . To apply this theorem to our problem we choose  $h(\mathbf{x}) = \mathbf{x}$  and  $g(\mathbf{y}) = \mathbf{y} - E[\mathbf{y}]$  to find

$$E[\mathbf{x}(\mathbf{y} - E[\mathbf{y}])] = E[E[\mathbf{x} | \mathbf{y}](\mathbf{y} - E[\mathbf{y}])], \tag{4.3.64}$$

recalling that “conditioned on  $\mathbf{y}$ ” refers to the situation where the measurement has been taken. Note that the remaining expectations are prior quantities. Now we substitute the basic approach given by Eq. 4.3.59 for  $E[\mathbf{x} | \mathbf{y}] = \hat{\mathbf{x}}$ . Thus

$$\begin{aligned}
E[\mathbf{x}(\mathbf{y} - E[\mathbf{y}])] &= E[(\mathbf{x}_{ap} + \mathbf{G}(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}]]) \\
&= E[\mathbf{x}_{ap}(\mathbf{y} - E[\mathbf{y}])] + \mathbf{G}E[(\mathbf{y} - E[\mathbf{y}])(\mathbf{y} - E[\mathbf{y}])^T] \\
&= E[\mathbf{x}_{ap}(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])] + \mathbf{G}E[(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])^T] + \mathbf{G}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] \\
&= \mathbf{x}_{ap}E[\mathbf{f}(\mathbf{x})] - E[\mathbf{x}_{ap}E[\mathbf{f}(\mathbf{x})]] + \mathbf{G}E[(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])^T] + \mathbf{G}\mathbf{S}_\varepsilon,
\end{aligned} \tag{4.3.65}$$

where  $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$  and  $E[\boldsymbol{\varepsilon}] = 0$  gives  $E[\mathbf{f}(\mathbf{x})] = E[\mathbf{y}]$ . Since the left hand side is zero, we can solve this equation for  $\mathbf{G}$ . Hence, with  $\mathbf{x}_{ap} = E[\mathbf{x}]$  we may write

$$\mathbf{G} = \{E[\mathbf{x}\mathbf{f}(\mathbf{x})] - E[\mathbf{x}]E[\mathbf{f}(\mathbf{x})]\} \cdot \{E[(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])^T] + \mathbf{S}_\varepsilon\}^{-1}. \tag{4.3.66}$$

The various expectations occurring in Eq. 4.3.66 are derived by first approximating the terms by Taylor series about  $\mathbf{x}_{ap}$  truncated to second order and then applying the expectation operator. We demonstrate these steps with the first term in Eq. 4.3.66,  $E[\mathbf{x}\mathbf{f}(\mathbf{x})]$ , starting with the Taylor expansion

$$\begin{aligned}
\mathbf{x}\mathbf{f}(\mathbf{x}) &= \mathbf{x}_{ap}\mathbf{f}(\mathbf{x}_{ap}) + \left[ \mathbf{f}(\mathbf{x}_{ap}) + \mathbf{x}_{ap} \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}_{ap}} \right] (\mathbf{x} - \mathbf{x}_{ap}) + \\
&\quad \frac{1}{2} \left[ 2 \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} + \mathbf{x}_{ap} \frac{\partial^2 \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^2} \Big|_{\mathbf{x}=\mathbf{x}_{ap}} \right] (\mathbf{x} - \mathbf{x}_{ap})(\mathbf{x} - \mathbf{x}_{ap})^T,
\end{aligned} \tag{4.3.67}$$

where the Leibniz' Rule to calculate higher derivatives of a product of two functions has been employed. Since the expectation of the second term in Eq. 4.3.67 vanishes, we get with  $\mathbf{S}_{ap} = E[(\mathbf{x} - \mathbf{x}_{ap})(\mathbf{x} - \mathbf{x}_{ap})^T]$ ,

$$\begin{aligned}
E[\mathbf{x}\mathbf{f}(\mathbf{x})] &= \mathbf{x}_{ap}\mathbf{f}(\mathbf{x}_{ap}) + \left[ \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} + \frac{1}{2} \mathbf{x}_{ap} \frac{\partial^2 \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}^2} \right] \mathbf{S}_{ap} \\
&= \mathbf{x}_{ap}\mathbf{f}(\mathbf{x}_{ap}) + \mathbf{S}_{ap} \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right)^T + \frac{1}{2} \mathbf{x}_{ap} \partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap}).
\end{aligned} \tag{4.3.68}$$

See Eq. 4.3.61 for the definition of  $\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})$ . Following the same steps we calculate the expectation of  $(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])^T$  to find

$$\begin{aligned}
E[(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])(\mathbf{f}(\mathbf{x}) - E[\mathbf{f}(\mathbf{x})])^T] &= \\
&= \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right)^T \mathbf{S}_{ap} \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right) - \frac{1}{4} [\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})][\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})]^T,
\end{aligned} \tag{4.3.69}$$

where we used Eq. 4.3.69 for  $E[\mathbf{f}(\mathbf{x})] = E[\mathbf{y}]$ . Inserting Eq. 4.3.68 and Eq. 4.3.69 together with Eq. 4.3.60 into Eq. 4.3.66 gives the following expression for  $\mathbf{G}$

$$\begin{aligned} \mathbf{G} &= \mathbf{S}_{ap} \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right)^T \left[ \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right)^T \mathbf{S}_{ap} \left( \frac{\partial \mathbf{f}(\mathbf{x}_{ap})}{\partial \mathbf{x}} \right) - \frac{1}{4} [\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})][\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})]^T + \mathbf{S}_\varepsilon \right]^{-1} \\ &= \mathbf{S}_{ap} \mathbf{K}_{ap}^T \left[ \mathbf{K}_{ap}^T \mathbf{S}_{ap} \mathbf{K}_{ap} - \frac{1}{4} [\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})][\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})]^T + \mathbf{S}_\varepsilon \right]^{-1}. \end{aligned} \quad (4.3.70)$$

If we compare this result with the gain function matrix in the linear case  $\mathbf{G} = \mathbf{S}_{ap} \mathbf{K}_{ap}^T (\mathbf{K}_{ap}^T \mathbf{S}_{ap} \mathbf{K}_{ap} + \mathbf{S}_\varepsilon)^{-1}$  (Eq. 4.3.33), we find that Eq. 4.3.70 has an additional term in the present in order to account for non-linearities in the considered problem.

The corresponding retrieval equation of second order can be stated as

$$\hat{\mathbf{x}} = \mathbf{x}_{ap} + \mathbf{G} \left[ \mathbf{y} - \mathbf{f}(\mathbf{x}_{ap}) + \frac{1}{2} \partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap}) \right] \quad (4.3.71)$$

with the gain function matrix

$$\mathbf{G} = \mathbf{S}_{ap} \mathbf{K}_{ap}^T \left[ \mathbf{K}_{ap}^T \mathbf{S}_{ap} \mathbf{K}_{ap} - \frac{1}{4} [\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})][\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})]^T + \mathbf{S}_\varepsilon \right]^{-1} \quad (4.3.72)$$

and the vector  $\partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})$  of length  $m$  described by Eq. 4.3.61. The retrieval error covariance  $\hat{\mathbf{S}}$  is given by the  $m$ -form formula (Eq. 4.3.34), that is,

$$\hat{\mathbf{S}} = \mathbf{S}_{ap} - \mathbf{G} \mathbf{K} \mathbf{S}_{ap}. \quad (4.3.73)$$

The term  $\frac{1}{2} \partial^2(\mathbf{f}(\mathbf{x}_{ap}), \mathbf{S}_{ap})$  is sometimes called bias correction term and accounts for the effect of the non-linearities. It should be noted that the associated increased computational burden can be severe. Nevertheless, second-order solutions provide in general considerably better performance than first-order equations due to the bias correction term, especially, when the problem is very non-linear. For further discussion on second-order filters see, e.g., Maybeck (1982).

In our IASI temperature retrieval work described in the next chapter we will make use of the iterated linearization approach discussed in Sect. 4.3.2.2.

---

## 5 IASI TEMPERATURE RETRIEVAL AND PERFORMANCE ANALYSIS

Our task is now to optimally estimate the state of the atmosphere when measurements corrupted by noise are given. Provided with the theoretical background given in the preceding chapters we present some results, their interpretation and a discussion on various topics associated with the retrieval. In Sect. 5.1 we give a description of the inputs required for the retrieval, such as a proper modeling of the *a priori* and measurement errors. In order to make the retrieval computationally efficient, we introduce methods to select a subset out of the total number of IASI channels (Sect. 5.2). We employ an iterative optimal estimation algorithm for moderately non-linear processes, as described in the previous chapter, Sect. 4.3.2. A comparison of retrieval results utilizing different channel sets is given in Sect. 5.3, performed with standard atmospheric profiles. In Sect. 5.4 we retrieve temperature fields and profiles to discuss special features and characteristics of the retrieval. Finally, we present temperature retrievals for a globally distributed ensemble of profiles by means of an empirical analysis based on an ensemble of retrieved temperature profiles (Sect. 5.5).

### 5.1 Retrieval Inputs

#### 5.1.1 Simulated Measurements and Measurement Error

Measurements are simulated with the fast radiative transfer model RTIASI described in Sect. 4.3. RTIASI requires an input profile vector containing temperature values in K, water vapor mixing ratio values in ppmv, and ozone mixing ratio values in ppmv at 43 pressure levels from surface (1013.25 mbar) to 0.1 mbar (see Appendix, Tab. A4). In addition, a few other quantities have to be specified including surface air temperature in K, surface water vapor mixing ratio in ppmv, surface pressure in mbar, u- and v- component of wind speed in m/s, surface skin temperature in K, cloud top pressure in mbar, and effective cloud coverage. For these quantities default settings in the RTIASI program have been used. Given this input profile vector RTIASI calculates the corresponding radiance or brightness temperature vector. If not otherwise noted, we will refer to

brightness temperature  $T_B$  as the measurements, which we denote by  $\mathbf{y}$ , of dimension  $m$ . In general,  $\mathbf{y}$  consists of  $m = 8461$  elements corresponding to the number of IASI channels. But  $m$  can be reduced, if certain specifications in the RTIASI program are made. As will be seen in Sect. 5.2 we select appropriate channels, therefore  $\mathbf{y}$  is generally of much smaller dimension. RTIASI also computes the Jacobian matrix  $\mathbf{K}$ , as described in Sect. 4.2, which is of dimension  $m$  (number of channels)  $\times n$  (number of levels). Since we do not have true measurements, we add a random noise factor  $\Delta\mathbf{y}$  to the simulated measurements in order to generate quasi-realistic data. For the noise modeling (recipe obtained from P. Schluessel, EUMETSAT, personal communications, 2000) we first create normally distributed random numbers with standard deviation values according to the Level 1c noise table (Tab. A5) interpolated to every wavenumber. Since RTIASI calculates apodized radiance or brightness temperatures, we apodize this noise with a Gaussian function of a full width at half maximum (FWHM) of  $0.5 \text{ cm}^{-1}$  (standard deviation of  $0.212 \text{ cm}^{-1}$ ). Furthermore, the noise is properly scaled, based on the Planck law, from the reference temperature ( $T_r = 280 \text{ K}$ ) to the actual brightness temperatures as given by the simulated measurements. The temperature factor to be multiplied by the noise values is evaluated in form of  $(z_1 / z_2)$  with

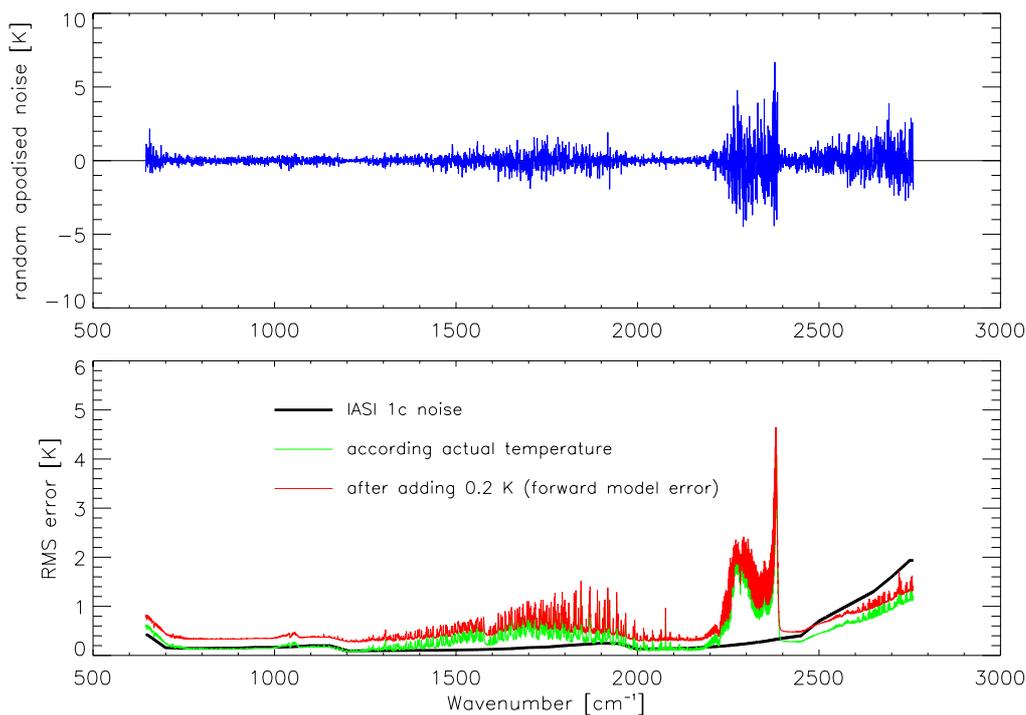
$$z_1 = \exp(c_2\nu/T_r)[T_B^2 \exp(c_2\nu/T_B)^2], \quad (5.1.1)$$

$$z_2 = T_r^2 \exp(c_2\nu/T_r)^2 \exp(c_2\nu/T_B), \quad (5.1.2)$$

where  $\nu$  is the wavenumber and  $c_2 = hc/k$  is the second radiation constant (Appendix, Tab. A1). The resulting values are shown in Fig. 1.5 (top).

In order to create an appropriate (and consistent) error covariance matrix  $\mathbf{S}_\epsilon$ , we assume the squared IASI 1c noise values to be our diagonal elements. Since the IASI 1c noise values are specified at a reference temperature  $T_r$  of 280 K as mentioned above, these values are modified according to the actual brightness temperature (see Eq. 5.1.1 and Eq. 5.1.2).

Finally, the temperature modified 1c noise values are superposed with an 0.2 K forward model error value to roughly account for errors in the forward model (Collard, 1998; J. Eyre, The Met. Office, personal communications, 2000). The impact of the RTIASI forward model error on the IASI retrieval accuracy is described in Sherlock (2000b). Fig. 5.1 (bottom) shows the raw IASI 1c noise values, and the modified values, according to a brightness temperature calculated for the U.S. standard mid-latitude summer profile.



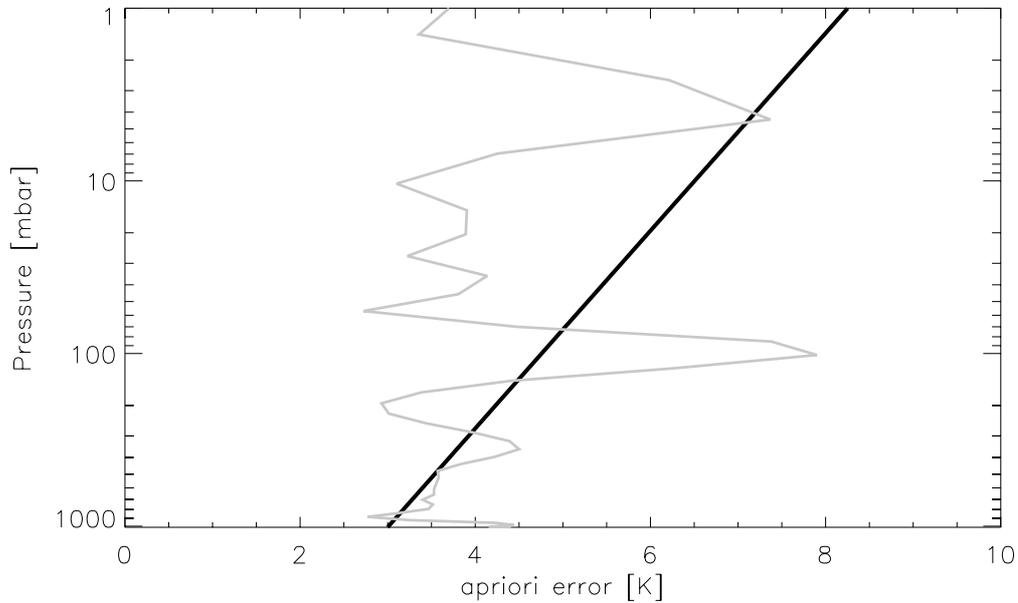
**Figure 5.1:** Top: Random apodized noise values versus wavenumber. Bottom: Square roots of the diagonal elements of the measurement error covariance matrix  $\mathbf{S}_e$ .

As mentioned above, RTIASI simulates radiance, which is apodized with a Gaussian function and this issue requires an apodization of the random noise as well. This procedure leads to a correlation between neighboring channels, which we also have to account for in  $\mathbf{S}_e$ . Hence, we generate non-diagonal elements of  $\mathbf{S}_e$  according to  $S_{ij} = c_{ij} / \sqrt{S_{ii}S_{jj}}$ , where correlation values  $c_{ij}$  of 0.71, 0.25, and 0.04, respectively, are assumed between a particular channel and its first, second and third neighbor. This produces a covariance matrix with a rather steep descent from the main diagonal (P. Schluessel, EUMETSAT, personal communications, 2000).

### 5.1.2 *A priori* Profiles and *a priori* Error

In most of the results shown below we utilize a standard background model, the climatological CIRA86aQ model (Kirchengast et al., 1999), to select *a priori* profiles  $\mathbf{x}_{ap}$ . Regarding temperature profiles the CIRA86aQ profiles are values just interpolated within the classical CIRA86 tabular model (Fleming et al., 1987). The “true“ temperature profiles  $\mathbf{x}$  are taken from a global ECMWF analysis field providing profiles at any desired latitude and longitude on a global grid. The CIRA86aQ model accounts for the latitudinal and seasonal variations of the atmospheric state, the ECMWF analysis is the operational analysis from Sept. 15, 1999, 12 UTC computed at T213L50

resolution (horizontal spectral resolution T213, 50 vertical levels up to 0.1 mbar). The *rms* (root mean square) error of an ensemble of the difference  $\mathbf{x} - \mathbf{x}_{ap}$  (about 100 profiles) is shown in Fig. 5.2 indicated as the gray line. We notice quite large errors at the tropopause and in the stratosphere near the 5 mbar level.



**Figure 5.2:** Root mean square error of  $\mathbf{x} - \mathbf{x}_{ap}$  (gray line) and assumed simplified *a priori* error (black line).

The *a priori* error covariance matrix  $\mathbf{S}_{ap}$  reflects the uncertainty how close the *a priori* profile is to the true profile. Since in reality we do not know the true profile, we assume as a baseline a linearly increasing error profile as indicated in Fig. 5.2 (black line) starting from 3 K at the surface to 10 K at 0.1 mbar. This specification fits the *rms* of the ensemble only in a simplified manner and may be too optimistic at the tropopause, but too strict in some regions of the stratosphere. The impact of this (over-) simplification will be analyzed and discussed in Sect. 5.4.3. Since we have quite thin layers especially in the lower atmosphere correlation between the levels exists. We assume non-diagonal elements of  $\mathbf{S}_{ap}$  with a correlation length  $L$  of 3 km obeying an exponential drop-off according to  $S_{ij} = \sigma_i \sigma_j \exp[-|z_i - z_j|/L]$  where  $\sigma_i = \sqrt{S_{ii}}$  is the standard deviation at level  $i$  and  $z$  denotes the height in kilometers at the particular level. We calculate the height  $z$  by implementing the hydrostatic equation  $z = -H \log(p/p_s)$  with a scale height  $H$  of 7 km and surface pressure  $p_s$  of 1013.25 mbar. The resulting matrix is illustrated in Fig. 5.3. Its correlations roughly reflect the correlations seen in background error covariance matrices derived at ECMWF from comparison of short-term forecasts with co-incident analyses (V. Sherlock, ECMWF, personal communications, 1999).

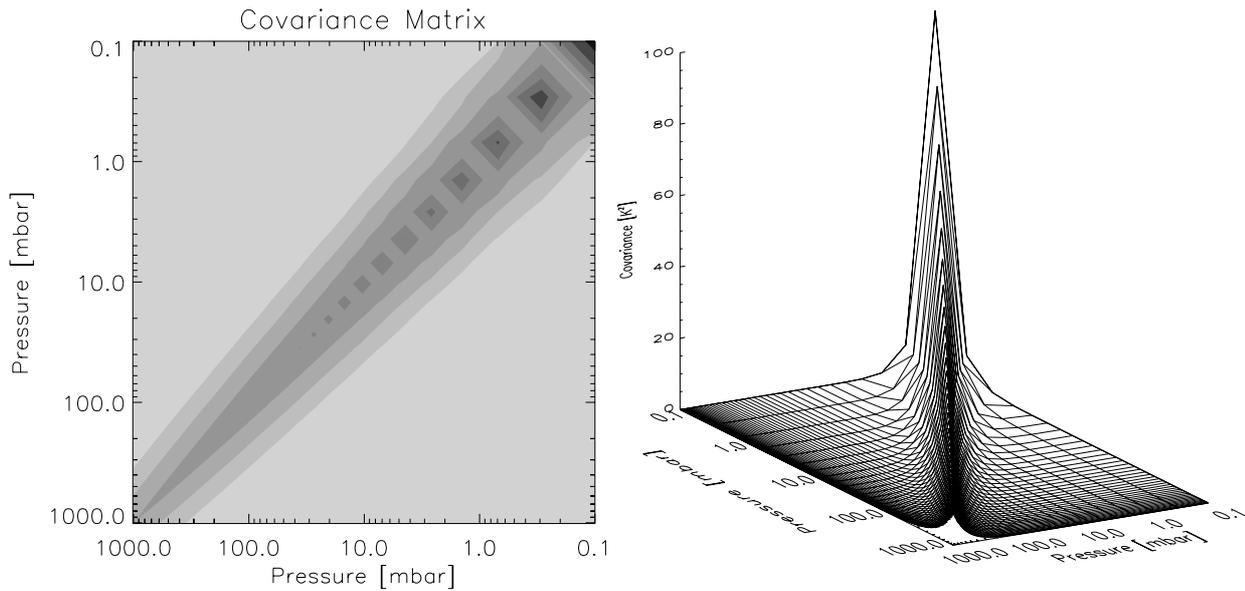


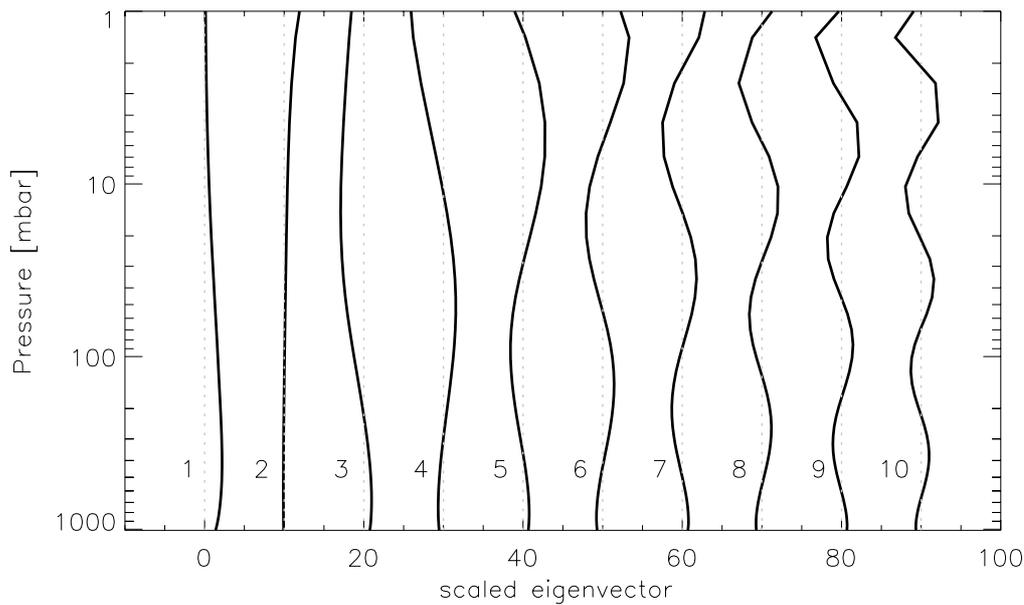
Figure 5.3: *A priori* error covariance matrix  $\mathbf{S}_{ap}$ , top view (left) and 3D view (right).

The *a priori* covariance matrix may be accurately known in some applications. To model this situation we calculate *a priori* profiles  $\mathbf{x}_{ap}$  which are consistent with  $\mathbf{S}_{ap}$  by construction as follows. Any covariance matrix  $\mathbf{S}$  can be decomposed as  $\mathbf{S} = \sum_j \mathbf{e}_j \mathbf{e}_j^T$ , where  $\mathbf{e}_j = \lambda_j^{1/2} \mathbf{l}_j$  are the error patterns (Rodgers, 2000), which can be regarded as eigenvectors  $\mathbf{l}_j$  scaled by the eigenvalues  $\lambda_j$  of  $\mathbf{S}$ . The random error in  $\mathbf{x}$  can then be expressed by

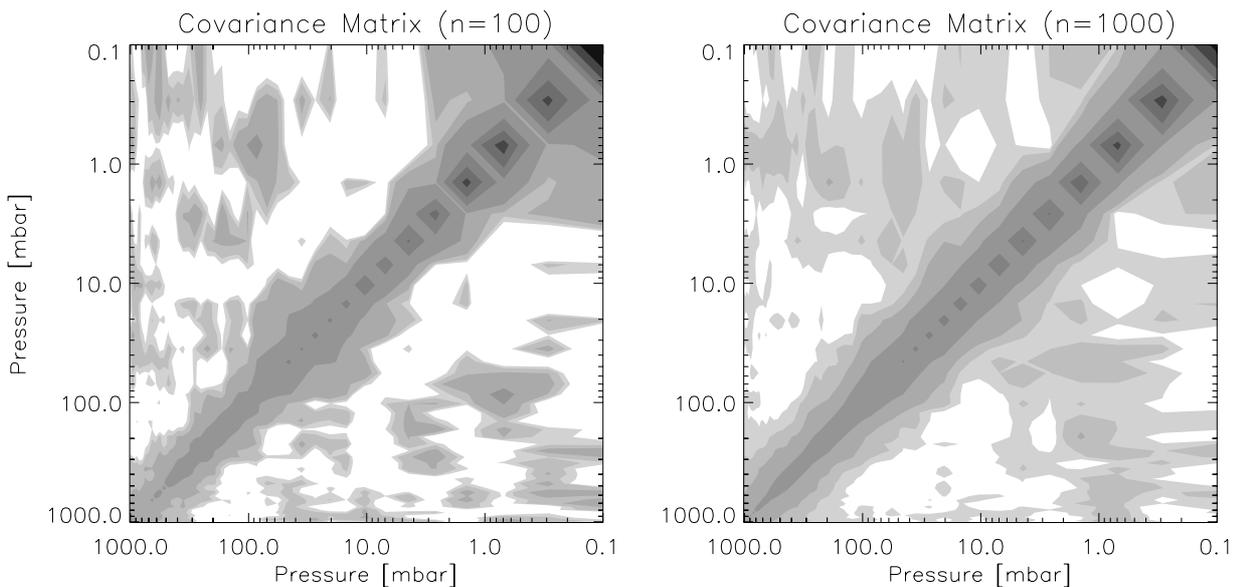
$$\boldsymbol{\varepsilon}_x = \sum_j a_j \mathbf{e}_j, \quad (5.1.3)$$

where the  $a_j$  are normally distributed random deviates with unit variance (computed by any standard normal random numbers generator). The first ten error patterns of  $\mathbf{S}_{ap}$  (Fig. 5.3) are depicted in Fig. 5.4.

To generate *a priori* profiles consistent to  $\mathbf{S}_{ap}$  we simply can add  $\boldsymbol{\varepsilon}_x$  to the “true” state vector  $\mathbf{x}$ . We can verify that the empirical covariance of an ensemble of realizations  $\boldsymbol{\varepsilon}_x$ , Eq. 5.1.3, indeed reconstructs the original  $\mathbf{S}_{ap}$  as shown in Fig. 5.3. This is illustrated in Fig. 5.5, where the empirical covariance  $E[\boldsymbol{\varepsilon}_x \boldsymbol{\varepsilon}_x^T]$  of an ensemble of 100 as well as of 1000 random vectors  $\boldsymbol{\varepsilon}_x$  has been calculated. It is seen that with an increasing number of ensemble members, the empirical matrix approaches the actual covariance matrix.



**Figure 5.4:** First ten scaled eigenvectors (i.e., error patterns) of the covariance matrix shown in Fig. 5.3.



**Figure 5.5:** Empirical covariance matrix using an ensemble of 100 (left) and an ensemble of 1000 (right) random vectors.

The process described by Eq. 5.1.3 can also be implemented to generate humidity realizations to gain more realistic retrievals than with using perfect humidity knowledge. For this purpose we assumed a covariance matrix  $\mathbf{S}_{ap}$  with constant diagonal elements of 0.15 in  $\ln(q)$  throughout the range to 200 mbar. This value corresponds to a percentage error of approximately 15% in specific humidity  $q$ , which approximately corresponds to the error expected for IASI humidity retrievals (e.g., Collard, 1998; Lerner et al., 2000a). The humidity can be specified in units parts per million

by volume (ppmv) (as in RTIASI) or as specific humidity (as in our inversion scheme). The latter is the ratio of the mass of water vapor to the total air mass (i.e., moist air mass), usually expressed in g/kg. In this work we use the (slightly simplified) conversion formula  $q[\text{g/kg}] = q[\text{ppmv}] \times (M_{\text{H}_2\text{O}}/M_{\text{d}}) \times 10^{-3}$ , where the ratio of the molar mass of water vapor,  $M_{\text{H}_2\text{O}}$ , to that of dry air,  $M_{\text{d}}$ , is 0.622.

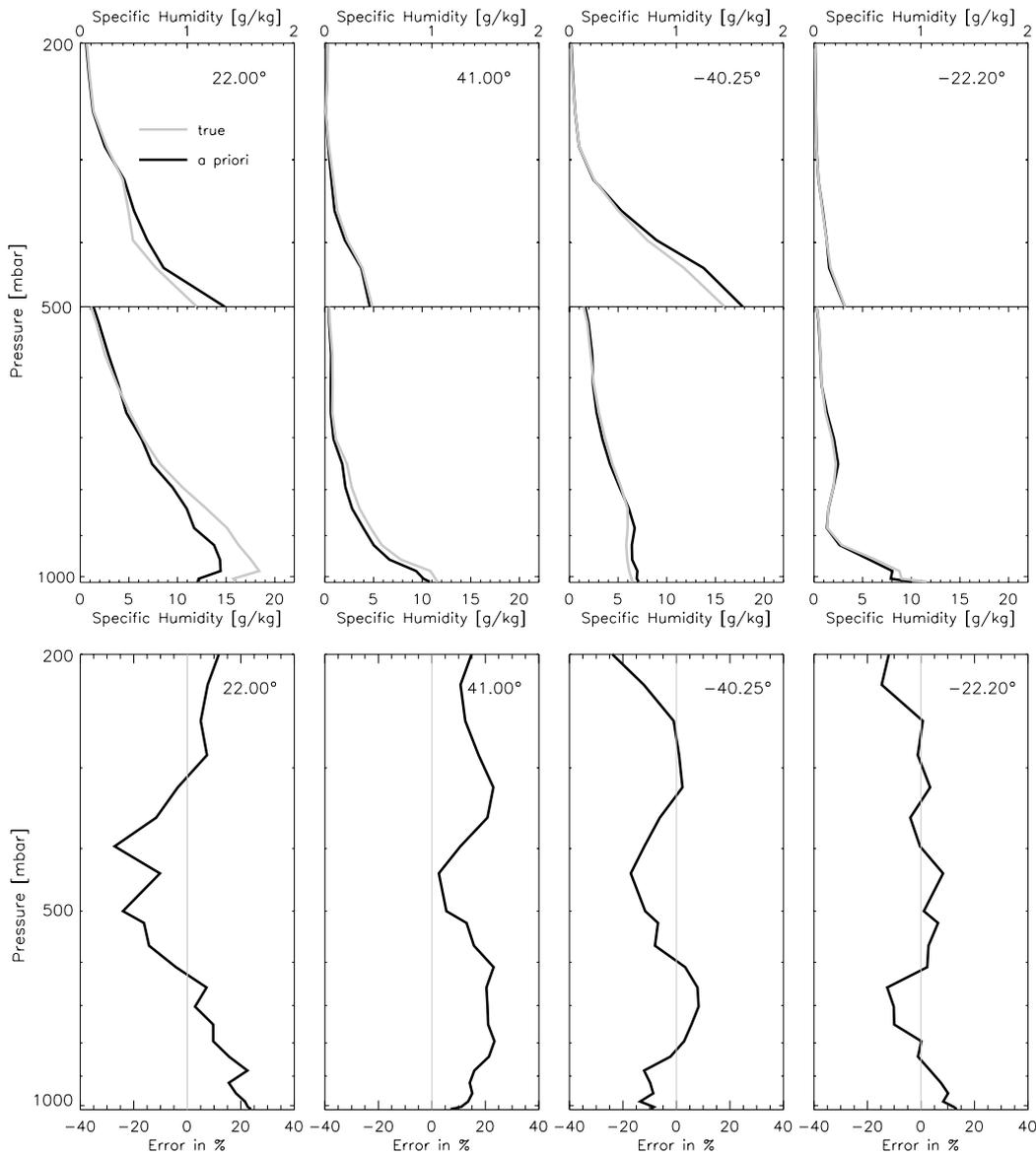


Figure 5.6: Top: *A priori* humidity profiles (black lines) and “true“ (ECMWF analysis) profiles (gray lines) at different latitudes (in degrees). Bottom: Corresponding deviation from the “true“ profiles in percentage.

The non-diagonal elements of the covariance matrix  $S_{ap}$  were calculated according an exponential drop-off with a correlation length of 6 km. Fig. 5.6 shows four exemplary profiles at different latitudes together with the “true“ profiles in specific humidity g/kg, as well as the deviation from

the “true“ profile in percents. The “true“ profiles have been taken from an ECMWF analysis. The profiles can be regarded as (unbiased) perturbed “true“ values and constitute the prescribed input profiles for humidity in our temperature retrievals. Note that in an extended scheme for simultaneous temperature and humidity retrieval these profiles would not be a fixed input but serve as *a priori* profiles (updated in the retrieval process together with the temperature profiles).

## 5.2 Channel Selection

Since the full IASI spectra contain 8461 channels, it is essential to reduce this number and somehow remove redundant information for practical and computational purposes. Hence, our task is to find a subset of channels, which is sufficiently sensitive to the retrieved variable.

We start with removing those channels at wave numbers above  $2500\text{ cm}^{-1}$ . The first reason to do this is that these channels have larger measurement errors (above 0.7 K) according to the IASI noise table (Tab. A5) and Fig. 5.1 (bottom), the second one is that in this spectral range ( $<4\text{ }\mu\text{m}$ ) residual solar contribution becomes important, which is not taken into account in the forward model, and the third is that this region comprises “atmospheric window” and  $\text{CH}_4$  channels, which can be ignored for temperature profiling.

Next we remove those channels whose “foreign” gas emissions contribute significantly to the measured brightness temperatures. This includes the following bands:  $825\text{--}1100\text{ cm}^{-1}$ ,  $1220\text{--}1370\text{ cm}^{-1}$ , and  $2085\text{--}2220\text{ cm}^{-1}$ . The associated trace gas constituents  $\text{O}_3$ ,  $\text{CO}$  and  $\text{CH}_4$  need to be modeled appropriately and considered as uncertainties in the temperature (or humidity) retrieval. With the luxury of high-spectral resolution measurements, we can exclude these subsets from our channel selection. The  $825\text{--}1100\text{ cm}^{-1}$  “foreign” band also excludes “atmospheric window” channels not required for the profiling.

At this point we still have about 5100 channels, which is not practical for most operational and climate applications. It is instructively shown in Rodgers (1996) that it is no advantage to utilize all pieces of information, since they are highly redundant for most purposes. In fact Rodgers (1996) gave the first description and application of Information Content Theory as discussed below to IASI measurements.

## 5.2.1 Information Content Theory

In information theory one seeks to know how much information is contained in a possible outcome by knowing it. If an outcome  $x_i$  has a probability density  $p(x_i)$ , then  $-\log p(x_i)$  defines the information gained by knowing the value of the outcome. Hence, if  $p(x_i)=1$ , then the information is zero implying that we gain no information when knowing the certain state. The entropy defined as

$$S(p(x)) = -\sum_i p(x_i) \log p(x_i) \quad (5.2.1)$$

represents the mean information provided per possible outcome  $x_i$  of a system  $x$  such that the entropy defines the information of an event as a whole. In information theory the logarithm is usually taken to base 2 to get units of bits. Now the mutual information between  $x$  and  $y$  can be expressed by

$$H = S(p(x)) - S(p(x|y)), \quad (5.2.2)$$

where  $p(x|y)$  is the probability density function of  $x$  conditioned on  $y$ .  $H$  measures the amount of information that  $x$  conveys about  $y$ , or in other words,  $H$  determines the reduction in uncertainty in  $x$  by learning the value of  $y$ .

Following the aforementioned issues we describe the information content  $H$  gained by making a measurement as the change in entropy according to

$$H = S(p(\mathbf{x})) - S(p(\mathbf{x}|\mathbf{y})), \quad (5.2.3)$$

where  $p(\mathbf{x})$  and  $p(\mathbf{x}|\mathbf{y})$  represent the probability density functions of the state  $\mathbf{x}$ , and of  $\mathbf{x}$  conditioned on the measurements  $\mathbf{y}$ , respectively. The entropy for normally distributed probability densities with covariance  $\mathbf{S}$  is given by  $S(p) = \frac{1}{2} \log_2 |\mathbf{S}|$  (Rodgers, 1996; Rodgers, 2000). If we select the channels sequentially by retaining the channel with highest  $H$  and removing it from subsequent calculations we may write

$$\begin{aligned} H_i &= \frac{1}{2} \log_2 |\hat{\mathbf{S}}_{i-1}| - \frac{1}{2} \log_2 |\hat{\mathbf{S}}_i| \\ &= \frac{1}{2} \log_2 |\hat{\mathbf{S}}_{i-1} \hat{\mathbf{S}}_i^{-1}|. \end{aligned} \quad (5.2.4)$$

The matrix  $\hat{\mathbf{S}}_i$  is the retrieval error covariance given by

$$\hat{\mathbf{S}}_i = \hat{\mathbf{S}}_{i-1} - \hat{\mathbf{S}}_{i-1} \mathbf{K}^T (\mathbf{K}^T \hat{\mathbf{S}}_{i-1} \mathbf{K} + \mathbf{S}_\varepsilon)^{-1} \mathbf{K} \hat{\mathbf{S}}_{i-1} \quad (5.2.5)$$

(see Chapt. 4, Eq. 4.3.34, with  $\hat{\mathbf{S}}_{i-1} = \mathbf{S}_{ap}$ ). We notice that  $\mathbf{S}_i$  is updated with the covariance from the previous step. We utilize the scaling of the Jacobian matrix  $\tilde{\mathbf{K}} = \mathbf{S}_\varepsilon^{-\frac{1}{2}} \mathbf{K} \mathbf{S}_{ap}^{\frac{1}{2}}$  such that the measurement error covariance matrix becomes the unity matrix. The matrices  $\mathbf{S}_{ap}$  and  $\mathbf{S}_\varepsilon$  equal those described in Sect. 5.1, except that all non-diagonal elements are assumed to be zero. Thus Eq. 5.2.5 may be rewritten as

$$\hat{\mathbf{S}}_i = \hat{\mathbf{S}}_{i-1} \left\{ \mathbf{I}_n - \frac{\mathbf{k}_i (\hat{\mathbf{S}}_{i-1} \mathbf{k}_i)^T}{1 + (\hat{\mathbf{S}}_{i-1} \mathbf{k}_i)^T \mathbf{k}_i} \right\}, \quad (5.2.6)$$

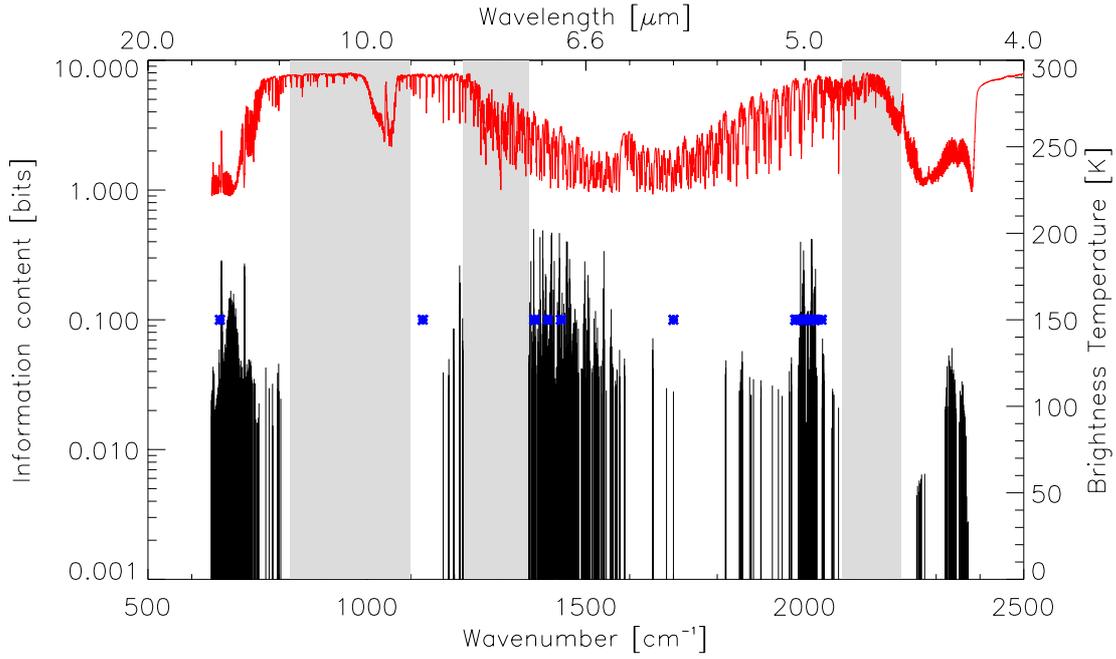
where the vectors  $\mathbf{k}_i$  of length  $n$  are the weighting functions and  $\mathbf{I}_n$  is the identity matrix of dimension  $n$ . Eq. 5.2.4. can be expressed by

$$H_i = \frac{1}{2} \log_2 (1 + \mathbf{k}_i^T \hat{\mathbf{S}}_{i-1} \mathbf{k}_i), \quad (5.2.7)$$

given Eq. 5.2.6.

The selection starts with  $\mathbf{S}_0 = \mathbf{S}_{ap}$ . The total number of IASI channels is pre-sorted according to the pressure level, where the weighting function of a channel peaks. The information content  $H$  is calculated using Eq. 5.2.7 for every available channel at each specific pressure level. The channel with the highest  $H$  is retained and removed from subsequent calculations, and  $\hat{\mathbf{S}}$  is updated according to Eq. 5.2.6. In the example shown below (Fig. 5.7) the number of selected channels per level is determined by 10% of the total number of peaking channels. In addition, we assume a maximum threshold of 30 and 50 channels for levels above 100 mbar and below 100 mbar, respectively. A number of 866 channels has been selected in the case of Fig. 5.7 for the U.S. standard mid-latitude summer profile. Shaded areas in Fig. 5.7 denote the absorption bands of  $\text{O}_3$ ,  $\text{CO}$  and  $\text{CH}_4$  as well as “atmospheric window” channels removed before the channel selection, the asterisk symbols indicate pseudo channels (to be discussed in Sect. 5.2.3) and the IASI brightness temperature spectrum is plotted for reference. The same example is illustrated in different form in Fig. 5.8a, while Fig. 5.8b depicts the case where only two channels per level have been allowed leading to 72 selected channels only.

It should be mentioned that, according to Fig. 4.4 in Sect. 4.2, most channels peak between 1 mbar and 950 mbar. For instance, about 700 channels peak around level 38 (922.46 mbar). There are no channels peaking between the surface and 950 mbar, or above 1 mbar. Hence, almost no surface channels have been selected.



**Figure 5.7:** Information content in bits of selected channels, foreign gas absorption bands (light gray), pseudo channels (asterisk) and IASI brightness temperature spectrum calculated for the standard mid-latitude summer profile.

## 5.2.2 Maximum Sensitivity Approach

As a simplified and faster alternative to using the selection approach of Sect. 5.2.1 we introduce an approach, which is solely based on the weighting function matrix scaled by the measurement errors. It is desirable to selectively choose those channels whose instrument noise is small or measurement sensitivity to temperature and humidity perturbations is high. This is achieved by using the following channel selection criterion which maximizes the sensitivity-to-error ratio, a matrix denoted by  $\mathbf{H}$ ,

$$\mathbf{H} = \mathbf{S}_\varepsilon^{-\frac{1}{2}} \mathbf{K}. \quad (5.2.8)$$

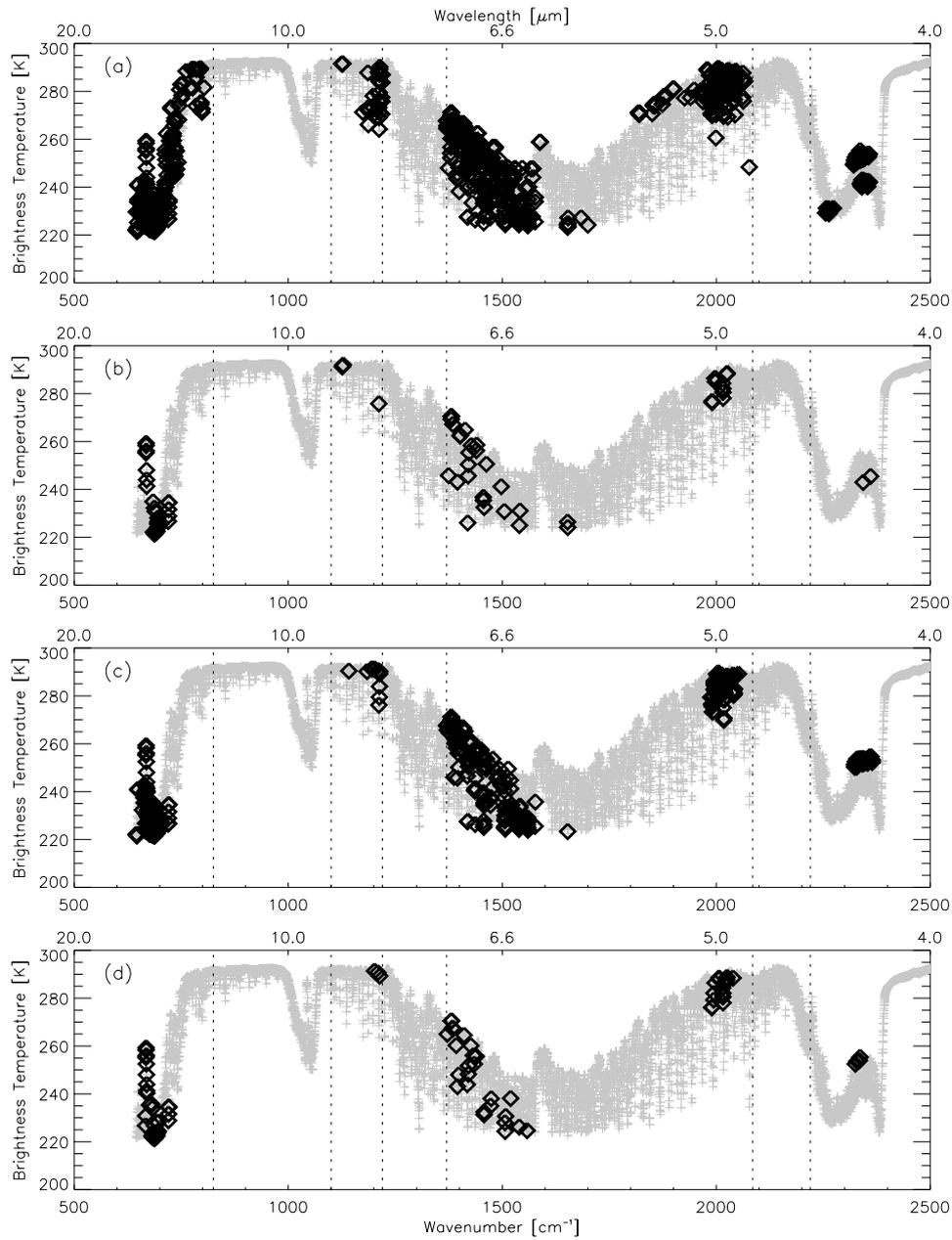
Again we take a diagonal measurement error covariance matrix  $\mathbf{S}_\varepsilon$  only (i.e., ignore the off-diagonal elements for this purpose). Thus, the parameter to be multiplied with the Jacobian matrix  $\mathbf{K}$  are the inverse standard deviations (inverse square roots of the diagonal elements of  $\mathbf{S}_\varepsilon$ ) of each measurement expressing the value or the uncertainty of a measurement. At each pressure level we select those channels with highest values  $H_{ij}$  (elements of  $\mathbf{H}$ ) and retain them from subsequent calculations. The results of selecting 10 and 2 channels per level are depicted in Fig. 5.8c and Fig. 5.8d, leading to 430 and 86 channels, respectively. A similar approach (“Jacobian method”) utilizes  $\mathbf{H} = \mathbf{S}_\varepsilon^{-\frac{1}{2}} \mathbf{K} \mathbf{S}_{ap}^{\frac{1}{2}}$  instead of Eq. 5.2.8 and was recently described by Rabier et al. (2000).

### 5.2.3 Channel Clustering

After a certain number of channels has been selected either according to Sect. 5.2.1 or Sect. 5.2.2, we average over a cluster, if such exist, to reduce measurement errors. We define a cluster as a group of four adjacent channels, all at the same level. For each such cluster occurring in the set of selected channels, we retain the first wavenumber, the averaged brightness temperatures  $T_B$  and  $\mathbf{K}$  values. The noise corresponding to such a pseudo channel is evaluated by summing all variances and covariances (since neighboring channels are correlated), taking the square root and dividing by 4. Assume, as an example, that four channels between 2021.00 and 2021.75  $\text{cm}^{-1}$  have the following  $T_B$  values: 286.39, 286.67, 286.71 and 286.76 K. The measurement errors are between 0.13210 and 0.13218 K. If we consider correlation between these channels as described in Sect. 5.2.1, the pseudo channel at 2021.0  $\text{cm}^{-1}$  with  $T_B = 286.63$  K has a measurement error of 0.10 K only.

In the case illustrated in Fig. 5.7 we found 8 clusters, denoted by asterisk symbols, and the number of channels is thereby reduced from 866 to 806. In the case of the maximum sensitivity approach, where we selected 10 channels per level leading to 430 channels (Fig. 5.8c), 10 clusters were found yielding finally 400 channels. In the remaining examples shown in Fig. 5.7b and Fig. 5.7d no clusters could obviously appear since only 2 channels per level had been allowed.

It can be seen in Fig. 5.8 that the regions of selected channels are quite similar in all four cases. Most of the channels are selected from the water vapor absorption band between 1200 and 1600  $\text{cm}^{-1}$ . Further “source bands” of channels are found in the 600–740  $\text{CO}_2$  band, a lower-troposphere  $\text{H}_2\text{O}$  region near 2000  $\text{cm}^{-1}$ , and in the strongly absorbing  $\text{CO}_2$  band near 2400  $\text{cm}^{-1}$ . Note that only a few surface channels are selected, which is adequate for our purpose, since the forward model is presumably not sufficiently accurate for satisfactory retrievals of surface parameters. A detailed discussion on retrieval results employing these four channel sets (Fig. 5.8, a–d) is given in the following section.



**Figure 5.8:** Channels selected via (a) information content (IC) theory (final number 806 channels), (b) IC theory (72), (c) “Maximum sensitivity“ (MS) approach (400), and (d) MS approach (86). Channels are indicated as black diamonds overlaid on the brightness temperature spectrum calculated for the U.S. standard mid-latitude summer profile.

### 5.3 Standard Temperature Profile Retrieval

To estimate temperature we utilize the iterative (iteration index  $i$ ) inverse algorithm described in Sect. 4.3.2, Eq. 4.3.48 and Eq. 4.3.49. Thus we have

$$\mathbf{x}_{i+1} = \mathbf{x}_{ap} + \mathbf{S}_i \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} [(\mathbf{y} - \mathbf{y}_i) - \mathbf{K}_i (\mathbf{x}_{ap} - \mathbf{x}_i)] \quad (5.3.1)$$

and

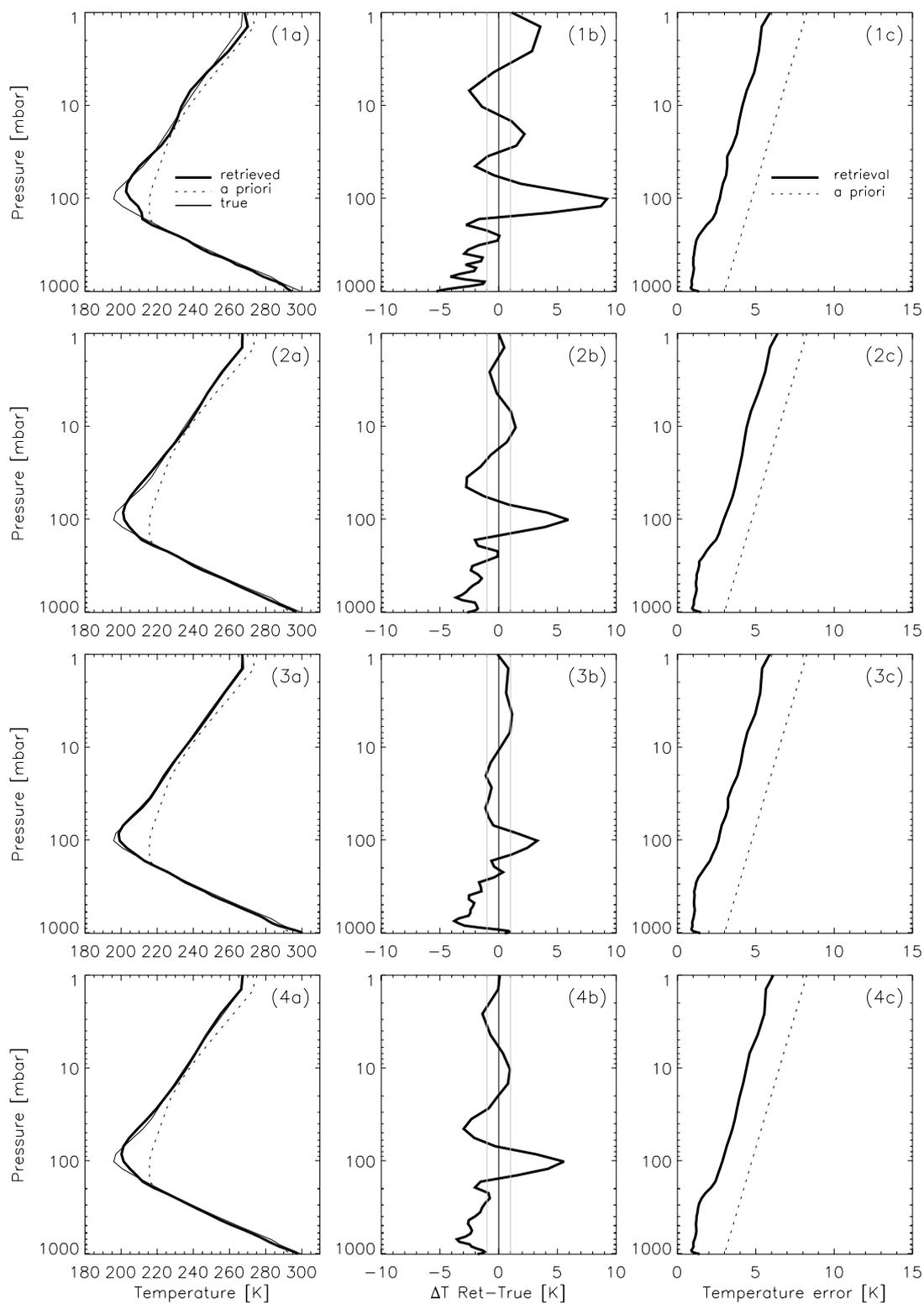
$$\mathbf{S}_i = (\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1}, \quad (5.3.2)$$

respectively. The *a priori* error covariance matrix  $\mathbf{S}_{ap}$  and the measurement error covariance matrix  $\mathbf{S}_\epsilon$  are implemented as described in Sect. 5.1.1 and Sect. 5.1.2, except that the measurement error covariance is modified in its diagonal according to

$$S_\epsilon(j, j) = \max \left[ \frac{1}{\alpha} (y(j) - y_i(j))^2, \sigma^2(j) \right], \quad (5.3.3)$$

where  $j$  is the channel index,  $i$  the iteration index,  $\sigma^2(j)$  is the variance of the measurement noise (i.e., the original diagonal element of  $\mathbf{S}_\epsilon$ ) and  $\alpha$  is a control parameter set to 4 in this study (Liu et al., 2000). At each iteration step we compare the difference of the actual measurements to the “true“ measurements with the measurement error and retain the larger value of these two quantities. This procedure, termed the “D-rad“ method (Liu et al., 2000), aids convergence when the *a priori* profile is poor, since in that case the measurement error grows and the retrieval algorithm, Eq. 5.3.1, does not attach too much weight on that particular measurement. This process usually effects the retrieval only in the very first iteration step, simply because the first factor in Eq. 5.3.3 decreases very quickly.

RTIASI supplies the measurements  $\mathbf{y}$  computed for an assumed “true“ input vector  $\mathbf{x}$ , as well as  $\mathbf{K}_i$  and  $\mathbf{y}_i$  corresponding to the actual estimate  $\mathbf{x}_i$ . In the example shown in Fig. 5.9 we take the U.S. standard mid-latitude summer profile to be our *a priori* profile  $\mathbf{x}_{ap}$ , whereas the “true“ profile  $\mathbf{x}$  is represented by the U.S. standard tropical profile.



**Figure 5.9:** Temperature retrieval for the U.S. standard tropical profile (true) given the U.S. standard mid-latitude summer profile (*a priori*). Panel 1(a-c): 806 channels used as in Fig. 5.7a, panel 2(a-c): 72 channels used as in Fig. 5.7b, panel 3(a-c): 400 channels used as in Fig. 5.7c, and panel 4(a-c): 86 channels used as in Fig. 5.7d.

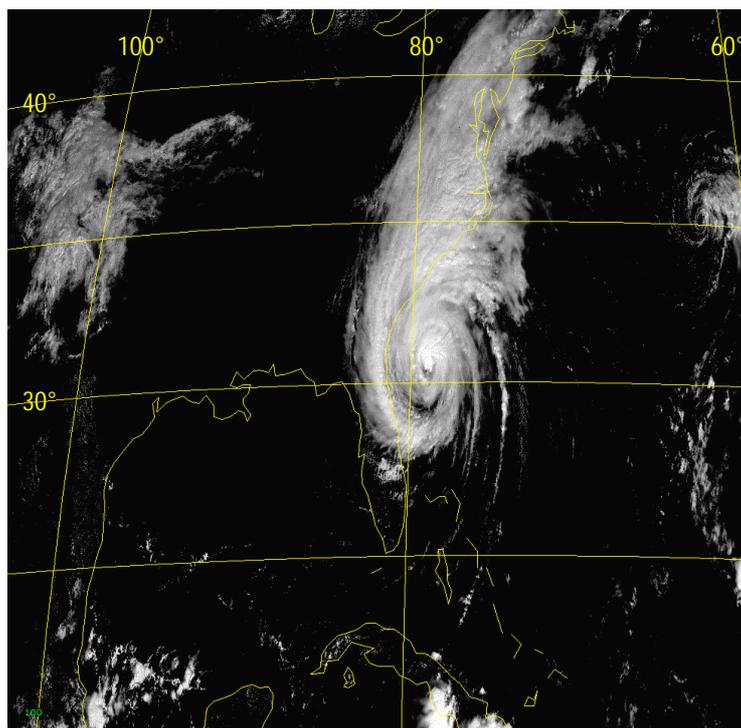
The number of iterations of Eq. 5.3.1 is determined by the criterion  $\chi^2 \leq m$  ( $m$  denotes the number of channels) employing the cost function quantity

$$\chi^2 = [\mathbf{y} - \mathbf{y}_i]^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y} - \mathbf{y}_i] + [\mathbf{x}_i - \mathbf{x}_{ap}]^T \mathbf{S}_{ap}^{-1} [\mathbf{x}_i - \mathbf{x}_{ap}] \quad (5.3.4)$$

If this criterion is not met, then convergence is obtained when  $\chi_i^2 \geq \chi_{i-1}^2$ . In the present case (Fig. 5.9) the convergence criterion is reached after 1 or 2 iterations. The first feature worth mentioning when inspecting Fig. 5.9 is that the retrieval is somewhat sensitive to the subset of selected channels. A high number of channels does not necessarily yield better retrievals as seen in panel 1(a-c), where about 800 channels have been processed. Concentrating on the second panels, denoted by (b), it seems, that especially at the tropopause temperature is more difficult to retrieve. The large errors originate mainly from a too optimistic assumption of the *a priori* error (see Fig. 5.2) and to a lesser extent from an insufficient sensitivity of  $\mathbf{K}$ . In the particular case of Fig. 5.9 the “maximum sensitivity” approach (400 channels) illustrated in panel 3(a-c) shows the best results in the tropopause region as well in the stratosphere. Retrievals with a small number of channels as in panel 2(a-c) and 4(a-c), where we employed 72 and 86 channels, respectively, seem quite satisfactory for many applications, except there are difficulties around the tropopause due to the previously mentioned reasons. In the third panels, denoted by (c), of Fig. 5.9 the standard deviation of  $\hat{\mathbf{S}}$ , Eq. 5.3.2, is indicated as a black line, while the dotted line represents the *a priori* standard deviation (square root of diagonal elements of  $\mathbf{S}_{ap}$ ). The retrieval error represents the theoretically estimated uncertainty of the retrieved profile and depends in great parts on  $\mathbf{S}_{ap}$ . In particular in the stratosphere, where measurements have insufficient weight, the estimated retrieval error tends towards the *a priori* error. The troposphere is well-measured by the instrument (large  $\mathbf{K}$  values) and the retrieval error is small. Since we use the same  $\mathbf{S}_{ap}$  in all four cases, the four estimated retrieval errors are almost identical, the small differences originate from different channels having different weighting functions and measurement errors.

## 5.4 The “Hurricane Floyd” Case Study: Performance Analysis

In this section we present the retrieval of a latitudinal temperature slice in the region of Hurricane Floyd, which threatened the U.S. Southeast coast on September 15, 1999 (Fig. 5.10). Note, though, that this is just selected as a representative arbitrary example featuring some reasonable temperature variability; we treat the situation as clear-sky situation and ignore for the purpose of this baseline case study any potential complication related to clouds. The “true” field comprises an ECMWF analysis (T213L50, Sep. 15 1999, 12 UTC, at longitude 79°W) interpolated to our 43-level pressure grid yielding the temperature field as shown in Fig. 5.11. For the retrieval we assume latitude steps of 0.475°, which corresponds to the IASI scan separation of 52.7 km. Thus we simulate an IASI flight along the aforementioned latitudinal slice and consider only Earth views at nadir.



**Figure 5.10:** Satellite image of Hurricane Floyd on 15 September 1999.

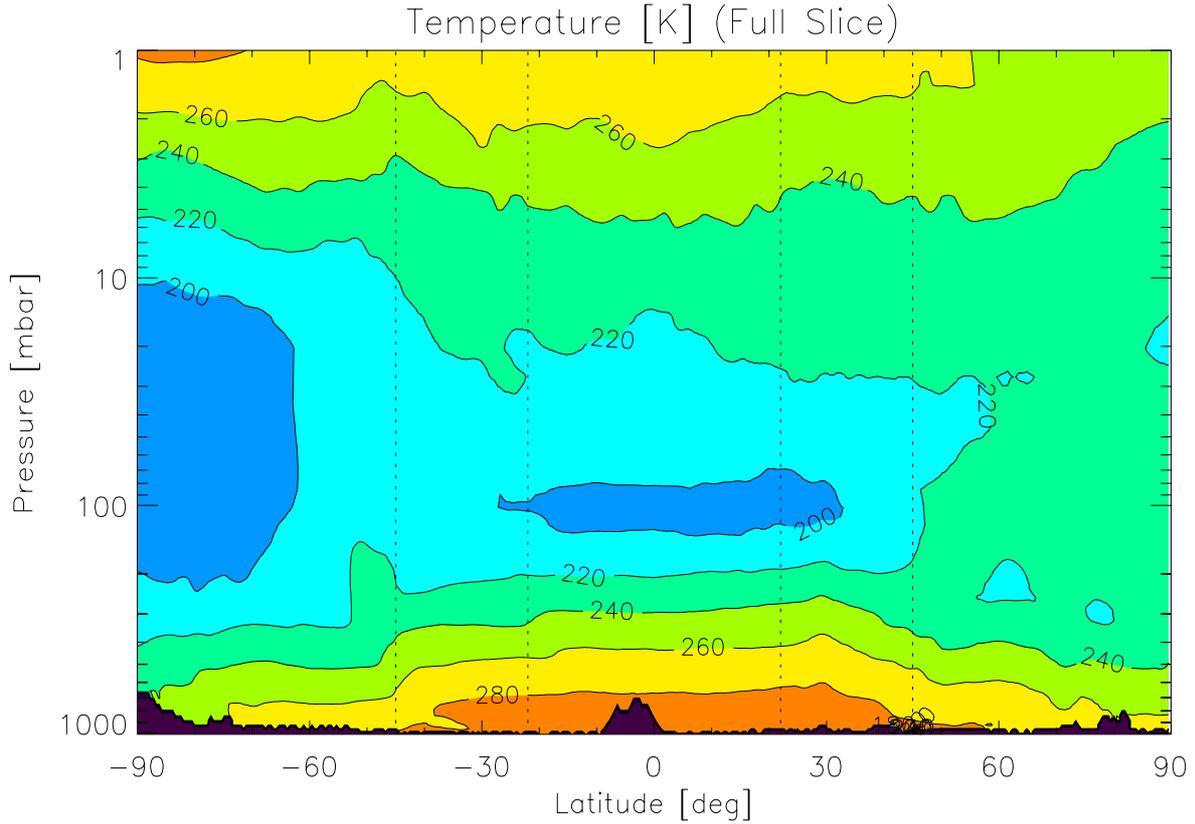


Figure 5.11: “True“ latitude-height temperature slice at longitude 79°W (ECMWF analysis T213L50, 15 Sept. 1999, 12 UTC). The bands bounded by dotted verticals are specifically selected (see Fig. 5.12).

### 5.4.1 Temperature Field Retrieval

We will concentrate on two latitude ranges, indicated as dashed lines in Fig. 5.11, namely the southern part from 45°S to 22°S, and the northern part from 22°N to 45°N, the latter containing the “Hurricane Floyd” region. The corresponding temperature fields are plotted individually in Fig. 5.12. With a latitude step of 0.475° we obtain 49 profiles per range.

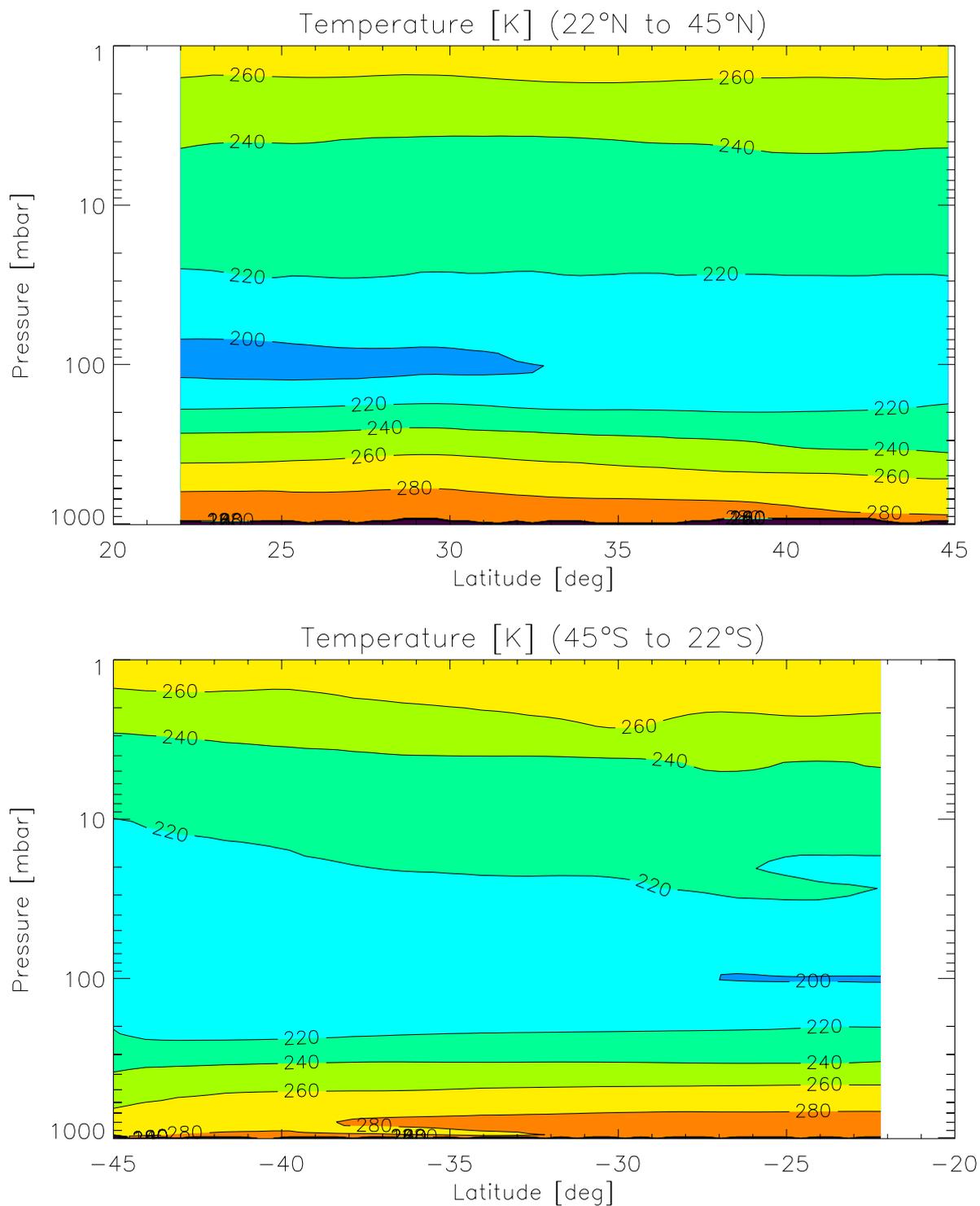
For the temperature retrieval we follow the algorithm described in Sect. 5.3 with

$$\mathbf{x}_{i+1} = \mathbf{x}_{ap} + \mathbf{S}_i \mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y} - \mathbf{y}_i - \mathbf{K}_i (\mathbf{x}_{ap} - \mathbf{x}_i)] \quad (5.4.1)$$

and

$$\mathbf{S}_i = (\mathbf{K}_i^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}_i + \mathbf{S}_{ap}^{-1})^{-1}, \quad (5.4.2)$$

where  $i$  is the iteration index. The measurements  $\mathbf{y}_i$  and the Jacobian  $\mathbf{K}_i$  are calculated by RTIASI for each iteration step, until the convergence criterion, Eq. 5.3.4, is reached. We will denote the final estimate with  $\hat{\mathbf{x}}$  and its error covariance with  $\hat{\mathbf{S}}$ , although strictly speaking,  $\hat{\mathbf{x}}$  is the solution when the number of iterations becomes infinity, (Rodgers, 2000).



**Figure 5.12:** "True" temperature slice at longitude 79°W (ECMWF analysis T213L50, 15 Sept. 1999, 12 UTC) . Top: Northern range from 22°N to 45°N. Bottom: Southern range from 45°S to 22°S.

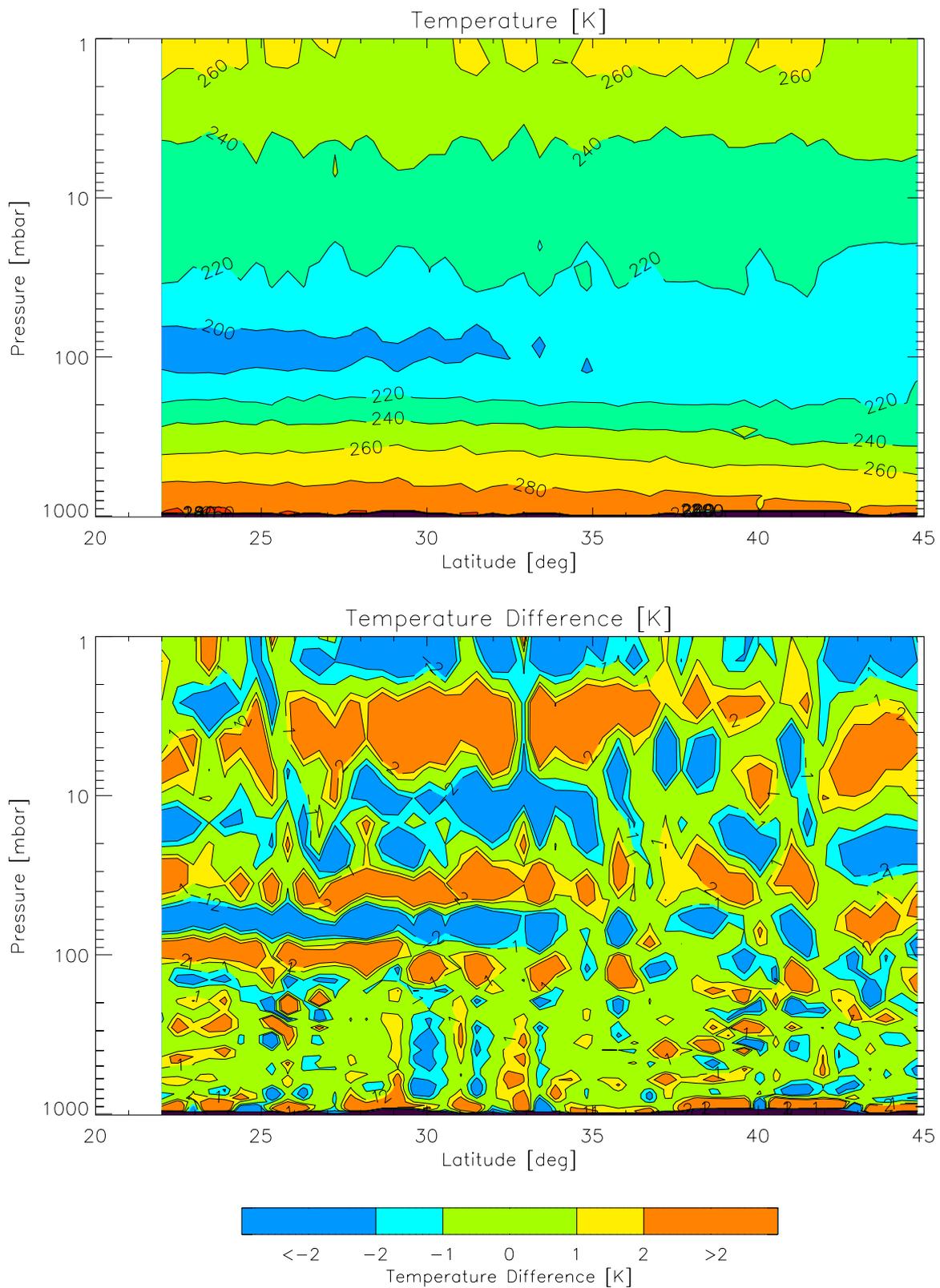
The measurement error covariance  $\mathbf{S}_\varepsilon$  has been described in Sect. 5.1.1. For the *a priori* profiles selection we employ the climatological background model CIRA86aQ (Kirchengast et al., 1999). The CIRA86aQ values are interpolated to the same latitude steps where the “true” temperature field is available.

Due to the varying meridional temperature structure it seems useful to determine the channel set according to the different latitude ranges. We pre-compute channels at every 10 degrees according to the CIRA86aQ climatology. This step can also be included in the retrieval program. But additional computation time due to the calculation of more than 8000 weighting functions, required for the channel selection procedure, has to be taken into account. To retrieve the temperature profile at, e.g., latitude 24.85°N we use the channel set created by processing the CIRA86aQ profile for 20°N and as the *a priori* profile we take the CIRA86aQ profile interpolated to 24.85°N. The channels have been selected by employing information content (IC) theory and the “maximum sensitivity” (MS) approach described in Sect. 5.2.1 and Sect. 5.2.2, respectively.

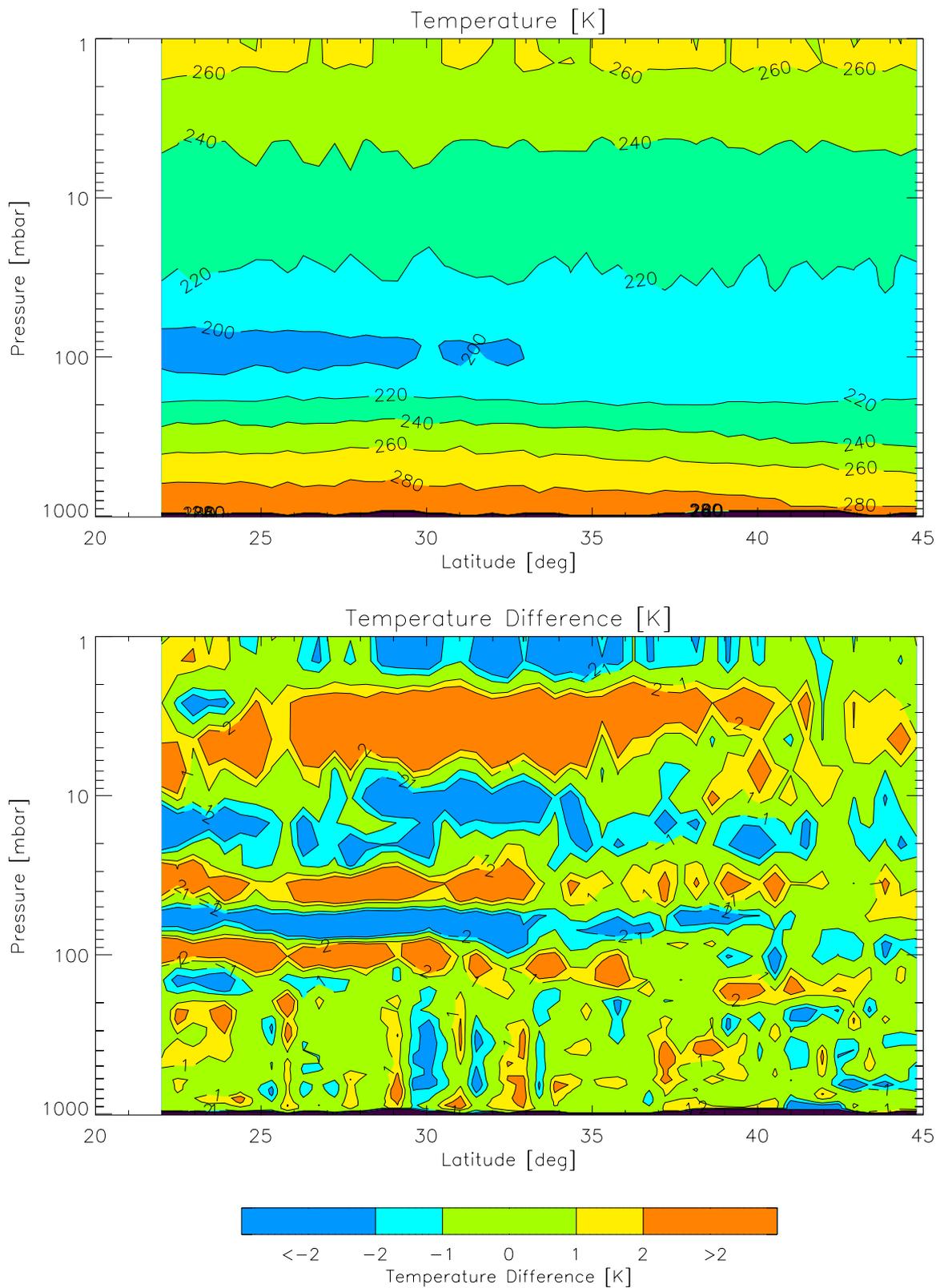
With these assumptions and inputs we present the retrieved temperature slices and the difference field (retrieved minus “true”) in Fig. 5.13 to Fig. 5.16 for the northern latitude range from 22°N to 45°N. We employ three different channel sets per case corresponding to 20°, 30° and 40° latitude.

In general, tropospheric temperature retrievals turned out quite satisfactory, whereas retrieved values in the boundary layer, near the tropopause or in parts of the stratosphere region show larger deviations from the “true” temperature field. The primary cause of these features is the shape and magnitude of the weighting functions, which exhibit largest values in the troposphere, but only a few channels peak, e.g., in the boundary layer. The large differences in the stratosphere also depend on the *a priori* error profile, which was assumed to have quite large values in the stratosphere (about 8 K at 1 mbar, see Fig. 5.2). At the heights where the sensitivity of  $\mathbf{K}$  is reduced, the estimated state tends towards the *a priori* state because there is not enough information about the true state (comprised in  $\mathbf{K}$ ) from the measurements. Fig. 5.13 has been created after employing information content (IC) theory as described in Sect. 5.2.1 and using the channel set depicted in Fig. 5.8a. Compared to the second experiment (Fig. 5.14) with IC theory but with fewer channels (as in Fig. 5.8b) we notice larger differences especially in the troposphere. Again, it seems that the allowance for more channels does not improve the retrieval. The same conclusions hold, when comparing Fig. 5.15 and Fig. 5.16. Thus, additional channels, selected by using the MS approach, do not provide better retrievals. To find out whether the simple MS approach (Sect. 5.2.2) does worse than the more sophisticated IC theory (Sect. 5.2.1) or not, we examine Fig. 5.14 and Fig. 5.16. The most significant distinctions may be found at the surface (in

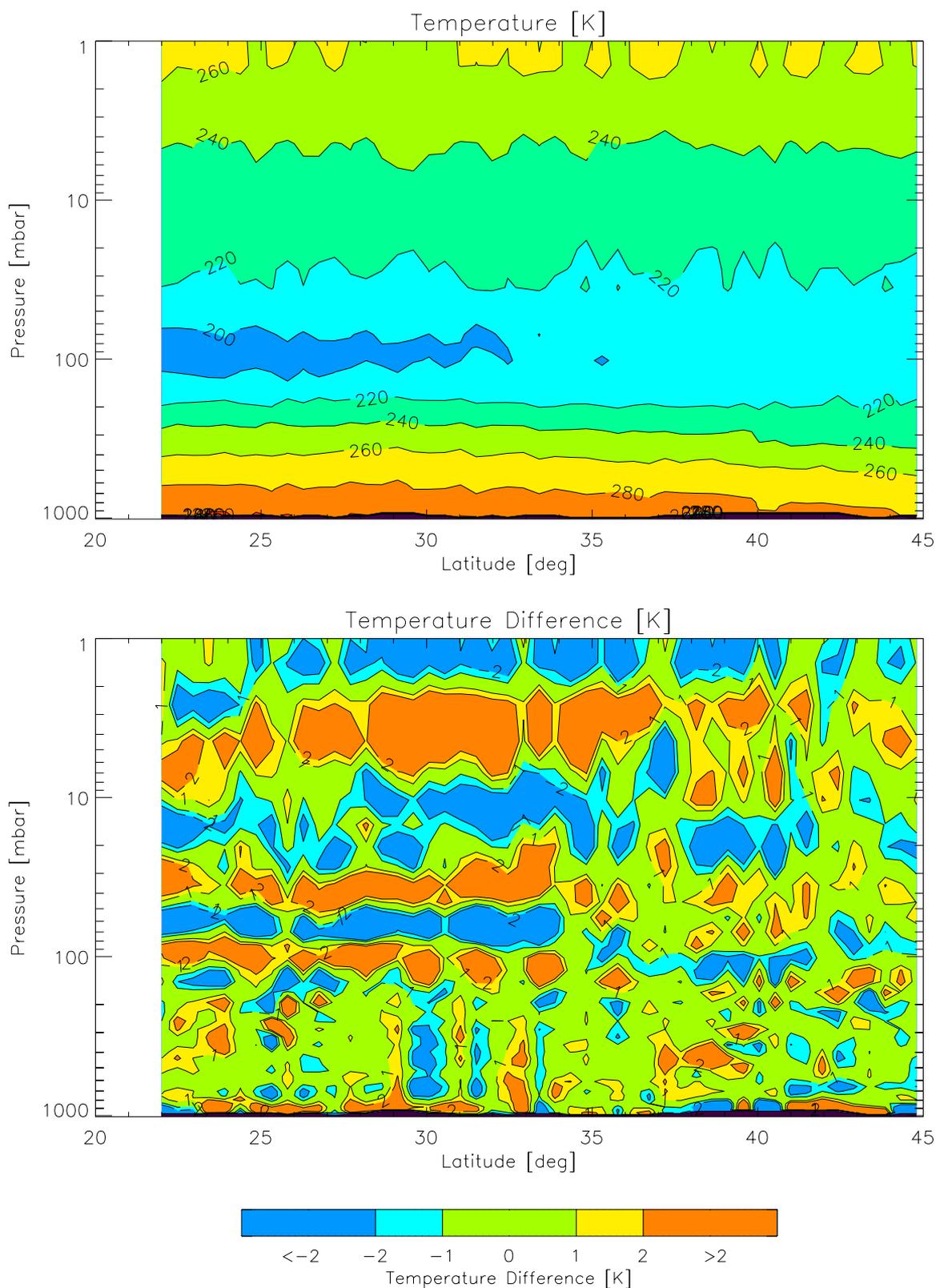
some areas), where the IC technique generally does a better job. At higher levels, retrieval results are similar. Since the “maximum sensitivity” approach requires a significantly simpler equation to implement, it is found to be an efficient substitute for the information content method for many applications.



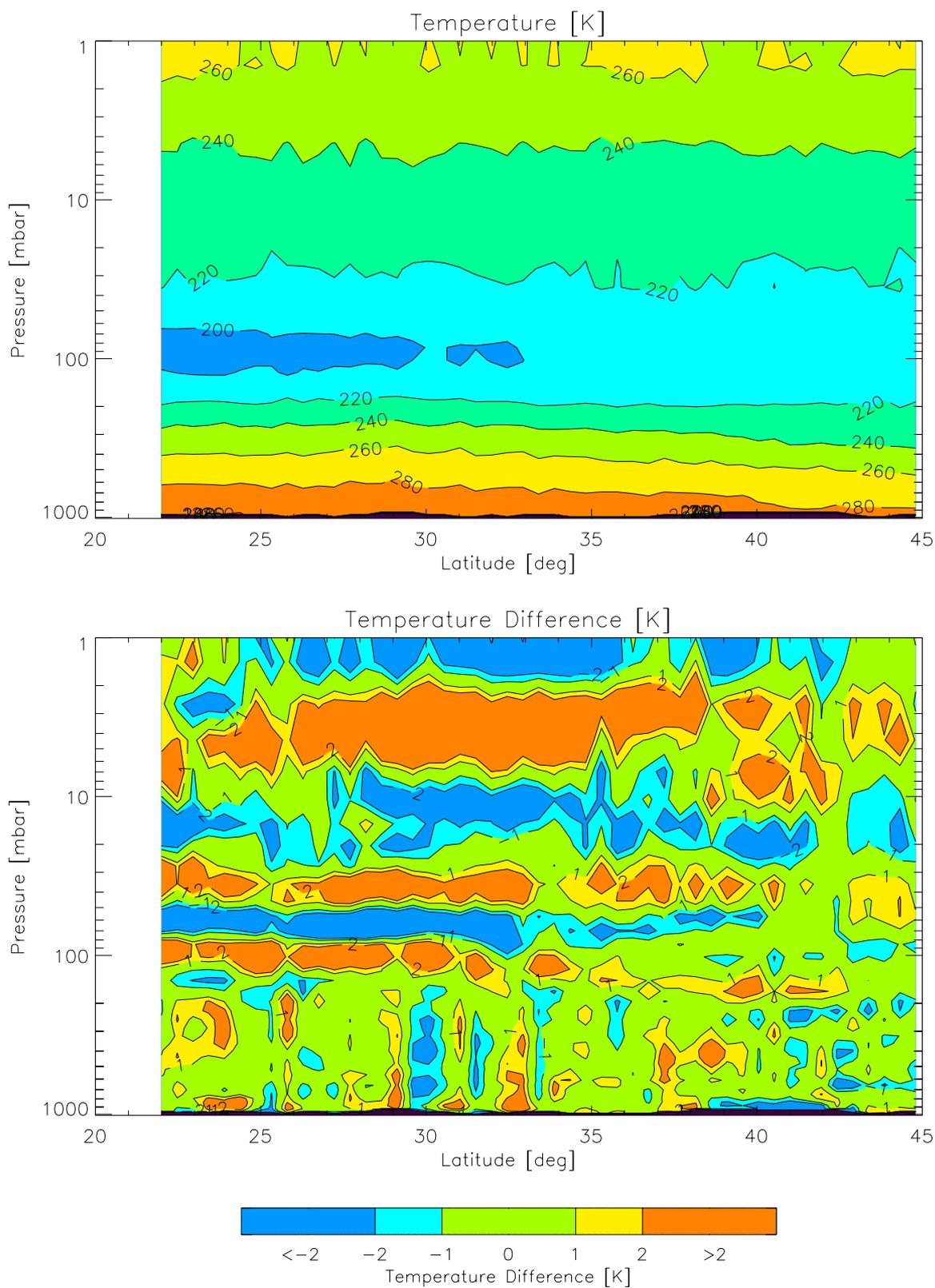
**Figure 5.13:** Retrieved temperature field (top) and “retrieved-minus-true” temperature difference (bottom) in the latitude range 22°N to 45°N. Information content theory and channel clustering applied as in Fig. 5.8a yields 803, 814 and 830 channels at 20°, 30° and 40° latitude, respectively.



**Figure 5.14:** Retrieved temperature field (top) and “retrieved-minus-true” temperature difference (bottom) in the latitude range 22°N to 45°N. Information content theory and channel clustering applied as in Fig. 5.8b yields 71, 71 and 72 channels at 20°, 30° and 40° latitude, respectively.



**Figure 5.15:** Retrieved temperature field (top) and “retrieved-minus-true” temperature difference (bottom) in the latitude range 22°N to 45°N. “Maximum sensitivity” approach and channel clustering applied as in Fig. 5.8c yields 403, 403 and 400 channels at 20°, 30° and 40° latitude, respectively.



**Figure 5.16:** Retrieved temperature field (top) and “retrieved-minus-true” temperature difference (bottom) in the latitude range 22°N to 45°N. “Maximum sensitivity” approach applied as in Fig. 5.8d yield 86 channels at 20°, 30° and 40° latitude.

## 5.4.2 Temperature Profile Retrieval Characteristics

In this section we concentrate on single vertical temperature profiles extracted out of the total field illustrated in Fig. 5.16 in order to discuss some characteristics of the retrieval results and the algorithm. In the following we employ the retrieval algorithm using 86 channels selected by the “maximum sensitivity” approach as described in Sect. 5.2.2.

### ITERATION AND CONVERGENCE

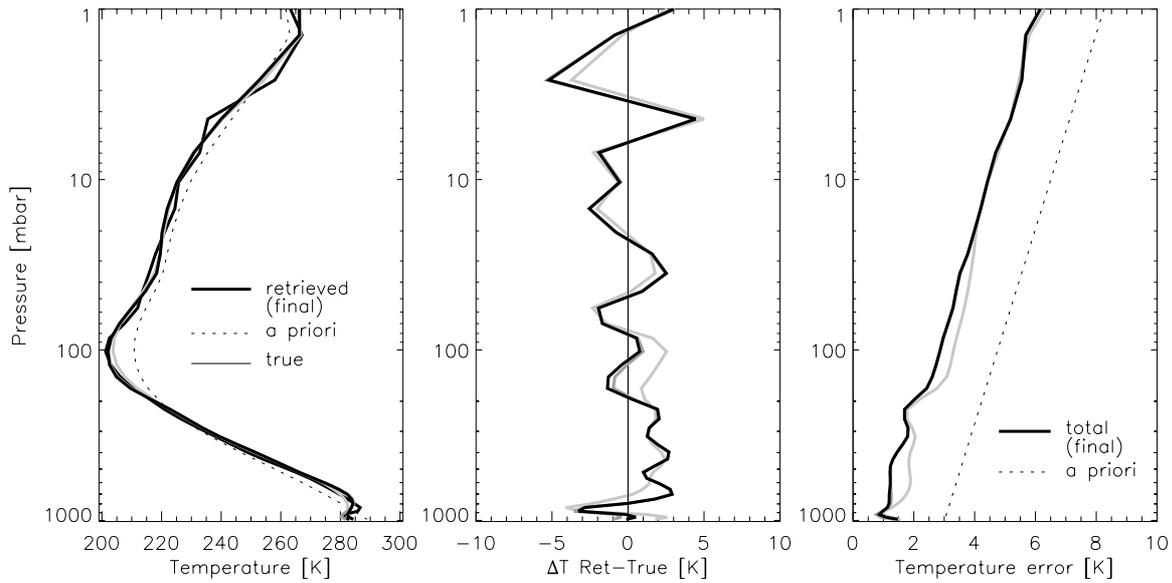
As mentioned in Sect. 5.3 we utilize the convergence criterion  $\chi^2 \leq m$  ( $m$  is the number of channels) where the cost function is given by

$$\chi^2 = [\mathbf{y} - \mathbf{y}_i]^T \mathbf{S}_\varepsilon^{-1} [\mathbf{y} - \mathbf{y}_i] + [\mathbf{x}_i - \mathbf{x}_{ap}]^T \mathbf{S}_{ap}^{-1} [\mathbf{x}_i - \mathbf{x}_{ap}]. \quad (5.4.3)$$

In the “northern range” we found that at most latitudes convergence is reached after one iteration, in two events three iterations were necessary. In the southern part we found 6 profiles, where the number of iterations equaled 3 or 4. In Fig. 5.17 we show the retrieval at latitude 31.7°S, which needed four iterations. The gray line indicates the first run, and the black line represents the final retrieval. We notice that the biggest step towards the “true” profile is accomplished in the first iteration. The following iterations do not improve the first iteration significantly. The fact, that the subsequent retrieval improves on the first step indicates that a purely linear approach would be insufficient. Nevertheless, the small number of necessary iterations suggests that the process is only modestly non-linear. The  $\chi^2$  values for that case are 140.9, 132.5, 112.6, and 75.5.

### ACCURACY AND RESOLUTION

Now we take a closer look at one typical vertical profile, extracted out of Fig. 5.16, namely at latitude 24.85°N. The retrieved temperature profile and associated features are shown in Fig. 5.18 and Fig. 5.19, respectively. The retrieved, *a priori* and “true” temperature profile are shown in Fig. 5.18a, where the pressure levels are indicated as horizontal gray lines for convenience. As mentioned before, temperatures at the tropopause and in some regions in the stratosphere are most difficult to retrieve. This can be seen in the Fig. 5.18, a and b. The large differences at the tropopause (~100 mbar) originate mainly from a too optimistic *a priori* error assumption (about 4.5 K at 100 mbar, see Fig. 5.2), whereas the actual *a priori* profile deviates from the “true” profile by almost 8 K. The  $\pm 1$  K limits (vertical dotted lines) in panel c should emphasize the IASI objective to retrieve temperature with an average accuracy of 1 K.

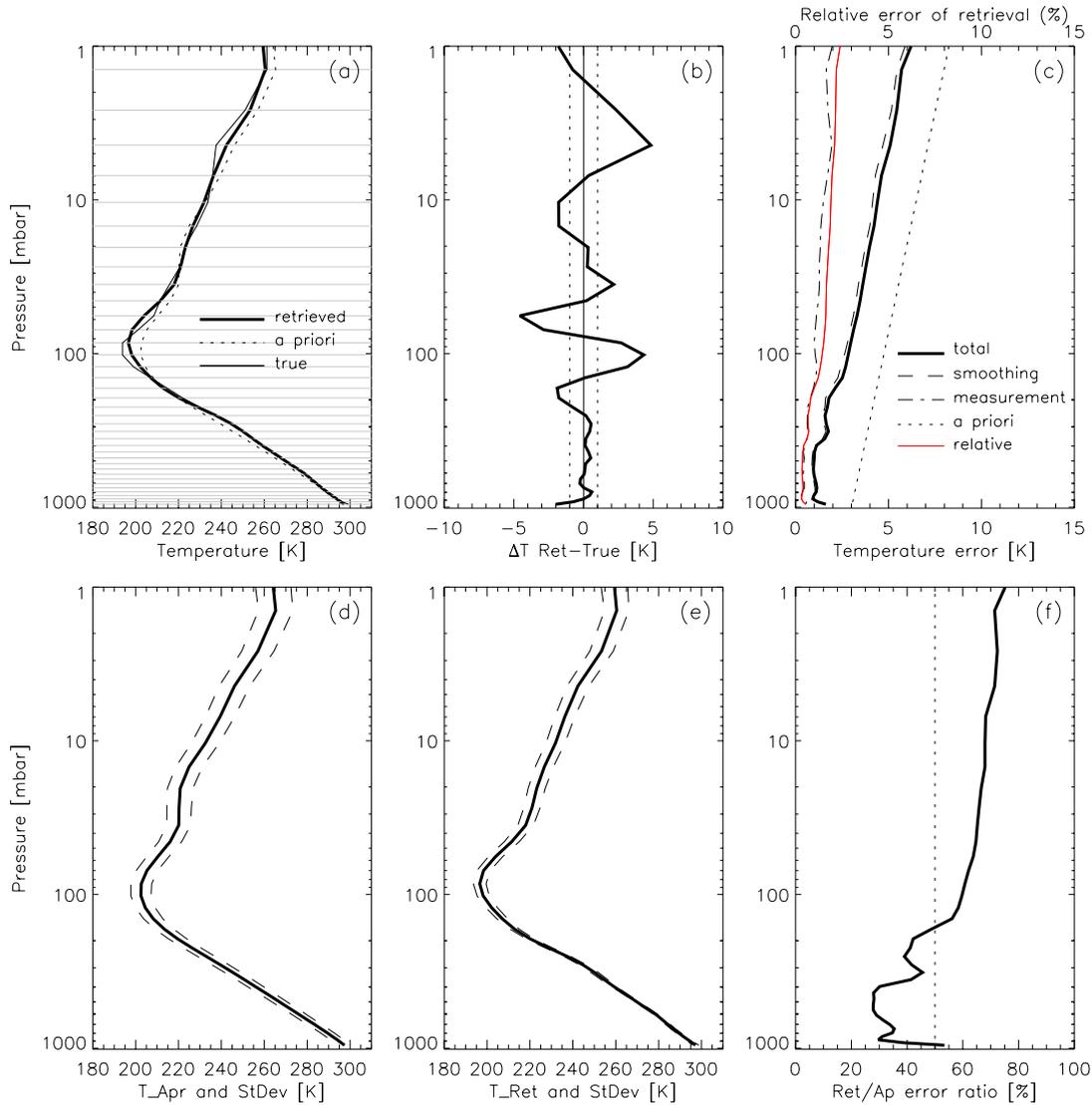


**Figure 5.17:** Retrieved temperature (black) for latitude 31.7°S after the first (light gray), second (gray), and fourth (black) iteration.

In this example tropospheric temperature retrievals meet the expectations. The expected total standard deviation (or *rms* error) of the retrieved profile (i.e., the square root of the diagonal elements of  $\hat{\mathbf{S}}$ ) depends on the shape of the weighting functions as well as on the *a priori* error profile (Fig. 5.18c). The retrieval error covariance matrix  $\hat{\mathbf{S}}$ , Eq. 5.4.2, can be decomposed as

$$\hat{\mathbf{S}} = (\mathbf{A} - \mathbf{I}_n) \mathbf{S}_{ap} (\mathbf{A} - \mathbf{I}_n)^T + \mathbf{G} \mathbf{S}_\epsilon \mathbf{G}^T, \quad (5.4.4)$$

where the iteration index has been dropped, and  $\mathbf{I}_n$  is the identity matrix of dimension  $n$ . The averaging kernel matrix  $\mathbf{A}$  is given by  $\mathbf{G}\mathbf{K}$  (see Sect. 4.3.1, Eq. 4.3.36), where the gain function matrix  $\mathbf{G}$  (Sect. 4.3.1, Eq. 4.3.32) is described by  $\hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\epsilon^{-1}$ . The first term on the right hand side is known as the smoothing error covariance matrix, while the second term is the measurement-based error covariance matrix. The total error (Fig. 5.18c) is mainly determined by the smoothing error (the square roots of the diagonal elements of the smoothing error covariance), which is largest in the stratosphere, where the retrieval is mostly specified by the *a priori* profile. The measurement-based error (the square root of the diagonal elements of the measurement-based error covariance) depends on the shape of  $\mathbf{G}$  (to be investigated below) and is small in regions where the weighting functions are small. Both quantities and their contribution to the total retrieval error will be discussed in more detail in the subsequent section.



**Figure 5.18:** Temperature profile and associated features for latitude 24.85°N. See text for explanation and discussion of panels (a) to (f).

Fig. 5.18d shows the assumed error of the *a priori* profile, while Fig. 5.18e visualizes the estimated retrieval error of the retrieved profile. The uncertainty in the retrieved profile decreases (starting from the initial guess, panel d), most significantly at tropospheric heights. The impact of the *a priori* error can be best viewed in Fig. 5.18f. The “retrieved-to-*a priori*” error ratio illustrates, by indication of the 50% limit (dotted line), that the error in  $\hat{\mathbf{x}}$  comes mainly from  $\mathbf{x}_{ap}$  at heights above 200 mbar; the measurements improve most (more than a factor of 2) upon the *priori* errors at tropospheric heights.

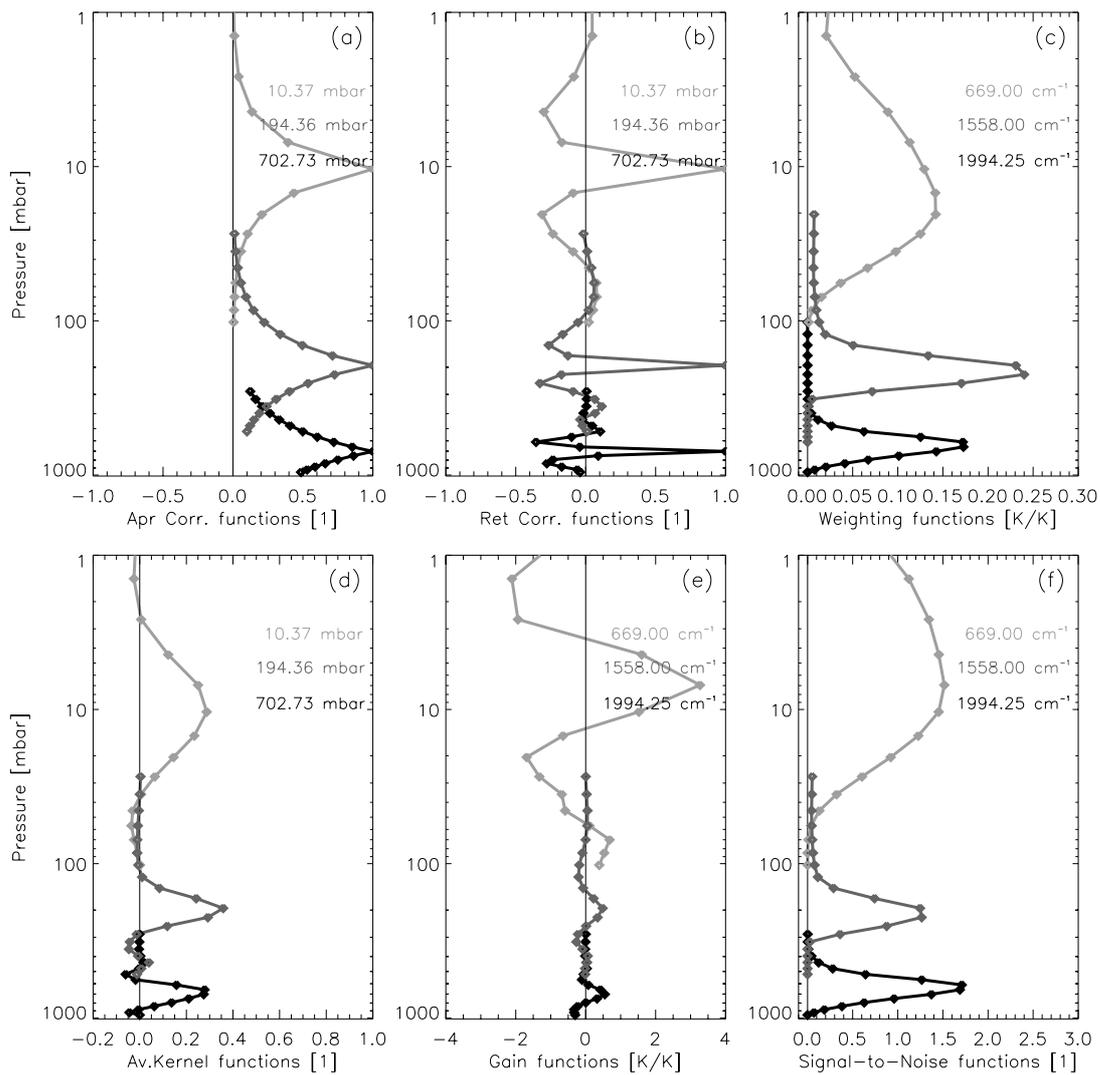


Figure 5.19: Temperature profile and associated features for latitude 24.85°N. See text for explanation and discussion of panels (a) to (f).

In Fig. 5.19a and Fig. 5.19b the correlation functions of the *a priori* error covariance matrix  $S_{ap}$  and of the retrieval error covariance matrix  $\hat{S}$  (given by Eq. 5.4.2) for three selected pressure levels (10.37, 194.36 and 702.73 mbar) are shown. These three levels are thought to be representative of the stratosphere, tropopause region, and troposphere, respectively. The square symbols indicate the pressure levels and remind on the use of the log-pressure coordinates. The correlation matrix is given by  $C_{ij} = S_{ij} / \sigma_i \sigma_j$ , where  $\sigma$  is the standard deviation of  $S$ ,  $i, j$  are level indices, and  $C_{ij}$  are the correlation coefficients, which are normalized measures of the correlation between neighboring levels. Since we assumed a covariance structure of  $S_{ap}$  with a linear increase from 3 K to about 8 K at 1 mbar and an exponential decrease in the non-diagonals with a

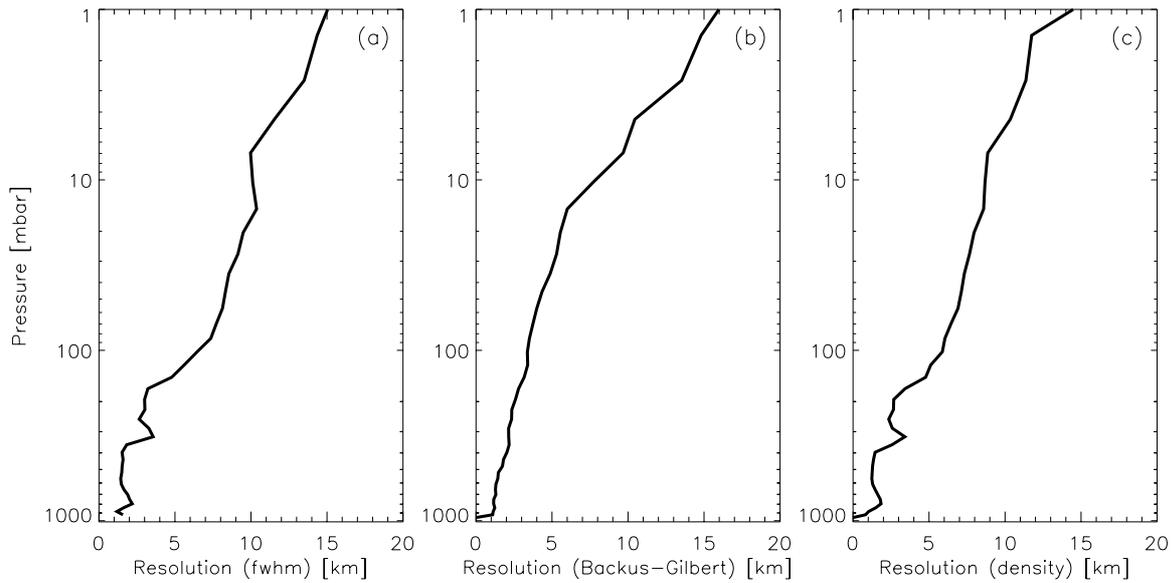
correlation length of 3 km (according Sect. 5.1.2), we find a similar but sharpened structure for the resulting retrieval error covariance matrix  $\hat{\mathbf{S}}$  (Fig. 5.19b). At heights where prior information is indispensable the shapes of the correlation functions are very similar (e.g., at 10.37 mbar). But at lower levels the contribution of the measurements becomes more important as reflected in the narrow correlation structure of the lower levels. The characteristic shape of the weighting functions can be seen in Fig. 5.19c, where we chose channels, which have their maximum value at the three selected pressure levels mentioned above; we investigate channels at  $669.0 \text{ cm}^{-1}$ ,  $1558.0 \text{ cm}^{-1}$ , and  $1994.25 \text{ cm}^{-1}$ . Each function indicates the contribution of the particular channel from different height levels to the radiance observed at the instrument. For example, measurements at  $1558.0 \text{ cm}^{-1}$  stem from tropospheric regions around 200 mbar. The atmospheric state in that region can be regarded as well-measured, whereas the upper stratospheric states are more difficult to observe by the instrument because of broad weighting functions. The rows of the matrix  $\mathbf{A} = \mathbf{G}\mathbf{K}$ , with  $\mathbf{G} = \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}$ , are referred to as the averaging kernel functions or just averaging kernels, displayed in Fig. 5.19d. Since the iterative algorithm, Eq. 5.4.1, is linear, we may consider

$$\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A})\mathbf{x}_{ap} + \mathbf{A}\mathbf{x} + \mathbf{G}\boldsymbol{\varepsilon} \quad (5.4.5)$$

(see Eq. 4.3.37), where  $\mathbf{x}$  is the true profile and  $\boldsymbol{\varepsilon}$  is the measurement error. Hence, the averaging kernels reflect which elements of the true state vector  $\mathbf{x}$  contribute to the retrieval value  $\hat{\mathbf{x}}$  at a certain height level. The averaging kernels may be used to define the vertical resolution of the instrument in various ways. The full width at half maximum (FWHM) of each averaging kernel can be regarded as a measure of vertical resolution at a particular height level. In the case considered, the resolution (in FWHM) is less than 3 km in most of the troposphere and increases to about 15 km at 1 mbar (Fig. 5.20a). The resolution  $r_i$  at level  $i$ , where  $\Delta z_j$  is the height interval at level  $j$ , is defined by Backus and Gilbert (1970), as

$$r_i = 12 \frac{\sum_j (z_i - z_j)^2 A_{ij}^2 / \Delta z_j}{\left(\sum_j A_{ij}\right)^2}. \quad (5.4.6)$$

In the case of negative sidelobes (as in Fig. 5.19d) we may substitute  $A_{ij}$  by  $|A_{ij}|$  in the denominator otherwise the calculated resolution would be too large (Collard, 1998). Another way is to express the vertical resolution as  $1/\rho_i$ , where the data density is defined as  $\rho_i = A_{ii} / \Delta z_i$  (Purser and Huang, 1993). Both methods have been applied to our retrieval and are shown in Fig. 5.20b and Fig. 5.20c yielding quite similar values as the FWHM, i.e., the resolution is 1–3 km below 200 mbar, and then decreases to about 15 km at 1 mbar.



**Figure 5.20:** Vertical resolution according to three measures. See text for explanation of panels (a) to (c).

The gain functions (i.e., columns of  $\mathbf{G} = \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}$ ) quantify how a particular measurement (channel) contributes to the retrieved state vector (Fig. 5.19e). In general we can say that these functions peak at the same altitude regions as the corresponding weighting functions and show large values, if the measurement error of the considered channel is small. Note that the large values in the stratosphere (i.e., for wavenumber  $669.0\text{ cm}^{-1}$ ) originate from the high values of the assumed *a priori* covariance, which appears in  $\hat{\mathbf{S}} = (\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1}$ . The  $\mathbf{S}_\varepsilon$  values (less than 1 K for most channels, see Fig. 5.1) are concealed by the large values of  $\mathbf{S}_{ap}$ . The functions of the signal-to-noise matrix  $\mathbf{S}_\varepsilon^{-\frac{1}{2}}\mathbf{K}\mathbf{S}_{ap}^{\frac{1}{2}}$ , calculated with diagonal  $\mathbf{S}_\varepsilon$  and  $\mathbf{S}_{ap}$ , reflect the relative influence of the measurement and the *a priori* uncertainties at different height levels (Fig. 5.19f). Their shape largely resembles that of the weighting functions (Fig. 5.19c) and their magnitude is determined by the *a priori* and retrieved standard deviations. The comparatively small value of channel  $1558.0\text{ cm}^{-1}$  results from a  $0.78\text{ K}$  measurement error, whereas the measurement error of channel  $1994.25\text{ cm}^{-1}$  is only  $0.34\text{ K}$ .

### RETRIEVAL ERROR

As previously mentioned the retrieval error covariance matrix can be considered as the sum of two factors according to Eq. 5.4.4. The smoothing error covariance matrix  $\mathbf{S}_s$  defined by

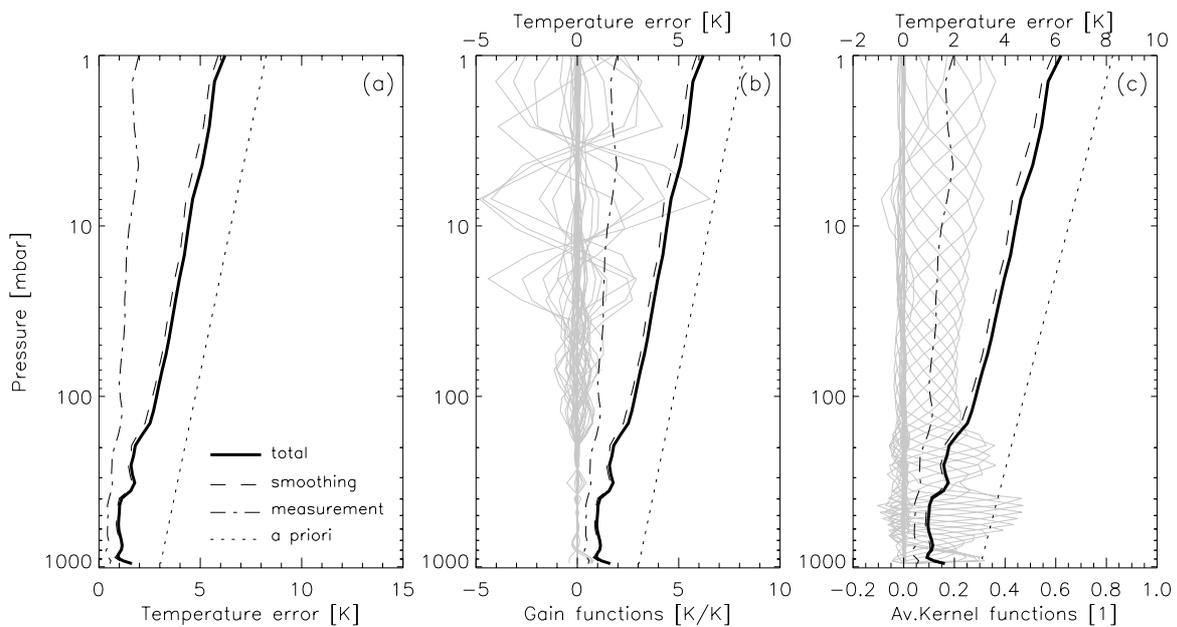
$$\mathbf{S}_s = (\mathbf{A} - \mathbf{I})\mathbf{S}_{ap}(\mathbf{A} - \mathbf{I})^T \quad (5.4.7)$$

reflects the impact of the *a priori* error on the retrieval error. The measurement-based error covariance matrix  $\mathbf{S}_m$ , given by

$$\mathbf{S}_m = \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}^T, \quad (5.4.8)$$

describes the impact of the measurement error on the retrieval error (Rodgers, 2000). The total retrieval standard deviations and its components (standard deviations of  $\mathbf{S}_s$  and  $\mathbf{S}_m$ ) are displayed in Fig. 5.21, and correspond to the retrieval illustrated in Fig. 5.18 and Fig. 5.19; Fig. 5.21a equals Fig. 5.18c. The measurement-based error (dash-dotted line) is determined by the shape of the gain functions (Fig. 5.21b), which are large in the stratosphere due to large  $\mathbf{S}_{ap}$ . In the troposphere,  $\mathbf{G}$  is primarily shaped by  $\mathbf{S}_\varepsilon$  and  $\mathbf{K}$ , and the measurement-based error  $\mathbf{S}_m$  is large when  $\mathbf{G}$  is large, i.e., the retrieved vector is mainly determined by the measurements and not by the *a priori* vector. That can be best seen if we rewrite Eq. 5.4.5 as  $\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A})\mathbf{x}_{ap} + \mathbf{D}\mathbf{y}$ , since  $\mathbf{y} = \mathbf{K}\mathbf{x} + \varepsilon$ . If the true vector  $\mathbf{x}$  is reproduced well by the measurements then the contribution of the measurement-based error to the total retrieval error is large.

The smoothing error (dashed line) is large in regions where  $\mathbf{A}$  is small (i.e.,  $\mathbf{A} - \mathbf{I}$  is large, Eq. 5.4.7) and  $\mathbf{S}_{ap}$  is large. Hence, in the stratosphere the retrieval error is primarily determined by the *a priori* error and the smoothing error tends to the *a priori* error profile (Fig. 5.21c).



**Figure 5.21:** Retrieval error and associated features. See text for explanation and discussion of panels (a) to (c).

COMPARISON OF FINE AND COARSE RETRIEVAL GRID

We investigate the impact of the chosen pressure grid. So far we employed 43 pressure levels. In Fig. 5.22 we show the retrieval (after the first step, without further iteration, 43-level case is already converged) calculated on every second pressure level leading to 22 levels as indicated by horizontal gray lines in Fig. 5.22a. The inputs and assumptions are the same as in the retrieval described in Fig. 5.18, to which Fig. 5.22 may be directly compared. Figure 5.23 allows a direct comparison of the 43-level (black line) retrieval to the 22-level (gray line) retrieval. We employ the first run, that is  $\hat{\mathbf{x}} = \mathbf{x}_{ap} + \mathbf{S}\mathbf{K}^T\mathbf{S}_\epsilon^{-1}[\mathbf{y} - \mathbf{y}_{ap}]$  derived from Eq. 5.4.1. The retrieval from 22 levels (Fig. 5.23, a and b) shows larger deviations from the truth in the troposphere but smaller errors at the tropopause. These two regions are represented by two levels, namely at 102 mbar and at 396 mbar, which will be investigated in more detail in Fig. 5.24 and Fig. 5.25, respectively.

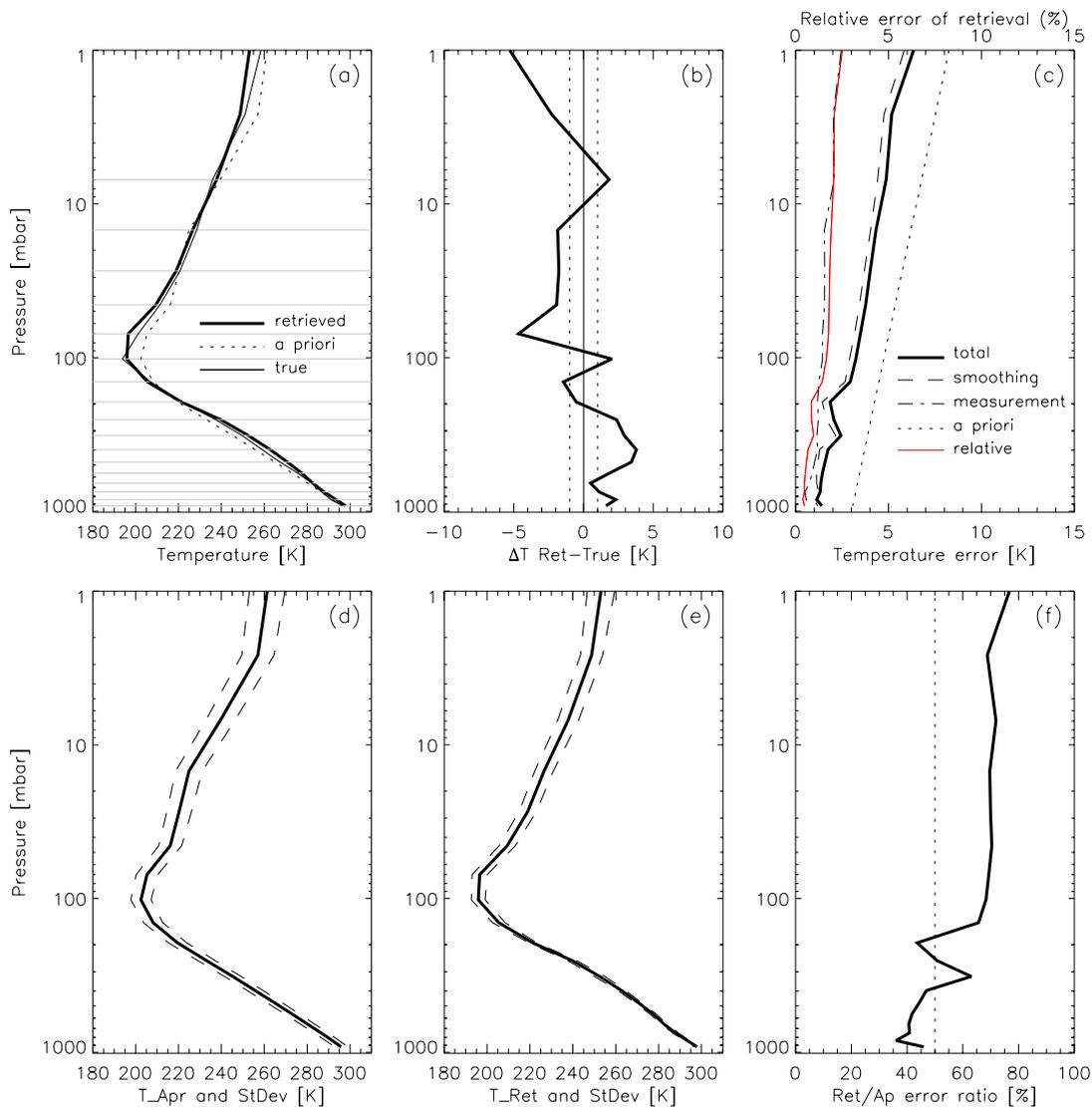
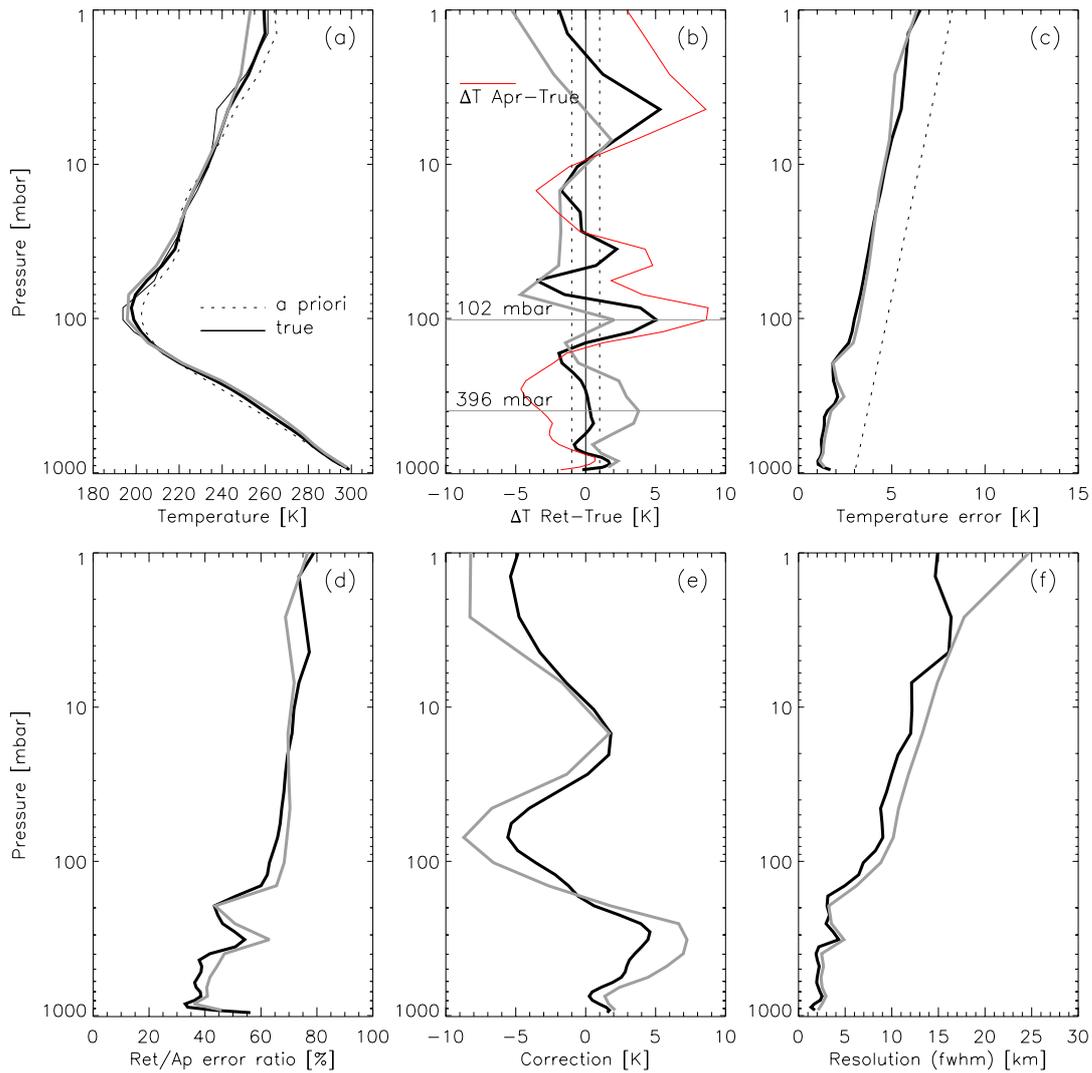


Figure 5.22: Temperature Profile and associated features for latitude 24.85°N on a coarser grid (22 levels).



**Figure 5.23:** Gray line: 22-level retrieval, black line: 43-level retrieval. See text for explanation and discussion of panels (a) to (f).

The retrieval error in Fig. 5.23c is quite similar for both cases, as well as the “retrieved-to-*a priori*” error ratio (panel d), although the retrieval at 22 levels is slightly degraded at most heights. Fig. 5.23e depicts the correction vector  $\mathbf{G}(\mathbf{y} - \mathbf{y}_{ap})$ , where different  $\mathbf{G} = \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}$  lead to the different retrieval results for the two cases as shown in panels a and b.

The resolution, defined as the full width at half maximum (FWHM) of the averaging kernels is displayed in Fig. 5.23f; it is worse for the coarser grid as expected. The difference in the retrieval originates from  $\mathbf{G}(\mathbf{y} - \mathbf{y}_{ap})$  as shown in (Fig. 5.23e), more precisely speaking, from the gain function matrix  $\mathbf{G}$  alone, since the difference “true-minus-*a priori* measurement vector” is the same for both cases. To find the main cause we investigate the components of  $\mathbf{G}$  at the tropopause

pressure level 102 mbar (Fig. 5.24) and the tropospheric level 396 mbar (Fig. 5.25). Recall that the gain function matrix is given by

$$\begin{aligned} \mathbf{G} &= \hat{\mathbf{S}}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1} \\ &= (\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}\mathbf{K} + \mathbf{S}_{ap}^{-1})^{-1}\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}. \end{aligned} \tag{5.4.9}$$

The factor  $\mathbf{K}^T\mathbf{S}_\varepsilon^{-1}$  in  $\mathbf{G}$  was found to be very similar for the 22-level and the 43-level retrieval. Note that the  $\mathbf{K}$  values differ slightly in that the 22-level case shows smaller values in most regions (Fig. 5.24c, Fig. 5.25c).

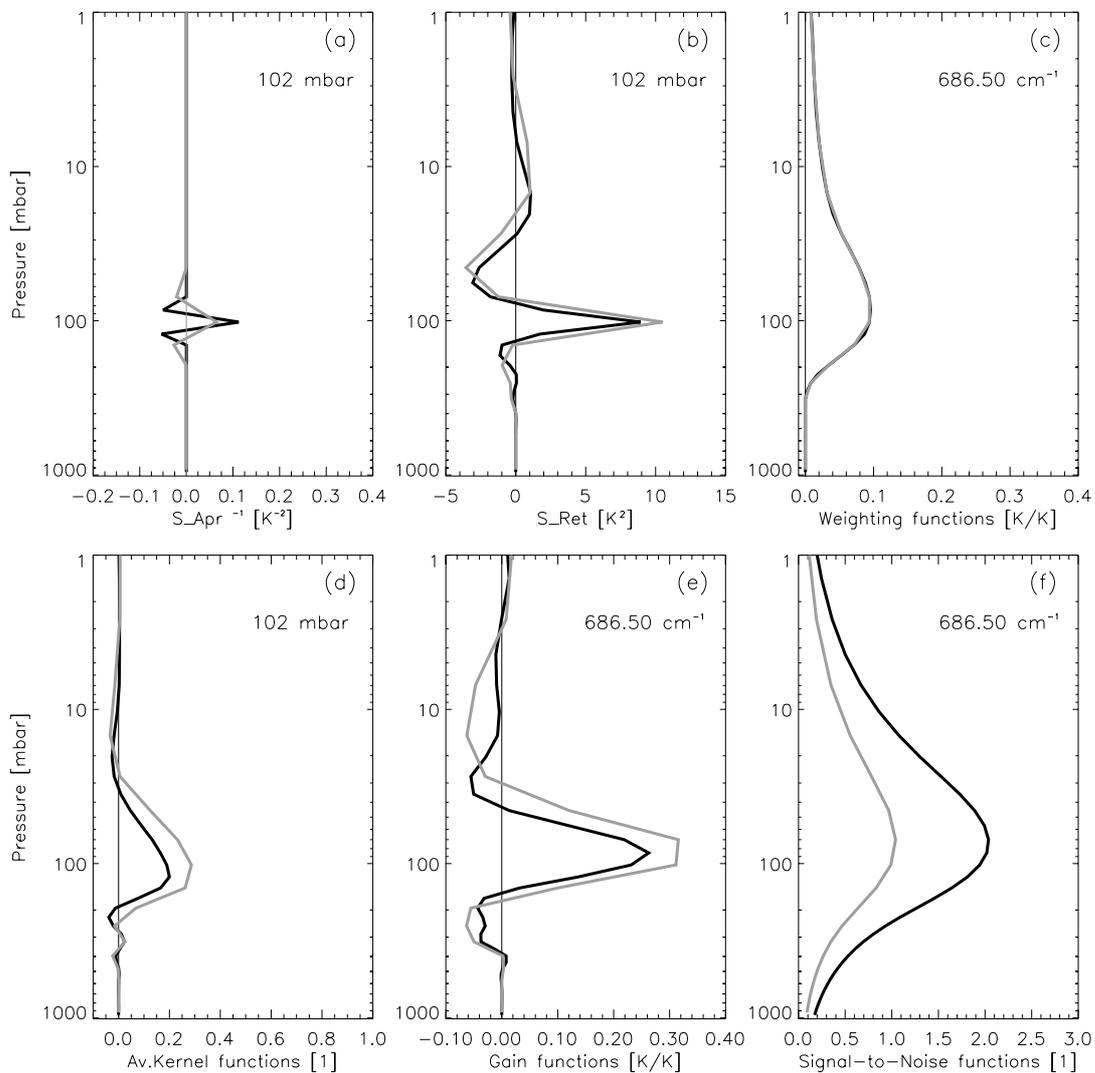
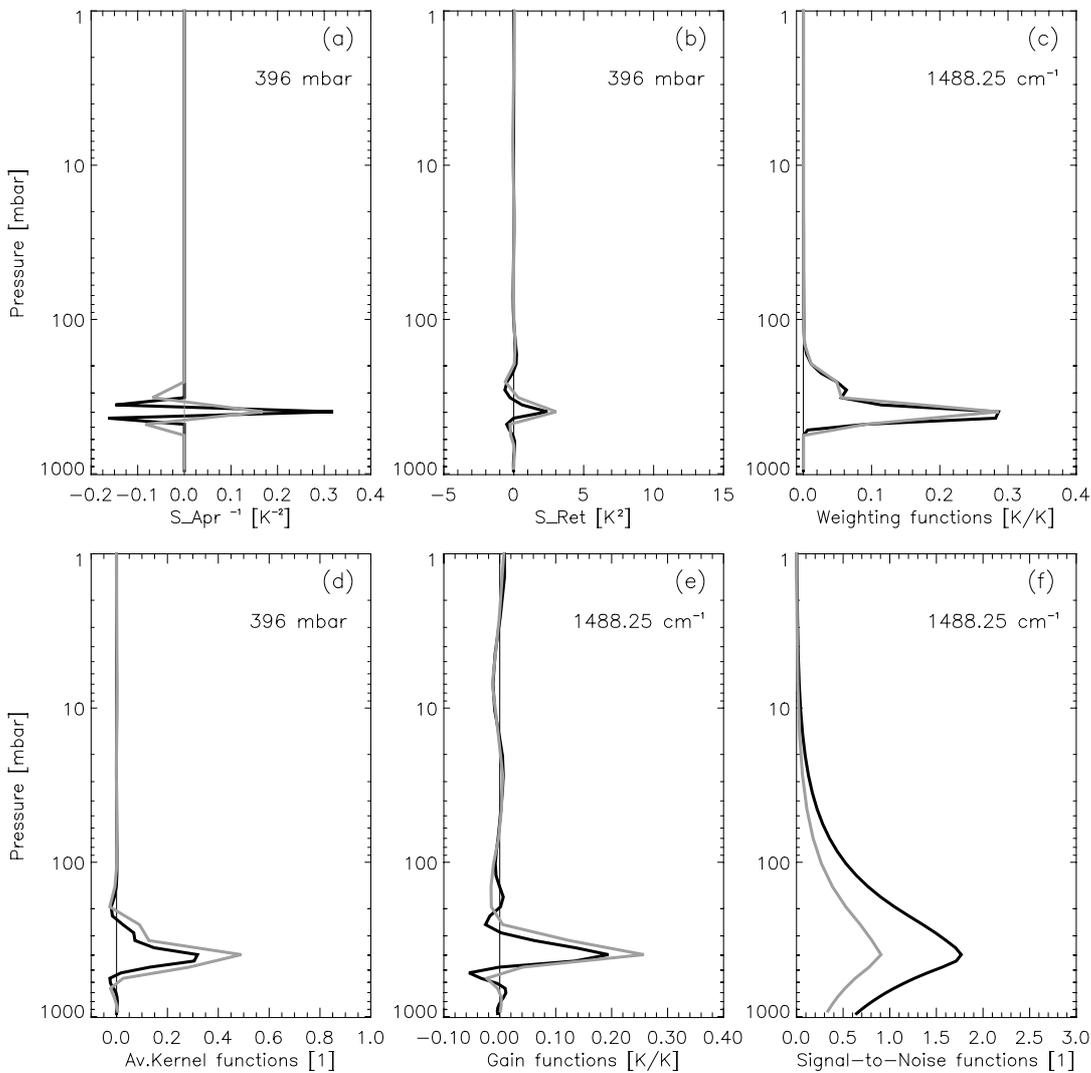


Figure 5.24: Gray line: 22-level retrieval, black line: 43-level retrieval. The channel at 686.5 cm<sup>-1</sup> peaks at the 102 mbar level. See text for explanation and discussion of panels (a) to (f).

The retrieval error covariance matrix  $\hat{\mathbf{S}}$  (Fig. 5.24b, Fig. 5.25b) shows more differences between the gray and black line. Panel a shows the inverse *a priori* error covariance diagonal elements. For

both pressure levels the 22-level retrieval shows smaller values, but a broader shape. The matrix  $\hat{\mathbf{S}}^{-1}$  is primarily shaped by  $\mathbf{S}_{ap}^{-1}$ , and only to a lesser extent by the factor  $\mathbf{K}^T \mathbf{S}_\varepsilon^{-1} \mathbf{K}$ . Note that inversion of matrices of different dimension but equal values at the same levels yields different results. That is the reason why  $\mathbf{S}_{ap}^{-1}$  has different values, which influence  $\hat{\mathbf{S}}$ . We have large  $\hat{\mathbf{S}}$ , and hence large  $\mathbf{G}$  (panels e), due to smaller  $\mathbf{S}_{ap}^{-1}$  values in the 22-level retrieval. This fact also causes larger averaging kernel (panels d), but smaller signal-to-noise ratios (panels f) according to the matrix  $\mathbf{S}_\varepsilon^{-\frac{1}{2}} \mathbf{K} \mathbf{S}_{ap}^{\frac{1}{2}}$ . The latter originates in the somewhat different  $\mathbf{S}_{ap}$  structure calculated for the coarser grid.



**Figure 5.25:** Gray line: 22-level retrieval, black line: 43-level retrieval. The channel at 1488.25 cm<sup>-1</sup> peaks at the 396 mbar level. See text for explanation and discussion of panels (a) to (f).

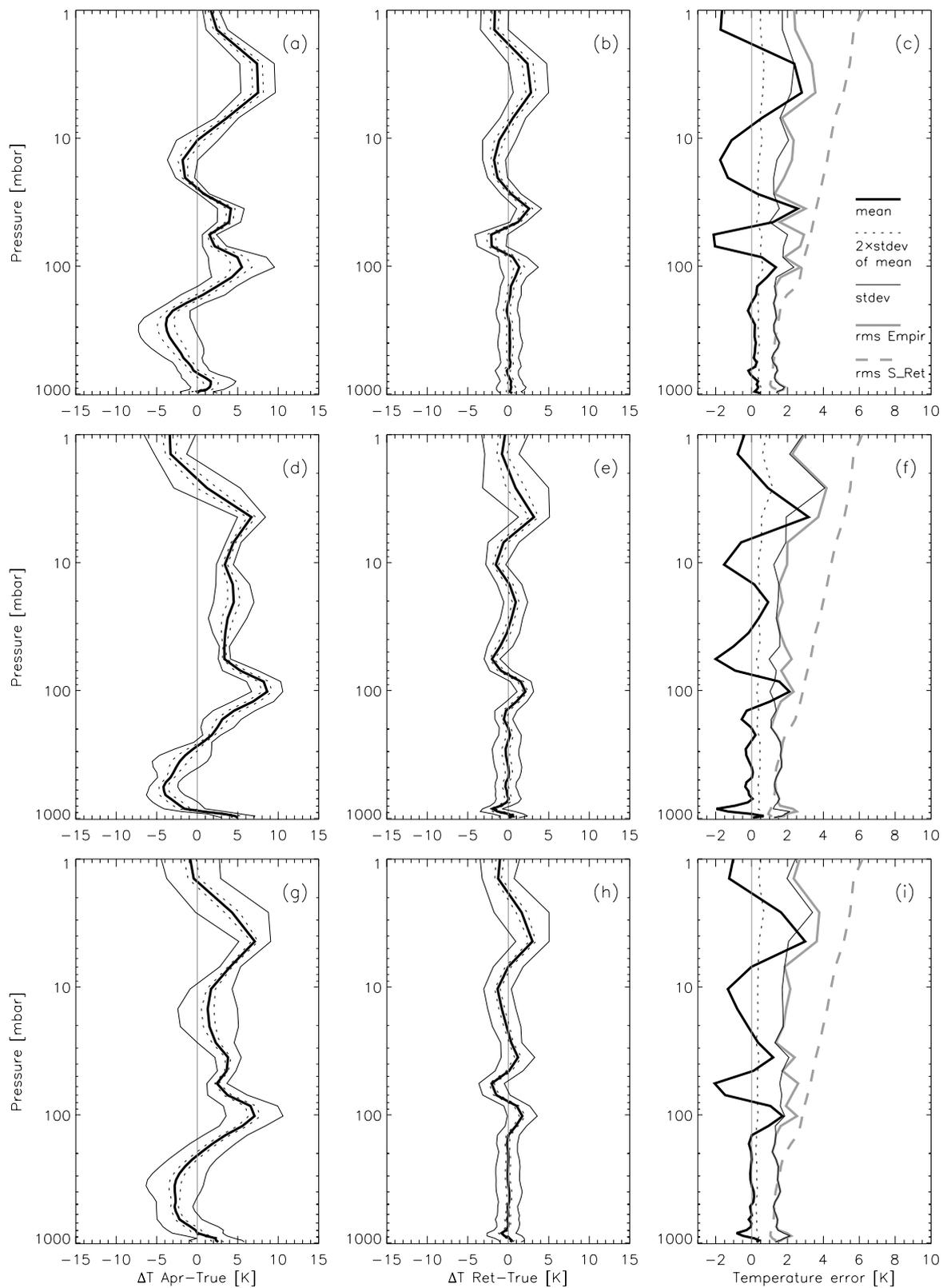
If we compare the retrieval results calculated at 22 levels at the two pressure levels (see again Fig. 5.23), we have to consider that the gain function for the particular level comprises all wavenumbers and is multiplied with the residual vector  $\mathbf{y} - \mathbf{y}_{ap}$ . The residual vector depends on the difference  $\mathbf{x} - \mathbf{x}_{ap}$ , which is positive in the troposphere but negative at the tropopause region. This results in positive values of the correction vector in the troposphere, but in negative values at the tropopause (Fig. 5.23e). In any case the retrieved 22-level profiles move further away from the *a priori* profile (Fig. 5.23b) because the measurements are trusted more due to large  $\mathbf{G}$  values.

In summary, in order to utilize the potential of the instrument to sound the atmosphere at high resolution, which is the case in particular in the troposphere (where many weighting functions have narrow peaks), the grid spacing has to be of an adequate dimension. Otherwise the retrieval performance degenerates in that region as illustrated in Fig. 5.23.

### 5.4.3 Empirical Error Analysis

As discussed in Sect. 5.1.2 we utilized *a priori* profiles taken from a climatological background field. The associated covariance matrix  $\mathbf{S}_{ap}$  has been obtained relatively *ad hoc* in a simplified manner, since the actual errors of the *a priori* profiles (i.e., deviation from the “true” profile) are not known. How does this “simplification” affect the quality of the retrieval results? Fig. 5.26 depicts the statistical errors of an ensemble consisting of 49 profiles of the northern latitude range (panels a-c), 49 profiles of the southern range (panels d-f) and the total 98 profiles consisting of both the northern and southern part (panels g-i). The latitude ranges are the ones indicated in Fig. 5.11, i.e., the same ones as exploited in Sect. 5.4.1.

The mean (or bias) of an ensemble is indicated as a heavy black line and the dotted lines display the doubled standard deviation of the mean, i.e., the mean lies with 95% certainty within the dotted limits. The thin black lines directly illustrate the standard deviation (*stdev*) of an ensemble. The dashed heavy gray line in the right most panels shows the average of the retrieval standard deviation, as theoretically estimated by the inversion algorithm, Eq. 5.4.2, while the solid heavy gray line is the empirical *rms* retrieval error ( $rms^2 = bias^2 + stdev^2$ ).

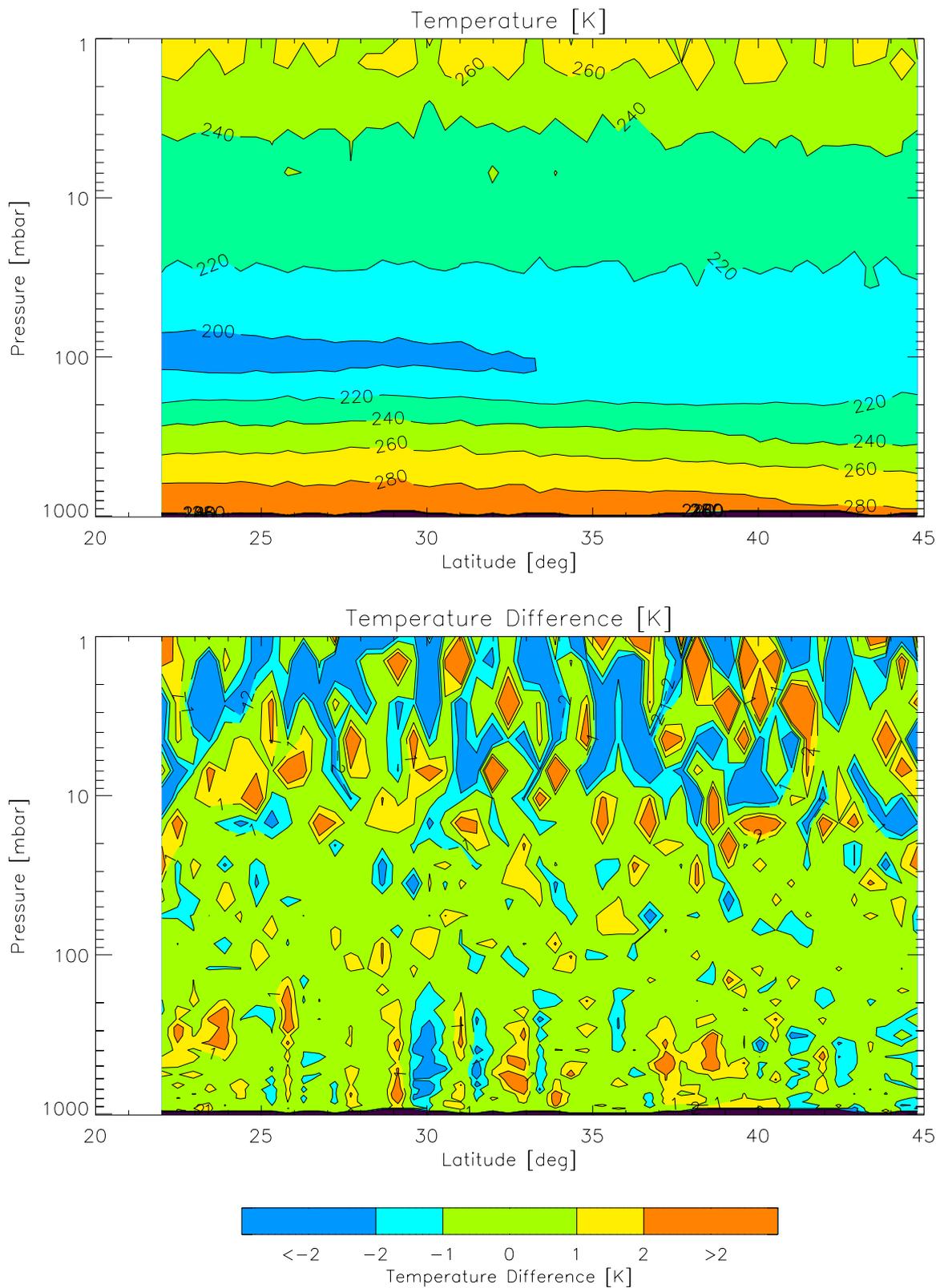


**Figure 5.26:** Empirical error analysis using an ensemble of (a-c) 49 profiles (“northern latitude range”), (d-f) 49 profiles (“southern latitude range”) and (g-i) 98 profiles (northern and southern range). *A priori* profiles were taken from the CIRA86aQ climatological model. See text for description of panels (a) to (i).

Fig. 5.26a indicates a quite strong bias of the ensemble “*a priori*-minus-true”. Recall that the *a priori* profiles and “true” profiles were taken from the CIRA86aQ model and the ECMWF analysis field, respectively. This *a priori* bias leads to a significant bias of the “retrieved-minus-true” in the stratosphere (Fig. 5.26b). All errors are summarized in Fig. 5.26c. In the troposphere the *rms* retrieval error (gray) equals the standard deviation (thin black) due to small bias (heavy black), whereas above about 150 mbar the *rms* retrieval error is additionally determined by the bias. The theoretically estimated *rms* retrieval error (dashed gray), while incidentally close to the empirical *rms* retrieval error (gray) in the troposphere, diverges in the stratosphere in that the former is overly conservative. The discrepancies in the stratosphere stem from the biasedness of  $\mathbf{x}_{ap}$  and the over-simplification of  $\mathbf{S}_{ap}$  and thus from inconsistencies in the inversion process. Fig. 5.26 (d-f) for the southern part shows similar results as Fig. 5.26 (a-c) (with different bias structure), which, in turn, in large parts determines the result depicted in Fig. 5.26 (g-i) for the northern and southern part together.

We summarize that in general the absence of consistency between *a priori* profiles and the associated covariance matrix can introduce significant retrieval errors, in particular in the stratosphere.

Now, for verification purposes, we utilize *a priori* profiles, which are consistent with the assumed  $\mathbf{S}_{ap}$  by construction, following Eq. 5.1.3 of Sect. 5.1.2. We test them as alternatives to the climatological ones. Running this test first for the “northern slice”, we obtain temperature retrievals as displayed in Fig. 5.27. These results may be directly compared to Fig. 5.16, since we use the same channels, and the same matrices  $\mathbf{S}_{ap}$  and  $\mathbf{S}_e$ . The differences are quite large, in particular in the tropopause region where the present case exhibits almost no errors exceeding  $\pm 1$  K. Recall that in the example described in Fig. 5.26 the *a priori* bias is most obvious in this regions, which affects the “retrieved-minus-true” difference quite a bit. Performing the same empirical error analysis as illustrated in Fig. 5.26 for the case with consistent *a priori* profiles we obtain the results illustrated in Fig. 5.28. The bias is mostly gone (i.e.,  $rms = stdev$ ) and the overlap of the two *rms* error profiles (empirical and theoretically estimated) is rather close (Fig. 5.27, panels c, f, i). The smaller values of the theoretically estimated *rms* error, compared to the empirical *rms* error, from surface to about 200 mbar originate from the fixed prescription of humidity profile realizations with 15% assumed uncertainty; the theoretically estimate cannot “know” this uncertainty in temperature-only retrievals.



**Figure 5.27:** Retrieved temperature field (top) and “retrieved-minus-true” temperature difference (bottom) in the northern latitude range 22°N to 45°N.

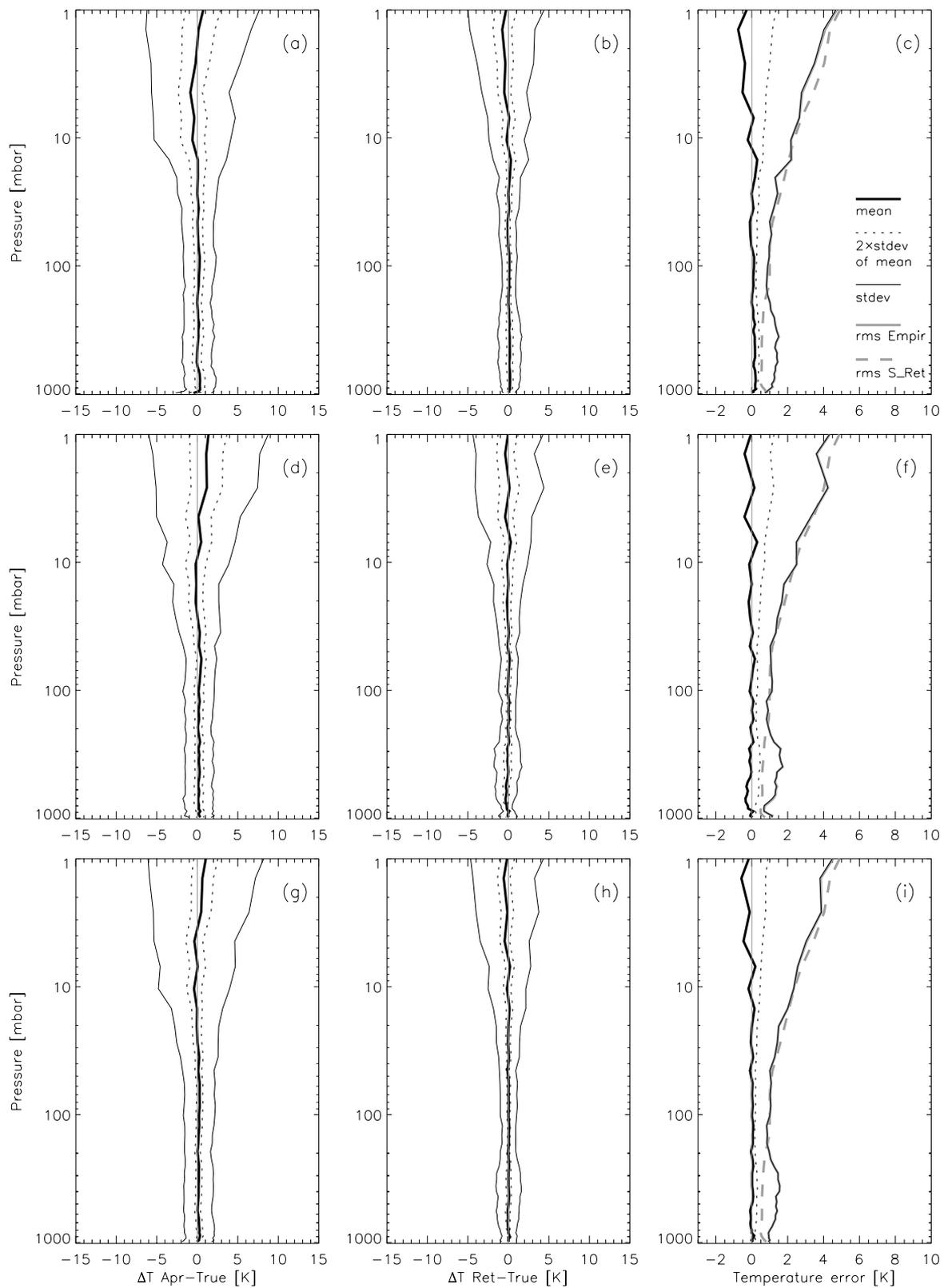


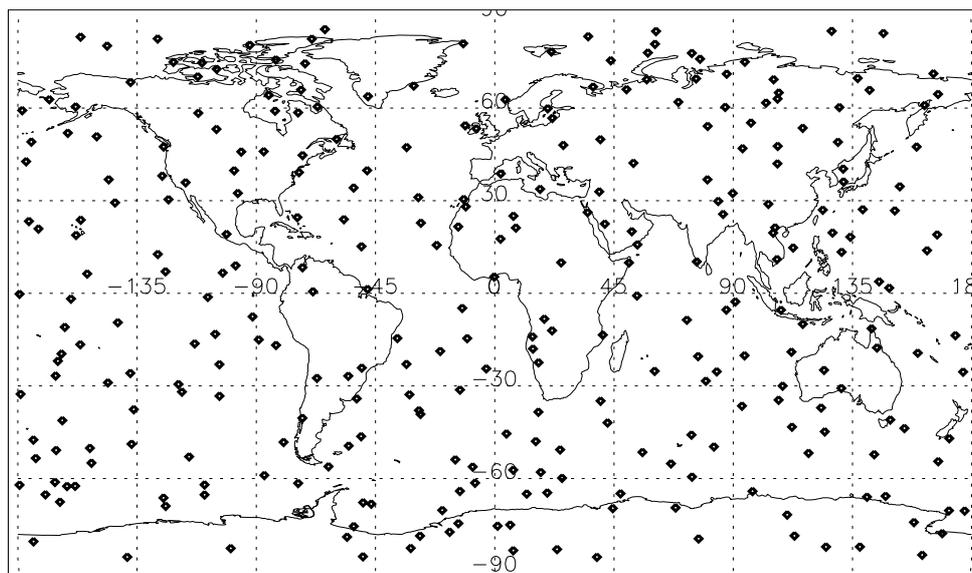
Figure 5.28: Empirical error analysis using an ensemble of (a–c) 49 profiles (‘‘northern latitude range’’), (d–f) 49 profiles (‘‘southern latitude range’’) and (g–i) 98 profiles (northern and southern range). The *a priori* profiles used are consistent with the *a priori* error covariance matrix since constructed via Eq. 5.1.3 in Sect. 5.1.2.

The first main conclusion which may be drawn from an empirical analysis as described by Fig. 5.26 and Fig. 5.28 is the necessity of a proper choice of a suitable sample of *a priori* profiles. These profiles should be statistically unbiased against the true ones and as close to the true profile as possible. From Eq. 5.4.5 we find the sensitivity of the retrieval to an incorrect *a priori* state as  $\partial\hat{\mathbf{x}}/\partial\mathbf{x}_{ap} = \mathbf{I}_n - \mathbf{A}$ . Hence, a large averaging kernel yields small  $(\mathbf{I}_n - \mathbf{A})$  values implying that the retrieved state is only to a small extent determined by the *a priori* state in this case, as discussed in Sect. 5.4.2. If this is not the case, such as for the stratosphere in this context, the retrieval starts to become vulnerable to the inconsistencies discussed.

The second main conclusion is that the corresponding covariance matrix  $\mathbf{S}_{ap}$  should be statistically consistent with the ensemble of *a priori* profiles used. For many applications, error covariance matrices can be constructed. Sometimes climatological models are used to calculate a large number of temperature profiles for various atmospheric conditions and the covariance is evaluated. For example, ECMWF can provide such background error covariance matrices to be used as *a priori* error covariance matrices. The background error covariance based on short-term weather forecast errors constitutes another possibility to create a suitable  $\mathbf{S}_{ap}$  matrix. In general, the latter background field is more adequate than the field solely based on climatological variations. Nevertheless, both background fields may yield similar retrieval results as described by Collard (1998). In the “model world” studies, once  $\mathbf{S}_{ap}$  has been constructed, associated prior profiles can be evaluated by decomposing the covariance in eigenvectors and adding the error described by  $\boldsymbol{\varepsilon}_x = \sum_j a_j \mathbf{e}_j$  (see Eq. 4.1.3) to a “true” profile. This can be useful to properly assess the performance limits of a retrieval scheme. Unfortunately, there is no such handy prerequisite as a “true” profile in the “real world”; performance assessment based on actual data requires careful and elaborated validation.

## 5.5 Retrievals for a Globally Distributed Ensemble

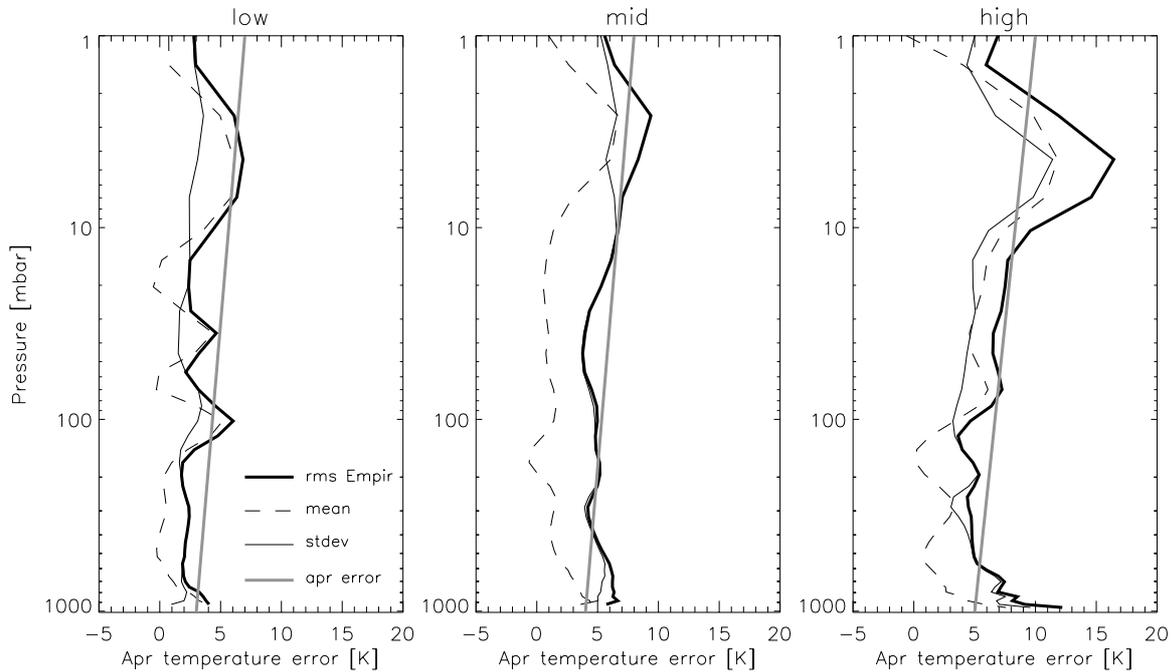
In this section we consider 300 events globally distributed as illustrated in Fig. 5.29. The global range can be divided into low ( $< \pm 30^\circ$ ), mid ( $\pm 30^\circ$  to  $\pm 60^\circ$ ) and high ( $> \pm 60^\circ$ ) latitudes. The temperature, humidity, and pressure profiles at these points have been calculated based on the same ECMWF analysis field as used above. The locations can be regarded as event locations of GNSS (Global Navigation Satellite System) radio occultation measurements. Briefly, in the latter method a LEO satellite uses GNSS signals to measure the angle through which the signal is refracted when passing tangentially through the atmosphere. The refractivity gradient, determining the angle, depends on the gradient of density, temperature, and humidity. The lowest height down to which temperature and humidity profiles are obtained by the occultation technique changes due to various factors, such as humidity variability. In our case all observations reach down to at least 500 mbar.



**Figure 5.29:** 300 profile locations with global distribution. The locations represent a selected subset of GNSS radio occultation profile locations.

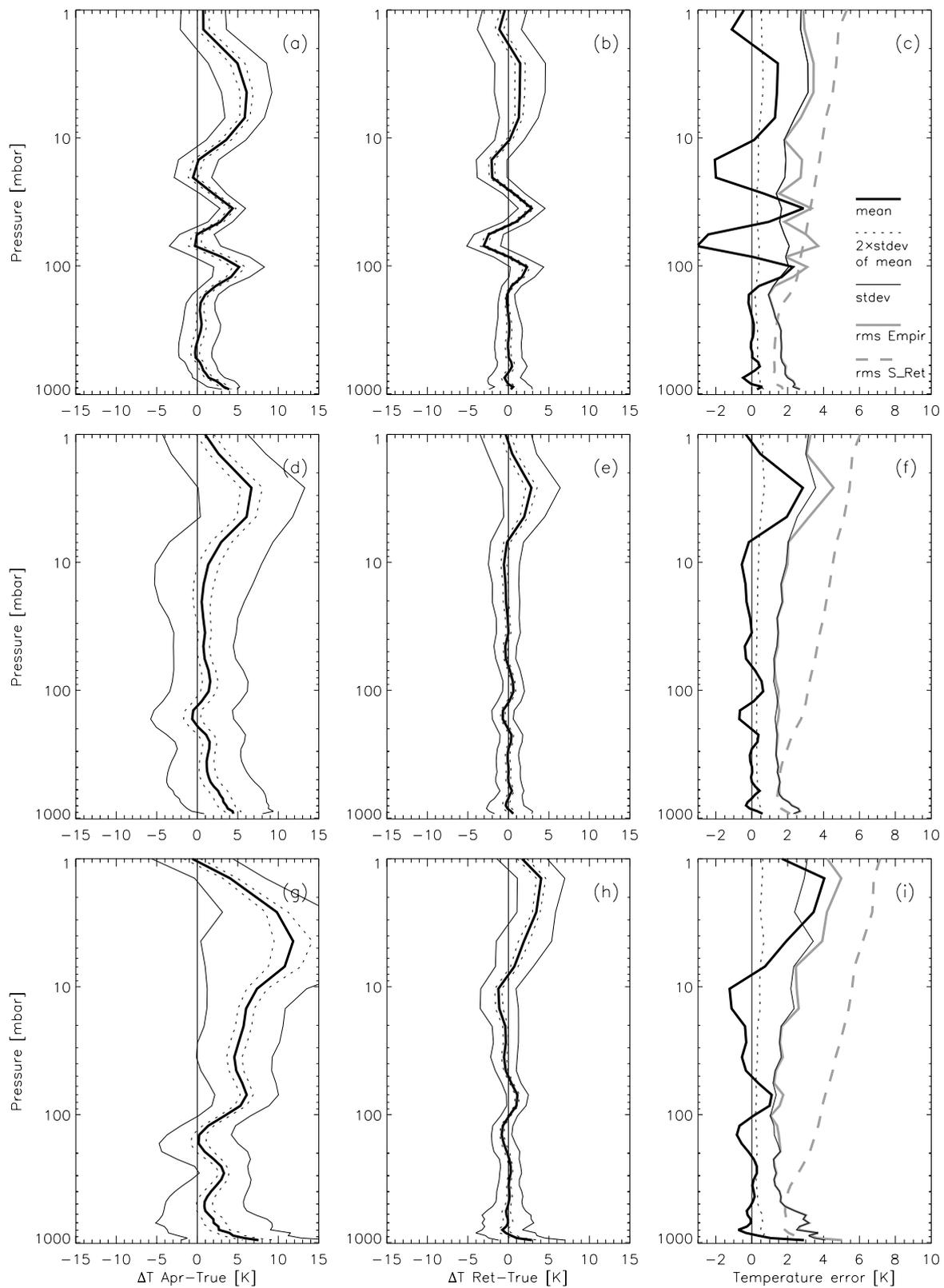
We estimate temperature profiles as described in Sect. 5.4 by employing CIRA86aQ *a priori* profiles, interpolated to the specific latitude. We present the corresponding empirical analysis (in the same way as done above in Sect. 5.4.3) for the low, mid, and high latitude bands, as well as for the total field. The *a priori* errors were determined by investigating the “*a priori*-minus-true” differences of each particular ensemble of 100 profiles per latitude range (Fig. 5.30). The straight gray lines present the square roots (standard deviation) of the finally assumed values for the

diagonal elements of  $S_{ap}$ . Correlation between levels were modeled, as earlier, by an exponential drop-off with a correlation length of 3 km.

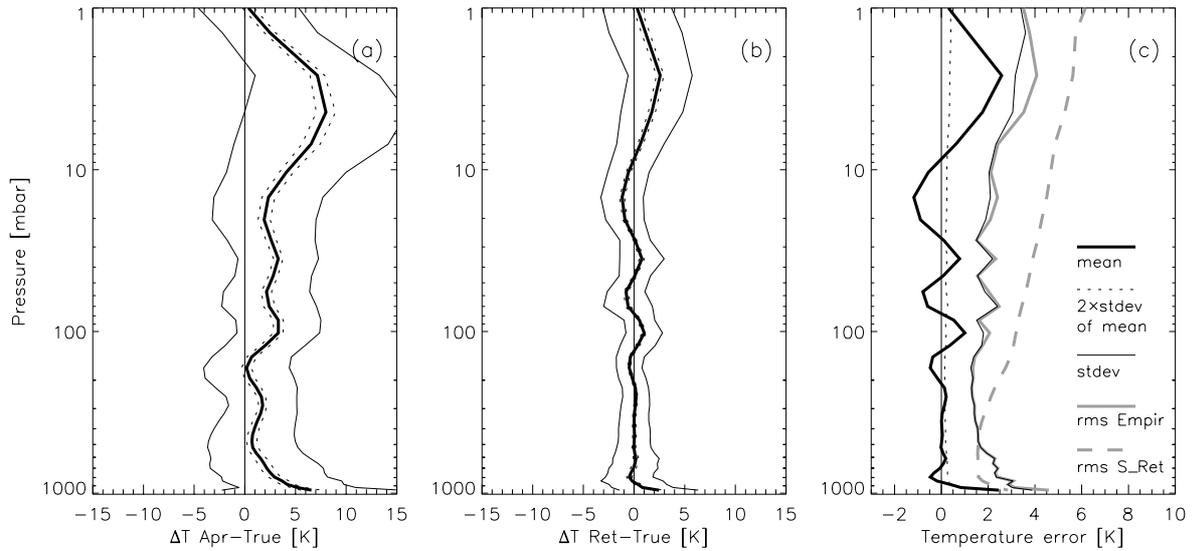


**Figure 5.30:** *A priori* error estimation for low, mid, and high latitudes.

The empirical error analysis results when using the CIRA86aQ *a priori* profiles are shown in Fig. 5.31 (in the same format as in Fig. 5.26). In Fig. 5.31, the bias (heavy black) of the “*a priori*-minus-true” difference profiles (panels a, d, g) is again quite pronounced. Almost all *a priori* (CIRA86aQ) profiles are warmer than the “true” (ECMWF analysis) data. The low latitude results (panels a-c) exhibit the smallest bias and standard deviation range, but show sharp features at the tropopause, which leads to the distinct shape of the empirical *rms retrieval* error (heavy gray line, panel c). From surface to about 40 mbar the theoretically estimated *rms retrieval* error (dashed heavy gray line) matches the empirical error roughly. Apparently the assumed *a priori* error profile (Fig. 5.30a), which is the main contributor to the retrieval error, emerged to be a relatively reasonable first guess. The mid and high latitude results (Fig. 5.31 d-f, g-i) comprise profiles of a large variety as indicated by the large standard deviation range of the “*a priori*-minus-true” difference profiles. This affects the “retrieved-minus-true” difference primarily in the boundary layer and in the upper stratosphere. The empirical *rms retrieval* error (gray) is mostly determined by the standard deviation, particular for the mid latitude case (Fig. 5.31f). The theoretically estimated *rms retrieval* error (dashed gray) is again a too conservative guess, especially at heights above 200 mbar.



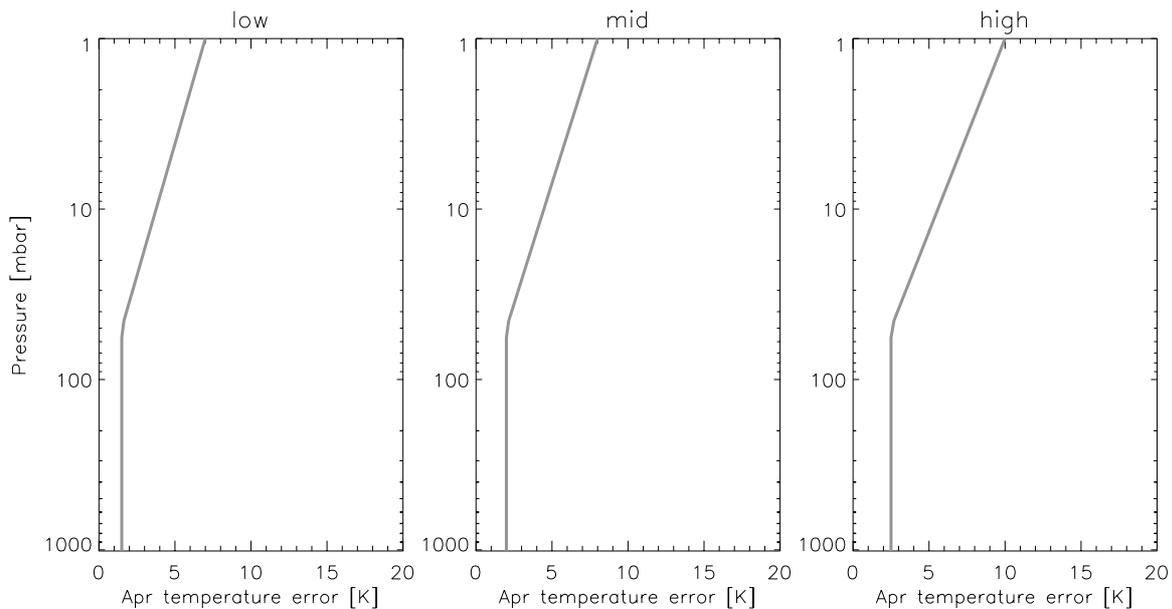
**Figure 5.31:** Empirical analysis for low (a-c), mid (d-f), and high (g-i) latitudes, each with 100 profiles. See text for discussion of all panels. *A priori* profiles were taken from the CIRA86aQ model.



**Figure 5.32:** Empirical error analysis for low, mid and high latitudes together (300 profiles). See text for description.

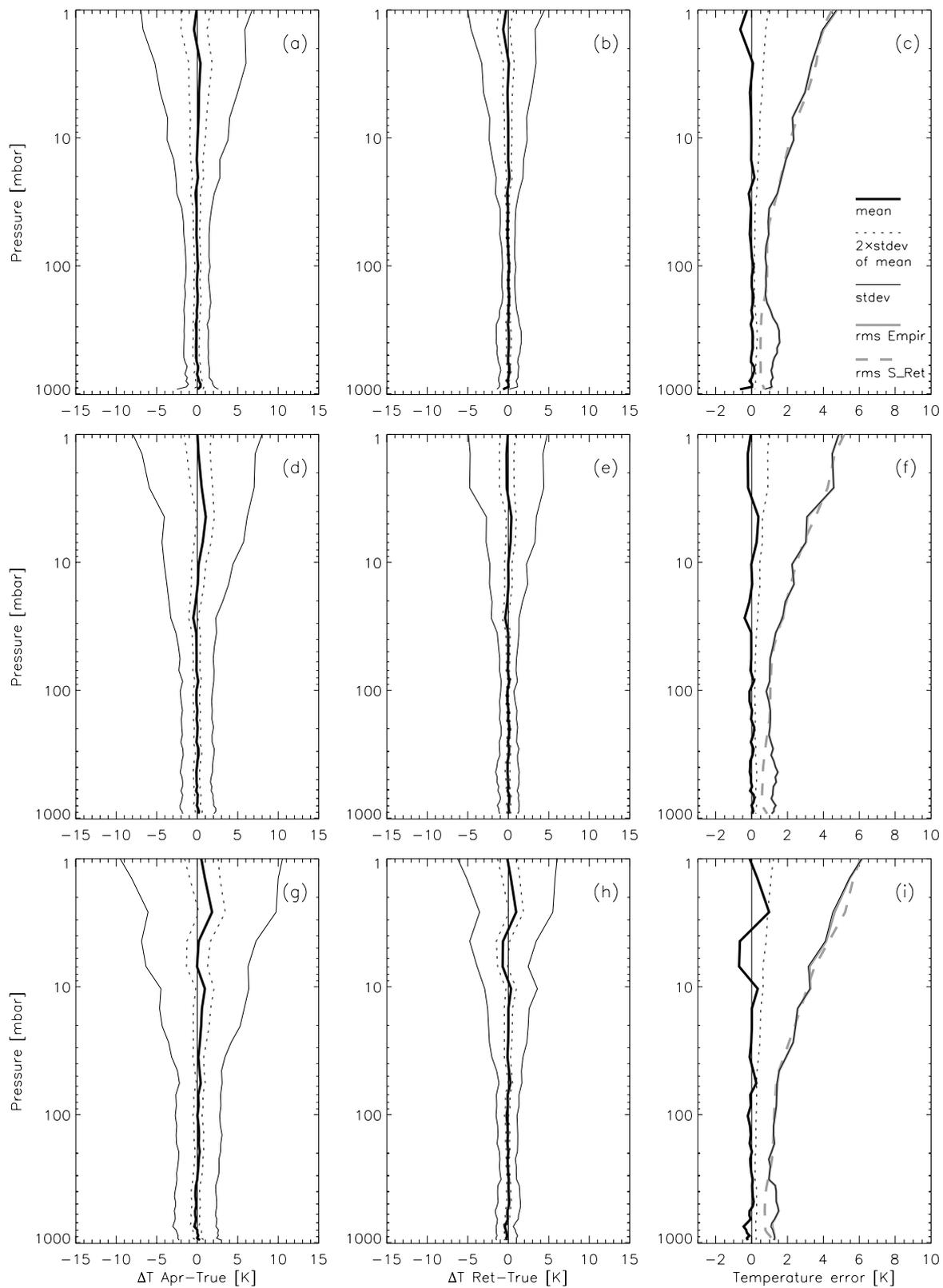
Overall, it should be mentioned that the retrieval algorithm does a fairly good job in improving by the measurements upon the *a priori* information as obvious from comparing the left and the middle panels. The analysis results for the total of 300 profiles are depicted in Fig. 5.32, which can be regarded as a mixture of features from the individual latitude ranges.

For the following and last experiment we assume that the *a priori* error covariance matrix with the standard deviation as illustrated in Fig. 5.33 is known.

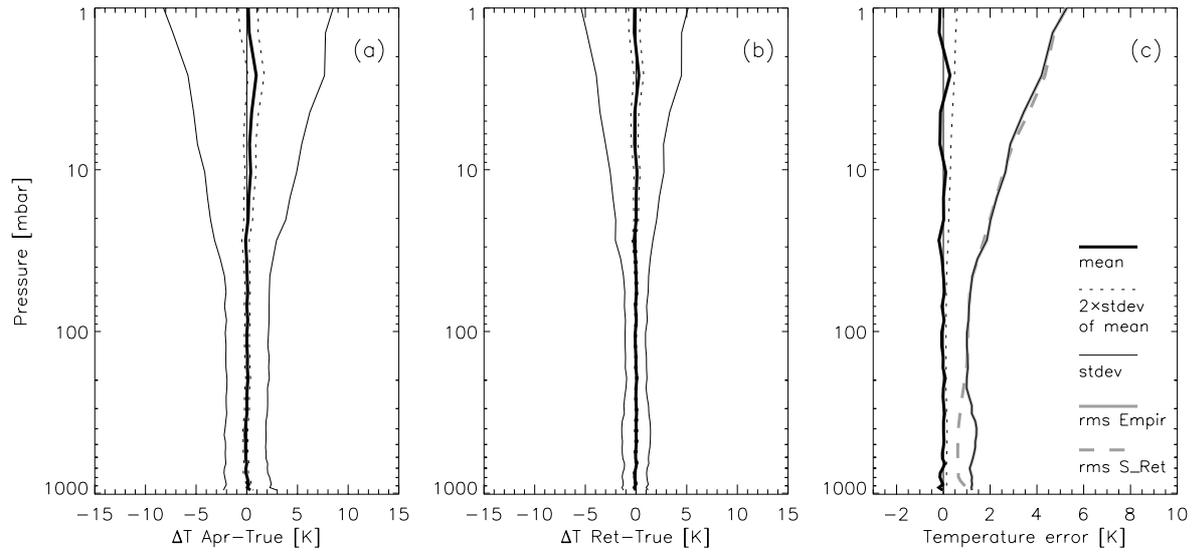


**Figure 5.33:** *A priori* error specification for low, mid, and high latitudes. The specification is roughly based on the short-range forecast uncertainty shown by Palmer et al. (2000).

*A priori* profiles  $\mathbf{x}_{ap}$  consistent with  $\mathbf{S}_{ap}$  are generated again by adding random vectors  $\boldsymbol{\epsilon}_x$  (Eq. 5.1.3) to the “true” profiles. The corresponding empirical error analysis results are shown in Fig. 5.34 and Fig. 5.35 and the conclusions drawn in Sect. 5.4.3 apply here as well. The bias is virtually removed and the resulting empirical *rms* retrieval error (gray) is to a full extent determined by the standard deviation (thin black) of the ensemble. Furthermore, since the *a priori* profiles have been generated consistently with  $\mathbf{S}_{ap}$ , the theoretically estimated *rms* retrieval errors (dashed gray) coincide closely with the empirical *rms* retrieval errors, except in the troposphere due to the uncertainty in the prescribed humidity as explained in Sect. 5.4.3.



**Figure 5.34:** Empirical error analysis for low (a-c), mid (d-f), and high (g-i) latitudes. See text for discussion of all panels. *A priori* profiles used are consistent with the *a priori* error covariance matrix as described in Sect. 5.1.2.



**Figure 5.35:** Empirical error analysis for low, mid and high latitudes together (300 profiles). See text for description.



## 6 SUMMARY AND CONCLUSIONS

After an introduction on radiative transfer theory and operational satellite meteorology, with the emphasis placed on the IASI instrument, a detailed description of temperature retrieval and associated problems was given.

Physics-based retrieval algorithms highly depend on the forward model, i.e., on our ability to model the physical-functional dependence of measurements on the atmospheric parameters to be retrieved. The forward problem may be reduced to the problem of computing atmospheric transmittance to evaluate the radiance as a function of the atmospheric state, which is measured by a sensor above the top of the atmosphere. The desired forward model algorithm needs to be not only accurate but in particular fast, in the light of the high number of IASI channels to be processed. Several models are currently developed for IASI measurements; most of them constitute fast radiative transfer models combining line-by-line calculations with regression schemes to generate radiance (or brightness temperature) as well as analytic Jacobians. In this work we utilized the fast radiative transfer model RTIASI (Matricardi and Saunders, 1999), which calculates optical depth on specified pressure levels. This forward model has been found suitable in calculating radiance and temperature Jacobians for our purpose to retrieve atmospheric temperature above the boundary layer. RTIASI provides satisfactory forward model error characteristics, and a good reproduction of temperature Jacobians, as stated by Sherlock (2000a), who performed a comparison of RTIASI with other fast models for IASI.

The large number of channels leads to a large computational burden, which is neither desirable nor practical in many applications. After removing channels, which contain significant contributions from “foreign” atmospheric constituents ( $O_3$ ,  $CH_4$ ,  $CO$ ) not relevant for temperature retrievals, we presented two approaches to reduce the total number of IASI channels for the sake of computational efficiency and retrieval robustness. In both cases, the retrieval is somewhat sensitive to the subset and number of channels selected. The simpler channel reduction procedure, the so-called “maximum sensitivity” approach, turns out to be sufficient in preserving most of the information for retrievals of temperature in most of the tropo- and stratosphere. The information

content technique, which is more expensive in selecting the appropriate channels, tends to perform better in the boundary layer region (below the 800 mbar level).

We employed an optimal estimation approach to solve the inverse problem. The inputs for this were RTIASI-simulated measurements (with reasonable noise superposed) and temperature Jacobians, *a priori* temperature profiles, and parameters of statistical models of the uncertainties (*a priori* error covariance matrix and measurement error covariance matrix). One major advantage of optimal estimation methodology (as described comprehensively in Rodgers, 2000) is that it naturally allows a rigorous analysis of errors and an overall characterization of the retrieval performance. The non-linear radiative transfer problem can be modeled by an iterative inversion algorithm employing a Taylor series expansion about a first guess value. In our IASI temperature retrieval work, the number of iterations needed to reach convergence was only 1 or 2 in the majority of tested cases.

According to the IASI Mission Rationale and Requirements (Diebel et al., 1998), IASI measurements shall provide information on temperature profiles in the troposphere and lower stratosphere with an accuracy of 1 K and a vertical resolution of 1 km in the lower troposphere. In this region IASI should, in addition, measure humidity profiles with an accuracy of 10% and a vertical resolution of 1-2 km. Regarding the latter, since this work did not address IASI humidity retrievals, the interested reader is referred to other qualified literature. For example, Collard (1999) reports that the requirements will be (marginally) met when retrieving relative humidity (not for specific humidity, though). Regarding temperature, one usually finds that the requirements, as stated above, are considered to be met when the profile can be retrieved with an accuracy of 1 K on a grid sampling at 1 km or finer. In this work, we employed different - as we think more meaningful - resolution measures based on the averaging kernel matrix  $\mathbf{A}$ . These measures specified in terms of FWHM of  $\mathbf{A}$ , Backus-Gilbert spread or data density (see, e.g., Rodgers, 2000) yield, for 1 K accuracy, a resolution of >1 km to 3 km in the troposphere. This finding is consistent with the results of Collard (1999) and constitutes a significant improvement on vertical resolution performance compared to current infrared sounders, though the formal requirement of 1 km is failed to be met. As shown by a specific experiment in this work, this vertical resolution and the accuracy of 1 K in the troposphere cannot be maintained when using a grid appreciably coarser than the RTIASI grid.

We studied various sensitivity matrices, such as the averaging kernel matrix (sensitivity of the retrieved profile to the true profile), the gain matrix (sensitivity of the retrieved profile to the measurements) as well as the impact of the *a priori* data on the retrieval. In addition, we

investigated the retrieval error, correlation functions, and signal-to-noise ratios. The main conclusions found by investigating these characteristics may be summarized as follows:

- The retrieval depends on the shape and the width of the weighting functions, and therefore relies partly on the ability of the fast model to characterize the transmissions and weighting functions at high spectral resolution. The difficulty in retrieving temperature at levels, where the sensitivity of the Jacobian matrix  $\mathbf{K}$  is insufficient, is reflected in retrievals in the boundary layer and in some regions in the stratosphere. If the prescribed humidity is well-known (within 15%, the expected accuracy of IASI specific humidity retrievals), the retrieval of temperature performs quite well in the troposphere (less than 1 K difference from the “true”). We found that a few (typically 1–3) iterations are required to improve the first retrieval step, indicating that a purely linear retrieval process would be insufficient.
- In general, optimal estimation provides retrieved temperature profiles, which improve highly throughout the entire height range upon the prior knowledge of atmospheric temperature.
- The standard deviation estimate of the retrieval error covariance matrix depends for small scales on the weighting functions, but mainly on the shape of the assumed *a priori* standard deviation. In the troposphere, where measurements dominate the *a priori* profiles in terms of information, it is less than 2 K, whereas at greater heights the retrieval error approaches the *a priori* error, which was assumed to be ~8 K at 1 mbar (stratopause) in most of the cases.
- From the upper troposphere up into the stratosphere, the retrieval is increasingly influenced by the *a priori* data, whereas in the troposphere the contributions of the measurements are dominating. Thus, the tropospheric region can be considered as well-measured by the IASI instrument, implying that the true state will be well reproduced by the measurements in this region, where *a priori* data almost become dispensable. These statements are expressed by the profiles of the retrieval-to-*a priori* error ratios and the averaging kernels. The signal-to-noise profiles also reveal the necessity for suitable *a priori* data at heights above the tropopause.
- The sensitivity of the retrieved profiles to the measurements (gain matrix  $\mathbf{G}$ ) indicates that channels peaking in the stratosphere still contribute significantly to the information in the retrievals, but to a high extent this is due to the presence of large *a priori* errors in that height region.

In most of the experiments presented, the climatological background model CIRA86aQ (Kirchengast et al., 1999) was used to supply *a priori* profiles accounting for the latitudinal and seasonal variations of the atmospheric situation. The error incurred by using prior profiles taken

from CIRA86aQ but utilizing a simplified covariance matrix only very roughly reflecting the actual errors has been investigated. A systematic “*a priori*-minus-true” difference introduces a bias in this case, which cannot be modeled by the inverse algorithm. Thus, the empirical *rms* (root mean square) retrieval error estimated a posteriori does not match the theoretically estimated *rms* retrieval error as calculated by the inverse algorithm. This is particularly apparent in the stratosphere, where the prior information exerts most influence. The retrieval performance is improved, as should be expected, by employing *a priori* profiles, which are evaluated by eigenvector-decomposition of the *a priori* covariance matrix and therefore consistent with the latter by construction. The results illustrate the virtual elimination of biases in the latter case and stress the importance of adequate modeling of the *a priori* data and associated errors. In other words, only a consistent specification of *a priori* profiles and their associated covariance matrix enables retrievals with “optimally” minimized retrieval errors. This also implies that any ensemble of *a priori* profiles should be statistically unbiased against the corresponding ensemble of “true” states. Similar arguments apply to the adequate statistical modeling of measurement errors.

Referring to future work, a more consistent *a priori* model and a more suitable scheme to retrieve below 800 mbar, in the case of interest in the boundary layer, seem necessary. The latter aim may be supported by channel selection procedures, which specifically aid reliable accuracy below 800 mbar. The present algorithm may be applied to a wider variety of tested scenery, as well as to fully realistic IASI sounding geometry; in this work only nadir viewing along a latitudinal slice at a certain longitude has been considered. Investigations of retrieving further parameters in addition to temperature, especially humidity, shows additional advantages of IASI measurements (e.g., Lerner et al., 2000a). Inversion algorithms to simultaneously retrieve temperature and water vapor need to be established. This may be accomplished with formally the same optimal inverse algorithm as described in this work. For example, Palmer et al. (2000), applied such an approach for radio occultation measurements and Lerner et al. (2000b) for IASI measurements, respectively.

The promising results obtained in this work for IASI confirm its potential for significantly improving upon atmospheric information available from current meteorological sounding measurements. If further advancements, such as mentioned above, will be achieved in the future, the capabilities of the IASI sensor for supporting operational meteorology and climatology as well as atmosphere and climate research can be exploited to their full extent.

## APPENDIX

Table A1. Physical Constants

Constant	value	uncertainty	unit
Speed of light in free space ( $c$ )	299 792 458	(exact)	$\text{m s}^{-1}$
Planck constant ( $h$ )	$6.62606876 \times 10^{-34}$	$0.00000052 \times 10^{-34}$	J s
	$4.13566727 \times 10^{-15}$	$0.00000016 \times 10^{-15}$	eV s
Boltzmann constant ( $k$ )	$1.3806503 \times 10^{-23}$	$0.0000024 \times 10^{-23}$	$\text{J K}^{-1}$
Stefan-Boltzmann constant ( $\sigma$ )	$5.670400 \times 10^{-8}$	$0.000040 \times 10^{-8}$	$\text{W m}^{-2} \text{K}^{-4}$
First radiation constant ( $c_1$ )	$1.191042722 \times 10^{-16}$	$0.000000093 \times 10^{-16}$	$\text{W m}^2 \text{sr}^{-1}$
Second radiation constant ( $c_2$ )	$1.4387752 \times 10^{-2}$	$0.0000025 \times 10^{-2}$	m K
Permittivity of free space ( $\epsilon_0$ )	$8.854187817 \times 10^{-12}$	(exact)	$\text{F m}^{-1}$
Permeability of free space ( $\mu_0$ )	$12.566370614 \times 10^{-7}$	(exact)	$\text{N A}^{-2}$
Newtonian gravitational constant ( $G$ )	$6.673 \times 10^{-11}$	$0.010 \times 10^{-11}$	$\text{m}^3 \text{kg}^{-1} \text{s}^{-2}$
Molar gas constant ( $R^*$ )	8.314472	0.000015	$\text{J mol}^{-1} \text{K}^{-1}$
Avogadro constant ( $N_A$ )	$6.02214199 \times 10^{23}$	$0.00000047 \times 10^{23}$	$\text{mol}^{-1}$

[All values according to the NIST reference on constants, units and uncertainty available at <http://physics.nist.gov/> ]

**Table A2. Parameters of the Earth and Atmosphere**

constant	value	unit
Solar irradiance (solar constant)	1380	W m <sup>-2</sup>
Mean earth-sun distance	1.4956 × 10 <sup>11</sup>	m
Volume of earth	1.070 × 10 <sup>21</sup>	m <sup>3</sup>
Mass of earth ( $m_e$ )	5.97370 × 10 <sup>24</sup>	kg
Mean density of earth	5.52 × 10 <sup>3</sup>	kg m <sup>-3</sup>
Standard acceleration of gravity	9.80665	m s <sup>-2</sup>
Gravitational parameter ( $Gm_e$ )	3.986005 × 10 <sup>14</sup>	m <sup>3</sup> s <sup>-2</sup>
Quadrupole gravitational	1.08263 × 10 <sup>-3</sup>	
Oblateness (Ellipticity)	0.003527	
Equatorial radius of earth	6.378137 × 10 <sup>6</sup>	m
Polar radius of earth	6.356752 × 10 <sup>6</sup>	m
Equatorial circumference	40075	km
Tropical year	3.15569259747 × 10 <sup>7</sup>	s
Orbital eccentricity	0.0167	
Obliquity of orbit	23.45	deg
Mean orbital velocity	29.79	km s <sup>-1</sup>
Angular velocity	7.29221 × 10 <sup>-5</sup>	rad s <sup>-1</sup>
Specific gas constant of dry air	287.06	J kg <sup>-1</sup> K <sup>-1</sup>
Mean molecular mass of dry air	2.8964 × 10 <sup>-2</sup>	kg mol <sup>-1</sup>
Standard pressure	1013.25	mbar (or hPa)
Standard temperature	273.16	K
Density of air (at $p_0$ and $T_0$ )	1.273 × 10 <sup>-3</sup>	g cm <sup>-3</sup>

[Values adopted from Hanel et al., 1992, Kidder and Vonder Haar, 1995, and Liou, 1992.]

**Table A3. Atmospheric constituents**

Constituent	Mixing ratio	Vertical distribution	Controlling processes
N <sub>2</sub> Molecular nitrogen	0.78708	Homogenous	Vertical mixing
O <sub>2</sub> Molecular oxygen	0.2095	Homogenous	Vertical mixing
H <sub>2</sub> O Water vapor	≤ 0.030	Decreases sharply in troposphere; increases in stratosphere; highly variable	Evaporation, condensation, transport; production by CH <sub>4</sub> oxidation
A Argon	0.0093	Homogenous	Vertical mixing
CO <sub>2</sub> Carbon dioxide	~370 ppmv	Homogenous	Vertical mixing; production by surface and anthropogenic processes
O <sub>3</sub> Ozone	~10 ppmv	Increases sharply in stratosphere; highly variable	Photochemical production in stratosphere; destruction at surface transport
CH <sub>4</sub> Methane	~1760 ppbv	Homogenous in troposphere; decreases in middle atmosphere	Production by surface processes; oxidation produces H <sub>2</sub> O
N <sub>2</sub> O Nitrous oxide	~320 ppbv	Homogenous in troposphere; decreases in middle atmosphere	Production by surface and anthropogenic processes; dissociation in the middle atmosphere; produces NO transport
CO Carbon monoxide	~70 ppbv	Decreases in troposphere; increases in stratosphere	Production anthropogenically and by oxidation of CH <sub>4</sub> transport
NO Nitric oxide	~0.1 ppbv	Increases vertically	Production by dissociation of N <sub>2</sub> O, catalytic destruction of O <sub>3</sub>
CFCl <sub>3</sub> (CFC-11) Trichlorofluoromethane	~270 pptv	Homogenous in troposphere; decreases in stratosphere	Industrial production; mixing in troposphere; photodissociation in stratosphere
CF <sub>2</sub> Cl <sub>2</sub> (CFC-12) Dichlorodifluoromethane	~530 pptv	See CFCl <sub>3</sub>	See CFCl <sub>3</sub>

[Table after from Liou, 1992, updated. Volume mixing ratios representative of troposphere or stratosphere. Mixing ratio of ozone represents stratospheric values. H<sub>2</sub>O, CO<sub>2</sub>, O<sub>3</sub>, CH<sub>4</sub>, N<sub>2</sub>O and CFCs are radiatively active.]

**Table A4. RTIASI Pressure Levels**

Level nr	Pressure [mbar]	Level nr	Pressure [mbar]
1	0.1	23	253.71
2	0.29	24	286.60
3	0.69	25	321.50
4	1.42	26	358.28
5	2.61	27	396.81
6	4.41	28	436.95
7	6.95	29	478.54
8	10.37	30	521.46
9	14.81	31	565.54
10	20.40	32	610.60
11	27.26	33	656.43
12	35.51	34	702.73
13	45.29	35	749.12
14	56.73	36	795.09
15	69.97	37	839.95
16	85.18	38	882.80
17	102.50	39	922.46
18	122.04	40	957.44
19	143.84	41	985.88
20	167.95	42	1005.43
21	194.36	43	1013.25
22	222.94		

[Information from Matricardi, 1999.]

**Table A5. IASI 1c Noise Values**

Wavenumber [cm <sup>-1</sup> ]	Noise [K]	Wavenumber [cm <sup>-1</sup> ]	Noise [K]
650	0.419	1750	0.170
700	0.157	1800	0.200
750	0.145	1850	0.224
800	0.145	1900	0.250
850	0.150	1950	0.240
900	0.150	2000	0.130
950	0.165	2050	0.135
1000	0.165	2100	0.141
1050	0.176	2150	0.151
1100	0.200	2200	0.172
1150	0.200	2250	0.200
1200	0.095	2300	0.239
1250	0.096	2350	0.287
1300	0.098	2400	0.351
1350	0.100	2450	0.400
1400	0.105	2500	0.700
1450	0.105	2550	0.900
1500	0.111	2600	1.100
1550	0.116	2650	1.300
1600	0.125	2700	1.600
1650	0.137	2750	1.935
1700	0.160		

[Table from P. Schlüssel, personal communications, ISSWG/EUMETSAT, 2000; values every 50 cm<sup>-1</sup>.]



## LIST OF SYMBOLS

$a$ .....	semi-major axis; random factor
$A$ .....	Ampere
$A$ .....	albedo
$A$ .....	averaging kernel matrix
$B$ .....	Planck function
$c$ .....	speed of light
$^{\circ}\text{C}$ .....	degrees Celsius
$c_1$ .....	first Planck constant
$c_2$ .....	second Planck constant
$\text{cm}$ .....	centimeter
$\text{deg}$ .....	degrees
$e$ .....	specific energy; eccentric anomaly
$\mathbf{e}$ .....	errorpattern
$E$ .....	energy; expectation value operator
$\text{eV}$ .....	electron Volt
$E$ .....	electric field amplitude
$f$ .....	arbitrary function
$\mathbf{f}$ .....	forward model function
$F$ .....	Farad; force
$F$ .....	Fourier Transform
$g$ .....	gram
$g$ .....	acceleration of gravity

$G$ .....	Universal Gravitation Constant
$\mathbf{G}$ .....	gain function matrix
GHz .....	Giga Hertz
$h$ .....	Planck constant
$H$ .....	scale height
$\mathbf{H}$ .....	maximum-sensitivity matrix
Hz .....	Hertz
hPa .....	hecto Pascal
$i$ .....	iteration index; inclination
$\mathbf{I}$ .....	identity matrix
$I_\nu$ .....	monochromatic radiance (intensity)
$j$ .....	iteration index; level index
J .....	Joule
$J$ .....	cost function
$k_\nu$ .....	absorption coefficient
$k$ .....	Boltzmann constant
$\mathbf{k}$ .....	weighting functions
K .....	Kelvin (degrees)
$\mathbf{K}$ .....	weighting function matrix
km .....	kilometer
kPa .....	kilo Pascal
$\mathbf{l}$ .....	eigenvectors
$L$ .....	angular momentum; correlation length; likelihood function
$m$ .....	dimension (length of measurement vector); mass
$M$ .....	molecular mass; mean anomaly
mbar .....	millibar

---

$mm$	millimeter
$n$	dimension (length of state vector); mean motion
$N$	Newton
$p$	pressure, phase function
$p_s$	surface pressure
$pdf$	probability density function
ppbv	parts per billion by volume
ppmv	parts per million by volume
pptv	parts per trillion by volume
$q$	humidity
$Q$	predictor
$r$	(Earth) radius
$R$	specific gas constant; radiance
$R_v$	monochromatic radiance
rad	radians
$rms$	root mean square
rpm	revolutions per minute
s	seconds
$S$	arbitrary covariance matrix
$S$	solar constant
$S_{ap}$	<i>a priori</i> error covariance matrix
$S_{\varepsilon}$	measurement error covariance matrix
$S_m$	measurement-based error covariance matrix
$S_s$	smoothing error covariance matrix
$\hat{S}$	retrieval error covariance matrix
sr	steradian

---

$T_v$ .....	monochromatic transmittance
$T$ .....	temperature, orbital period
$T_B$ .....	brightness temperature
$u$ .....	path length
$v$ .....	velocity
$\nu$ .....	wavenumber
$W$ .....	Watt
$W$ .....	atmospheric weighting functions
$\mathbf{x}$ .....	state vector
$\mathbf{x}_{ap}$ .....	<i>a priori</i> state vector
$\mathbf{x}_0$ .....	initial guess vector
$\hat{\mathbf{x}}$ .....	estimate of state vector
$\mathbf{y}$ .....	measurement vector
$z$ .....	height coordinate
$\alpha$ .....	control parameter; absorptance
$\delta$ .....	density; optical path difference
$\delta_D$ .....	delta Dirac function
$\boldsymbol{\varepsilon}$ .....	measurement error
$\boldsymbol{\varepsilon}$ .....	emissivity; eccentricity; shape error index
$\varepsilon_0$ .....	permittivity of free space
$\phi$ .....	azimuth angle
$\Delta T$ .....	temperature difference
$\lambda$ .....	wavelength
$\mu$ .....	gravitational parameter; cosine of zenith angle
$\mu\text{m}$ .....	micrometer
$\nu$ .....	true anomaly

---

$\theta$ .....	zenith angle
$\rho$ .....	density; reflectivity
$\sigma$ .....	Stefan-Boltzmann constant; standard deviation
$\sigma_a$ .....	absorption cross section
$\sigma_s$ .....	scattering cross section
$\tau_\nu$ .....	monochromatic optical depth
$\nu$ .....	frequency
$\omega$ .....	scattering albedo; argument of perigee
$\Omega$ .....	solid angle; right ascension of ascending node; scattering angle



---

## LIST OF ACRONYMS

A/D .....	Analogue/Digital
AIRS .....	Advanced Infrared Sounder
AMSU .....	Advanced Microwave Sounding Unit
APT .....	Automatic Picture Transmission
ASCAT .....	Advanced Scatterometer
ATOVS .....	Advanced TIROS Operational Vertical Sounder
AU .....	Astronomical Unit
AVHRR .....	Advanced Very High Resolution Radiometer
CDA .....	Command and Data Acquisition Station
CGMS .....	Coordination Group for Meteorological Satellites
CMA .....	Chinese Meteorological Administration
CNES .....	Centre National d'Etudes Spatiale
DMSP .....	Defense Meteorological Satellite Program
DPT .....	Delayed Picture Transmission
ECMWF .....	European Center for Medium-Range Weather Forecasts
ECT .....	Equator Crossing Time
EPS .....	EUMETSAT Polar System
EOS .....	Earth Observation System
EM .....	Electromagnetic
EUMETSAT .....	European Organization for the Exploitation of Meteorological Satellites
FOV .....	Field of View

FT .....	Fast Transmittance; Fourier Transform
FWHM.....	Full Width at Half Maximum
GAC.....	Global Area Coverage
GENLN2 .....	General Line-by-line atmospheric transmittance and radiance model
GMS .....	Geostationary Meteorological Satellite
GNSS.....	Global Navigation satellite System
GOES.....	Geostationary Operational Environmental Satellite
GOME .....	Global Ozone Monitoring Experiment
GRAS .....	GNSS Receiver Atmospheric Sounder
HIRS .....	High Resolution Infrared Radiation Sounder
HITRAN .....	High Resolution Transmittance model
HRPT.....	High Resolution Picture Transmission
IASI .....	Infrared Atmospheric Sounding Interferometer
IFOV.....	Instantaneous Field of View
IJPS.....	Initial Joint Polar System
IR .....	Infrared
ISRF.....	Instrument Spectral Response Function
ISRO .....	Indian Space Research Organization
ISSWG.....	IASI Sounding Science Working Group
IC .....	Information Content
JMA.....	Japan Meteorological Agency
LEO .....	Low Earth Orbit
LAC .....	Local Area Coverage
LRIT .....	Low Rate Information Transmission
LRPT .....	Low Rate Picture Transmission
LW .....	Long Wave

---

MHS .....	Microwave Humidity Sounder
MSG .....	Meteosat Second Generation
MSU .....	Microwave Sounding Unit
MTSAT .....	Multipurpose Transportation Satellite
NASA .....	National Aeronautics and Space Administration
NEAT .....	Noise Equivalent Temperature Difference
NESDIS .....	National Environmental Satellite, Data and Information Service
NIST .....	National Institute of Standards and Technology
NOAA .....	National Oceanographic and Atmospheric Administration
NWP .....	Numerical Weather Prediction
OPD .....	Optical Path Difference
POES .....	Polar-Orbiting Operational Environmental Satellite
RMS.....	Root Mean Square
RS .....	Remote Sensing
RT .....	Radiative Transfer
SEM.....	Space Environment Monitor
SI .....	Système International (d'Unités)
SMS .....	Synchronous Meteorological Satellite
SSM.....	Special Sensor Microwave
SW .....	Short Wave
TIGR .....	TOVS Initial Guess Retrieval
TOA.....	Top of the Atmosphere
TIROS .....	Television and Infrared Observation Satellite
TOVS.....	TIROS Operational Vertical Sounder
UV .....	Ultraviolet
VHF .....	Very High Frequency

VHRR.....	Very High Resolution Radiometer
VIS.....	Visible
VISSR.....	Visible and Infrared Spin Scan Radiometer
WEFAX.....	Weather Facsimile
WMO.....	World Meteorological Organization
WWW.....	World Weather Watch
ZPD .....	Zero Path Difference

---

## REFERENCES

**Anderson, T. W.**, An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, New York, 1984.

**Brown, R. G., P. Y. C. Hwang**, Introduction to Random Signals and Applied Kalman Filtering, John Wiley & Sons, New York, 1997.

**Camy-Peyret, C., J. Eyre**, The IASI Science Plan, ISSWG report, 1998.

Available on-line via <http://sads.cnes.fr:8060/IAS/>

**Collard, A. D.**, Notes on IASI Performance, Forecasting Research Technical Report No.256, The Met. Office, U.K., December 1998.

**Diebel, D., F. Cayla, T. Phulpin**, IASI Mission Rationale and Requirements, EUMETSAT/CNES Technical Report IA-SM-0000-10-CNE/EUM, Issue 4b, 1996.

**Eyre, J.**, A Fast Radiative Transfer Model for Satellite Sounding Systems, ECMWF, Research Department, Technical Memorandum No. 176, March 1991.

**Fleming, E. L., S. Chandra, M. R. Schoeberl, J. J. Barnett**, Monthly Mean Global Climatology of Temperature, Wind, Geopotential Height and Pressure for 0–120 km, NASA Techn. Memorandum No. 100697, Nat. Aeronaut. and Space Admin., Washington, D.C., U.S., 1988.

**Gelb, A.**, Applied Optimal Estimation, M.I.T Press, Cambridge, MA, 1974.

- Hanel, R. A., B. J. Conrath, D. E. Jennings, R. E. Samuelson**, Exploration of the Solar System by Infrared Remote Sensing, Cambridge University Press, 1992.
- Hannon, S., L. L. Strow, W. W. McMillan**, Atmospheric Infrared Fast Transmittance Models: A Comparison of two Approaches, *SPIE*, 2830, 1996.
- Hiebler, S. D.**, IASI – New Aspects in Infrared Atmospheric Sounding, Diplomarbeit, Universität Graz, 1998.
- Kidder, S. Q., T. H. Vander Haar**, Satellite Meteorology - An Introduction, Academic Press, San Diego, 1995.
- Kirchengast, G., J. Hafner, W. Poetzi**, The CIRA86aQ\_UoG model: An extension of the CIRA-86 monthly tables including humidity tables and a Fortran95 global moist air climatology model, Techn. Report for ESA/ESTEC No. 8/1999, 18p., Inst. for Geophys., Astrophys., and Meteorol., Univ. of Graz, Austria, 1999.
- Lerner, A. J., E. Weisz, G. Kirchengast**, Temperature and Humidity Retrieval from Simulated Infrared Atmospheric Sounding Interferometer (IASI) Measurements, *J. Geophys. Res.*, revised, 2000a.
- Lerner, A. J., E. Weisz, G. Kirchengast**, Retrieval of Water Vapor from IASI Measurements: Methodology and Case Study Application, in Proceedings of the ITSC-XI conference, Budapest, Hungary, in press, 2000b.
- Liou, K. N.**, Radiation and Cloud Processes in the Atmosphere, Oxford University Press, Oxford, 1992.

---

**Liu, X., T. S. Zaccheo, J-L. Moncet**, Comparison of Different Non-Linear Inversion Methods for Retrieval of Atmospheric Profiles, in Proceedings of the 10<sup>th</sup> Conference of Satellite Meteorology, Long Beach, California, 293-295, 2000.

**Lorenc, A. C.**, Analysis Methods for Numerical Weather Prediction, *Q.J.R.Met.Soc.*, 112, 1177-1194, 1986.

**Lorenc, A. C.**, A Practical Approximation to Optimal Four-dimensional Objective analysis, *Mon.Wea.Rev.*, 116, 730-745, 1988.

**Mardia, K. V., J. T. Kent, J. M. Bibby**, Multivariate Analysis, Academic Press, London, 1979.

**Matricardi, M.**, RTIASI-1 User's Guide, ECMWF Report, U.K., November 1999.

**Matricardi, M., R. Saunders**, A Fast Radiative Transfer Model for Simulation of Infrared Atmospheric Sounding Interferometer Radiances, *J.Appl.Optics*, 38, 5679-5691, 1999.

**Maybeck, P.**, Stochastic Models, Estimation and Control, Volume I, Academic Press, New York, 1979.

**Maybeck, P.**, Stochastic Models, Estimation and Control, Volume II, Academic Press, New York, 1982.

**Palmer, P. I., J. J. Barnett, J. R. Eyre, S. B. Healy**, A Non-linear Optimal Estimation Inverse Method for Radio Occultation Measurements of Temperature, Humidity, and Surface Pressure, *J.Geophys.Res.*, 102, 17513-17521, 2000.

**Priestley, M. B.**, Spectral Analysis and Time Series, Academic Press, London, 1981.

**Purser, R. J., H.-L. Huang,** Estimating Effective Data density in a Satellite Retrieval or an Objective Analysis , *J.Appl. Meteor.*, 1092-1107, 1993.

**Rabier, F., D. Chafaï, R. Abboubi,** Comparison of Channels Selection Methods for IASI, in Proceedings of the ITSC-XI conference, Budapest, Hungary, 2000.

**Rodgers, C. D.,** Information Content and Optimization of High Spectral Resolution Measurements, *SPIE*, 2830, Optical Spectroscopic Techniques and Instrumentation for Atmospheric Space Research II (P. B. Hays and J. Wang, eds.), 136–147, 1996.

**Rodgers, C. D.,** Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publ., Singapore, 2000.

**Salby, M. L.,** Fundamentals of Atmospheric Physics, Academic Press, San Diego, 1996.

**Sellers, J. J.,** Understanding Space – An Introduction of Aeronautics, editor: W. Larson, McGraw-Hill, Inc., 1994.

**Sherlock, V. J.,** Results from the first U.K.MO IASI Radiative Transfer Model Intercomparison, Forecasting Research Technical Report No. 287, The Met. Office, U.K., 2000a.

**Sherlock, V. J.,** Impact of RTIASI Fast Radiative Transfer Model Error on IASI Retrieval Accuracy, Forecasting Research Technical Report No. 319, The Met. Office, U.K., 2000b.

**Susskind, J., J. Rosenfeld , and D. Reuter,** An Accurate Radiative Transfer Model for use in the Direct Physical Inversion of HIRS2 and MSU Temperate Sounding Data, *J. Geophys. Res*, 88, 8550-8568, 1983.

**Talagrand, O.**, Four-dimensional Variational Assimilation, Laboratoire de Météorologie Dynamique, Paris, France, 1989.

**Thepaut, J-N., P. Moll**, Variational Inversion of Simulated TOVS Radiances using the Adjoint Technique, *Q.J.R.Met.Soc.*, 116, 1425-1448, 1990.

**Weisz, E.**, Kalman Filter Analysis of Meteorological Profiling Data, Diplomarbeit, Universität Graz, 1998.

**Wells, N.**, The Atmosphere and Ocean, John Wiley and Sons, Chichester, England, 1997.

**Zuev, V. E.**, Propagation of Visible and Infrared Radiation in the Atmosphere, John Wiley & Sons, New York, 1974.

#### SUMMARY OF WEBLINKS REFERRED TO IN THE REPORT

The EUMETSAT homepage: <http://www.eumetsat.de/>

The CNES / IASI mission server: <http://sads.cnes.fr:8060/IASI/>

The CGMS Directory of Meteorological Satellite Application: <http://www.wmo.ch/hinsman/>

The Geostationary Satellite / NOAA Server: <http://www.goes.noaa.gov/>

The Office of Systems Development / NOAA server: <http://www.osd.noaa.gov>

The National Geophysical Data Center / DMSP server: <http://www.ngdc.noaa.gov/dmsp/>

The NOAA Satellite Information System web site:

<http://psbsgi1.nesdis.noaa.gov:8080/naaasis.html>

The ISRO homepage: <http://www.isro.org/>

The JMA homepage: <http://www.kishou.go.jp/english/>

The National Meteorological Center of China homepage: <http://www.ncw.gov.cn/>

The Russian Sputnik server: [http://sputnik\\_infospace.ru/](http://sputnik_infospace.ru/)

The NIST homepage: <http://physics.nist.gov/>