

Institut für Meteorologie und Geophysik, Universität Graz

IMG/UoG

Institute for Meteorology and Geophysics, University of Graz



Wissenschaftlicher Bericht No. 10

Institut für Meteorologie und Geophysik, Karl-Franzens-Universität Graz
Halbärthgasse 1, A-8010 Graz, Austria

**Tropospheric water vapor imaging by
combination of ground-based and
spaceborne GNSS sounding data**

by
Ulrich Foelsche

Graz, June 1999

This report contains the Ph.D. thesis of Dr. Ulrich Foelsche, which he worked out during the years 1996 - 1999 at the IMG/UoG.

ACKNOWLEDGMENTS

I am very grateful to my thesis supervisor Prof. Dr. G. Kirchengast for his distinguished overall support and a great number of fruitful scientific discussions. I am furthermore much indebted to him for furnishing the economic basis of the work by a doctoral position based on his ATFERN discretionary funds at the IMG and his START-Programm Nr. Y103-CHE (START-ATCHANGE) research award of the Bundesministerium für Wissenschaft und Verkehr, Vienna, Austria. I am also very grateful to Prof. Dr. R. Leitinger for actively supporting my scientific work and my academic education at the Institute. I continue to express my appreciation to Prof. Dr. S.J. Bauer for supporting my academic efforts and for many interesting discussions.

I am most grateful to the entire staff of the IMG for contributing so much to the excellent working atmosphere at the Institute. I continue to honestly thank all my colleagues and friends within the Institute.

Thanks are due to Prof. Dr. A.E. Niell, Massachusetts Institute of Technology, for providing valuable information on the "art of mapping function construction". The European Centre for Medium-range Weather Forecasts (ECMWF) kindly provided the atmospheric analysis fields used in this work.

Most of all I want to thank my girlfriend Rosi for being that patient with me throughout all the years (and specially during the last weeks). My warmest thanks are due to my parents for their understanding and their mental and emotional support whenever it was needed. Last, but not least, I wish to thank all my friends, who accompanied me during the last years.

TABLE OF CONTENTS

INTRODUCTION	1
1 WATER VAPOR IN THE CLIMATE SYSTEM	5
1.1 PROPERTIES OF WATER VAPOR	5
1.2 THE HYDROLOGICAL CYCLE.....	9
1.3 CLOUDS AND RADIATION	11
1.4 WATER VAPOR IN THE ATMOSPHERE.....	13
1.5 PRECIPITATION AND EVAPORATION	21
1.6 OCEAN FLUXES AND LAND SURFACE PROCESSES.....	23
1.7 THE ROLE OF WATER VAPOR IN THE CLIMATE SYSTEM	25
2 PHYSICS OF REFRACTIVITY	27
2.1 BASIC PROPERTIES OF THE ELECTROMAGNETIC FIELD.....	27
2.2 REFRACTIVE INDEX AND POLARIZABILITY	32
2.3 THE REFRACTIVITY OF "DRY" AIR	36
2.4 THE REFRACTIVITY OF WATER VAPOR.....	39
2.5 EXPERIMENTAL VALUES OF THE REFRACTIVITY	43
3 GROUND-BASED GPS METEOROLOGY	47
3.1 THE GLOBAL POSITIONING SYSTEM.....	47
3.2 PHYSICS OF THE PROPAGATION DELAY IN THE NEUTRAL ATMOSPHERE.....	54
3.3 THE HYDROSTATIC DELAY ALONG RAY PATHS WITH ARBITRARY ELEVATION ANGLES.....	61
3.4 ESTIMATION OF INTEGRATED WATER VAPOR FROM OBSERVED WET DELAYS	65
3.5 RESULTS FROM FIELD EXPERIMENTS	71
4 THE RADIO OCCULTATION TECHNIQUE.....	75
4.1 HISTORY OF THE RADIO OCCULTATION TECHNIQUE	75
4.2 PRINCIPLES OF THE RADIO OCCULTATION TECHNIQUE	77
4.3 DERIVATION OF ATMOSPHERIC PARAMETERS.....	86
4.4 THE GPS/MET EXPERIMENT	89
5 SPACE-TIME SAMPLING CONDITIONS FOR ATMOSPHERIC IMAGING	91
5.1 INTRODUCTION	91
5.2 OBSERVATIONAL REQUIREMENTS	93
5.3 LEO SATELLITE CONSTELLATIONS	95
5.4 DISTRIBUTION OF OCCULTATION EVENTS.....	96
5.5 COMBINATION WITH GROUND-BASED MEASUREMENTS.....	106

6	TROPOSPHERIC WATER VAPOR IMAGING.....	115
6.1	FORWARD MODEL - GEOMETRY	115
6.2	FORWARD MODEL - SYNTHETIC REFRACTIVITY FIELDS	117
6.3	FORWARD MODEL - ECMWF-DERIVED REFRACTIVITY FIELDS	118
6.4	DISCRETE INVERSE THEORY	122
6.5	RETRIEVAL GEOMETRY	125
6.6	RETRIEVAL RESULTS - SYNTHETIC DATA.....	126
6.7	RETRIEVAL RESULTS - ECMWF-DERIVED DATA	142
	SUMMARY AND CONCLUSIONS.....	151
	REFERENCES	153

INTRODUCTION

The global water or hydrological cycle traces the largest movement of any substance on earth, and is a consequence of the conservation of water substance. The exchanges of moisture and heat between the atmosphere and the Earth's surface fundamentally affect the dynamics and thermodynamics of the climate system. About 50% of surface cooling results from evaporation, the release of latent heat of condensation in clouds provides ~30% of the thermal energy that drives the Earth's circulation [Chahine, 1992].

Water vapor plays a fundamental role in the transport of energy in the atmosphere on a wide range of temporal scales, it is the atmospheric constituent with the greatest effect on the dynamics and thermodynamics of the atmosphere, caused by the unusually large latent heat associated with water's change of phase.

Water vapor is furthermore the most abundant of the greenhouse gases in the atmosphere and the most important in establishing the Earth's climate. Without the abundance of water and the variety of its forms in the Earth system, the average global temperature would reside far below 0°C [cf. Houghton *et al.*, 1996]. Water vapor has a greater effect than any other substance on the radiation balance of the atmosphere, it is the dominant absorber of infrared radiation.

The water vapor-greenhouse feedback, which means warming from increased greenhouse gases leading to an increase of water vapor and therefore even more warming, is considered to be the strongest positive feedback in the climate system [Gaffen *et al.*, 1992a]. This positive feedback could amplify the temperature change due to a doubling of CO₂ by some 60% - 70% [cf. Houghton *et al.*, 1996] and it is an important feature of all General Circulation Model (GCM) simulations to date.

Model results show [see, e.g., Del Genio *et al.*, 1991 and 1994] that substantial increase in the concentration of water vapor should be expected at all latitudes as climate warms due to increasing greenhouse gases.

A deeper understanding of the mechanisms distributing water vapor through the atmosphere and of water vapor's effects on atmospheric radiation and circulation is vital to estimate long-term changes in climate. This understanding is hampered by the fact, that water vapor is the most variable of the major constituents of the atmosphere.

Our ability to measure time varying global and regional water vapor distributions is severely limited. The traditional method for measuring the amount of atmospheric water vapor uses radiosondes, which are balloon-borne instruments sending temperature, humidity and pressure data to the ground by radio signals. These instruments yield good information on the vertical distribution of moisture, but their geographic sparseness and the low frequency of their launches (typically once or twice per day) are not suitable for measuring the water vapor field with high resolution. Limitations in the analysis of water are therefore the major error source in short term (0-24 hours) forecasts of precipitation.

Emerging networks of continuously operating Global Positioning System (GPS) receivers provide a new and powerful approach to the remote sensing of atmospheric water vapor [see Bevis *et al.*, 1992]. The GPS consists of a constellation of 24 satellites that transmit L-band radio signals to large numbers of users engaged in navigation, time transfer, and relative positioning. These signals are delayed by the atmospheric water vapor as they propagate from GPS satellites to ground-based receivers. This "geodetic noise" can be used as valuable meteorological signal, as the so-called *wet delay* is nearly proportional to the quantity of water vapor integrated along the signal path, which in turn can be transformed into an estimate of precipitable water, i.e., the height of an equivalent column of liquid water. Given accurate delay measurements, precipitable water can be recovered with an accuracy of ~ 1 mm. In principle it is possible to use signals from the Russian GLObal Navigation Satellite System (GLONASS) in exactly the same way as described above, though the number of GLONASS receivers in use is still small compared with the number of GPS receivers. We will therefore generally use the term GNSS (Global Navigation Satellite System), comprising GPS and GLONASS.

Another GNSS based remote sensing application is the radio occultation technique, which uses a satellite-to-satellite limb sounding concept to probe the Earth's atmosphere. From the standpoint of the receiver onboard a Low Earth Orbit (LEO) satellite an occultation occurs whenever a GPS or GLONASS satellite rises or sets and the ray path from the transmitter traverses the Earth's atmosphere towards the LEO receiver.

The influence of the atmospheric refractivity field on the propagation of the electromagnetic signal enables, in turn, the retrieval of atmospheric refractivity profiles, and the derivation of further parameters such as density, pressure, and temperature. In the lower troposphere (especially at low latitudes), refractivity variations are dominated by water vapor, and high vertical resolution humidity profiles can be retrieved, given independent temperature data.

The combination of these two (complementary) information sources, the ground-based and the spaceborne GNSS sounding data, allows the retrieval of two-dimensional, height resolving water vapor density fields employing tomographic reconstruction techniques. The main emphasis of this work was to develop a new technique for this type of "water vapor imaging". Such a data product can help to provide the much needed information on regional and global water vapor distributions.

In Chapter 1 we demonstrate why accurate measurements and knowledge of the distribution of water vapor are fundamental for a deeper understanding of the hydrological cycle and for the assessment of the effects caused by anthropogenic climate change. This well indicates the urgent need for such measurements.

In Chapter 2 we adopt a theoretical approach in order to understand the physics behind the atmospheric refractivity, the cornerstone for the remote sensing techniques proposed in this work for water vapor imaging. We will describe and explain how the atmospheric refractivity relates to microscopic quantities like the molecular polarizability and the permanent dipole moment of the water molecule.

In Chapter 3 we show (after a short description of the GPS) how ground-based GPS measurements are used to retrieve integrated water vapor or precipitable water . We focus, furthermore, on the dominant error sources and present a new "geometric" mapping function (mapping functions are needed for the accurate retrieval of the "wet delay"), which was developed within this work.

In Chapter 4 we give an overview of the radio occultation technique and the theoretical background of the standard retrieval methodology, used in order to derive atmospheric profiles like water vapor profiles from spaceborne GNSS measurements.

In Chapter 5 we present the results of a geometry study, undertaken in order to demonstrate how frequently radio occultation events can be expected to occur in the vicinity of a ground receiver network given different LEO satellite constellations. The results of this study were used to develop a realistic scenario for the new imaging technique.

In Chapter 6 we introduce the new water vapor imaging technique, which is based on ground-based measurements of integrated water vapor under different elevation angles, that are combined with the occultation-derived water vapor density profile using a pixel geometry and employing optimal estimation. The water vapor density profile could also come from radiosondes or other profiling instruments. The performance of the retrieval algorithm is tested with simulated data, derived from different simple synthetic refractivity fields and from realistic European Centre for Medium-range Weather Forecasts (ECMWF) atmospheric analysis data.

Finally, a summary and conclusions on the main results of the work are given.

1 WATER VAPOR IN THE CLIMATE SYSTEM

Water vapor is the gaseous form of dihydrogen oxide and is the only atmospheric constituent that may change its state of aggregation, as the temperatures occurring in the atmosphere are near the triple point of water. Although water exists in the atmosphere as vapor, liquid, and solid, the vapor phase dominates. Water vapor is the atmospheric constituent with the greatest effect on the dynamics, thermodynamics, and radiation balance of the atmosphere, it is the most abundant of the greenhouse gases in the atmosphere and the most important in establishing the Earth's climate.

1.1 Properties of water vapor

1.1.1 Water vapor terminology

Meteorologists have defined several different terms to express the amount of water vapor in air, a practice which can cause confusion. Some expressions refer to the actual amount, or concentration, of water vapor in the air, and some relate the actual amount to the amount that would saturate the air [cf. Peixoto and Oort, 1992; Bucci Mockler 1995; Liljequist and Cehak, 1984].

Vapor pressure (e) is the partial pressure of a given sample of moist air that is attributable to water vapor, measured in hectopascal (hPa) or millibar (mbar), where 1 mbar = 1 hPa = 100 N/m².

Absolute humidity or **vapor density** (ρ_w) is the mass of vapor in a given volume, usually expressed in kg/m³ or g/m³. Vapor pressure and absolute humidity are linked via the ideal gas law (equation of state):

$$e = \rho_w \frac{R^*}{m_w} T, \quad (1.1)$$

where R^* is the universal gas constant (8.314510 ± 0.000070 Pa m³ K⁻¹ mol⁻¹ or 8314.510 ± 0.070 J kmol⁻¹K⁻¹), T is the absolute temperature in K, and m_w is the molar mass of water vapor (18.01528 kg/kmol) [CRC, 1993].

Specific humidity (q) is the ratio of the mass of water vapor in a moist air parcel to the total mass of the moist air, expressed in kg/kg or g/kg:

$$q = \frac{\rho_w}{\rho} = \frac{m_w}{m_d} \frac{e}{p - \left(1 - \frac{m_w}{m_d}\right)e}, \quad (1.2)$$

where ρ is the density of moist air, m_d is the molar mass of dry air (28.9644 ± 0.0014 kg/kmol) [CRC, 1993], $m_w/m_d = 0.62198$, and p is the total pressure of moist air.

Mixing ratio by mass (m) is the mass of water vapor divided by the mass of dry air in a given volume, measured in kg/kg or g/kg.

$$m = \frac{\rho_w}{\rho_d} = \frac{m_w}{m_d} \frac{e}{p - e}, \quad (1.3)$$

where ρ_d is the density of dry air, and $p - e = p_d$, the pressure of dry air.

Mixing ratio by volume (m_v), defined as the number of H₂O molecules divided by the number of molecules from dry air in a given volume.

Relative humidity (r) is the ratio of the actual vapor pressure to the saturation vapor pressure (e_s , see section 1.1.3) at the air temperature expressed as a percentage:

$$r = \frac{e}{e_s} 100\% . \quad (1.4)$$

Dew point is the temperature the air would have if it were cooled, at constant pressure and water vapor content, until saturation (or condensation) occurred. The difference between actual temperature and dew point is called the *dew point depression*.

Wet-bulb temperature is the temperature an air parcel would have if it were cooled to saturation at constant pressure by evaporating water into the parcel. The term comes from the operation of a psychrometer, a widely used instrument for measuring humidity, in which a pair of thermometers, one of which has a wetted piece of cotton on the bulb, is ventilated. The difference between the temperatures of the two thermometers is a measure of humidity. The wet-bulb temperature is the lowest air temperature that can be achieved by evaporation [Bucci Mockler, 1995].

Within this work we will confine ourselves (mostly) to the use of the terms vapor pressure, vapor density, and specific humidity.

Integrated water vapor (IWV) is the amount of water vapor in a unit area integrated over a column that extends the full height of the atmosphere:

$$IWV = \int_0^{\infty} \rho_w(z) dz. \quad (1.5)$$

Precipitable water (PW) is the integrated water vapor expressed as the height of an equivalent column of liquid water, usually measured in mm. PW can be written as [Peixoto and Oort, 1992]:

$$PW = \frac{1}{\rho_l} \int_0^{p_0} q \frac{dp}{g}, \quad (1.6)$$

where ρ_l is the density of liquid water.

Total column water includes water in all phases but is usually dominated by water vapor.

1.1.2 Physical properties of H₂O

Water vapor is the only common atmospheric constituent that possesses a permanent dipole moment, caused by an asymmetric distribution of charge in the water molecule. The large dipole moment ($6.184 \cdot 10^{-30}$ Cm) and the resulting strong intermolecular forces are directly responsible for the unusually large latent energy associated with phase changes, which in turn significantly impacts the vertical stability of the atmosphere, the structure and evolution of storm systems, and the energy balance of the earth-atmosphere system. The large heat capacity of liquid water (only surpassed by liquid ammoniac) has a strong influence on the climate of Earth. Oceans store vast quantities of heat both by possessing a higher heat capacity than the land areas and by being able to mix heat downward through convective and turbulent processes [cf. Webster, 1994]. Values of latent and specific heat at 0°C are [Laube and Höller, 1988]:

Latent heat of vaporization of pure water	$L_v = 2.5008 \cdot 10^6$ J/kg
Latent heat of melting	$L_m = 0.3337 \cdot 10^6$ J/kg
Latent heat of sublimation ($L_s = L_v + L_m$)	$L_s = 2.8345 \cdot 10^6$ J/kg
Specific heat of liquid water	$c_w = 4218$ J/kgK
Specific heat of water vapor at constant pressure	$c_{pv} = 1850$ J/kgK
Specific heat of water vapor at constant volume	$c_{vv} = 1390$ J/kgK

All values are functions of the absolute temperature T , the temperature dependence of L_v , e.g., can be approximated (for temperatures between 233 K and 313 K) by the formula [Laube and Höller, 1988]:

$$L_v = 2.501 \cdot 10^6 \left(\frac{T_0}{T} \right)^b [\text{J/kg}], \quad b = 0.167 + 3.67 \cdot 10^{-4} T, \quad (1.7)$$

where $T_0 = 273.15$ K.

1.1.3 The Clausius-Clapeyron equation

The maximum vapor pressure of water (at thermodynamic equilibrium) is limited by the temperature of the vapor. However, because the rest of the dry air makes up the bulk of the mass of the atmosphere, the temperature of the water vapor is determined largely by the temperature of the air with which it coexists. The Clausius-Clapeyron equation relates changes in the saturation vapor pressure of water (e_s) to the absolute temperature T :

$$\frac{de_s}{dT} = \frac{e_s L_v m_w}{R^* T^2}. \quad (1.8)$$

Under the assumption that the latent heat is constant (L_v and L_s vary by less than 5% over temperatures of meteorological interest), Eq. (1.8) can be integrated to yield:

$$e_s = e_s(T_0) \exp \left[\frac{L_v m_w}{R^*} \left(\frac{1}{T_0} - \frac{1}{T} \right) \right], \quad (1.9)$$

where $e_s(T_0) = 6.11$ hPa for $T_0 = 273.15$ K. The saturation vapor pressure therefore increases exponentially with increasing temperature, this is the reason why warm air generally contains much more water vapor than cold air. Although the integrated Clausius-Clapeyron equation is a good approximation of the maximum vapor pressure near the surface of water bodies, it cannot easily be applied to the equilibrium of water droplets in clouds and the surrounding water vapor. Corrections for meteorological application can be found, e.g., in Laube and Höller [1988]. If the air is very clean, supersaturation of several hundred percent is possible. On the other hand, condensation can occur on hygroscopic nuclei at relative humidities much lower than 100% - for example, at about 75% for sodium chloride. In practice, the available supply of condensation nuclei allows condensation to begin very close to the saturation vapor pressure at thermodynamic equilibrium [cf. Handel, 1996].

1.2 The hydrological cycle

Water vapor is constantly cycling through the atmosphere, evaporating from the surface, condensing to form clouds, and subsequently returning to the earth as precipitation. The global water or hydrological cycle traces the largest movement of any substance on earth, and is a consequence of the conservation of water substance. The exchanges of moisture and heat between the atmosphere and the Earth's surface fundamentally affect the dynamics and thermodynamics of the climate system. About 50% of surface cooling results from evaporation, the release of latent heat of condensation in clouds provides ~30% of the thermal energy that drives the Earth's circulation [Chahine, 1992].

1.2.1 Reservoirs and fluxes

Estimates of contents of major reservoirs and transfer rates between them are shown in Fig. 1.1. The oceans are the dominant reservoir in the global water cycle, holding almost 96% ($1.4 \cdot 10^{21}$ kg), about 4% ($59.1 \cdot 10^{18}$ kg) reside on land (predominantly in the form of ice caps), whereas the atmosphere holds only 0.001% ($15.5 \cdot 10^{15}$ kg).

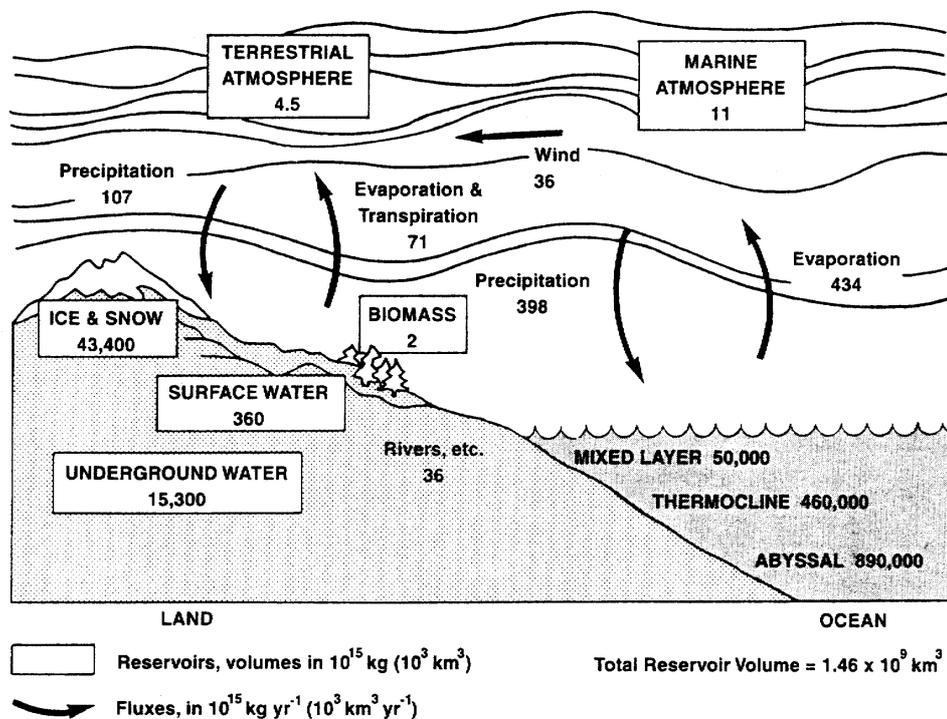


Fig. 1.1: The global hydrological cycle, showing estimates of contents of major reservoirs and fluxes (rates of transfer) between them [from Bucci Mockler, 1995].

The total amount of water substance on earth does not vary appreciably, although some new water is being produced by volcanoes and hot springs. However, most steam ejected by volcanoes is actually either rainwater that in earlier times had saturated the upper layers of rock, or seawater that was trapped at the time the marine sediments were deposited. A certain amount of water vapor is destroyed in the upper atmosphere through photodissociation by solar radiation, on the other hand, water vapor is produced by oxidation of methane, and a small amount enters the atmosphere by impacts of comets. On the time scales up to millennia, all these effects are negligible so that the total amount of water (in all phases) on earth can be considered to be constant [Peixoto and Oort, 1992].

1.2.2 Mean residence times

The global water content in the atmosphere is equivalent to a uniform layer of about 30 mm covering the globe. As the global average precipitation is about 1 m annually, the atmosphere recycles its entire water content ~ 33 times per year, yielding a mean atmospheric residence time of ~ 11 days. This rapid turnover, combined with the variation of temperature with height and latitude, causes water vapor to be distributed unevenly in the atmosphere. In contrast, the mean residence time for the oceans as a whole is over 3000 years (increasing from weeks in the surface layer to millennia in the abyssal). The mean residence time of water on land can range from a few years (in vegetation and soils) to more than 10 000 years (in aquifers and icecaps). These two domains of residence time, days-to-weeks and decades-to-millennia, control the Earth's climate in two distinct ways. The fast regime determines the amplitude and regional patterns of climate (and climate change). The slow regime modulates the transient response of the climate system and introduces considerable delay [cf. Chahine, 1992].

The continuous exchange of water among the reservoirs, as shown in Fig. 1.1, occurs mainly through evaporation and precipitation. The driver for this exchange is the Sun's differential heating, which varies with latitude, but the exchange pathways are controlled by surface properties and atmospheric and ocean circulation. When an energy imbalance occurs in the atmosphere or at the surface, the atmosphere-surface system reacts to re-establish the balance. In the atmosphere, balance is most efficiently re-established by means of transport of latent heat created by evaporation and released by condensation.

Evaporation occurs mainly at the surface of the oceans and continents leading to a transfer of latent heat into the atmosphere, that is released as sensible heat where water condenses. In the subtropical and polar regions evaporation exceeds precipitation. The water vapor, released mainly over the subtropical oceans is continuously transported equatorward and poleward to regions with excess of precipitation over evaporation, associated with the intertropical convergence zone (ITCZ) and with baroclinic perturbations along the polar front [see, e.g., Peixoto and Oort, 1992].

1.2.3 The Classic equation of hydrology

The classic equation of hydrology is obtained as a water balance requirement for the terrestrial branch of the hydrological cycle. Applying the principle of continuity to a specific region, the balance equation can be written as [Peixoto and Oort, 1992]:

$$S = P - E - R_o - R_u, \quad (1.10)$$

where S is the rate of storage of water, P is the precipitation rate (in liquid and solid phase), E is the evaporation rate (including evapotranspiration over land and sublimation over snow and ice), R_o is the surface runoff, and R_u is the subterranean runoff. For a large land region, the net subterranean runoff is usually small, for long periods of time S tends to be small compared to the other terms, so that equation (1.10) can be written as:

$$\{\bar{E}\} = \{\bar{P}\} - \{\bar{R}_o\}, \quad (1.11)$$

where $\{ \}$ denotes a space average and the overbar a time average.

Within the conceptual framework of climate research, the hydrological cycle can be divided [following Chahine, 1992] into five components: clouds and radiation, atmospheric moisture, precipitation, ocean fluxes, and land surface processes.

1.3 Clouds and radiation

Water vapor has the most significant influence on the transfer of infrared radiation and clouds have the most significant influence on the transfer of sunlight. Clouds play an equally important role as water vapor in the radiation budget even though they represent only $\sim 1\%$ of the water in gaseous phase [cf. Rossow, 1996]. Although globally there is a near balance in the net radiation at the top of the atmosphere when averaged over seasons, the regional distribution shows nonuniformities on all scales. These nonuniformities

represent sources and sinks of energy that modulate the circulation of the atmosphere and the oceans. Clouds play multiple roles as scatterers and absorbers of radiation. They reflect incoming solar radiation, reducing the direct solar energy input to the environment, but at the same time trap part of the Earth's emitted energy, reducing the net outflow of heat energy to space. Clouds therefore influence evaporation from the surface, which in turn influences cloud formation. Precipitation from clouds, on the other hand, determines soil moisture and evaporation rates. There is general agreement that the current, annual, global mean effect of clouds is to cool the climate system [cf. Chahine, 1992].

1.3.1 Cloud climatology

The average cloud cover of the entire Earth is about 60%. Cloud cover in the Southern Hemisphere is slightly greater than in the Northern Hemisphere, and over oceans it is greater than over land [Coakley, 1996]. Cloud cover is high in the equatorial belt, associated with the strong convection in the ITCZ, several typical cloud streamers extend in the western Pacific and northeast of Australia. Intense high level clouds above India, Indonesia, northern Australia and the western equatorial Pacific define the *stratospheric fountain*, where could tropospheric air enters the stratosphere with energy, provided by intense convective tropical rain [Chahine, 1992].

Pronounced minima in cloudiness in subtropical regions are caused by predominantly downward motion of air. The mean cloud cover increases markedly from the subtropics poleward, reaching maximum values (of more than 80% in the annual mean) near 50° - 60° latitude associated with the polar fronts. Relatively large amounts of clouds occur in the stratus regimes on the eastern sides of the Pacific and Atlantic oceans. The lowest values of cloud cover are observed over the continental deserts [Peixoto and Oort, 1992]. In general, there is a strong correlation between the distribution of water vapor and clouds.

1.3.2 Cloud-climate feedback

The role of clouds in climate change is a subject of intense scientific research and debates. The net effect of clouds on the radiation budget depends on a wide range of cloud properties including their microphysical characteristics. Further important factors are cloud height and albedo. Model results indicate that an increase in the effective cloud top height (with cloud cover and albedo fixed) leads to an increase in surface temperatures at all

latitudes [cf. Chahine, 1992]. Thus, if doubling CO₂ leads to higher clouds, then clouds will produce a positive feedback. If, however, the resulting clouds have larger water content, their albedo will be higher and they will reflect more solar radiation, thereby providing a negative feedback leading to a cooling of the surface. The magnitude, and even the sign, of the cloud feedback on the greenhouse warming are not known at present.

1.4 Water vapor in the atmosphere

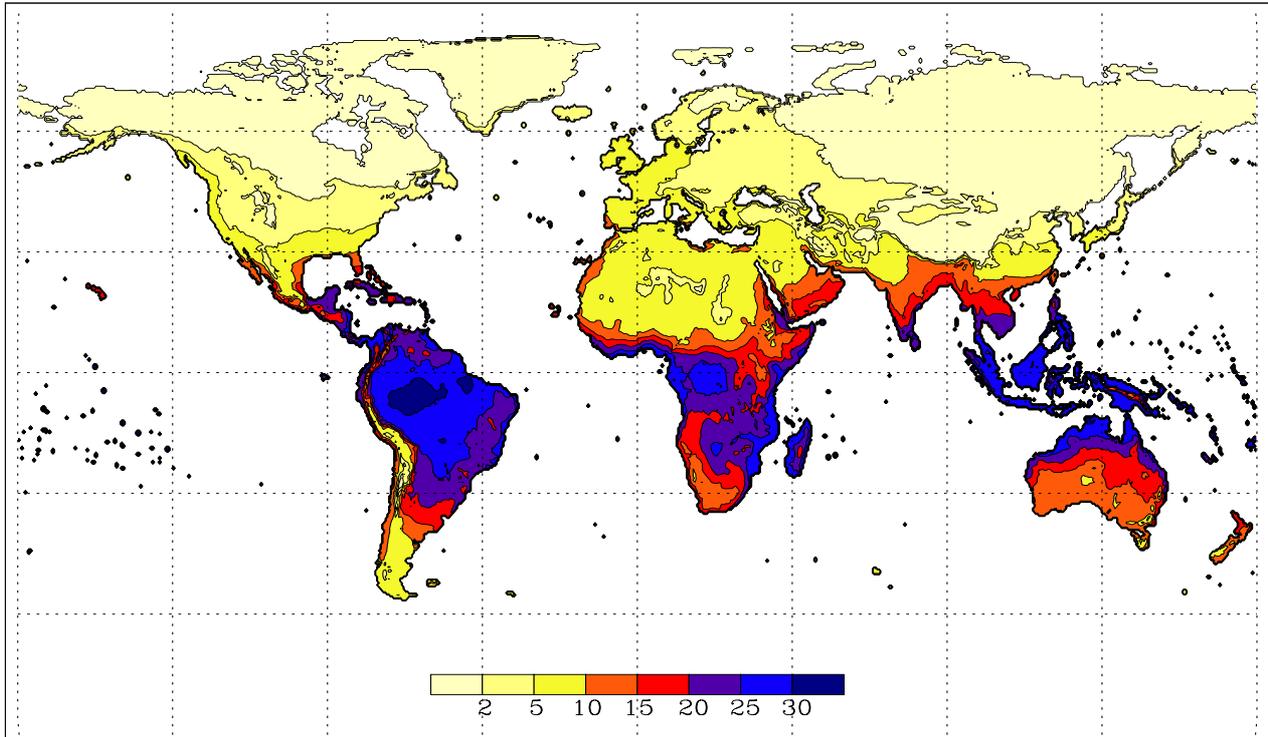
The composition of the atmosphere is quite uniform up to very high altitudes with respect to all major constituents except water vapor. In the tropics near the ocean surface, water vapor makes up almost 4% of the atmosphere per volume. Near the tropopause, the molecular fraction drops to only few parts per million [cf. Handel, 1996].

1.4.1 Mean spatial distribution of water vapor

The spatial distribution of the water vapor partial pressure over land areas in winter and summer is shown in Fig. 1.2. Since the capacity of the atmosphere to retain water depends strongly on temperature (cf. Eq. (1.9)), the global pattern of vapor pressure resembles the temperature pattern, with major exceptions in the desert regions. As expected the highest values (corresponding to humidities of $\sim 20\text{g/kg}$) are found in the equatorial regions, following the annual movement of the ITCZ. With exception of the desert regions, there is a continuous decrease of vapor pressure with latitude down to very low values of less than 1 hPa (corresponding to humidities of less than 1g/kg) over the polar regions (see Fig. 1.4) [cf. Peixoto and Oort, 1992].

Over the oceans, the relative humidity shows an almost constant value of $\sim 80\%$, the absolute humidity therefore closely follows the spatial distribution of temperature, governed by the Clausius-Clapeyron equation [cf. Liljequist and Cehak, 1984]. The most humid region corresponds to the so-called *oceanic warm pool* in the western equatorial Pacific, where the highest sea surface temperatures are found.

January mean surface water vapor partial pressure (1961–1990) [hPa]



July mean surface water vapor partial pressure (1961–1990) [hPa]

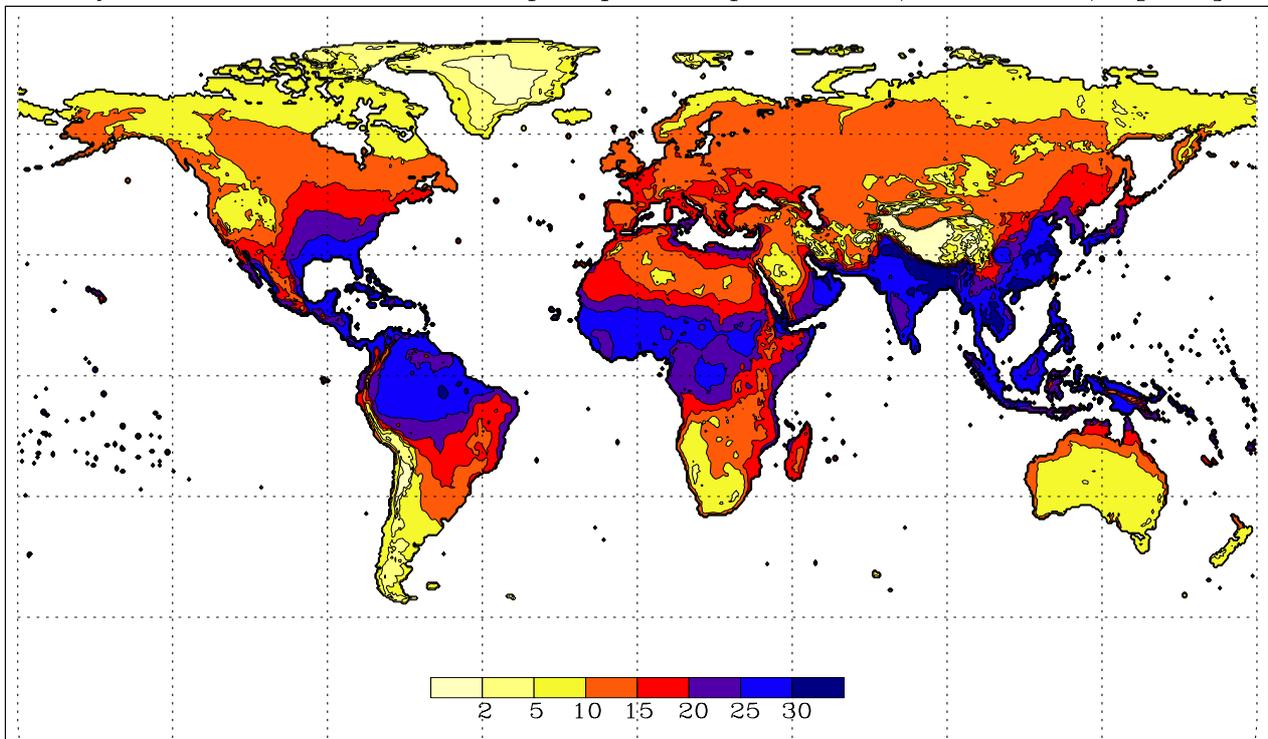


Fig. 1.2: Mean surface water vapor partial pressure for the period 1961–1990, IPCC climatological data for global land areas excluding Antarctica [data from <http://ipcc-ddc.cru.uea.ac.uk>].

The relationship between precipitable water and sea surface temperature or surface air temperature T_s can be expected to resemble an exponential form, due to the *Clausius-Clapeyron equation* (see Eq. (1.9)), therefore a relation of the form

$$\ln PW = a + bT_s \quad (1.12)$$

should apply. Analyzing radiosonde data from 50 globally distributed stations for the period 1973 - 1990, Gaffen *et al.* [1992a] found that the linear correlation coefficient between $\ln PW$ and T_s has a value of 0.94.

The spatial distribution of the annual-mean precipitable water is shown in Fig. 1.3. The general decrease from equator to the poles is a reflection of the global distribution of temperature. Departures from the zonal symmetry are associated with the physiography of the Earth's surface, and are apparent in both hemispheres. Precipitable water is generally higher over the oceans than over the continents.

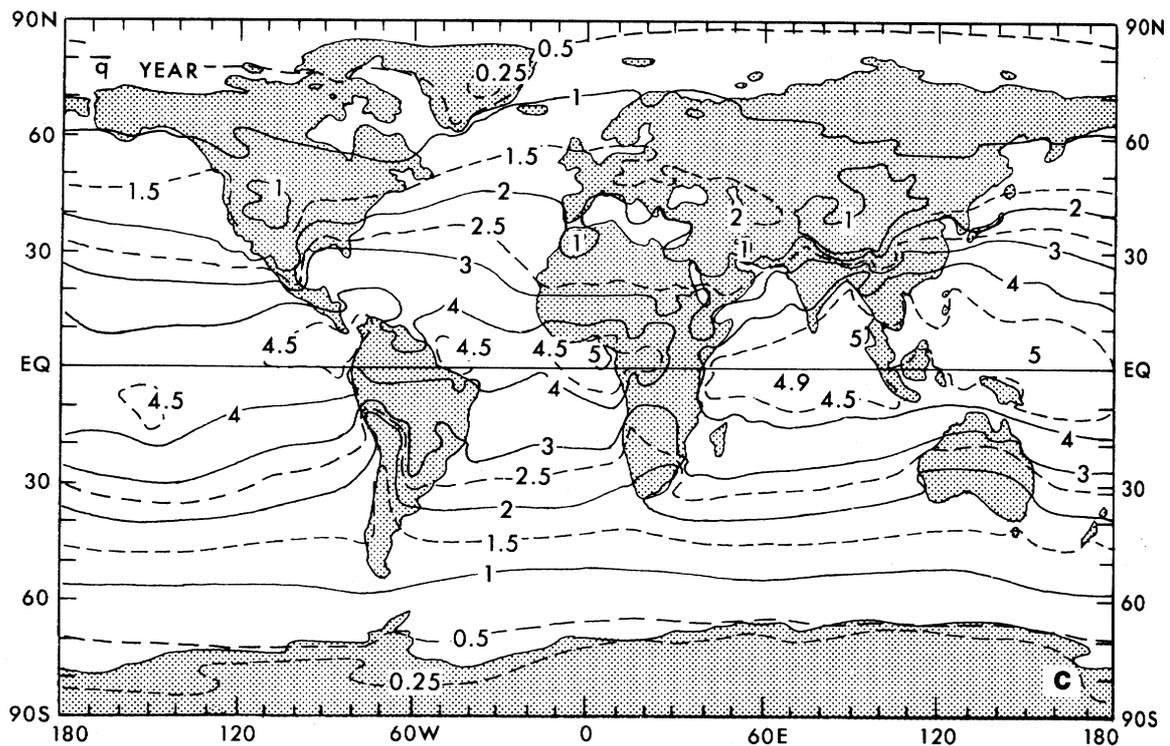


Fig. 1.3: Global distribution of the annual-mean and vertical-mean specific humidity in g/kg, which is equivalent to precipitable water in units 10 kg/m^2 or 1 cm [from Peixoto and Oort, 1992].

The deflection of the isolines near the western and eastern coasts of the continents is reinforced by the topography and the presence of warm and cold ocean currents. The distribution over the Southern hemisphere is practically zonal, since the area covered by oceans exceeds by far that covered by continents. The lowest values occur over the subpolar and polar regions (with $PW < 5 \text{ kg/m}^2$). The precipitable water over desert areas is considerably smaller than the corresponding zonal average, mainly due to subsidence in the descending part of the Hadley-cell.

The effect of orography on precipitable water is illustrated by relatively dry areas (frequently with $PW < 10 \text{ kg/m}^2$) over the major mountain regions, such as the highlands of Tibet and Bolivia (see also Fig. 1.2). The effects of topography and the land-sea contrast in the Southern Hemisphere are shown by the dipping of the 20 kg/m^2 isoline towards lower latitudes [cf. Peixoto and Oort, 1992].

Relative humidity is on the order of 75% in the equatorial region and has values of 30% - 40% in the dry subtropical regions.

1.4.2 Mean height distribution of water vapor

Mixing ratio of water vapor is nearly constant in the so-called *mixed* or *Prandtl layer*, which is some tens of m's thick. Above, the mixing ratio decreases rapidly with height, almost following an exponential law (see Fig. 1.4). More than 50% of the water vapor is concentrated below the 850 hPa surface, while more than 90% is confined to the layer below 500 hPa [Peixoto and Oort, 1992]. The upper troposphere in middle and high latitudes is a factor 2 - 4 drier during austral winter than during boreal winter [Kelly *et al.*, 1991]. In the lower stratosphere, the mixing ratio is usually in the range of 3 to 8 parts per million [Handel, 1996]. Above the lower stratospheric minimum, the mixing ratio tends to increase slightly with height, due to oxidation of methane, a process which is most effective in the upper stratosphere [cf. Webster, 1994].

Relative humidity also decreases with height but weaker than specific humidity, as water vapor concentration closely follows temperature. The relative humidity ranges from values on the order of 70% - 80% near the surface to 30% - 50% near the 400 hPa level. The latitude gradients tend to increase with height, reflecting the influence of the rising and sinking branches of the mean meridional circulation. Thus in the annual mean, strong sinking motions near 20° - 30° latitude lead to a minimum in relative humidity at those

latitudes in each hemisphere, whereas the ascending motions in the ITCZ give rise to the maximum just north of the equator. The seasonal range has a maximum of 15% - 20% near 15° latitude in each hemisphere at the 500 hPa level [Peixoto and Oort, 1992].

The vertical profile of water vapor furthermore influences the seasonal variation of the local greenhouse effect [see Bony and Duvel, 1994].

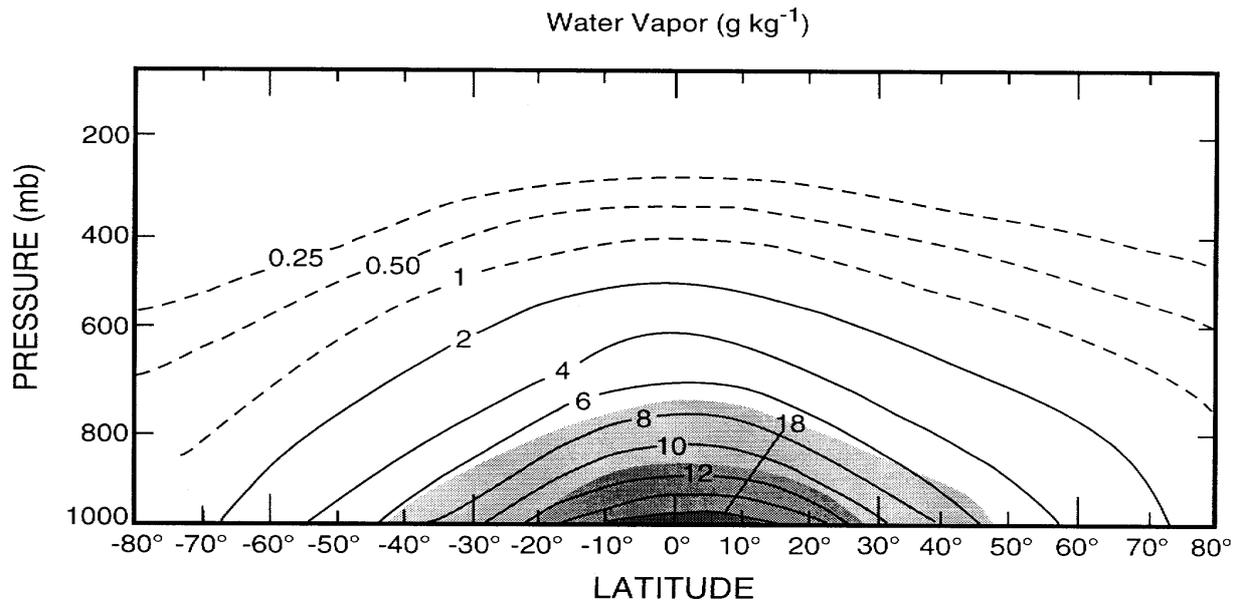


Fig. 1.4: Zonal-mean mixing ratio of water vapor (contoured) and water vapor density (shaded) as functions of latitude and pressure (height). The shaded levels correspond to 20, 40, and 60% of the maximum value [from Salby, 1996].

1.4.3 Zonal transport of water vapor

Movement of water vapor in the atmosphere represents movement of energy in the form of latent heat. The zonal transport of water vapor reflects the planetary behavior of the general circulation in the lower half of the troposphere, since the specific humidity acts as a weighting factor for the wind field. The general pattern shows eastward transport in midlatitudes (with a maximum around 40° latitude) and westward transport in the tropical belt with a bimodal annual distribution, the dominant maximum being in the winter hemisphere. The zero isoline coincides with the mean location of the centers of the subtropical anticyclones in both hemispheres. The gaseous hydrosphere as a whole moves eastward, faster than the earth, with a relative flux rate of about $0.5 \text{ m s}^{-1} \text{ g kg}^{-1}$ for the annual mean [Peixoto and Oort, 1992].

1.4.4 Meridional transport of water vapor

The meridional transport of water vapor over midlatitudes is predominantly poleward within both hemispheres throughout the year (with maxima near 40° latitude), mainly accomplished by baroclinic lows associated with the polar front and by stationary eddies, such as subpolar lows and subtropical anticyclones. The lower branches of the Hadley-cell are very effective in transporting moisture into the ITCZ and give the largest contribution to the total meridional flux of water vapor. During Northern summer, there occurs a cross equatorial water vapor flux into the Northern Hemisphere of about $19 \cdot 10^8$ kg/s, during Northern winter, there occurs a corresponding net flux into the Southern Hemisphere of about $14 \cdot 10^8$ kg/s. For the year as a whole, there is a net flux into the Northern hemisphere of $3.2 \cdot 10^8$ kg/s, balancing an annual excess of precipitation over evaporation of 39 mm/year within the Northern hemisphere [Peixoto and Oort, 1992].

1.4.5 Vertical transport of water vapor

The vertical transport of water vapor in the atmosphere plays an essential role in the hydrological cycle, as it links the terrestrial and the atmospheric branches. Maximum upward transport occurs in the equatorial region, associated with the ascending branches of the Hadley-cell, whereas the upward flux in middle and high latitudes is connected with quasi-stationary low pressure systems. The centers of maximum downward flux occur mainly in the eastern parts of the subtropical anticyclones over the oceans where subsidence prevails [cf. Peixoto and Oort, 1992].

Deep precipitating cumulus updrafts lift moisture from near the surface and inject some of it into the upper troposphere in the form of vapor or ice, a process known as *detrainment*. The clear environment surrounding the updrafts slowly sinks in response, this compensating subsidence dries the atmosphere by bringing air with low humidity down from above. Large-scale eddies transfer water vapor upwards from low altitudes, especially in mid-latitude storms where moist air raises near warm fronts and drier air sinks behind cold fronts [cf. Del Genio *et al.*, 1991]. In the subtropics, drying by Hadley-cell subsidence is partly offset by eddy-moistening. Elsewhere, both mean meridional circulation and eddies moisten the upper troposphere and are balanced primarily by stratiform condensation drying [Del Genio *et al.*, 1994].

1.4.6 Water vapor variations

Variations in the atmospheric water vapor field occur on timescales from a few minutes to decades. Seasonal variations are larger in the northern hemisphere than in the Southern Hemisphere, as the corresponding temperature fluctuations are stronger. The latter are stronger because of the Northern Hemisphere's larger fraction of land, which has a lower heat capacity than the oceans and thus responds faster to variations. The largest variations occur at relatively low latitudes over the continents, variations associated with the Asian monsoon are also very pronounced (see Fig. 1.2).

Relative humidity is approximately constant throughout the year at middle and high latitudes, except at locations influenced by the monsoon circulation. In low-latitude regions, horizontal and vertical moisture advection, combined with low-amplitude temperature changes, allows for substantial seasonal variations in relative humidity, especially above the planetary boundary layer. The annual cycles of precipitable water are consistent with the relative humidity patterns. In middle and high latitudes PW follows the temperature variations. In the tropics the annual cycle is closer related to mid-tropospheric variations in relative humidity and reaches maximum values during the local rainy season when deep convection is strong [for details see Gaffen *et al.*, 1992b].

On multiyear timescales, large changes of sea surface temperatures in the tropical Pacific associated with the El Niño-Southern Oscillation (ENSO) cause shifts in the water vapor distribution [Bucci Mockler, 1995].

1.4.7 Water vapor trends

The most recent global estimates of long-term changes in tropospheric water vapor show a general increase in precipitable water during the period 1973-1990, with the largest trends in the tropics, where increases as large as 13% per decade are found. Regions of moisture increase are associated with regions of rising temperature over the same period [Bucci Mockler, 1995].

Making measurements with balloon-borne frost-point hygrometers over a 14-year period (1981-1994), Oltmans and Hofmann [1995] found an increase in water vapor within the lower stratosphere over Boulder, Colorado, of up to 0.8% per year. Doubling of water vapor in the stratosphere can be expected to lead to a 1°C increase in surface temperature.

1.4.8 Current water vapor measurements

A detailed description of various measurement systems currently in use can be found in Bucci Mockler [1995] and Rossow [1996]. A brief summary is given below. Future systems, in particular future spaceborne systems, will certainly improve water vapor data quality.

- Routine **surface meteorological observations** by hygrometers provide long-term records of high quality data, but measurements are restricted to the Earth's surface and the spatial coverage is nonuniform.
- Routine **radiosonde balloons** are launched once or twice per day from about 800 stations, providing data with relative good vertical resolution and quality within the lower troposphere, but poor quality at higher altitudes. The geographic coverage of balloon launch sites is relatively poor, particularly over the oceans and in the Southern Hemisphere.
- **Research soundings** using, e.g., frost point hygrometers, deliver high quality data that extend beyond the limit of radiosondes, but they are made infrequently at limited locations [see also Oltmans and Hofmann, 1995].
- Spaceborne **infrared sensors** provide total column water vapor (and some profile information with poor vertical resolution), but data are limited to cloud-free regions and can exhibit regional biases.
- Spaceborne **microwave sensors** provide total column water vapor data which are not highly influenced by clouds, the data are here restricted to ice-free ocean regions.
- Spaceborne **radio occultation** measurements provide accurate profiles of temperature and humidity with high vertical resolution and global coverage, but have still problems in delivering useful data in the planetary boundary layer (see chapter 4 for details). The method is still in the development stage.
- Spaceborne **solar occultation** measurements (e.g., SAGE II) deliver global humidity data at very high altitudes with high accuracy and vertical resolution. The coverage is limited by clouds and sampling is poor in tropical regions.
- **Ground-based GPS** measurement have the potential to provide accurate estimates of integrated water vapor under all-weather conditions (see chapter 3 for details).

1.5 Precipitation and evaporation

1.5.1 Precipitation

The average annual global precipitation is equivalent to 950 - 1150 mm per year or about 3 mm per day. Precipitation is highly variable in space, with about two thirds occurring between latitudes of 30° N and 30° S. The spatial distribution of precipitation is shown in Fig. 1.5 for January and July, the most significant feature being the strong rainfall at low latitudes associated with strong convection in the ITCZ. The Bismarck sea, e.g., which is fairly typical for the tropical western Pacific, receives about 5 m of precipitation per year resulting in an excess of precipitation over evaporation of about 2 m per year. This precipitation excess reduces the surface water density, maintaining the stable layer of the tropical *oceanic warm pool*, which in turns is responsible for the high precipitation rate [cf. Webster, 1994].

Extreme rainfall like the global maximum in Cherrapunji, India, with more than 12 m precipitation per year, is always associated with advection and orographic effects [see, e.g., Liljequist and Cehak, 1984]. During the course of the year, the ITCZ migrates north and south in phase with the solar insolation which explains the shift of the maxima in Fig. 1.5. The effect of the monsoon circulation can be clearly seen by comparing the January and July maps.

Most of the precipitation falling over North America is supplied by water vapor originating from warm waters of the Gulf of Mexico, with a deep northward intrusion of water vapor during all seasons, as the existence of the Rocky Mountains parallel to the west coast does not allow moisture from the Pacific Ocean to penetrate deeply into the American continent. An analogous situation occurs in South America with respect to the Andes. Absence of large mountains along the Atlantic coast of Europe, on the other hand, favors deep penetration of moisture and consequent precipitation into the Eurasian continent and the Mediterranean region [cf. Peixoto and Oort, 1992].

The atmospheric forcing caused by spatial variation in the release of latent heat of condensation is the main driver of the dynamics of the interactions between the atmosphere, oceans, and land.

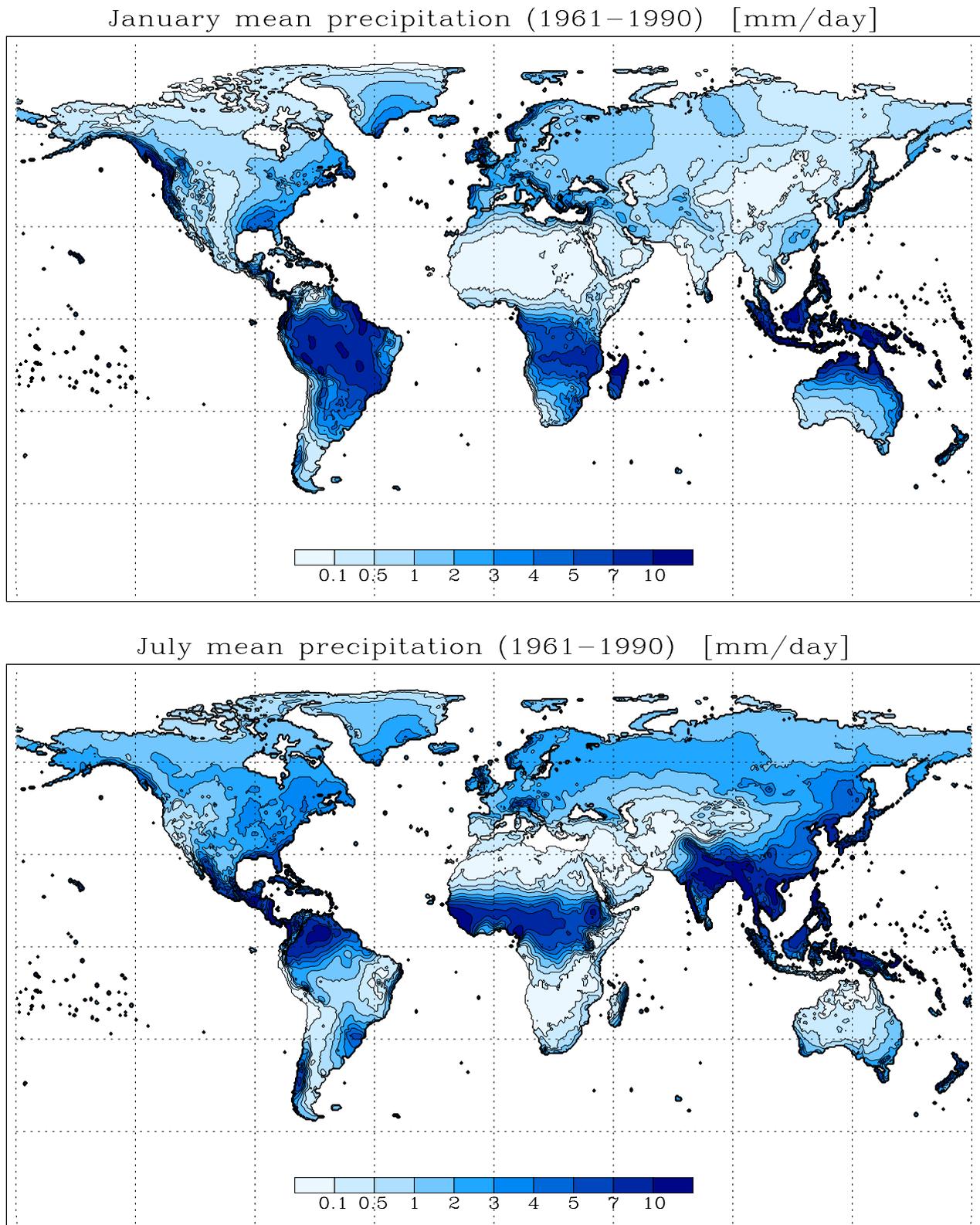


Fig. 1.5 Mean precipitation for the period 1961–1990, IPCC climatological data for global land areas excluding Antarctica [data from <http://ipcc-ddc.cru.uea.ac.uk>].

The annual variability of tropical rainfall is strongly related to the annual variability of the sea surface temperature, reflecting the strong coupling between ocean and atmosphere [cf. Chahine, 1992].

1.5.2 Evaporation

The global mean annual precipitation of about 1 m is balanced on average by the same amount of evaporation (cf. equation (1.11)). Evaporation from the oceans is higher with a mean annual value of 1.2 m it exceeds precipitation and the difference contributes to precipitation over land [see Webster, 1994 for details]. The highest values (more than 2 m per year) occur over the subtropical oceans, where the oceanic "deserts" are found. Evaporating 1 m of water per year corresponds to a mean energy sink of $\sim 80 \text{ W/m}^2$, more than 20 % of the average solar irradiance at the top of the atmosphere or one third of the average solar irradiance reaching the surface. On average, almost half of the net local energy loss from the Earth's surface to the atmosphere is by evaporation of water.

In the mean, there is excess evaporation in the Southern Hemisphere and excess precipitation in the Northern Hemisphere. The northward atmospheric transport must be balanced by southward transport in the oceans because the two major rivers that cross the equator, the Amazon and the Nile, have their headwaters in the Southern hemisphere.

1.6 Ocean fluxes and land surface processes

1.6.1 Ocean fluxes

The oceans contribute to the global hydrological cycle in two ways. First, they provide long-term memory by storing heat and releasing water vapor. Second, in the subtropics the oceans transport heat poleward at a rate equal to the heat transported by the atmosphere. Wind driven circulation is restricted to the mixed layer above the ocean thermocline, whereas in the lower three quarters of the ocean mass, motions are generated from density variations [see, Webster, 1994]. The latter result from differential heating and changes in salinity which can be generated by evaporation, precipitation, runoff, and ice melting. Salinity modulates the interactions of the oceans with the hydrological cycle.

In the thermohaline circulation of the northern Atlantic Ocean, warm surface water flows northward, thereby releasing enormous amounts of heat (about 10^{15} W) to the overlying

eastward moving air masses, which in turn greatly ameliorate winter temperatures in northern Europe. The upper ocean water becomes more dense through increased evaporation and cooling up to the point, where it sinks to the bottom and flows southward. There is increasing evidence that changes in the thermohaline circulation may have caused substantial abrupt climate changes during the last glacial period [see, e.g., Broecker, 1997]. Model simulations indicate that small changes in the hydrological cycle which cause changes in the salinity and therefore in the thermohaline circulation could have prolonged effects on the whole climate system [see, e.g., Rahmstorf, 1994].

Present estimates of the heat transport by the oceans are deficient, resulting in a discrepancy known as the *mystery of the missing 1.6 PW* ($1 PW = 10^{15} W$) in the heat balance of the climate system. The poleward transport of heat at $24^\circ N$ by the Pacific and Atlantic oceans has been estimated to be $2.0 PW$. This is comparable to the estimated value of $1.7 PW$ for the heat transport by the atmosphere. The oceans and the atmosphere together thus transport $3.7 PW$ northwards, far less than the $5.3 PW$ required to balance the Earth radiation budget according to satellite data [for details see Chahine, 1992].

1.6.2 Land surface hydrological processes

Evaporation (or evapotranspiration, if plants are considered) and precipitation over land are major components of the global hydrological cycle. Energy is needed to convert soil water into vapor. Most of the energy comes from radiation absorbed by the surface, which depends on the surface albedo. Surface albedo itself is determined by vegetation and bare soil conditions. Changes in vegetation and soil moisture alter the partition between evaporation and runoff which, in turn, changes surface conditions.

The role of plants in controlling evaporation adds to the complexity of the situation. About 65% (in the Amazon basin even more than three quarter) of the precipitation falling onto the land's surface is returned to the atmosphere via evapotranspiration and is ultimately recycled into subsequent precipitation. Evapotranspiration is intricately associated with CO_2 intake and vegetation growth and also serves to protect plants from overheating [see, e.g., Chahine, 1992].

1.7 The role of water vapor in the climate system

1.7.1 Influence on the radiation balance

Without the abundance of water and the variety of its forms in the Earth system, the average global temperature would reside far below 0°C [cf. Houghton *et al.*, 1996]. Water vapor has a greater effect than any other substance on the radiation balance of the atmosphere, it is the dominant absorber of infrared radiation.

As a triatomic molecule, water vapor is capable to undergo simultaneous vibration-rotation transitions, which produce a clustering of absorption lines at infrared wavelengths, appearing as continuous bands at coarse resolution. Vibrational bands occur at higher energies than rotational bands, as the energies of vibrational quanta are about two orders of magnitude larger than those of rotational quanta. The absorption spectrum of water vapor has strong bands centered near 1.4, 1.8, 2.7, and 6.3 μm respectively, caused by vibration-rotation modes. A series of pure rotational absorption bands begins at 11 μm [see Salby, 1996, for details].

Although most of the water vapor resides in the lower levels of the atmosphere (almost 50% below 1500 m), the small amounts that exist at higher altitudes have a disproportional large effect on the radiation balance. A 5% increase in the relative humidity at 10 km has about the same effect on the radiation balance of the atmosphere as a 5 % increase in the relative humidity at 1 km, even though the latter changes involves much more water vapor [cf. Handel, 1996, and Peixoto and Oort, 1992].

1.7.2 The water vapor - greenhouse feedback

Independent of seasons, the water content of the atmosphere increases throughout the troposphere as a function of increased sea surface temperature. The water vapor-greenhouse feedback, which means warming from increased greenhouse gases leading to an increase of water vapor and therefore even more warming, is considered to be the strongest positive feedback in the climate system [Gaffen *et al.*, 1992a]. This positive feedback could amplify the temperature change due to a doubling of CO₂ by some 60% - 70% [cf. Houghton *et al.*, 1996] and it is an important feature of all General Circulation Model (GCM) simulations to date.

Model results show [see, e.g., Del Genio *et al.*, 1991 and 1994] that substantial increase in the concentration of water vapor should be expected at all latitudes as climate warms due to increasing greenhouse gases. The greatest fractional enhancement occurs in the upper troposphere, although the greatest absolute increase occurs near the surface. The upper troposphere moistens with warming, mostly due to increased eddy transport. The strength of the eddies themselves declines in warmer climate because a smaller meridional temperature gradient reduces the available potential energy that drives baroclinic instability. Although the eddies are weaker, they transport more moisture upward because the vertical gradient of specific humidity is stronger.

A deeper understanding of the mechanisms distributing water vapor through the atmosphere and of water vapor's effects on atmospheric radiation and circulation is vital to estimate long-term changes in climate.

2 PHYSICS OF REFRACTIVITY

Information accessible to a distant observer is almost exclusively sensed by means of electromagnetic radiation, either by passive methods in the form of exploiting thermal or non-thermal emissions or of reflected solar radiation, or by active methods involving a natural or artificial localized radiation source generating electromagnetic radiation. The GNSS-based remote sounding techniques employed for this work belong to the latter category of active methods and rest on radiation-matter-interaction processes the understanding of which requires good knowledge of electromagnetics. It seems therefore pertinent to start with reviewing the basics of electromagnetic wave propagation [for details see, e.g., Jackson, 1985; Jones, 1986, Hanel *et al.*, 1992; Born and Wolf, 1993].

2.1 Basic properties of the electromagnetic field

The state of excitation which is established in space by the presence of electric charges and currents constitutes the electromagnetic field. It is represented, within the framework of classical electrodynamic theory [e.g., Jackson, 1985], by the *electric vector field* \mathbf{E} [V/m] and the *magnetic vector field* \mathbf{H} [A/m]. To describe the effect of the field on material objects, it is necessary to introduce a second set of vectors: the electric current density \mathbf{j} [A/m²], the *electric displacement field* \mathbf{D} [As/m²], and the *magnetic induction field* \mathbf{B} [Vs/m²].

2.1.1 Maxwell's equations

The space and time derivatives of the vectors field are related by *Maxwell's equations*, which hold at every point in the neighborhood of which the physical properties of the medium are continuous. In differential form, Maxwell's equations read:

$$\operatorname{div} \mathbf{D} = \rho_{el} , \quad (2.1)$$

$$\operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} , \quad (2.2)$$

$$\operatorname{div} \mathbf{B} = 0 , \quad (2.3)$$

$$\operatorname{curl} \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{j} , \quad (2.4)$$

where ρ_{el} is the electric charge density and t is the time. Equation(2.1) may be regarded as definition equation for the electrical displacement \mathbf{D} , while Eq. (2.3) states the nonexistence of magnetic charges or monopoles. Equation (2.2) represents *Faraday's law* of induction and Eq. (2.4) represents *Ampère's law* of magnetic field generation.

2.1.2 The material equations

To allow a unique determination of the vector fields from a given distribution of currents and charges, Maxwell's equations must be supplemented by relations for \mathbf{D} , \mathbf{B} , and \mathbf{j} , which describe the behavior of substances co-existing with an electromagnetic field. These relations are known as *material equations* or *constitutive relations*.

Imposing an electric field on matter tends to separate the electrons and atomic nuclei and creates therefore an induced electric dipole moment, \mathbf{p} (matter is considered in this case as a *dielectric medium*). The total charge density ρ_{tot} thus consists of free charges (ρ_{el}) and bounded charges (ρ_{bound}), caused by polarization,

$$\rho_{tot} = \rho_{el} + \rho_{bound} . \quad (2.5)$$

ρ_{bound} can be expressed as [see, e.g., Jackson, 1985]:

$$\rho_{bound} = -\operatorname{div} \mathbf{P} . \quad (2.6)$$

where \mathbf{P} is the *electric polarization*, which is the mean electric dipole moment per volume. From the *Gauss theorem* follows the *Poisson equation*:

$$\operatorname{div} \mathbf{E} = \frac{\rho_{tot}}{\epsilon_0} , \quad (2.7)$$

where ϵ_0 is the *permittivity of vacuum* ($\epsilon_0 = 8.854\,187\,817 \cdot 10^{-12} \text{ AsV}^{-1}\text{m}^{-1}$) [CRC, 1993].

With Eqs. (2.5) and (2.6), Eq. (2.7) becomes:

$$\operatorname{div}(\epsilon_0 \mathbf{E} + \mathbf{P}) = \rho_{el} , \quad (2.8)$$

which can be seen as the definition of the electric displacement \mathbf{D} (see Eq. (2.1)):

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} . \quad (2.9)$$

Analogous considerations for \mathbf{H} and \mathbf{B} yield:

$$\mathbf{B} = \mu_0 \mathbf{H} + \mathbf{M} , \quad (2.10)$$

where μ_0 is the *permeability of vacuum* ($\mu_0 = 4\pi \cdot 10^{-7} \text{ VsA}^{-1}\text{m}^{-1}$) [CRC, 1993], and \mathbf{M} is the *magnetic polarization* (the mean magnetic dipole moment per volume).

In free space, \mathbf{D} and \mathbf{B} are thus given by:

$$\mathbf{D} = \varepsilon_0 \mathbf{E}, \quad (2.11)$$

$$\mathbf{B} = \mu_0 \mathbf{H}. \quad (2.12)$$

In a *linear* medium (like the Earth's atmosphere at electrical field strengths of practical interest), the induced polarization is a linear function of the imposed electric field strength.

\mathbf{P} is therefore related to \mathbf{E} via:

$$\mathbf{P} = \varepsilon_0 \tilde{\chi}_e \cdot \mathbf{E} \quad (2.13)$$

In general, the *electric susceptibility* $\tilde{\chi}_e$ is a symmetric tensor of second rank, and \mathbf{P} and \mathbf{E} have different directions. Non-linear effects have only to be expected for exceptionally strong fields (like Laser fields at short range). In an isotropic medium (i.e. when the physical properties at each point are independent of direction), $\tilde{\chi}_e$ degenerates to a scalar, χ_e , and Eq. (2.9) can be written as:

$$\mathbf{D} = \varepsilon_0 (1 + \chi_e) \mathbf{E} = \varepsilon_0 \varepsilon_r \mathbf{E} = \varepsilon \mathbf{E}, \quad \text{with} \quad (2.14)$$

$$\varepsilon = \varepsilon_0 \varepsilon_r, \quad (2.15)$$

where ε is the *permittivity* (or *dielectric constant* in a homogenous medium) and $\varepsilon_r = 1 + \chi_e$ is the dimensionless *relative permittivity*, which depends on the material and is unity for vacuum.

Analogous considerations for \mathbf{M} and \mathbf{H} yield:

$$\mathbf{M} = \mu_0 \tilde{\chi}_m \cdot \mathbf{H}, \quad (2.16)$$

and

$$\mathbf{B} = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 \mu_r \mathbf{H} = \mu \mathbf{H}, \quad \text{with} \quad (2.17)$$

$$\mu = \mu_0 \mu_r, \quad (2.18)$$

where χ_m is the *magnetic susceptibility*, $\mu_r = 1 + \chi_m$ is the dimensionless *relative permeability*, and μ is the (absolute) *permeability*.

The electric current density \mathbf{j} is related to electric field strength \mathbf{E} via the electric conductivity σ ($\Omega^{-1} \text{m}^{-1}$), which is a scalar for isotropic media, but in general a tensor of second rank:

$$\mathbf{j} = \vec{\sigma} \cdot \mathbf{E}. \quad (2.19)$$

Equation (2.19) is the differential form of *Ohm's law*.

2.1.3 Restrictions for the case of the Earth's lower atmosphere

The lower atmosphere (troposphere and stratosphere, at least up to ~ 50 km) is a neutral ($\rho_{el} = 0$), and isotropic medium, and has a negligible electric conductivity ($\sigma = 0$) yielding $\mathbf{j} = 0$. The lower atmosphere has also negligible magnetic susceptibility ($\chi_m = 0$, $\mu_r = 1$), but this simplification is, for didactical purposes, not introduced before the end of subsection 2.1.4.

With the simplifications concerning $\rho_{el} (= 0)$ and $\mathbf{j} (= 0)$, the material equations become:

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (2.20)$$

$$\mathbf{B} = \mu \mathbf{H}, \quad (2.21)$$

$$\mathbf{j} = 0, \quad (2.22)$$

and the Maxwell equations can be written (using now explicitly the *Nabla operator* ∇) as:

$$\nabla \cdot \mathbf{E} = 0, \quad (2.23)$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \quad (2.24)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (2.25)$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}. \quad (2.26)$$

2.1.4 The wave equation and the refractive index

Maxwell's equations relate the vector fields by means of simultaneous differential equations. On elimination we can obtain differential equations, which each of the vectors must separately satisfy. Applying the curl operator ($\nabla \times$) on Eq. (2.24) and interchanging the order of differentiation with respect to space and time (which can be done for a slowly varying medium like the atmosphere is one at frequencies of practical interest) yields:

$$\nabla \times \nabla \times \mathbf{E} = -\mu \frac{\partial}{\partial t} (\nabla \times \mathbf{H}). \quad (2.27)$$

Insertion of Eq. (2.26) and application of the vector identity (Δ denotes the *Laplace operator*, $\Delta \equiv \nabla^2$)

$$\nabla \times \nabla \times \mathbf{E} \equiv \nabla (\nabla \cdot \mathbf{E}) - \Delta \mathbf{E}, \quad (2.28)$$

leads (as $\nabla \cdot \mathbf{E} = 0$) to the equation:

$$\Delta \mathbf{E} = \epsilon \mu \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad (2.29)$$

Analogous considerations for \mathbf{H} yield:

$$\Delta \mathbf{H} = \epsilon \mu \frac{\partial^2 \mathbf{H}}{\partial t^2}. \quad (2.30)$$

The partial differential equations (2.29) and (2.30) are standard *wave equations*. Considering plane waves, the solutions have the form:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \text{ and } \mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (2.31)$$

where \mathbf{k} is the wave number vector, pointing in the direction of wave propagation. The angular frequency, ω (rad/s) and the angular wave number, k (rad/m), are defined as (with $k = |\mathbf{k}|$):

$$\omega = 2\pi f, \quad (2.32)$$

$$k = \frac{2\pi}{\lambda}, \quad (2.33)$$

where f is the frequency (Hz) and λ is the wavelength (m). Inserting Eq. (2.31) into the Maxwell equations ((2.23) - (2.26)) yields

$$i \mathbf{k} \cdot \mathbf{E}_0 = 0, \quad (2.34)$$

$$i \mathbf{k} \times \mathbf{E}_0 = i \mu \omega \mathbf{H}_0, \quad (2.35)$$

$$i \mathbf{k} \cdot \mathbf{H}_0 = 0, \quad (2.36)$$

$$i \mathbf{k} \times \mathbf{H}_0 = -i \epsilon \omega \mathbf{E}_0, \quad (2.37)$$

which shows that the field vectors \mathbf{E} and \mathbf{H} are perpendicular to each other and that both are perpendicular to \mathbf{k} . Electromagnetic waves are thus transverse waves.

Inserting (2.31) into the wave equation (2.29) using the vector identity

$$\mathbf{k} \times \mathbf{k} \times \mathbf{E} = (\mathbf{k} \cdot \mathbf{E})\mathbf{k} - (\mathbf{k} \cdot \mathbf{k})\mathbf{E}, \quad (2.38)$$

and the orthogonality of \mathbf{k} and \mathbf{E} (Eq. (2.34)) yields:

$$k^2 = \epsilon \mu \omega^2. \quad (2.39)$$

With the definition of the phase velocity c ,

$$c = \frac{\omega}{k}, \quad (2.40)$$

we see that monochromatic electromagnetic waves (as, e.g., the GPS carrier signals L1 and L2) propagate in a medium with the velocity:

$$c = \frac{1}{\sqrt{\epsilon\mu}} = \frac{1}{\sqrt{\epsilon_0\mu_0\epsilon_r\mu_r}}. \quad (2.41)$$

In vacuum, the wave velocity c_0 is therefore given by:

$$c_0 = \frac{1}{\sqrt{\epsilon_0\mu_0}}, \quad (2.42)$$

and $c_0 = 299\,792\,458$ m/s [CRC, 1993].

In geometric optics, the refractive index of a medium (n) is defined as the ratio of the speed of light in vacuum (c_0) to that in the medium (c):

$$n = \frac{c_0}{c} = \sqrt{\epsilon_r\mu_r}, \quad (2.43)$$

which is known as *Maxwell's relation*. The Earth's atmosphere has unity relative permeability and Eq. (2.43) becomes therefore:

$$n = \sqrt{\epsilon_r}. \quad (2.44)$$

It should be noted that n and ϵ_r are functions of frequency, ω , the Maxwell relation is thus, more generally, written in the form:

$$n(\omega) = \sqrt{\epsilon_r(\omega)}. \quad (2.45)$$

However, only for some materials the optical value of n and the radio frequency value of ϵ_r are obeying Eq. (2.45) (see section 2.3.3 for details).

2.2 Refractive index and polarizability

In this section we will derive, for furnishing a good understanding of the physics at work, a relation between the macroscopic material constants ϵ_r and n , and the microscopic atomic or molecular polarizability, α , which we can compare with experimental results on refractivity. Molecules are polarizable so that under the influence of an external electric field they show electric moments, as the positive and negative charges are impelled to move in opposite directions. The influence of permanent dipole moments as exhibited by water vapor molecules will be discussed in section 2.4.

2.2.1 Induced dipole moments

It can be reasonably assumed that the components of the induced dipole moments are linear functions of the components of the field. This assumption is fulfilled as long as the external field is weak compared to the internal electric field between the nucleus and the electron cloud.

The time-averaged charge distributions of atoms in the ground state are always spherically symmetric.

In this case, the induced electric dipole moment \mathbf{p} (the product of charge and the separation distance of positive and negative charges) is given by:

$$\mathbf{p} = \alpha \mathbf{E}, \quad (2.46)$$

where α is the scalar atomic polarizability.

Molecules are generally not symmetric (and thus not isotropic), the induced dipole moment is therefore not parallel to the electric field. In general, the polarizability is therefore a symmetric tensor of second rank, which can be fully described, after proper diagonalization, with the three main polarizabilities α_x , α_y , and α_z . In gases however, the orientations of the molecules are randomly distributed. As long as the effect of many molecules is considered, it is not necessary that each *individual* molecule is isotropic. α will then be regarded as presenting the *mean* polarizability given by [cf. Born and Wolf, 1993]:

$$\alpha = \frac{1}{3} (\alpha_x + \alpha_y + \alpha_z). \quad (2.47)$$

The polarization \mathbf{P} is then:

$$\mathbf{P} = n_m \alpha \mathbf{E}, \quad (2.48)$$

where n_m is the number density of the molecules.

2.2.2 The Clausius-Mossotti relation

The *effective* field \mathbf{E} , which is responsible for the polarization of an individual molecule, is the superposition of the *external* electric field \mathbf{E}_e and the *internal* field \mathbf{E}_i , caused by the other dipoles in the medium. Furthermore we must distinguish between the effective field \mathbf{E} and the mean observed field \mathbf{E}_m , obtained by averaging over a region which contains a great number of molecules.

We now consider a particular molecule and imagine it to be surrounded by a small sphere whose radius r is nevertheless large compared with the linear dimension of the molecule. Outside the sphere we may clearly neglect the molecular structure and treat the substance as continuous with a constant polarization \mathbf{P} produced by the mean electric field \mathbf{E}_m . The molecules inside the sphere do not produce any resulting field at the central molecule [cf., e.g., Born and Wolf, 1993, and Liou, 1980].

Besides the mean field \mathbf{E}_m we have therefore to consider the field \mathbf{E}_s produced by the polarization charge on the inside of the sphere, which is simply given by [cf., e.g., Jackson, 1985]:

$$\mathbf{E}_s = \frac{\mathbf{P}}{3\epsilon_0}. \quad (2.49)$$

The total field acting on the central molecule is therefore:

$$\mathbf{E} = \mathbf{E}_m + \mathbf{E}_s = \mathbf{E}_m + \frac{\mathbf{P}}{3\epsilon_0}. \quad (2.50)$$

With

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}_m, \quad (2.51)$$

we get

$$\mathbf{P} = \epsilon_0 \frac{\chi_e}{\left(1 + \frac{\chi_e}{3}\right)} \mathbf{E}. \quad (2.52)$$

By comparison with Eq. (2.48) we see that

$$\frac{n_m \alpha}{\epsilon_0} = \frac{\chi_e}{\left(1 + \frac{\chi_e}{3}\right)}, \quad (2.53)$$

which becomes, expressed in terms of relative permittivity $\epsilon_r (= 1 + \chi_e)$, the equation known as *Clausius-Mossotti relation*:

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{n_m \alpha}{3\epsilon_0}. \quad (2.54)$$

Equation (2.54) can be seen as bridge which connects Maxwell's phenomenological theory with the atomistic theory of matter.

2.2.3 The Lorentz-Lorenz formula

In a medium with negligible magnetic susceptibility, like the Earth's atmosphere, Eq. (2.54) can be written in terms of the refractive index n , using the Maxwell relation, Eq. (2.44):

$$\frac{n^2 - 1}{n^2 + 2} = \frac{n_m \alpha}{3 \epsilon_0}. \quad (2.55)$$

Equation (2.55) is also known as *Lorentz-Lorenz formula*.

It should be noted, though, that air has some small non-zero susceptibility component due to molecular oxygen. At radio frequencies, the relative permeability μ_r has a value of 1.000 00037 [cf. Thayer, 1974].

As n is very close to unity, the right-hand side of Eq. (2.55) can be expanded into a power series [cf., Saastamoinen, 1972]:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{2}{3}(n - 1) \left[1 - \frac{1}{6}(n - 1) + \frac{4}{9}(n - 1)^2 - \dots \right]. \quad (2.56)$$

It is usually sufficient to take the linear term in $(n - 1)$, which yields:

$$\frac{n^2 - 1}{n^2 + 2} = \frac{2}{3}(n - 1). \quad (2.57)$$

With a typical value of $n = 1.000\ 288\ 04$ (at radio frequencies), using Eq. (2.57) instead of Eq. (2.56) corresponds to a relative error of $\sim 5 \cdot 10^{-5}$ only. Keeping the quadratic term in Eq. (2.56) even reduces the relative error to $\sim 3 \cdot 10^{-8}$. On the other hand, for the refractive index of liquid water in the visible part of the spectrum ($n = 1.33$), Eq. (2.57) would introduce a relative error of almost 8%.

2.2.4 Definition of the refractivity

As the refractive index of air is very close to unity, it is convenient to use in this context the *refractivity* N , defined as [see, e.g., Doviak and Zrnic, 1984]:

$$N \equiv (n - 1) \cdot 10^6. \quad (2.58)$$

With Eqs. (2.57) and (2.58), the Lorentz-Lorenz formula can be written as:

$$N = 10^6 \frac{n_m \alpha}{2 \epsilon_0}. \quad (2.59)$$

Using the ideal gas law in the form

$$\frac{pV}{T} = \frac{p_0 V_0}{T_0}, \quad (2.60)$$

where p_0 is 1013.25 hPa, T_0 is 273.15 K, and V_0 , the molar volume of an ideal gas under standard conditions, is 22.41410 m³/kmol [CRC, 1993], and assuming conservation of mass, we get:

$$n_m = \frac{n_L T_0}{p_0} \frac{p}{T}, \quad (2.61)$$

where n_L is the *Loschmidt number* ($n_L = 2.686\,7627 \cdot 10^{25} \text{ m}^{-3}$), defined as:

$$n_L = \frac{N_a}{V_0}, \quad (2.62)$$

where N_a is the *Avogadro constant* ($N_a = 6.022\,1367 \cdot 10^{26} \text{ kmol}^{-1}$) [CRC, 1993].

Combining Eq. (2.59) with Eq. (2.61) we can then relate the refractivity to pressure and temperature:

$$N = \frac{10^6 n_L T_0}{2\varepsilon_0 p_0} \alpha \frac{p}{T}. \quad (2.63)$$

2.3 The refractivity of "dry" air

Using Eq. (2.63) we shall now derive a numerical value of the proportionality between N and p/T . In the case of the Earth's atmosphere, we have to consider in this context the atomic and molecular polarizabilities of the most important constituents N₂, O₂, Ar, CO₂, and Ne (H₂O will be treated later). As the mixing ratios of the "dry" constituents of the atmosphere are almost perfectly constant up to a height of more than 100 km (see Table 2.1 for volume mixing ratios), it is possible to combine their contributions (for our purpose) to a single term, the *dry refractivity*, denoted N_d .

2.3.1 Composition of the standard atmosphere

The volume mixing ratios of the gases listed above and their molecular (or atomic in the case of Ar and Ne) polarizabilities are tabulated in Table 2.1. In SI units, the polarizability is measured in (Cm^2/V), but polarizability values are almost exclusively tabulated in cgs units (cm^3), thus the expression *polarizability volume* is frequently encountered. The conversion from cgs units to SI units is: $\alpha (\text{Cm}^2/\text{V}) = 4\pi \epsilon_0 10^{-6} \alpha (\text{cm}^3)$ [CRC, 1993].

Gas species	Fractional volume	variability	polarizability [$10^{-40} \text{Cm}^2/\text{V}$]
N ₂	0.78084	0.00004	1.93634
O ₂	0.20948	0.00002	1.75932
Ar	0.00934	0.00001	1.82597
CO ₂	0.000314	0.00001	3.23892
Ne	0.00001818	0.0000004	0.44016

Table 2.1: Primary constituents of the U.S. Standard atmosphere (1976), their volume mixing ratios (dimensionless) and static molecular (atomic) polarizabilities [CRC, 1993]. Values of the variability (1σ) are from Davis et al. [1985].

2.3.2 The dry refractivity

Using the values of Table 2.1, we can compute a mean weighted static polarizability of dry air:

$$\bar{\alpha} = 1.89859 \cdot 10^{-40} \text{Cm}^2/\text{V}. \quad (2.64)$$

The dry refractivity can now be written as

$$N_d = k_1 \frac{p_d}{T}, \quad (2.65)$$

where p_d is the partial pressure of dry air and k_1 is given by:

$$k_1 = \frac{10^6 n_L T_0}{2\epsilon_0 p_0} \bar{\alpha} = 77.65 \text{K/hPa}. \quad (2.66)$$

In spite of the approximations used in the course of the derivation of Eq. (2.66), the excellent agreement with the experimental result of $77.60 \pm 0.05 \text{K/hPa}$ (see section 2.5) is remarkable.

2.3.3 The dielectric function

So far we have considered a static electric field. In order to determine the dependence of the polarization (and consequently of the refractivity) on the frequency of the electric field, we must formulate the displacement \mathbf{r}_0 of each charged particle from its equilibrium position. The displacement of the nuclei is small compared with the displacement of the electrons and will therefore be neglected. As long as the velocity of the electron is small compared with the velocity of light, the contribution of the magnetic field may be neglected in the formulation of the Lorentz force [cf., e.g., Fließbach, 1997].

The electron moves therefore under the force

$$\mathbf{F} = q_e \mathbf{E}(\mathbf{r}, t) = q_e \mathbf{E}_0 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)], \quad (2.67)$$

where q_e is the charge of the electron ($q_e = 1.602\,177\,33 \cdot 10^{-19} \text{C}$ [CRC, 1993]).

The rigorous determination of the effective displacement of the nuclei and the electrons is a complicated problem of quantum mechanics. It is however plausible, and actually confirmed by the rigorous theory [see Born and Wolf, 1993] that the electron behaves, to a good approximation, as if it were bound to an equilibrium position by a quasi-elastic restoring force \mathbf{F}_r ,

$$\mathbf{F}_r = -\kappa \mathbf{r}_0, \quad (2.68)$$

where κ is the restoring coefficient. If m_e denotes the mass of the electron, its equation of motion is given by (as $\exp(i \mathbf{k} \cdot \mathbf{r}_0) \approx 1$):

$$m_e \frac{d^2 \mathbf{r}_0}{dt^2} + \gamma m_e \frac{d \mathbf{r}_0}{dt} + \kappa \mathbf{r}_0 = q_e \mathbf{E}_0 \exp(-i\omega t). \quad (2.69)$$

The damping force $-\gamma m_e d\mathbf{r}_0/dt$ describes all processes that involve loss of energy, γ is the mass-specific damping factor. From this *Lorentz-model* of a damped oscillator follows the time-dependent dipole moment $\mathbf{p}(t)$:

$$\mathbf{p}(t) = -q_e \mathbf{r}_0(t) = \frac{q_e^2}{m_e(\omega_0^2 - \omega^2 - i\gamma\omega)} \mathbf{E}_0 \exp(-i\omega t), \quad (2.70)$$

and the frequency-dependent polarizability $\alpha(\omega)$:

$$\alpha(\omega) = \frac{q_e^2}{m_e(\omega_0^2 - \omega^2 - i\gamma\omega)}, \quad (2.71)$$

where ω_0 is the resonance frequency of the oscillator ($\omega_0^2 = \kappa/m_e$). Generally, the molecule has a number of resonances that correspond to the frequencies ω_{0i} of the electromagnetic radiation that the substance can emit or absorb, but the resonance frequencies of the atmospheric constituents lie almost exclusively in the UV and IR range of the spectrum ($\omega^2 \ll \omega_0^2$) [see Jeske, 1988]. The relative permittivity can now be written as (Eq.(2.54) with $\epsilon_r + 2 \approx 3$):

$$\epsilon_r = 1 + \frac{n_m q_e^2}{\epsilon_0 m_e} \frac{1}{(\omega_0^2 - \omega^2 - i\gamma\omega)}. \quad (2.72)$$

In a transparent medium, like the Earth's atmosphere, $\gamma \approx 0$ leads to a real value of ϵ_r (and consequently n). In a lossy medium both are complex numbers.

Induced polarization is the result of the deformation of the molecular electron cloud, a process which involves only minor mass movement and which is accomplished quasi instantaneously (within a typical period of $\sim 10^{-16}$ s). Unpolar substances show therefore almost exactly the same dielectric behavior from constant fields up to frequencies corresponding to visible light [cf. Purcell, 1984]. In the range of micro- and radiowaves with $\lambda > 2$ cm (including L-band, $15 \text{ cm} < 30 \text{ cm}$, used by GPS and GLONASS), dispersion effects can therefore be clearly neglected in the neutral atmosphere [cf. Jeske, 1988]. It is thus justified to use the Maxwell relation in the form of Eq. (2.44) with the static permittivity $\epsilon(0)$, as long as the contribution of water vapor to the refractivity can be neglected. The atmospheric refractivity in different frequency domains is discussed in detail by Hafner [1997].

2.4 The refractivity of water vapor

2.4.1 The permanent dipole moment of water vapor

So far we have only considered nonpolar molecules. Water vapor is the only principal atmospheric constituent with a permanent dipole moment \mathbf{p}_p , which is, moreover, exceptionally large and has the value: $p_p = |\mathbf{p}_p| = 1.854 \text{ Debye (D)}$ or $6.1843 \cdot 10^{-30} \text{ Cm}$ (where $1\text{D} \equiv 3.33564 \cdot 10^{-30} \text{ C}$) [CRC, 1993], corresponding to two elementary charges at a distance of $\sim 0.4 \text{ \AA}$ ($1\text{\AA} \equiv 10^{-10} \text{ m}$). In this section we will treat the contributions of atmospheric water vapor, which is a *paraelectric* medium, to refractivity.

2.4.2 The orientation polarization

Without an external electric field, the individual permanent dipoles of the water vapor molecules are randomly oriented, so that the resulting macroscopic field is zero. Under the influence of an external field however, the dipole moments tend to be oriented parallel to the vector of the electric field \mathbf{E} , as the electric field produces a moment of rotation \mathbf{T} , which is equal to [see, e.g., Jackson, 1985]:

$$\mathbf{T} = \mathbf{p}_p \times \mathbf{E}. \quad (2.73)$$

The potential energy U_{pot} of a dipole in an external electric field,

$$U_{pot} = -\mathbf{p}_p \cdot \mathbf{E}, \quad (2.74)$$

is consequently minimal, if $\mathbf{p}_p \parallel \mathbf{E}$. The orientation of the water molecules is however disturbed by the thermal motion of the other atmospheric constituents.

Let the z-axis of the coordinate system be parallel to \mathbf{E} , the potential energy of the dipole is then $U_{pot} = -p_p E \cos\theta$. In the statistical equilibrium, the probability $w(\theta)d\theta$ to find a dipole with an angle between θ and $\theta + d\theta$ is defined by the *Boltzmann factor* [see, e.g., Fließbach, 1997]:

$$w(\theta)d\theta = \frac{1}{z(T)} \exp\left(-\frac{U_{pot}}{k_B T}\right) d(\cos\theta), \quad (2.75)$$

where k_B is the *Boltzmann constant* ($k_B = 1.380\,658 \cdot 10^{-23}$ J/K [CRC, 1993]). $z(T)$ can be determined by the condition:

$$\int_{\pi}^0 w(\theta)d\theta = 1, \quad (2.76)$$

which yields:

$$z(T) = \int_{-1}^1 \exp\left(-\frac{U_{pot}}{k_B T}\right) d(\cos\theta). \quad (2.77)$$

The mean value of the dipole component in z-direction ($p_p \cos(\theta)$) is therefore:

$$\bar{p}_z = \frac{p_p}{z(T)} \int_{-1}^1 \exp\left(-\frac{U_{pot}}{k_B T}\right) \cos(\theta) d(\cos\theta), \quad \bar{p}_x = \bar{p}_y = 0. \quad (2.78)$$

Under normal conditions, U_{pot} is very small compared with $k_B T$. Even at temperatures below 173 K, an unrealistic electric field strength of more than $3.8 \cdot 10^6$ V/m is required to

let $U_{pot}/k_B T$ be larger than 0.01. The exponential function can consequently be expanded.

Setting $\cos(\theta) = u$, we get:

$$\begin{aligned} \bar{p} = \bar{p}_z &= \frac{p_p \int_{-1}^1 \left(1 + u \left(\frac{p_p E}{k_B T} \right) + \dots \right) u \, du}{\int_{-1}^1 \left(1 + u \left(\frac{p_p E}{k_B T} \right) + \dots \right) \, du} = \\ &= \frac{p_p \left[\frac{u^2}{2} + \left(\frac{p_p E}{k_B T} \right) \frac{u^3}{3} \right]_{-1}^{+1}}{\left[u + \left(\frac{p_p E}{k_B T} \right) \frac{u^2}{2} \right]_{-1}^{+1}} = \frac{p_p^2 E}{3k_B T} \end{aligned} \quad (2.79)$$

2.4.3 Langevin's formula

Introducing the *orientation polarizability* α_p as proportionality between \mathbf{p}_p and \mathbf{E} , we get:

$$\alpha_p = \frac{p_p^2}{3k_B T}, \quad (2.80)$$

a relation known as *Langevin's formula*.

If we consider the frequency dependence of Eq. (2.80), we see that [Jeske, 1988]:

$$\alpha_p = \frac{p_p^2}{3k_B T} \frac{1}{1 + i\omega\tau}, \quad (2.81)$$

where τ is the relaxation time required for external field-induced orientations of the molecules to return to random distribution after the field has been removed. For external fields with frequencies less than 100 GHz (microwaves, radiowaves), $\omega\tau \ll 1$ so that the *Langevin's formula*, Eq. (2.80) can be used. For radiowaves longer than about 2 cm (corresponding to frequencies < 15 GHz) there exist furthermore practically no dispersion effects [cf. Jeske, 1988].

Dispersion effects become however important at much lower frequencies than in the case of induced polarization, as orientation polarization involves the movement of the whole water molecule. As a consequence, the index of refraction of liquid water, which is still about 9 for UHF-waves (corresponding to $\epsilon_r \approx 80$), has only a value of 1.33 for visible

light. In the case of visible light, water behaves like an unpolar substance as $\omega\tau \gg 1$, and hence α_p becomes zero [see, e.g., Purcell, 1984].

2.4.4 The wet refractivity

Besides the orientation polarization we have also to consider the induced polarization of the water vapor molecules, the latter given by Eq. (2.63) introduced above. The part of the refractivity which is attributable to water vapor, termed *wet refractivity* (N_w), is thus in total given by:

$$N_w = \frac{10^6 n_L T_0}{2\varepsilon_0 p_0} \frac{e}{T} (\alpha_w + \alpha_p) = \frac{10^6 n_L T_0}{2\varepsilon_0 p_0} \frac{e}{T} \left(\alpha_w + \frac{p_p^2}{3k_B T} \right), \quad (2.82)$$

where e is the partial pressure of water vapor (cf. Eq. (2.63)) and α_w is the molecular polarizability of water ($\alpha_w = 1.6133 \cdot 10^{-40} \text{ Cm}^2/\text{V}$ [CRC, 1993]). Equation (2.82) can now be written as:

$$N_w = k_2 \frac{e}{T} + k_3 \frac{e}{T^2}, \quad (2.83)$$

where k_2 and k_3 are given by:

$$k_2 = \frac{10^6 n_L T_0}{2\varepsilon_0 p_0} \alpha_w = 65.99 \text{ K/hPa}, \quad (2.84)$$

and

$$k_3 = \frac{10^6 n_L T_0}{2\varepsilon_0 p_0} \frac{p_p^2}{3k_B} = 3.777 \cdot 10^5 \text{ K}^2/\text{hPa}. \quad (2.85)$$

The corresponding experimental values, e.g., in the Thayer formula (see section 2.5.2) are $64.8 \pm 0.08 \text{ K/hPa}$ and $(3.776 \pm 0.004) 10^5 \text{ K}^2/\text{hPa}$, respectively, again a very good agreement, indicating that our theoretical outline very well captures the essential physics. Under typical meteorological conditions, the polar term in the radio refractivity of water vapor is about twenty times larger than the induced polarization term.

2.5 Experimental values of the refractivity

2.5.1 The Smith-Weintraub formula

A frequently used representation of the refractivity at radio frequencies is the formula presented by Smith and Weintraub [1953]:

$$N = k_1 \frac{p_d}{T} + k_2 \frac{e}{T} + k_3 \frac{e}{T^2}, \quad (2.86)$$

The values of the empirical constants k_1 , k_2 , and k_3 and their assumed uncertainties (standard deviations) as given by that work are:

$$k_1 = 77.60 \pm 0.01 \text{ K/hPa}$$

$$k_2 = 72 \pm 8 \text{ K/hPa}$$

$$k_3 = (3.75 \pm 0.03) 10^5 \text{ K}^2/\text{hPa}$$

Utilizing the total pressure p and the slightly temperature dependent "constant" k_4 :

$$k_4 = (k_2 - k_1)T + k_3, \quad (2.87)$$

Eq. (2.86) can be written with little additional error in the two-term form:

$$N = k_1 \frac{p}{T} + k_4 \frac{e}{T^2}, \quad (2.88)$$

$$N = 77.6 \frac{p}{T} + 3.73 \cdot 10^5 \frac{e}{T^2},$$

where $3.73 \cdot 10^5 \text{ K}^2/\text{hPa}$ is the value of k_4 at 273.15 K. Equation (2.86) is considered to be accurate to 0.5% in wet atmospheres and accurate to $< 0.02\%$ in dry atmospheres [Smith and Weintraub, 1953]. It is, for example, widely used in GNSS radio occultation data processing.

2.5.2 The Thayer formula

An improved equation for the atmospheric refractivity at radio frequencies (below 20 GHz) was presented by Thayer [1974]:

$$N = k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1}, \quad (2.89)$$

where p_d is the partial pressure of "dry air" in hPa, e is the partial pressure of water vapor in hPa, and T is the absolute temperature in K.

The values of the empirical constants k_1 , k_2 , and k_3 and their assumed uncertainties as given by Thayer are:

$$k_1 = 77.60 \pm 0.014 \text{ K/hPa}$$

$$k_2 = 64.8 \pm 0.08 \text{ K/hPa}$$

$$k_3 = (3.776 \pm 0.004) \cdot 10^5 \text{ K}^2/\text{hPa}$$

Thayer developed an alternative and hybrid approach in which the value of k_2 was extrapolated from optical frequencies rather than measured using microwave techniques.

Z_d^{-1} and Z_w^{-1} are the inverse compressibility factors for dry air and water vapor, respectively. They describe the nonideal gas behavior and can be expressed as [see, e.g., Davis *et al.*, 1985]:

$$Z_d^{-1} = 1 + p_d \left[57.97 \cdot 10^{-8} \left(1 + \frac{0.52}{T} \right) - 9.4611 \cdot 10^{-4} \frac{T_C}{T^2} \right], \quad (2.90)$$

and

$$Z_w^{-1} = 1 + 1650 \frac{e}{T^3} \left(1 - 0.01317 T_C + 1.75 \cdot 10^{-4} T_C^2 + 1.44 \cdot 10^{-6} T_C^3 \right), \quad (2.91)$$

where T is the temperature in K and T_C is the temperature in °C. Using Eqs. (2.90) and (2.91), the ideal gas law for dry air and water vapor, respectively, can be written in the form:

$$\begin{aligned} p_d &= \rho_d \frac{R^*}{m_d} T Z_d, \\ e &= \rho_w \frac{R^*}{m_w} T Z_w, \end{aligned} \quad (2.92)$$

where R^* is the universal gas constant ($8.314510 \pm 0.000070 \text{ Pa m}^3 \text{ K}^{-1} \text{ mol}^{-1}$), ρ_d and ρ_w are the densities of dry air and water vapor, respectively, and m_d and m_w are the molar masses of dry air ($28.9644 \pm 0.0014 \text{ kg/kmol}$) and water vapor (18.01528 kg/kmol), respectively [CRC, 1993].

In the terrestrial atmosphere near the surface, the compressibilities differ from unity only by a few parts per thousand (for an ideal gas, $Z = 1$) and this difference decreases exponentially with height. Neglecting Z_w^{-1} can nevertheless lead to errors of as much as 0.1 N -units at high near-surface humidities (cf. Fig. 2.1 below). The expected accuracy of Eq. (2.89) is 0.05 N -units (0.018%) for dry air and 0.2 N -units (0.05%) for extremely moist air. The grouping together of all the dry constituents into one refractivity term is possible

because the relative mixing ratio of these gases remains nearly constant in time and in the vertical domain of interest.

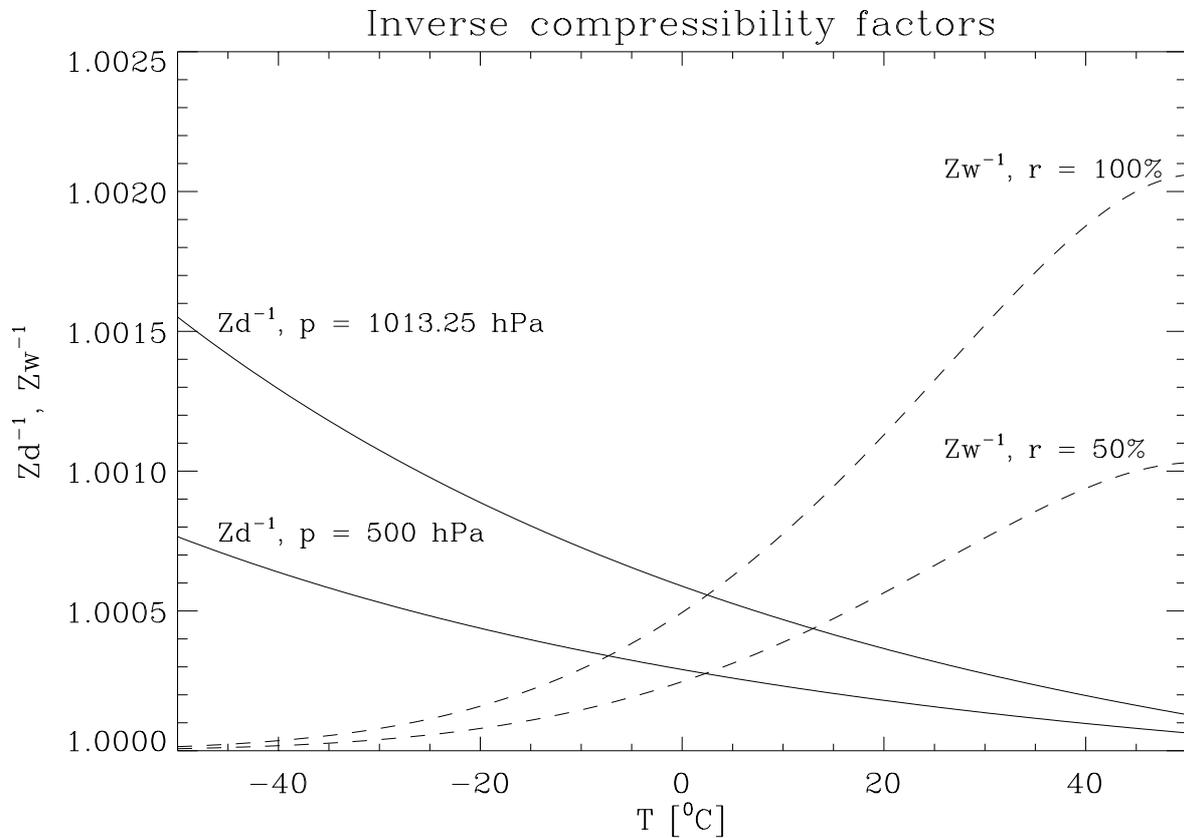


Fig. 2.1: Inverse compressibility factors of dry air and water vapor as a function of the temperature T . Z_d^{-1} (solid line) at an atmospheric pressure of 1013.25 and 500 hPa, respectively. Z_w^{-1} (dashed line) at relative humidities of 100% and 50%, respectively.

2.5.3 Review of the refractivity constants

Bevis *et al.* [1994] compiled published experimental determinations of the refractivity constants. This compilation included more than 20 direct measurements of k_1 and 7 direct measurements of k_2 and k_3 . The median values (as a robust estimators of the most likely values) are contained in the list below. Note that the nominal uncertainty is more conservative than that adopted by Thayer [1974]. Bevis *et al.* [1994] argue that extrapolating the value of k_2 across the infrared band is theoretically unjustifiable, they propose therefore to abandon Thayer's value for the less precise but more justifiable values

derived from direct measurements [for details see Bevis *et al.*, 1994]. The authors suggest to use the Thayer formula with the values:

$$k_1 = 77.60 \pm 0.05 \text{ K/hPa}$$

$$k_2 = 70.4 \pm 2.2 \text{ K/hPa}$$

$$k_3 = (3.739 \pm 0.012) \cdot 10^5 \text{ K}^2/\text{hPa}.$$

3 GROUND-BASED GPS METEOROLOGY

The dependence of atmospheric refractivity on the presence of water vapor (cf. chapter 2) enables ground-based Global Positioning System (GPS) receivers to provide time series of vertical integrated water vapor (IWV) above the receiving station [Bevis *et al.*, 1992] or of slant integrated water vapor (SIWV) along ray paths between GPS satellites and receivers [Ware *et al.*, 1997]. The estimation of integrated water vapor becomes feasible when the position of the receiver is accurately known, the ionospheric delay has been accounted for, and accurate surface temperature and pressure data are available. Knowledge of the vertical integrated water vapor is particularly useful, as IWV represents the total latent heat available in the column overlaying the receiving station [Businger *et al.*, 1996]. After a short description of the Global Positioning System [for details see, e.g., Dixon, 1991; Rothacher, 1992; and Hofmann-Wellenhof *et al.*, 1994], the retrieval method for integrated water vapor shall be discussed in this chapter.

3.1 The Global Positioning System

In 1967 an early prototype of a GPS satellite known as Timation 1 was launched into low Earth orbit (~900 km) as part of a military test program [Dixon, 1991]. The immediate predecessor of the GPS was the Navy Navigational Satellite System (NNSS), also called TRANSIT, composed of six satellites in nearly circular polar orbits with a height of about 1100 km [Hofmann-Wellenhof *et al.*, 1994]. With the advent of a satellite-based navigation system known as the Navigation Satellite Timing and Ranging Global Positioning System (NAVSTAR GPS, or simply GPS) operated by the U.S. Department of Defense, it became possible for a user with the proper receiver to obtain precise time tagged and almost instantaneous three-dimensional position information accurate to several meters.

3.1.1 The space segment

The first GPS satellite was launched in February 1978, it was part of eleven "Block I" satellites, which are no longer in use and have been replaced completely by 24 Block II and Block IIA satellites until 1994 ("A" denotes advanced). The actual, operational GPS

consists of 24 satellites (plus 3 operational spares), almost evenly distributed in six orbital planes with an inclination of 55° and a plane separation of 60° . Each orbit is circular, with an altitude of about 20 200 km, resulting in an orbital period of close to 12 sidereal hours.

Launch order	PRN	SVN	Launch Date	Frequency standard	Orbital plane
II-1	14	14	Feb 14, 89	Cs	E1
II-2	02	13	Jun 10, 89	Cs	B3
II-3	16	16	Aug 18, 89	Cs	E5
II-4	19	19	Oct 21, 89	Rb	A4
II-5	17	17	Dec 11, 89	Cs	D3
II-6	18	18	Jan 24, 90	Cs	F3
II-7	20	20	Mar 26, 90		no longer in service
II-8	21	21	Aug 02, 90	Cs	E2
II-9	15	15	Oct 01, 90	Cs	D2
IIA-10	23	23	Nov 26, 90	Cs	E4
IIA-11	24	24	Jul 04, 91	Rb	D1
IIA-12	25	25	Feb 23, 92	Cs	A2
IIA-13	28	28	Apr 10, 92		no longer in service
IIA-14	26	26	Jul 07, 92	Rb	F2
IIA-15	27	27	Sep 09, 92	Cs	A3
IIA-16	01	32	Nov 22, 92	Cs	F1
IIA-17	29	29	Dec 18, 92	Rb	F4
IIA-18	22	22	Feb 03, 93	Rb	B1
IIA-19	31	31	Mar 30, 93	Cs	C3
IIA-20	07	37	May 13, 93	Rb	C4
IIA-21	09	39	Jun 26, 93	Cs	A1
IIA-22	05	35	Aug 30, 93	Cs	B4
IIA-23	04	34	Oct 26, 93	Rb	D4
IIA-24	06	36	Mar 10, 94	Cs	C1
IIA-25	03	33	Mar 28, 94	Cs	C2
IIA-26	10	10	Jul 16, 96	Cs	E3
IIA-27	30	30	Sep 12, 96	Cs	B2
IIA-28	08	38	Nov 06, 96	Rb	A5
IIR-1		42	Jan 17, 97		unsuccessful launch
IIR-2	13	43	Jul 23, 97	Rb	F5

Table 3.1: Information on the current status of the GPS satellites (May 8, 1999), satellites are arranged in launch order, PRN stands for "Pseudo-Random-Noise code number" (explanation in the text) and SVN for "Space Vehicle Number". Frequency standards are Cesium (Cs) or Rubidium (Rb). The orbital position of the satellites is indicated by a letter for the orbital plane and a number for the position within the plane, data from [<http://tycho.usno.navy.mil/gpscurre.html>].

Block II satellites are equipped with four highly stable on-board time standards, two Rubidium (Rb) and two Cesium (Cs) clocks, whose frequency stability reaches 10^{-14} over one day. The cesium clock for example is based on the splitting of the ground electronic state of cesium 133, depending on whether the spin of the unpaired valence electron is parallel or antiparallel to the nuclear spin. The transition between these two hyperfine levels has a frequency of 9 192 631 Hz and is the basis for the currently accepted (SI) definition of the second [Dixon, 1991].

After March 1994, all subsequent launches are and will be to replace malfunctioning satellites with Block IIR satellites ("R" denotes replenishment). One successful IIR launch occurred so far, on July 23, 1997. Data on the current status of the GPS consisting of 27 Block II, IIA, and IIF satellites can be found in Table 3.1. Future Block IIR satellites shall have on-board hydrogen masers that are one order of magnitude more precise than the rubidium and cesium clocks. Actual information on the status of the GPS can be found on the U.S. Naval Observatory homepage (<http://tycho.usno.navy.mil/gpscurre.html>). Actual satellite ephemeris in NORAD two-line element sets [Kelso, 1988] can be found on the Celestrak homepage (<http://celestrak.com/NORAD/elements/index.html>), at least as long as the United States are not involved in a military conflict and the ephemeris of the GPS satellites (classified as "sensitive") are thus obviously removed.

With the full constellation, the space segment provides global coverage with four to eight simultaneously observable satellites above 15° elevation at any time. Occasionally up to 12 satellites are visible if the elevation mask is reduced to 5° [Hofmann-Wellenhof *et al.*, 1994].

3.1.2 The satellite signal

The key to the system's accuracy is the fact that all signal components are precisely controlled by atomic clocks. The oscillators on board of the satellites generate a fundamental L-band frequency f_0 . Two carrier signals in the L-band (L1 and L2) are generated by integer multiplication of f_0 (see Table 3.2). The dual frequency approach is essential for eliminating the ionospheric refraction, the major error source for neutral atmosphere sounding (see next section). The clock frequency is in fact set slightly lower (by ~ 0.005 Hz) to account for relativistic effects, so that an observer on the ground "sees" exactly 10.23 MHz [Dixon, 1991].

Two *pseudorandom noise* codes are superimposed onto the two L-band carrier frequencies:

- (1) The *Coarse/Acquisition-code* (C/A-code) with an effective wavelength of ~ 300 m is available for civilian use, it is modulated only upon L1 to deny full system accuracy to nonmilitary users.
- (2) The *Precision-code* (P-code) with an effective wavelength of ~ 30 m is modulated onto both carriers, but has been reserved for use by the U.S. military and other authorized users [Hofmann-Wellenhof *et al.*, 1994].

The codes can be considered as square waves with values of ± 1 (corresponding to the binary values 0 and 1). The so-called *biphase modulation* is performed by a 180° shift in the carrier phase whenever a change in the code state occurs. The codes are termed *Pseudo-Random Noise* (PRN) codes because they have sufficiently long repeat times (e.g., 37 weeks for the P-code) that they appear random to a user without knowledge of the code structure. Each satellite is assigned a 1-week portion of the code in order to be uniquely identifiable by a PRN number corresponding to the code portion transmitted [Dixon, 1991]. In addition to the PRN codes, a data message (the navigation message) is modulated onto the carriers comprising satellite ephemeris, satellite health, system time, and other information.

Component		Frequency [MHz]	Wavelength [m]
Fundamental frequency	f_0	10.23	29.305
Carrier frequency L1	$154 \cdot f_0$	1575.42	0.1903
Carrier frequency L2	$120 \cdot f_0$	1227.60	0.2442
Precision-code	f_0	10.23	29.305
Coarse/Acquisition-code	$f_0/10$	1.023	293.05
Encrypting W-code	$f_0/20$	0.5115	586.10
Navigation message	$f_0/204600$	$5 \cdot 10^{-5}$	$6 \cdot 10^6$

Table 3.2: Components of the GPS satellite signal [cf. Dixon, 1991; and Hofmann-Wellenhof *et al.*, 1994].

Block II satellites have additional security features for denying civilian users full use of the system:

- (1) *Selective availability (SA)* degrades the navigation accuracy by dithering the clock frequency and manipulating the broadcast ephemeris. Thus, the point positioning accuracy is reduced to ~ 100 m.
- (2) *Anti-spoofing (AS)* is the encryption of the P-code, normally required for high precision measurements, it is accomplished by the modulo 2 sum of the P-code and an encrypting W-code. The resulting code is termed Y-code. The rationale is to prevent adversaries from sending out false signals with the GPS signature. AS is either turned "on" or "off", a variable influence on the signal (as it is the case with SA) cannot occur [Hofmann-Wellenhof *et al.*, 1994].

3.1.3 Navigation with GPS

The GPS satellites are configured to provide the user with the capability of determining his position, this is accomplished by a simple resection process using the measured distances to the satellites. The coordinates of the satellites relative to the center of the earth can be computed from the ephemeris broadcast by the satellite.

If the ground receiver would be equipped with a clock that was set precisely to GPS system time, the true distance or *range* to each satellite could be accurately measured by recording the time required for the signal to reach the receiver. Each range defines the surface of a sphere with the satellite in the center. The intersection of three spheres defines two points, only one of them is usually located on the surface of the Earth. Using this technique (*ranging*), only three satellites would therefore be needed to determine the position of the receiver.

GPS receiver apply a slightly different technique as they are normally equipped with inexpensive crystal clocks. Due to clock errors, the measured distances (*pseudoranges*) differ from the "true" ranges, this problem can be solved by simultaneously measuring the distances to four satellites, the technique is consequently called *pseudorangeing*.

The pseudorange can be derived by standard methods with an accuracy of ~ 1 m for the P-code and 3-5 m for the C/A-code [Hofmann-Wellenhof *et al.*, 1994]. Observation geometry affects the quality of the resulting position. Satellites that appear close to one

another provide correlated range information, an effect known as *geometric dilution of precision* (GDOP) [Dixon, 1991].

It is also possible to obtain distance information (or change in distance, to be precise) through phase measurements on the carrier signal itself, keeping track of the number of cycles after signal acquisition (with a precision of less than 0.5% of a wavelength). Since the wavelength of the carrier is considerably shorter than that of the lower frequency code modulations, the resulting measurement, though ambiguous by the initial number of wavelengths, is considerably more precise than a pseudorange measurement and is one of the keys to high precision GPS measurements [Dixon, 1991]. The accuracy of the *phase pseudoranges* is $< 1\text{-}3$ mm [Hofmann-Wellenhof *et al.*, 1994].

3.1.4 Ionospheric correction

The free electrons in the ionosphere interact strongly with any electromagnetic signal in the frequency range of GPS. Ionospheric effects are proportional to the integrated or *Total Electron Content* (TEC) along the ray path and thus depend on the elevation angle of the satellite, latitude, time of day, and on the solar activity. The ionosphere is furthermore a dispersive medium, as the ionospheric refractive index depends on the frequency f . The ionospheric travel time t for a group delay measurement can be represented by [Dixon, 1991]:

$$t = \frac{S}{c_0} + \frac{A}{f^2} + \frac{B}{f^3} + \dots, \quad (3.1)$$

where S is the distance between transmitter and receiver and c is the velocity of light in vacuum. The A/f^2 term represents most of the ionospheric delay experienced at frequency f , A is a constant, and B is proportional to the average magnetic field strength. Ignoring the third- and higher-order terms, the differential group delay Δt between observations at f_1 (the L1 frequency) and f_2 (the L2 frequency) becomes:

$$\Delta t \cong A \left(\frac{1}{f_1^2} - \frac{1}{f_2^2} \right) = \frac{A}{f_1^2} \left(\frac{f_1^2 - f_2^2}{f_2^2} \right) = t_1 \left[\left(\frac{f_1}{f_2} \right)^2 - 1 \right], \quad (3.1)$$

where t_1 is the group delay at L1. Dual frequency observations therefore allow to eliminate major ionospheric effects (but not third- and higher-order terms).

3.1.5 The ground segment

There are five monitor stations located at: Colorado Springs, Ascension Island in the South Atlantic, Diego Garcia in the Indian Ocean, and Kwajalein and Hawaii in the North Pacific Ocean. Each of the monitor stations is equipped with a precise cesium time standard and receivers that continuously measure pseudoranges to all visible GPS satellites. Falcon Air Force Base, Colorado Springs, is also the location of the master control station, which collects the tracking data from the monitor stations and computes the satellite orbit and clock parameters. The satellite ephemeris and clock information calculated at the master control station and received via communication links (ground control stations Ascension, Diego Garcia, and Kwajalein) are uploaded to each GPS satellite [Hofmann-Wellenhof *et al.*, 1994].

From within and for serving the civilian user community, in particular the geodetic community, a global system of currently more than 200 satellite tracking stations was established by the International GPS Service for Geodynamics (IGS).

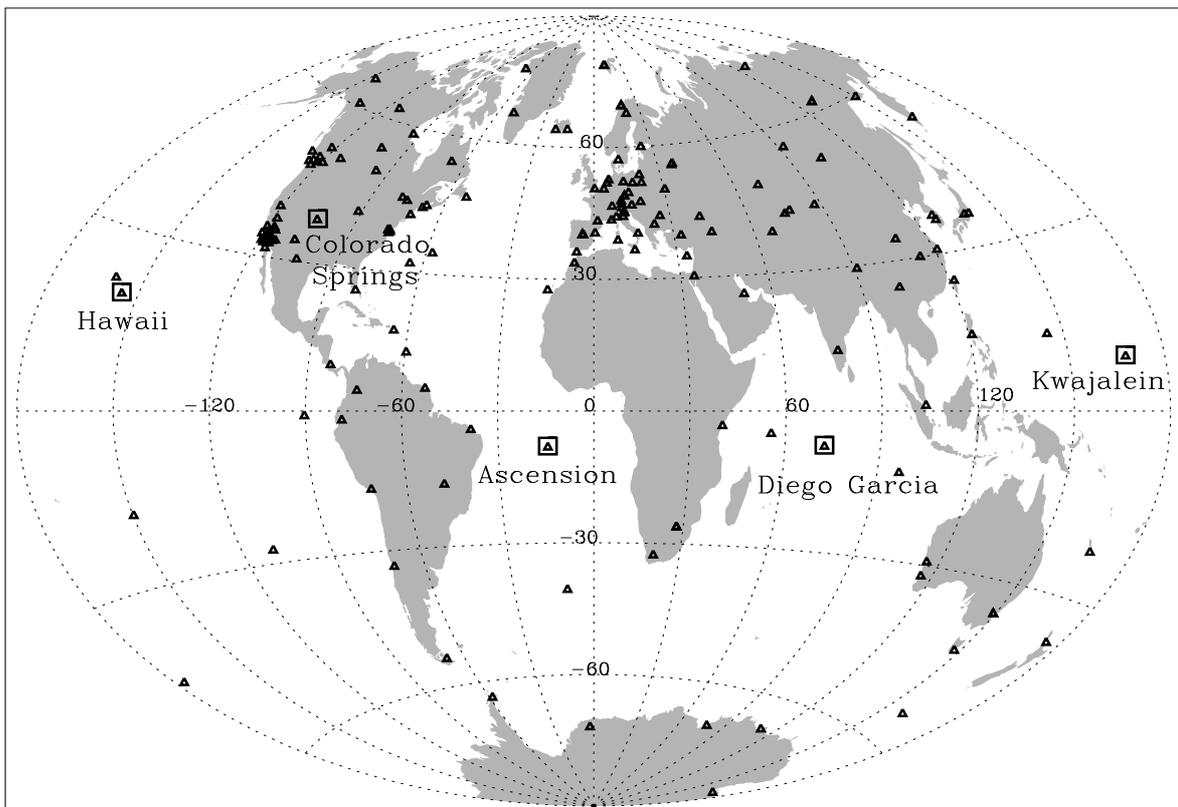


Fig. 3.1: Actual IGS network of 205 tracking stations for GPS satellites (May 10, 1999). Tracking stations are indicated with a triangle, monitor stations with an additional square; data from the IGS homepage [<http://igsb.jpl.nasa.gov/network/netindex.html>].

3.2 Physics of the propagation delay in the neutral atmosphere

The influence of the neutral atmosphere introduces errors into geodetic measurements, that cannot be removed using dual frequency observations, as the neutral atmosphere is essentially non-dispersive in the GPS frequency range. This geodetic "noise" can be used on the other hand to derive useful knowledge about the state of the atmosphere.

3.2.1 Definition of the excess propagation path

When electromagnetic waves propagate through a medium whose refractive index n is a function of the path, the waves undergo a change of direction (refractive bending) and a retardation of speed of propagation. The electrical path length L of a signal propagating along the ray-path S is defined as [cf. Born and Wolf, 1993]:

$$L = \int_S n(s) ds . \quad (3.1)$$

For GPS radio signals, the relative variations of n in the atmosphere are small over the distance of one wavelength, it is therefore possible to describe their propagation with the aid of geometrical optics.

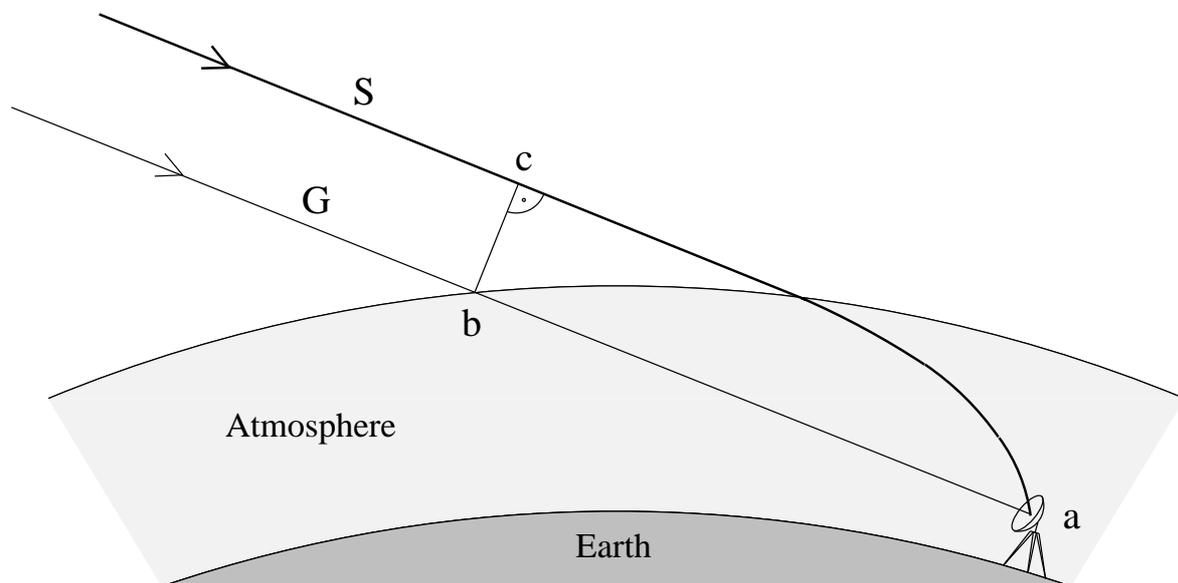


Fig. 3.2: The geometrical straight-line distance G is always shorter than the actual ray path S , but propagating along G , the signal would travel a longer distance with lower velocity close to the surface, where the index of refraction is larger, so that

$$\int_b^a n(s) ds > \int_c^a n(s) ds \quad \text{even though} \quad \int_b^a ds < \int_c^a ds.$$

In the geometric optics approximation, the actual curved path S is, by *Fermat's principle*, the one of least travel time: the signal will propagate along the path that gives the minimum value of L . The geometry is illustrated in Fig. 3.2. The delay in signal arrival time can be expressed in terms of an equivalent increase in travel path length. The difference between electrical path length and straight line distance, the excess propagation path is therefore usually called *path delay* [cf. Elgered, 1993]:

$$\Delta L = \int_S n(s) ds - \int_G ds = L - G, \quad (3.2)$$

which can be rewritten with $S = \int_S ds$:

$$\Delta L = \int_S (n(s) - 1) ds + (S - G), \quad (3.3)$$

where the first term on the right side is due to the slowing effect, and the second is due to bending. The bending term is usually called *geometric delay* (ΔL_g):

$$\Delta L_g = S - G. \quad (3.4)$$

In the absence of horizontal gradients, the bending term vanishes for rays in zenith direction, it is typically 3 cm at an elevation angle of 10° and 10 cm at 5° [Elgered, 1993]. As the corresponding total delays are at least 13 m and 23 m, respectively, ΔL_g can usually be neglected.

ΔL in zenith direction can also be written as:

$$\Delta L = c_0 (t - t_0), \quad (3.5)$$

where c_0 is the velocity of light in vacuum, t is the signal propagation time through the atmosphere, and t_0 is the corresponding propagation time in vacuum. The propagation time t is given by:

$$t = \frac{1}{c_0} \int_G n ds, \quad \text{since} \quad \frac{ds}{dt} = c = \frac{c_0}{n}. \quad (3.6)$$

Being precise, the term "delay" should refer to the time delay $t - t_0$ [Askne and Nordius, 1987]. Equation (3.3) is conveniently expressed in terms of atmospheric refractivity N (see chapter 2), which is a linear re-scaling of the refractive index n defined as:

$$N = 10^6 (n - 1). \quad (3.7)$$

We used the three-term formula for the total refractivity of moist air, provided by Thayer [1974]:

$$N = k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1}, \quad (3.8)$$

where p_d is the partial pressure of "dry air" in (hPa), e is the partial pressure of water vapor in (hPa), T is the absolute temperature in (K), and Z_d^{-1} and Z_w^{-1} are the inverse compressibility factors for dry air and water vapor, respectively (see chapter 2 for details), describing the nonideal gas behavior. The values of the empiric constants k_1 , k_2 , and k_3 and their uncertainties were taken from Bevis *et al.* [1994]:

$$k_1 = 77.60 \pm 0.05 \text{ K/hPa}$$

$$k_2 = 70.4 \pm 2.2 \text{ K/hPa}$$

$$k_3 = (3.739 \pm 0.012) 10^5 \text{ K}^2/\text{hPa}.$$

The grouping together of all the dry constituents of the atmosphere into one refractivity term is possible because the relative mixing ratio of these gases remain nearly constant in time and in the vertical domain of interest. At frequencies below 30 GHz, the delay due to the dispersive components of refractivity is much less than 1 mm in the zenith direction [Elgered, 1993], and will therefore not be further discussed. Using Eq. (3.7) and neglecting ΔL_g , Eq. (3.3) finally becomes

$$\Delta L = 10^{-6} \int_s N(s) ds. \quad (3.9)$$

3.2.2 The path delay due to condensed water

Even though the path delay is caused mainly by atmospheric gases, there is a small contribution from condensed water in the form of clouds and rain. In liquid water and ice, the hydrogen bonds between water molecules greatly reduce the contributions of the dipole moments to the delay [Businger *et al.*, 1996]. Therefore the contributions of liquid water vapor and ice are only 25% and 10%, respectively, of that of an equivalent mass of water vapor [Kursinski *et al.*, 1995]. Liquid water reaches furthermore only a small percentage of typical water vapor abundance, the contribution of liquid water to the total precipitable water is mostly smaller than 1% [Tregoning *et al.*, 1998]. Even very large amounts of 0.5 g/m^3 liquid in the upper part of stratiform clouds at the top of the marine boundary layer contribute only $\sim 1.5\%$ to the total water refractivity [Kursinski *et al.*, 1995].

As long as the cloud droplets are much smaller than the wavelength of the radio signal (which is fulfilled for L-band signals), and for the special case of a water temperature of 283 K, the following expression can be used to calculate ΔL_c , the non-dispersive delay due to clouds [Elgered, 1993]:

$$\Delta L_c = 0.0015 \rho_l S_c, \quad (3.10)$$

where ΔL_c is measured in (m), ρ_l is the liquid water density in (g/m^3), and S_c is the path length through the cloud in (km). Generally, ΔL_c can be neglected, as it almost never reaches more than 1 cm in zenith direction.

The delay due to rain is more difficult to assess. For extreme rainfall rates of 100 mm per hour, it can be expected to reach up to 0.6 cm per km path length within the rain cell, whereas a moderate rainfall of 10 mm per hour gives an approximate delay of 0.1 cm/km [Elgered, 1993]. Generally, the rain-induced delay can be neglected as well.

3.2.3 Definition of the zenith hydrostatic delay

The total atmospheric delay in zenith direction can be partitioned into a large quantity which depends only on surface pressure, the "hydrostatic delay", and a smaller quantity which depends only on the water vapor pressure profile, usually called "wet delay" [Saastamoinen, 1972; Davis *et al.*, 1985]. The hydrostatic delay is due to the induced dipole moment of all gaseous constituents of the atmosphere including water vapor, the frequently used term "dry delay" is therefore misleading.

As the hydrostatic equation is only valid for the total pressure and not for the partial pressure of dry air alone, it is convenient to rewrite Eq. (3.8), using *Dalton's law* for the total pressure p ,

$$p = p_d + e, \quad (3.11)$$

and the gas law for dry air and water vapor,

$$\begin{aligned} p_d &= \rho_d \frac{R^*}{m_d} T Z_d, \\ e &= \rho_w \frac{R^*}{m_w} T Z_w, \end{aligned} \quad (3.12)$$

where R^* is the universal gas constant ($8.314510 \pm 0.000070 \text{ Pa m}^3 \text{ K}^{-1} \text{ mol}^{-1}$), ρ_d and ρ_w are the densities of dry air and water vapor, respectively, and m_d and m_w are the molar

masses of dry air (28.9644 ± 0.0014 kg/kmol) and water vapor (18.01528 kg/kmol), respectively [CRC, 1993]. Equation (3.8) can now be rewritten as

$$\begin{aligned} N &= k_1 \frac{R^*}{m_d} \rho - k_1 \frac{R^*}{m_d} \rho_w + k_2 \frac{R^*}{m_w} \rho_w + k_3 \frac{e}{T^2} Z_w^{-1} \\ &= k_1 \frac{R^*}{m_d} \rho + \left(k_2 - k_1 \frac{m_w}{m_d} \right) \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1}, \end{aligned} \quad (3.13)$$

where ρ is the total density. Equation (3.13) can be simplified introducing the constant k_2' :

$$k_2' = \left(k_2 - k_1 \frac{m_w}{m_d} \right). \quad (3.14)$$

Under the assumption that the errors in k_1 and k_2 (σ_1 and σ_2) are uncorrelated, propagating these errors through (3.14) by evaluation of the total differential yields for σ , the error in the derived constant k_2' :

$$\sigma^2 = \sigma_2^2 + \left(\frac{m_w}{m_d} \right)^2 \sigma_1^2, \quad (3.15)$$

and $k_2' = 22.1 \pm 2.2$ K/hPa [cf. Bevis *et al.*, 1994] (in fact $\sigma_1 \ll \sigma_2$ so that $\sigma \cong \sigma_2$ holds rather precisely).

Equation (3.13) can thus be rewritten as:

$$N = k_1 \frac{R^*}{m_d} \rho + k_2' \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1}. \quad (3.16)$$

The first term on the right-hand side depends now only on the total density ρ , making it possible to obtain an integral expression that can be replaced by the surface pressure (or by the pressure at the height of the antenna, to be precise). The atmosphere normally satisfies the hydrostatic equation:

$$\frac{dp}{dz} = -\rho(z) g(z), \quad (3.17)$$

where $g(z)$ is the acceleration due to gravity at the vertical coordinate z . Integrating (3.17) yields:

$$p_0 = g_{eff} \int_{z_{antenna}}^{\infty} \rho(z) dz, \quad (3.18)$$

where p_0 is the total atmospheric pressure at the height of the antenna, $z_{antenna}$, and

$$g_{eff} = \frac{\int_{z_{antenna}}^{\infty} \rho(z) g(z) dz}{\int_{z_{antenna}}^{\infty} \rho(z) dz}. \quad (3.19)$$

The value of g can be derived from the approximate gravity formula [CRC, 1993]:

$$\begin{aligned} g(H, \varphi) &= 9.7804(1 + 0.00529 \sin^2 \varphi) - 0.00309 H \\ &= 9.7804(1 + 0.00529 \sin^2 \varphi - 0.000316 H) \\ &= 9.8063(1 - 0.00264 \cos 2\varphi - 0.000315 H) \end{aligned} \quad (3.20)$$

where g is measured in (m/s^2), φ is the geographic latitude, and H is the height above the geoid in (km). Equation (3.19) very nearly represents the acceleration due to gravity at the center of mass of the vertical column of air [for details see Saastamoinen, 1972]. For the height of the center of mass, H_c , Saastamoinen [1972] found the expression

$$H_c = 0.9H + 7.3 \text{ km}, \quad (3.21)$$

which is, according to the author accurate to within ± 0.4 km for all latitudes and seasons. The expression for local gravity at the center of mass of the atmospheric column becomes therefore:

$$g_{eff} = g_m [f(\varphi, H) \pm 0.0001], \quad (3.22)$$

with $g_m = 9.7837 \text{ m/s}^2$, and

$$f(\varphi, H) = 1 - 0.00265 \cos 2\varphi - 0.000285 H. \quad (3.23)$$

Using these relations, Eq. (3.9) becomes for rays in zenith direction (indicated by the superscript "0"):

$$\Delta L^0 = 10^{-6} \left[k_1 \frac{R^* p_0}{m_d g_{eff}} + \int_{z_{antenna}}^{\infty} \left(k_2' \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1} \right) dz \right]. \quad (3.24)$$

The first term on the right side is called the *zenith hydrostatic delay*, ΔL_h^0 :

$$\Delta L_h^0 = \frac{k_1 R^* p_0}{10^6 m_d g_m f(\varphi, H)}. \quad (3.25)$$

3.2.4 Error budget of the zenith hydrostatic delay

The constants in Eq. (3.25) can be combined to a new constant ΔL_{hc}^0 :

$$\Delta L_{hc}^0 = \frac{k_1 R^*}{10^6 m_d g_m} = 0.0022768 \text{ m hPa}^{-1}. \quad (3.26)$$

Under the assumption that the errors in k_1 , R^* , m_d , and g_m , (σ_1 , σ_R , σ_m , and σ_g) are uncorrelated, propagating these errors through Eq.(3.26) yields for the relative error $\sigma_L/\Delta L_{hc}^0$:

$$\frac{\sigma_L}{\Delta L_{hc}^0} = \left[\frac{\sigma_1^2}{k_1^2} + \frac{\sigma_R^2}{R^{*2}} + \frac{\sigma_m^2}{m_d^2} + \frac{\sigma_g^2}{g_m^2} \right]^{1/2} = 6.55 \cdot 10^{-4}. \quad (3.27)$$

and for ΔL_h^0 :

$$\Delta L_h^0 = \left[(0.0022768 \pm 0.0000015) \text{ m hPa}^{-1} \right] \frac{p_0}{f(\varphi, H)}. \quad (3.28)$$

Davis *et al.* [1985] come to a slightly different error estimation, as they assume a smaller uncertainty in k_1 .

The troposphere is responsible for $\sim 3/4$ of the hydrostatic delay. Under average meteorological conditions, ΔL_h^0 reaches about 2.3 m. A 1 hPa error in the measurement of the surface pressure leads to a ~ 2.3 mm error in the zenith delay. The error introduced by the assumption of hydrostatic equilibrium depends on the wind profile above the site, it is typically of the order of 0.01%, which corresponds to 0.2 mm in the zenith delay [Elgered, 1993]. Only under extreme weather conditions like thunderstorms, vertical accelerations can be expected to reach 1% of gravity, leading to an error of ~ 20 mm [Davis *et al.*, 1985].

3.2.5 Definition of the zenith wet delay

The remainder of (3.24) is denoted *zenith wet delay*, ΔL_w^0 :

$$\Delta L_w^0 = 10^{-6} \int_{z_{antenna}}^{\infty} \left(k_2' \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1} \right) dz. \quad (3.29)$$

The zenith wet delay can be as small as a few centimeters in arid regions and as large as 40 cm in humid areas [Bevis *et al.*, 1992].

3.3 The hydrostatic delay along ray paths with arbitrary elevation angles

For a plane parallel model of the earth and the atmosphere (neglecting the curvature of the earth and azimuthal variation of atmospheric parameters), the total atmospheric delay in an arbitrary direction is given by the simple cosecant law:

$$\Delta L = \frac{1}{\sin(\epsilon)} \Delta L^0 = \csc(\epsilon) \Delta L^0, \quad (3.30)$$

where ϵ is the elevation angle of the radio source [Davis *et al.*, 1985].

3.3.1 Definition of the mapping function

Generally, ΔL is given by:

$$\Delta L = m(\epsilon, \mathbf{p}) \Delta L^0. \quad (3.31)$$

The function $m(\epsilon, \mathbf{p})$ depends on the elevation angle ϵ and on the vector \mathbf{p} , which is some parameterized representation of the refractivity in the atmosphere [Davis *et al.*, 1985]. The function $m(\epsilon, \mathbf{p})$ is usually called *mapping function*, the dependence on the vector \mathbf{p} will be suppressed for the sake of simplicity by writing $m(\epsilon)$. The mapping function is therefore a dimensionless factor that relates the line of sight delay to the zenith hydrostatic delay.

3.3.2 The Niell mapping function

Niell [1996; and personal communications, 1998] has developed a mapping function whose coefficients depend on the latitude ϕ in ($^\circ$), the height above sea level H in (km), and on the day of year (*doy*). The mapping function does not depend on surface weather conditions on the site, it is formulated as follows:

$$m(\epsilon) = \frac{1 + \frac{a}{1 + \frac{b}{1 + c}}}{\sin \epsilon + \frac{a}{\sin \epsilon + \frac{b}{\sin \epsilon + c}}}. \quad (3.32)$$

At each latitude the coefficients a , b , and c are modeled as sinusoids in time with a period of 365.25 days, the phase is defined by $doy = 28$, corresponding to the winter extremum. For coefficient a , for example, the model is:

$$a(\varphi, doy) = a_{avg}(\varphi) - a_{ampl}(\varphi) \cos\left(2\pi \frac{doy - 28}{365.25}\right). \quad (3.33)$$

The same procedure is followed for the coefficients b and c . Average and amplitude values for a , b , and c are listed in Table 3.3.

Coefficient	Latitude				
	15°	30°	45°	60°	75°
a_{avg}	1.2769934e-3	1.2683230e-3	1.2465397e-3	1.2196049e-3	1.2045996e-3
b_{avg}	2.9153695e-3	2.9152299e-3	2.9288445e-3	2.9022565e-3	2.9024912e-3
c_{avg}	62.610505e-3	62.837393e-3	62.721774e-3	63.824265e-3	64.258455e-3
a_{ampl}	0.0	1.2709626e-5	2.6523662e-5	3.4000452e-5	4.1202191e-5
b_{ampl}	0.0	2.1414979e-5	3.0160779e-5	7.2562722e-5	11.723375e-5
c_{ampl}	0.0	9.0128400e-5	4.3497037e-5	84.795348e-5	170.37206e-5

Table 3.3: Coefficients of the hydrostatic Niell mapping function.

The coefficients were derived from profiles of the Standard Atmosphere at 15, 30, 45, 60, and 75° northern latitude. Since no data from the Southern Hemisphere were used to develop the mapping function, the inversion of the seasons has been accounted for by adding half a year to the phase for southern latitudes.

Values for different latitudes are obtained by linear interpolation between the nearest tabulated values. The height correction is given by:

$$\Delta m(\varepsilon) = \frac{dm(\varepsilon)}{dz} H, \quad (3.34)$$

with

$$\frac{dm(\varepsilon)}{dz} = \frac{1}{\sin \varepsilon} - f(\varepsilon, a_h, b_h, c_h), \quad (3.35)$$

where H is measured in (km) and $f(\epsilon, a_h, b_h, c_h)$ is the continued fraction Eq. (3.32), now used with the coefficients:

$$\begin{aligned} a_h &= 2.53e-5 \\ b_h &= 5.49e-3 \\ c_h &= 1.14e-3 \end{aligned}$$

3.3.3 The "geometric" mapping function

We have developed a simple mapping function, where the only free parameter (besides the elevation angle) is the climatological scale-height within the troposphere. The value of $m(\epsilon)$ is given by the ratio of the actual ray path length within the first scale-height above the surface (S_H) and the tropospheric scale-height (H_{trop}):

$$m(\epsilon) = \frac{S_H}{H_{trop}}. \quad (3.36)$$

It, alternatively, can be written as:

$$m(\epsilon) = \frac{1}{\sin \epsilon \frac{S_{flat}}{S_H}}, \quad (3.37)$$

to directly illustrate the deviation from the simple cosecant law. S_{flat} is the equivalent ray path length in a "flat" atmosphere, in which the "geometric" mapping function degenerates to the cosecant law. S_H can be readily determined by evaluation of the triangle formed by R_e , $R_e + H_{trop}$, and S_H (see Fig. 3.3), where R_e is the radius of the earth (or the local radius of curvature, depending on the desired accuracy). In general, it is sufficient to use a mean climatological scale-height value of 8.0 km.

Introducing the abbreviation

$$\tilde{r} = \frac{R_e}{R_e + H_{trop}}, \quad (3.38)$$

the path length within the first scale-height can be expressed as:

$$S_H = (R_e + H_{trop}) [\cos(\arcsin(\tilde{r} \cos \epsilon)) - \tilde{r} \sin \epsilon], \quad (3.39)$$

and Eq. (3.36), the "geometric" mapping function, becomes:

$$m(\epsilon) = \left(\frac{R_e}{H_{trop}} + 1 \right) [\cos(\arcsin(\tilde{r} \cos \epsilon)) - \tilde{r} \sin \epsilon]. \quad (3.40)$$

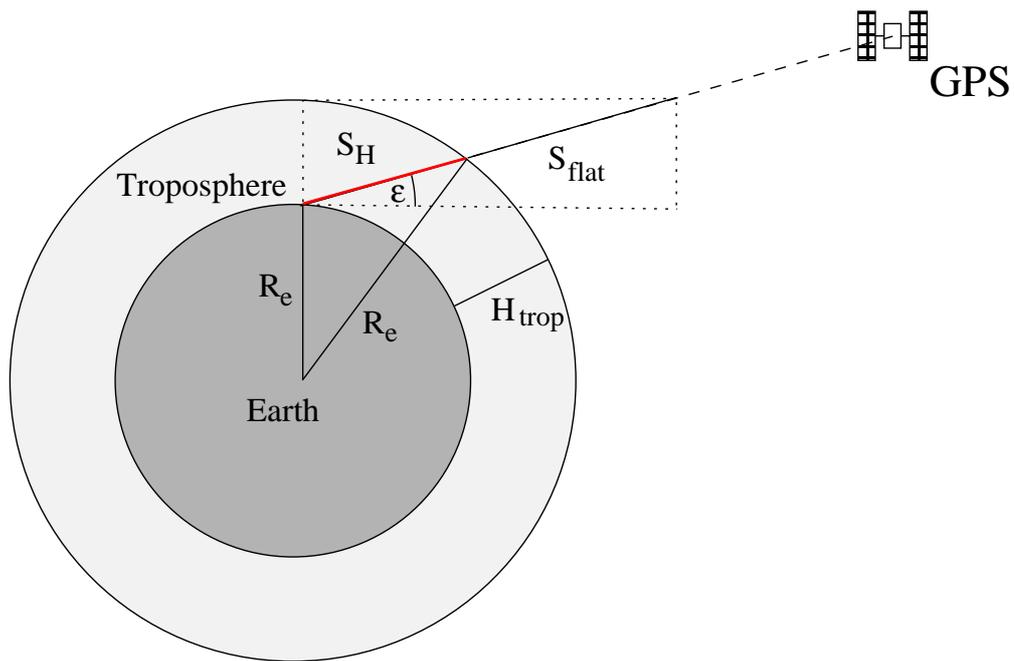


Fig. 3.3: Schematic illustration of the "geometric" mapping function, see text for explanation.

Values of the different mapping functions for elevation angles between 5° and 10° are plotted in Fig. 3.4 for the month October and for 60° northern latitude.

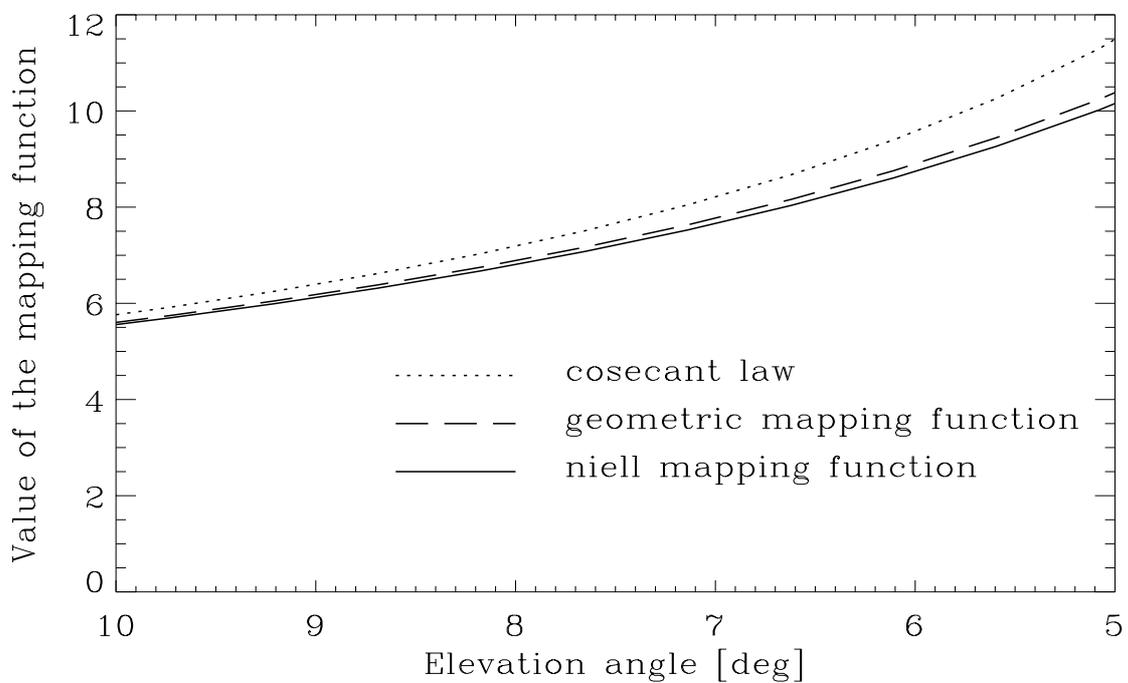


Fig. 3.4: Comparison of the Niell mapping function, the geometric mapping function, and the cosecant law as function of the elevation angle (low elevations) for October at 60° northern latitude.

3.4 Estimation of integrated water vapor from observed wet delays

The wet delay along a ray path with an arbitrary elevation angle can now be written as:

$$\Delta L_w = \Delta L - m(\epsilon)\Delta L_h^0. \quad (3.41)$$

3.4.1 From wet delay to integrated water vapor

By introducing a mean weighted temperature of the atmosphere, T_m [Davis *et al.*, 1985],

$$T_m = \frac{\int_S \frac{e}{T} Z_w^{-1} ds}{\int_S \frac{e}{T^2} Z_w^{-1} ds}, \quad (3.42)$$

Eq. (3.29) can be written in a form that is directly proportional to *the slant integrated water vapor (SIWV)* along the ray path S :

$$\Delta L_w = 10^{-6} \left(k_2' + \frac{k_3}{T_m} \right) \int_S \frac{e}{T} Z_w^{-1} ds, \quad (3.43)$$

which becomes, using the equation of state for water vapor:

$$\Delta L_w = 10^{-6} \left(k_2' + \frac{k_3}{T_m} \right) \frac{R^*}{m_w} \int_S \rho_w ds. \quad (3.44)$$

Equation (3.44) can be written in a simplified manner using the proportionality factor Π [cf. Bevis *et al.*, 1992]:

$$\Pi = \frac{10^6 m_w}{\left(k_2' + \frac{k_3}{T_m} \right) R^*}, \quad (3.45)$$

and the definition of the slant integrated water vapor, *SIWV* :

$$SIWV = \int_S \rho_w ds, \quad (3.46)$$

where *SIWV* is measured in (kg/m²) and Π has the dimension (kg/m³).

Equation (3.44) reads then simply as:

$$SIWV = \Pi \Delta L_w. \quad (3.47)$$

As k_2' is very small compared with k_3/T_m , Π is an almost linear function of T_m . Some Π values for selected mean temperatures are listed in Table 3.4, and Fig. 3.5 illustrates the quasi-linear dependence on the mean temperature.

T_m [°C]	Π [kg/m ³]
-20	144.54
-10	150.16
0	155.77
10	161.38
20	166.98
30	172.58

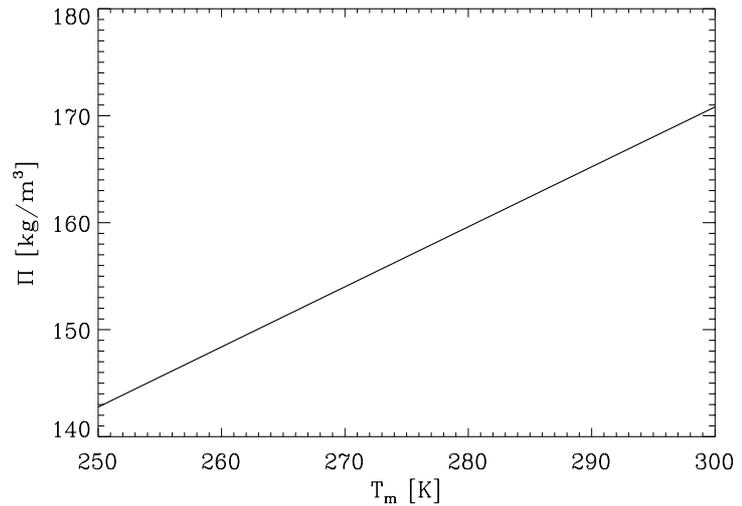


Table 3.4: Values of the factor Π for selected mean weighted temperatures of the atmosphere.

Fig. 3.5: The factor Π as a function of the mean temperature T_m .

The vertically integrated water vapor overlaying a station can also be stated in terms of *precipitable water (PW)*, the height of an equivalent column of liquid water:

$$PW = \frac{IWW}{\rho_l}, \quad (3.48)$$

where ρ_l is the density of liquid water, which has a value of, e.g., 999.84 kg/m³ at 0°C, 1000.00 kg/m³ at 3.98°C, and 998.21 kg/m³ at 20°C [CRC, 1993].

Introducing a dimensionless proportionality factor κ :

$$\kappa = \frac{\Pi}{\rho_l}, \quad (3.49)$$

Eq. (3.48) can be written as

$$PW = \kappa \Delta L_w^0, \quad (3.50)$$

where PW is measured in the same units as ΔL_w . The value of κ is typically ~ 0.16 , but it can vary by more than 15%, depending on latitude and season (cf. Table 3.4).

3.4.2 Estimation of the mean weighted temperature T_m

Equation (3.42) can be rewritten using the equation of state and the hydrostatic equation and neglecting the variation of g with height,

$$T_m = \frac{\int_0^{p_0} \frac{\rho_w}{\rho} dp}{\int_0^{p_0} \frac{\rho_w}{\rho T} dp} . \quad (3.51)$$

Using another representation of precipitable water PW

$$PW = \frac{1}{\rho_l g} \int_0^{p_0} \frac{\rho_w}{\rho} dp \quad (3.52)$$

and the definition of specific humidity q :

$$q = \frac{\rho_w}{\rho} , \quad (3.53)$$

Eq. (3.51) becomes:

$$T_m = \frac{\rho_l g}{\int_0^{p_0} \frac{q}{PW} \frac{dp}{T}} , \quad (3.54)$$

showing that the temperature profile is weighted by the fraction of the total water vapor in each layer [Ross and Rosenfeld, 1997]. Since about 50% of the total water vapor are usually found below 850 hPa, the weighting is strongest in the layers near the surface. T_m can therefore be expected to show a strong correlation with the surface temperature T_0 .

Analyzing more than 8700 radiosonde profiles from 13 sites in the United States with a latitude range of 27° to 65° (from West Palm Beach, Florida, to Fairbanks, Alaska) and a height range of 0 to 1.6 km over a two year period, Bevis *et al.* [1992] found the linear regression:

$$T_m = 70.2 + 0.72 T_0 , \quad (3.55)$$

with a root-mean-square (rms) deviation of 4.74 K, which corresponds to a relative error of less than 2% in the indicated latitude range.

Ross and Rosenfeld [1997] showed that the correlation between T_m and T_0 is weak for tropical stations with correlation coefficients generally less than 0.5. For low latitude stations it is therefore preferable to use climatological T_m values, the relative error can in this case be reduced to about 1% [for details see Ross and Rosenfeld, 1997]. The authors provide also tabulated station-specific values of regression slopes and intercepts for each month for 53 globally distributed stations, based on 23 years of radiosonde soundings. They showed furthermore that the relative errors in Eq. (3.55) range from 1% to 4%. Most errors are greater than 2% caused by a warm bias, probably introduced through the short data period.

Another approach is to model the height distribution of atmospheric water vapor. Water vapor usually decreases with height in a similar manner as the total pressure, but the scale height (the height at which the vapor pressure is reduced to 1/e of the surface value) is much smaller. This can be stated as [Askne and Nordius, 1987]:

$$r = r_0 \left(\frac{p}{p_0} \right)^\lambda, \quad (3.56)$$

where λ is typically in the order of 3 (depending on season and latitude) and r is the mass mixing ratio (subscript "0" indicates the surface value):

$$r = \frac{m_w}{m_d} \frac{e}{p_d} \approx \frac{m_w}{m_d} \frac{e}{p}, \quad (3.57)$$

With the approximation $p_d \approx p$, Eq. (3.56) can be written in terms of water vapor partial pressure:

$$e = e_0 \left(\frac{p}{p_0} \right)^{(\lambda+1)}. \quad (3.58)$$

The scale height of water vapor, H_w , is therefore given by [cf. Ross and Rosenfeld, 1997]:

$$H_w = \frac{H_{tot}}{(\lambda+1)}, \quad (3.59)$$

where H_{tot} is the scale height of the total pressure. Inserting a typical H_{tot} value of ~8 km, we see that H_w is in the range of 2 km.

In the height interval containing significant amounts of water vapor, the temperature normally decreases with height at an approximately constant rate (the lapse rate), denoted α and measured in (K/km):

$$T = T_0 - \alpha(z - H). \quad (3.60)$$

Inserting this simple but not unrealistic polytropic model for atmospheric stratification into the hydrostatic equation and integrating yields:

$$T = T_0 \left(\frac{p}{p_0} \right)^{\frac{\alpha R_d}{g_m}}, \quad (3.61)$$

which can be used to solve Eq. (3.42):

$$T_m = T_0 \left(1 - \frac{\alpha R_d}{(\lambda + 1)g_m} \right), \quad (3.62)$$

which can be rewritten using Eq. (3.59):

$$T_m \cong T_0 - \alpha H_w, \quad (3.63)$$

showing that T_m can be interpreted as the temperature of the atmosphere at the height equal to the effective thickness of the moist layer [Ross and Rosenfeld, 1997].

Askne and Nordius [1987] found, e.g., for Göteborg in May: $\alpha = 6.4$ K/km and $\lambda = 2.95$. For this special case, (3.62) reads with $T_0 = 285.2$ K:

$$T_m = T_0 (1 - 0.0475) = 271.7 \text{ K}, \quad (3.64)$$

whereas Eq. (3.55) yields a result of 275.5K.

3.4.3 Error budget of the factor Π

The uncertainties in the parameter Π derive from the uncertainties in T_m and in the constants k_3 and k_2' (in decreasing order of importance). Under the assumption that the errors in T_m , k_3 , and k_2' , (σ_T , σ_3 , and σ) are uncorrelated, propagating these errors through Eq. (3.45) yields for the relative error σ_Π/Π [cf. Bevis *et al.*, 1994]:

$$\frac{\sigma_\Pi}{\Pi} = \frac{1}{\left(\frac{k_3}{T_m} + k_2' \right)} \left[\frac{k_3^2}{T_m^4} \sigma_T^2 + \frac{\sigma_3^2}{T_m^2} + \sigma^2 \right]^{1/2}. \quad (3.65)$$

Equation (3.65) can be simplified, neglecting the small contribution of k_2' versus k_3/T_m (even for $T_m = 303\text{K}$, k_2' is only 1.8% of k_3/T_m):

$$\frac{\sigma_{\Pi}}{\Pi} \cong \left[\frac{\sigma_T^2}{T_m^2} + \frac{\sigma_3^2}{k_3^2} + \frac{T_m^2}{k_3^2} \sigma^2 \right]^{1/2}. \quad (3.66)$$

Assuming a T_m value of 283 K and a σ_T of 4.74 K [cf. Bevis *et al.*, 1992], the first term on the right side is $2.80 \cdot 10^{-4}$, the second is $1.03 \cdot 10^{-5}$, and the third is $2.77 \cdot 10^{-6}$. If σ_T exceeds $\sim 1\%$, the relative error in Π is clearly dominated by the relative error in T_m , as can be seen in Fig. 3.6:

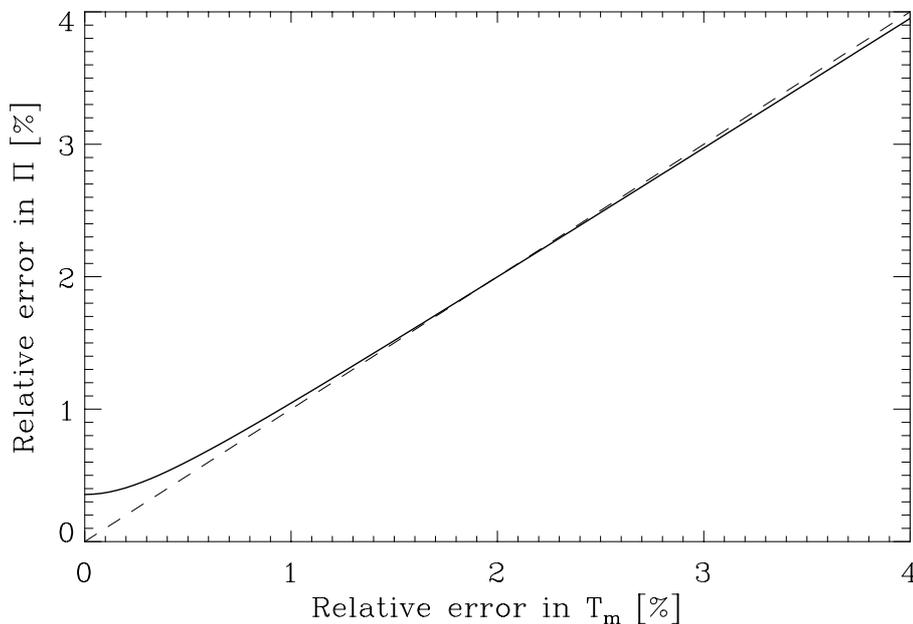


Fig. 3.6: The relative error in the factor Π as a function of the relative error in T_m assuming a mean temperature of 283K and the errors in the refractivity constants indicated in the text.

Figure 3.7 below shows the absolute error in the factor Π as a function of the mean temperature for different relative T_m errors. As the relative error in T_m can generally not be expected to be less than $\sim 1\%$, a minimum error of 1.5 kg/m^3 has to be assumed for the factor Π .

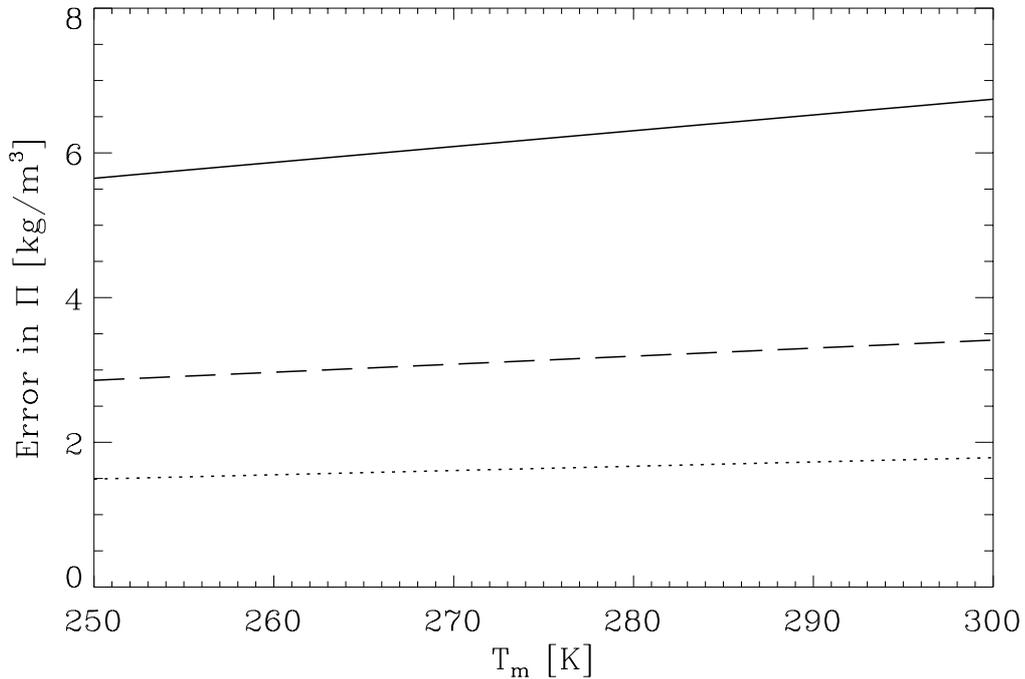


Fig. 3.7: Error in the factor Π as a function of T_m , assuming a relative error in T_m of 1% (dotted line), 2% (dashed line), and 4% (solid line).

3.5 Results from field experiments

In most experiments conducted so far, which have not aimed to determine the SIWV, a slightly different approach has been adopted. GPS measurements under different elevation angles were used to determine the zenith wet delay and consequently the precipitable water over the ground receiver [e.g., Rocken *et al.*, 1993, Duan *et al.*, 1996; Elgered *et al.*, 1997; and Emardson *et al.*, 1998].

3.5.1 Determination of precipitable water

Equation (3.41) becomes in this case

$$\Delta L = m_h(\epsilon) (\Delta L_h^0 + \Delta L_w^0), \quad (3.67)$$

or

$$\Delta L = m_h(\epsilon) \Delta L_h^0 + m_w(\epsilon) \Delta L_w^0, \quad (3.68)$$

if different mapping functions are used for the hydrostatic and the wet component of the neutral delay $m_h(\epsilon)$ and $m_w(\epsilon)$, respectively.

Based on Eq. (3.68) the zenith wet delay is given by:

$$\Delta L_w^0 = \frac{1}{m_w(\epsilon)} \left(\Delta L - m_h(\epsilon) \Delta L_h^0 \right), \quad (3.69)$$

and the precipitable water can subsequently be determined via Eq. (3.50):

$$PW = \kappa \Delta L_w^0.$$

The precipitable water is modeled using an average of all GPS observations above the elevation cut-off angle (usually chosen between 7° and 15°). This approach can introduce large errors, as the atmosphere frequently shows pronounced azimuthal (and temporal) water vapor variations [see, e.g., Treuhaft and Lanyi, 1987]. Davis *et al.* [1993], for example, found azimuthal variations of more than 20% using pointed radiometer measurements of slant integrated water vapor (SIWV) above 23° elevation.

3.5.2 Single and double differencing

Phase data from different stations of a receiver network can be used to eliminate clock errors, especially *Selective Availability* (SA) clock dithering errors (see section 3.1.2).

A *single difference* is the difference between measurements of the phase from one satellite as received at the same time at the two ends of a baseline (or at two stations in general) [Rocken *et al.*, 1993]. It eliminates GNSS transmitter clock errors.

Double differences are defined as the difference in carrier phase of one satellite as observed at the same time from two sites, differenced again with a similar observation of a second satellite:

$$\text{double difference} = (\varphi_{A1} - \varphi_{B1}) - (\varphi_{A2} - \varphi_{B2}), \quad (3.70)$$

where φ_{A1} is the carrier phase observation of satellite 1 from site A, φ_{B2} is the carrier phase observation of satellite 2 from site B, etc. Double differencing allows to eliminate GPS transmitter *and* GPS receiver clock errors [Ware *et al.*, 1997]. It is usually a necessary requirement to perform such differencing.

3.5.3 The problem of short baselines

For GPS networks with an inter-station spacing of less than several hundred kilometers, another significant problem arises in retrieving precipitable water data from differenced GPS observations (differential delays). The reason is that GPS satellite elevation angles are nearly identical at both ends of short baselines, and thus the zenith wet delay parameters for the two are highly correlated. The difference in the wet delays (ΔWD) associated with signals propagating from a single GPS satellite to two receivers (A and B):

$$\Delta WD = m(\epsilon_A) \Delta L_{wA}^0 - m(\epsilon_B) \Delta L_{wB}^0, \quad (3.71)$$

approaches

$$\Delta WD = m(\epsilon) (\Delta L_{wA}^0 - \Delta L_{wB}^0), \quad \text{with } \epsilon_A \cong \epsilon_B \equiv \epsilon, \quad (3.72)$$

as the stations A and B get closer together. In this case, the differential delay is sensitive only to the differential zenith delay ($\Delta L_{wA}^0 - \Delta L_{wB}^0$), and not to the absolute values (ΔL_{wA}^0 and ΔL_{wB}^0 , respectively) [cf. Duan *et al.*, 1996]. The obtained PW data are consequently correct except for an unknown bias that is common to all stations. The bias can be determined at one station using a collocated Water Vapor Radiometer (WVR). Using this technique (known as "WVR-levering"), Rocken *et al.* [1993] showed that it is possible to determine *PW* in near real time accurate to about 1 mm for a 50-km baseline [cf. also Ware *et al.*, 1993]. The usefulness of the "WVR-levering" technique is limited by the fact that WVR do not provide any useful data during periods in which the instruments are wetted by rainfall or heavy dew.

The "short baseline problem" can also be solved by incorporating a few remote global GPS tracking stations (and thus long baselines) into the geodetic solution. This approach eliminates any need for external comparison with WVR observations and delivers a pure GPS solution for precipitable water [Duan *et al.*, 1996].

3.5.4 The accuracy of PW and IWV measurements

Rocken *et al.* [1993] collected data from September 17 to November 28, 1992, at both ends of a 50-km baseline between Boulder and Platteville, Colorado. During this period, *PW* values ranged from 7 to 23 mm. Using broadcast orbit data for the GPS satellites, the WVR-measured and GPS-inferred values for the *PW differences* between the two ends of the baseline agreed to 1.0 mm (rms). Using precise orbit data, which are currently only

available with a delay of at least one day (about one week at the time of the Rocken *et al.* experiment), the overall rms difference could be reduced to 0.81 mm.

During a field experiment called GPS/STORM, data were collected at six stations in Oklahoma, Kansas, and Colorado for 22 h a day over a 30-day period in May 1993 [for a detailed description see, e.g., Businger *et al.*, 1996].

Rocken *et al.* [1995] used the "WVR-levering" method with one WVR at the reference site Platteville to derive PW values from the wet delay measurements. The measured PW values varied between ~ 10 mm and ~ 45 mm. At the secondary sites, the rms difference between GPS and WVR estimates was ~ 1.5 mm.

Duan *et al.* [1996] analyzed the data obtained during the GPS/STORM experiment with the pure GPS approach described above, incorporating a few remote tracking stations. The GPS-derived PW estimates were compared with WVR-derived PW estimates. The rms differences between the two sets of estimates vary between 1.14 and 1.45 mm, slightly better than those obtained by Rocken *et al.* [1995] with the "WVR-levering" method.

Elgered *et al.* [1997] used data acquired with the Swedish permanent GPS network during four days in December 1993. The PW data varied between 5 and 25 mm. Comparison with radiosonde and WVR data showed a rms difference of slightly less than 1 mm.

Emardson *et al.* [1998] estimated PW using three months of continuous data from permanent GPS receivers at 20 sites in Sweden and at 5 sites in Finland. The daily rms differences between GPS and WVR data ranged between 1 and 2 mm.

Ware *et al.* [1997] measured slant integrated water vapor (SIWV) along GPS ray paths. Comparing more than 17 000 GPS and pointed WVR observations, the authors found a rms agreement of 1.3 mm.

4 THE RADIO OCCULTATION TECHNIQUE

Radio occultation observations of atmospheres are made in an active limb scanning mode, where vertical scanning is provided by the relative motion between transmitter and receiver. From the standpoint of the receiver, an occultation event occurs when the spacecraft passes behind the planet and the ray path from the transmitter traverses the planetary atmosphere.

4.1 History of the radio occultation technique

The determination of the refractivity profile of a planetary atmosphere from radio occultation data represents basically the same problem as the determination of the seismic velocity profile in the earth from the observed travel times of seismic body waves [see, e.g., Phinney and Anderson, 1968].

4.1.1 The Wiechert-Herglotz method

The Wiechert-Herglotz method, developed 1910, is used to determine velocity-depth profiles (and consequently density profiles) in the earth from the observed variation of travel time with arc distance between source and receiver. The method is based on the assumptions of spherical symmetry, continuous refraction without reflection at interfaces, and monotonous increase of velocity with depth.

Phinney and Anderson [1968] adapted the Wiechert-Herglotz method for the radio occultation problem, the only constraints being those fundamental to geometric optics. In either case, a complete set of data can be inverted uniquely.

4.1.2 Planetary atmospheres

Initially the "bistatic radar occultation method" was used to probe the atmosphere of Mars during the Mariner IV flyby in 1965 ("Bistatic" means that transmitting and receiving terminals are at different locations). Fjeldbo and Eshleman [1965] determined scale height and density of the neutral atmosphere of Mars, analyzing atmospheric perturbations on the (single-frequency) tracking and telemetry signal received on earth. The authors furthermore proposed a two-frequency approach to allow the experimental separation of

dispersive ionospheric refraction effects from the (essentially) non-dispersive effects of the neutral atmosphere [see also Fjeldbo *et al.*, 1965 and Fjeldbo and Eshleman, 1968].

The proposed dual-frequency experiment was first conducted in 1967, during the Mariner V flyby at Venus and provided valuable information on the neutral atmosphere and ionosphere of the planet [Fjeldbo *et al.*, 1971].

The atmosphere of Mercury was explored by radio occultation of the Mariner X spacecraft [Fjeldbo *et al.*, 1976]. The atmosphere of Mars was continuously studied over one Martian year with the aid of radio occultation measurements during the Viking mission [Fjeldbo *et al.*, 1977; Lindal *et al.*, 1979].

In the following years, radio occultation measurements were an integral part of the spacecraft missions to the outer planets and their satellites. The atmospheres of Jupiter and Io were studied in 1973, when Pioneer 10 passed behind the planet and its moon [Kliore *et al.*, 1975]. In the course of the Voyager I mission, radio occultation experiments were carried out at Jupiter [see Lindal *et al.*, 1981] and Saturn. The Saturn system was explored in 1980 and occultation measurements were performed at the planet, its largest moon Titan, and also at the rings of the giant planet [see Lindal *et al.*, 1983 and 1985].

The atmospheres of Uranus [Lindal *et al.*, 1987 and Hinson; Magalhaes, 1991] and of Neptune [Lindal *et al.*, 1990 and 1992] were analyzed during the Voyager II mission.

First proposals to use the radio occultation technique also for remote sensing of the terrestrial atmosphere were made comparatively late [see, e.g., Lusignan *et al.*, 1969], when the method was fully developed for studying planetary atmospheres. Until recently, the comparatively low positioning accuracy of transmitter and receiver did not allow to derive atmospheric parameters with high quality (compared with other techniques). In the case of the planetary atmospheres, the situation was different, as only little was known about the height distribution of atmospheric parameters before the radio occultation experiments have been carried out.

With the advent of the U.S. Global Positioning System and the Russian counterpart, the Global Navigation Satellite System (GLONASS), information on satellite position and velocity could be obtained with the desired accuracy, making a terrestrial application of the radio occultation method feasible. Early proposals with transmitters on Navigation satellites and a receiver based on a Low Earth Orbit (LEO) satellite were presented by

Gurvich and Krasil'nikova [original publication 1987, English translation 1990] and Yunck *et al.* [1988].

The "proof of concept" was started with the launch of the MicroLab 1 satellite in 1995 and the onboard GPS Meteorology (GPS/MET) experiment (see section 4.4 for details), led by the University Corporation for Atmospheric Research (UCAR) [Ware *et al.*, 1996].

4.2 Principles of the radio occultation technique

4.2.1 The occultation geometry

In the geometrical optics approximation, a ray passing through the atmosphere is bent (refracted) according to Snell's law. The geometry of the occultation is summarized in Fig. 4.1. The refractivity N is a function of r in a shell $r_E \leq r \leq r_0$, where r_0 is an arbitrary level lying effectively above the mass of the atmosphere and r_E is the radius of Earth. Outside the Earth's atmosphere, the ray path is a straight line.

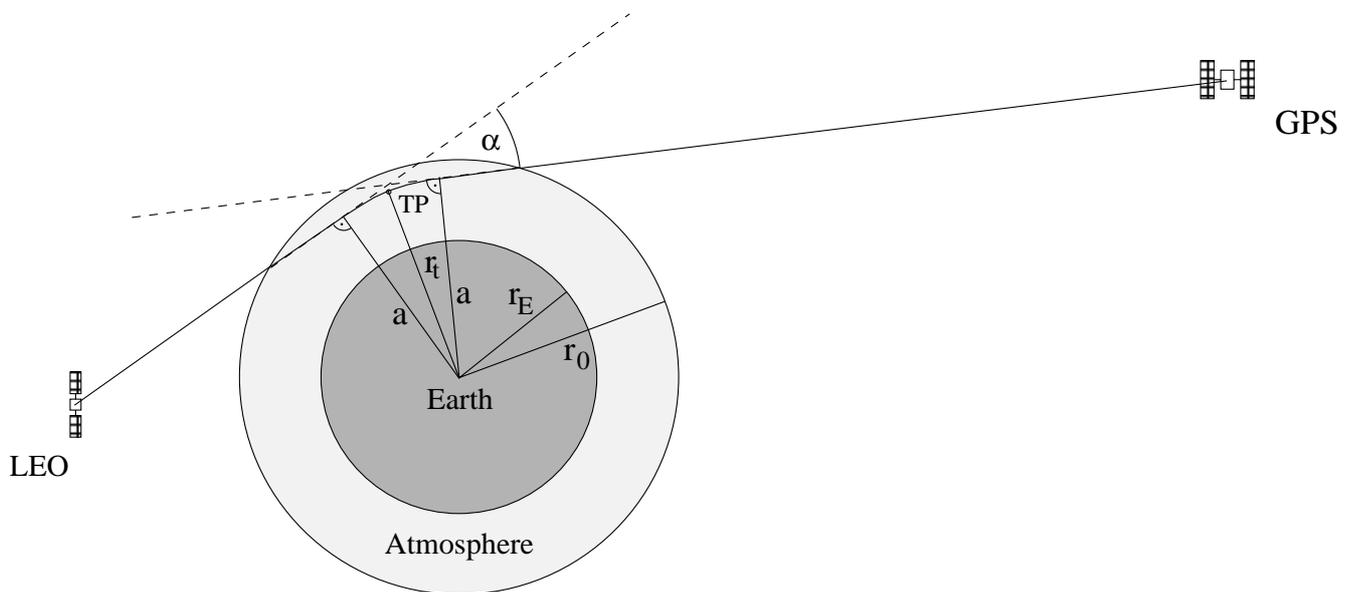


Fig. 4.1: Instantaneous geometry of the radio occultation experiment. The tangent point is indicated by "TP", the other symbols are explained in the text.

The overall effect of the terrestrial atmosphere can be characterized by the total bending angle α , the radius of the tangent point (or ray perigee) r_t , and the *impact parameter* a . The latter is defined as the perpendicular distance between the ray asymptote and the center of Earth (or the center of local curvature, depending on the desired accuracy). During an

occultation event, as the ray path descends or ascends through the atmosphere, the variation of α with a depends primarily on the vertical profile of refractivity. Assuming spherical symmetry, the time dependence of α and a can be derived from accurate measurements of the Doppler-shifted frequency of the occulted signal.

4.2.2 Atmospheric refraction

In the geometric optics approximation to the propagation of electromagnetic waves, the path of a ray passing through a region of varying refractive index n is governed globally by *Fermat's principle* of least time (see chapter 3 and [Born and Wolf, 1993]):

$$\int_S n(s) ds = \text{minimum}, \quad (4.1)$$

and locally by *Snell's law*, which states for plane layers that:

$$n \sin\beta = \text{constant}, \quad (4.2)$$

where β is the angle between the ray path and the gradient of the refractive index. Spherical symmetry implies furthermore that the ray path is lying in a plane [cf. Gorbunov and Sokolovskiy, 1993]. The situation in a spherically stratified medium is sketched in Fig. 4.2, indicated by concentric shells with constant refractive index (n_1 and n_2 , respectively) and therefore constant propagation velocity (c_1 and c_2 , respectively).

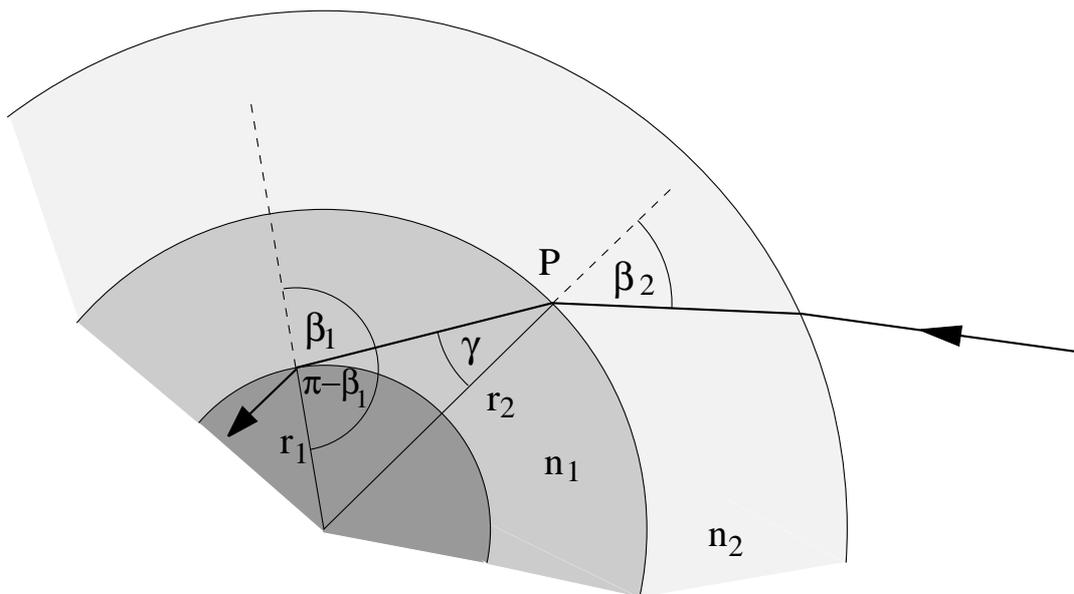


Fig. 4.2: Refraction in a spherically symmetrical medium, the gradient of n is directed radially ($n_1 > n_2$). The symbols are discussed in the text.

In a homogeneous medium, electromagnetic rays have the form of straight lines. At the point P, the law of refraction is valid in the original form (4.2), yielding:

$$\frac{\sin \gamma}{\sin \beta_2} = \frac{c_1}{c_2} = \frac{n_2}{n_1} . \quad (4.3)$$

On the other hand, $\sin \gamma$ can be expressed as:

$$\sin \gamma = \frac{r_1}{r_2} \sin(\pi - \beta_1) = \frac{r_1}{r_2} \sin \beta_1 , \quad (4.4)$$

and we see that:

$$n_1 r_1 \sin \beta_1 = n_2 r_2 \sin \beta_2 . \quad (4.5)$$

Letting the number of layers become infinitely large, we can apply this formula for an atmosphere where n is a continuous function of r . For this limiting case, we find:

$$n(r) r \sin \beta = \text{constant} = a , \quad (4.6)$$

where a is the *ray parameter* (or *impact parameter*). Equation (4.5) is also known as *Bouguer's formula* [cf. Born and Wolf, 1993]. At the tangent point, $\sin \beta = 1$, and $a = n(r_t) r_t$ (cf. Fig. 4.1).

4.2.3 The ray path equation

An expression for $d\beta$ may be obtained by differentiating Eq. (4.6). This operation gives:

$$d\beta = -\frac{a(ndr + rdn)}{nr \sqrt{(nr)^2 - a^2}} = -\frac{a \left(n + r \frac{dn}{dr} \right) dr}{nr \sqrt{(nr)^2 - a^2}} . \quad (4.7)$$

The ray path geometry and the polar coordinate system are illustrated in Fig. 4.3. We see that the angle of incidence, β , is related to the polar ray-path coordinates r and ϕ by:

$$\tan \beta = \frac{rd\phi}{dr} . \quad (4.8)$$

We see furthermore that:

$$\phi + \beta - \alpha = \frac{\pi}{2} . \quad (4.9)$$

The differentials of these parameters are therefore related by:

$$d\alpha = d\phi + d\beta . \quad (4.10)$$

The total bending angle α is given by integrating $d\alpha$ along the entire ray path. In a spherically symmetrical atmosphere, the ray path is symmetrical about the tangent point, therefore we get :

$$\alpha = \int_0^\alpha d\alpha' = 2a \int_{r=r_t}^{r=\infty} \frac{1}{\sqrt{(nr)^2 - a^2}} \frac{d \ln(n)}{dr} dr . \quad (4.13)$$

From (4.13) we see that α is a function of the impact parameter a , $\alpha = \alpha(a)$, and that it is caused by the radial variations of the refractive index.

4.2.4 The Abelian inversion

Eq. (4.13) has the form of an *Abelian integral equation* for $\ln(n)$, it can be inverted in order to express the refractive index as a function of the radius of the tangent point [cf. Bronstein and Semendjajew, 1989].

As an auxiliary variable, we define x [cf. Phinney and Anderson, 1968]:

$$x = r n(r) = \frac{a}{\sin \beta} , \quad (4.14)$$

which varies along the ray, taking the value r_0 at $r = r_0$ and the value a at $r = r_t$. x is the same function of r for all rays and is the value of the impact parameter of the ray which is horizontal at a given r . Equation (4.13) can now be written as:

$$\alpha = 2a \int_{x=a}^{x=\infty} \frac{1}{\sqrt{x^2 - a^2}} \frac{d \ln(n)}{dx} dx . \quad (4.15)$$

Let a_1 denote the parameter of a particular ray and r_1 the radius of the corresponding tangent point. We now multiply both sides of Eq. (4.15) with the kernel $(a^2 - a_1^2)^{-1/2}$ and integrate with respect to a from a_1 to infinity [cf. Fjeldbo *et al.*, 1971]:

$$\int_{a_1}^{\infty} \frac{\alpha da}{\sqrt{a^2 - a_1^2}} = \int_{a=a_1}^{a=\infty} \frac{2a}{\sqrt{a^2 - a_1^2}} \left[\int_{x=a}^{x=\infty} \frac{1}{\sqrt{x^2 - a^2}} \frac{d \ln(n)}{dx} dx \right] da . \quad (4.16)$$

Next we change the order of integration (and simultaneously the integration limits) on the right-hand-side and get:

$$\int_{a_1}^{\infty} \frac{\alpha da}{\sqrt{a^2 - a_1^2}} = \int_{x=a_1}^{x=\infty} \frac{d \ln(n)}{dx} \left[\int_{a=a_1}^{a=x} \frac{2a}{\sqrt{(a^2 - a_1^2)(x^2 - a^2)}} da \right] dx . \quad (4.17)$$

The inner integral (denoted I_1) can be solved with the aid of the following substitution:

$$u = \frac{a^2 - a_1^2}{x^2 - a_1^2} \quad \text{and} \quad du = \frac{2a da}{x^2 - a_1^2}. \quad (4.18)$$

Performing the substitution, we get:

$$I_1 = \int_{u=0}^{u=1} \frac{du}{\sqrt{u - u^2}} = \int_{u=0}^{u=1} \frac{2 du}{\sqrt{1 - (2u-1)^2}}. \quad (4.19)$$

Substituting $[(2u-1) = \sin t]$, we see that:

$$I_1 = \int_{t=-\frac{\pi}{2}}^{t=\frac{\pi}{2}} \frac{\cos t dt}{\sqrt{1 - \sin^2 t}} = \pi. \quad (4.20)$$

Equation (4.17) therefore becomes:

$$\int_{a_1}^{\infty} \frac{\alpha da}{\sqrt{a^2 - a_1^2}} = \pi \int_{x=n(r_1)r_1}^{x=\infty} \frac{d \ln(n)}{dx} dx = -\pi \ln n(r_1), \quad (4.21)$$

Rearranging Eq. (4.21), we get the classical Inverse Abel Transform:

$$n(r_1) = \exp \left[-\frac{1}{\pi} \int_{a_1}^{\infty} \frac{\alpha(a) da}{\sqrt{a^2 - a_1^2}} \right]. \quad (4.22)$$

Given $\alpha(a)$, Eq. (4.23) can be inverted numerically. To avoid the pole in the integrand, we may integrate Eq. (4.22) by parts and get [for details see, e.g., Steiner, 1998]:

$$n(r_1) = \exp \left[\frac{1}{\pi} \int_{\alpha=\alpha(a_1)}^{\alpha=0} \ln \left(\frac{a(\alpha)}{a_1} + \sqrt{\left(\frac{a(\alpha)}{a_1} \right)^2 - 1} \right) d\alpha \right]. \quad (4.23)$$

4.2.5 Ellipsoidal correction

The derivation of $n(r)$ from $\alpha(a)$ using the Abel Transform is only valid for a spherically symmetrical atmosphere. The ellipsoidal shape of the Earth introduces small departures from the spherical symmetry in the refractivity field. First order errors can be eliminated by replacing the Earth radius with the local *radius of curvature*, i.e., the radius of a sphere which is tangential to the ellipsoid at the location of the occultation event within the occultation plane. The center of local curvature replaces consequently the planetary center of mass [cf. Syndergaard, 1998].

4.2.6 The excess phase path

In practice, the bending angle itself is hardly measurable (only for near surface rays in a moist atmosphere it reaches near 2°), but we will show in this section that it can be derived from accurate measurements of Doppler shifted frequencies of the occulted, which are the time derivations of excess phase paths. The phases of the two GPS carrier frequencies (ϕ_1 and ϕ_2 , respectively) are measured at the receiver aboard the LEO satellite (up to a constant which, however, is cancelled out by time derivation). The corresponding excess phase paths (ΔL_1 and ΔL_2 , respectively) are given by:

$$\begin{aligned}\Delta L_1 &= \int_{GPS}^{LEO} n(s_1) ds_1 - S_0 = \Delta\phi_1 \lambda_1 \\ \Delta L_2 &= \int_{GPS}^{LEO} n(s_2) ds_2 - S_0 = \Delta\phi_2 \lambda_2 ,\end{aligned}\tag{4.24}$$

where λ_1 and λ_2 are the wavelengths of the carrier frequencies, and S_0 is the straight line distance between transmitter and receiver. It should be noticed that the two signals travel along slightly different paths (s_1 and s_2 , respectively) due to the dispersive effect of the ionosphere at radio frequencies.

4.2.7 Ionospheric correction

Exploiting the dispersion relation for radio waves in the ionosphere, we can remove the (first order) ionospheric contribution to bending by performing a linear combination (ΔL_c) of the excess phase paths (cf. chapter 3):

$$\Delta L_c = \frac{f_1^2 \Delta L_1 - f_2^2 \Delta L_2}{f_1^2 - f_2^2}.\tag{4.25}$$

The excess phase path ΔL_c contains only the delay due to the neutral atmosphere and a small (higher order) ionospheric residuum.

Instead of this approach known as "standard correction", the linear correction method can also be applied directly to bending [Vorob'ev and Krasil'nikova, 1994; Hocke, 1997]:

$$\alpha(a) = \frac{f_1^2 \alpha_1(a) - f_2^2 \alpha_2(a)}{f_1^2 - f_2^2}. \quad (4.26)$$

In this case the procedure described in sub-section 4.2.8 below is carried out for both ΔL_1 (yielding $\alpha_1(a)$), and ΔL_2 (yielding $\alpha_2(a)$), respectively, before Eq. (4.26) is applied. For occultations it is generally preferable to employ the "bending angle correction" rather than the "standard correction" as, e.g., shown by Steiner *et al.* [1999].

4.2.8 Derivation of bending angles from Doppler shift measurements

The atmospheric Doppler shift Δf in the carrier frequency f is given by the derivation of ΔL_c with respect to time:

$$\frac{\Delta f}{f} = \frac{1}{c} \frac{d(\Delta L_c)}{dt}. \quad (4.27)$$

Data from other GPS satellites can be used to determine the precise positions and velocities of the GPS and LEO satellites and to calculate the expected Doppler shift for a straight line signal path without atmospheric influence. Doppler shift is derived by the projection of spacecraft velocities onto the ray paths at the transmitter and receiver.

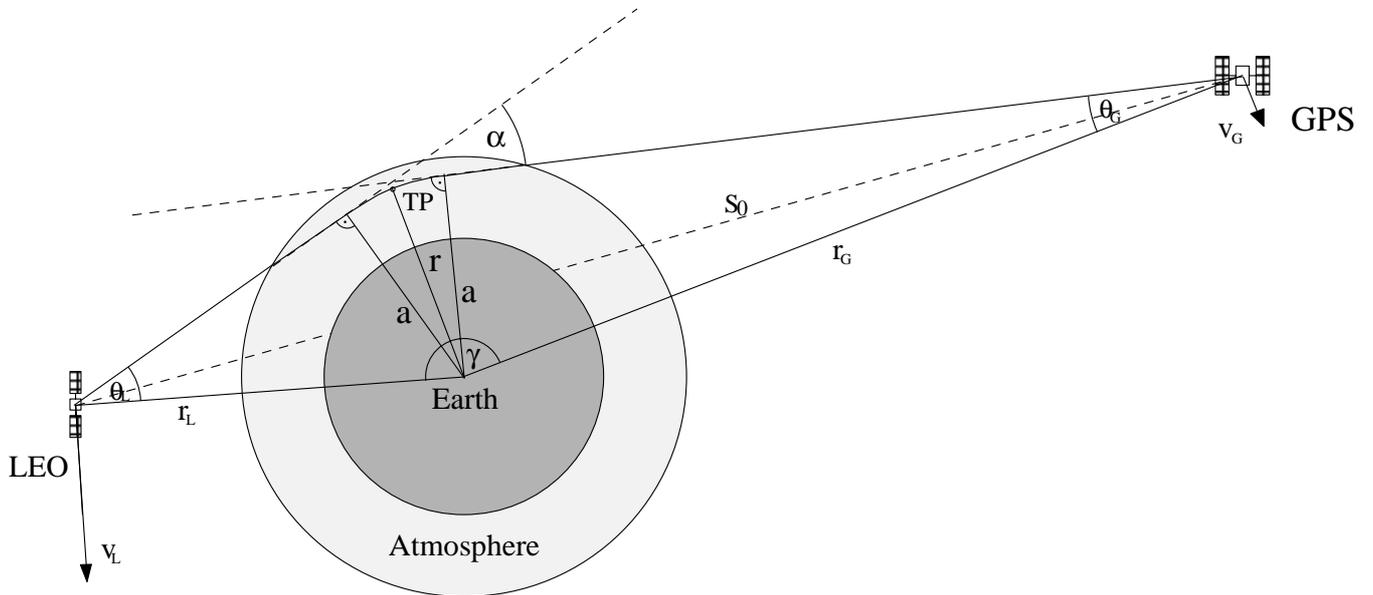


Fig. 4.4: Geometry of the occultation experiment. S_0 is the straight line distance between transmitter and receiver, the GPS satellite is characterized by the radius vector \mathbf{r}_G and the velocity (projected into the occultation plane) \mathbf{v}_G , the LEO satellite by \mathbf{r}_L and \mathbf{v}_L .

Using the geometry and notation of Fig. 4.4 and neglecting relativistic effects, we can relate Eq. (4.27) to kinematic quantities. Resolving the velocity vectors into radial and tangential components (denoted by the superscript r and ϕ , respectively) leads to:

$$\frac{\Delta f}{f} c = v_G^r \cos \theta_G + v_G^\phi \sin \theta_G + v_L^r \cos \theta_L - v_L^\phi \sin \theta_L . \quad (4.28)$$

As $n(r_L) = n(r_G) = 1$ (residual ionospheric refractivity at the height corresponding to r_G , is well canceled out by ionospheric correction) we can write Bouguer's law in the form:

$$r_G \sin \theta_G = r_L \sin \theta_L = a . \quad (4.29)$$

From the geometry of Fig. 4.4 we see furthermore that:

$$\alpha(a) = \theta_G + \theta_L + \gamma - \pi . \quad (4.30)$$

Thus, given measurements of transmitter and receiver position and velocity vectors, and of Doppler shift, $\alpha(a)$ can be derived iteratively from properly rearranged Eqs. (4.28) - (4.30) [cf., e.g., Kursinski, 1997].

4.2.9 Limits to vertical and horizontal resolution

The theory underlying section 4.2, geometric optics, assumes that transmitter signals follow a pencil-like ray path to the receiver, which is according to the Fermat principle (Eq. (4.1)) the path of minimum phase delay relative to neighboring, physically realizable paths. The geometric optics ray path is an approximation to Fresnel diffraction, discussed in the framework of wave optics, when the wavelength approaches zero. Considering Fresnel diffraction we see that the signal detected at the receiver corresponds to an effective cross beam sampling at the limb over the first Fresnel zone centered on the ray path [for details, see Born and Wolf, 1993]. The sampling volume is a cone formed by rays having the length of the distance from the receiver to the limb plus one-half the signal wavelength.

The vertical resolution ΔZ is therefore determined by the diameter of the first Fresnel zone. In the absence of significant atmospheric bending, the diameter of the first Fresnel zone (D_F) at the tangent point can be approximated by [cf., e.g., Kursinski *et al.*, 1997]:

$$\Delta Z = D_F = 2 \sqrt{\frac{\lambda R_G R_L}{R_G + R_L}} \approx 2 \sqrt{\lambda R_L} \quad (\text{as } R_G \gg R_L), \quad (4.31)$$

where λ is the signal wavelength, and R_G and R_L are the distance from the GPS and LEO satellite to the limb, respectively. For a receiver at 700 km altitude, R_L is ~ 3100 km, whereas R_G is about 28 500 km. At the GPS carrier frequency f_l ($\lambda_l \cong 19$ cm), Eq. (4.31) yields the result $D_F = 1.4$ km. In addition, the diffraction-limited vertical resolution is influenced by the vertical refractivity gradient. Due to the exponential increase of refractivity gradient with decreasing altitude, D_F decreases gradually from 1.4 km in the stratosphere to less than 0.5 km near the surface. The effect is particularly strong at the top of the marine boundary layer near 3.8 km, where the vertical resolution can reach 100 m [see, e.g., Kursinski *et al.*, 1997].

For a given vertical resolution, the along-track or horizontal resolution ΔH can be approximated by the length of the chord that is tangent to the inner of two concentric circles whose radii differ by ΔZ [cf. Melbourne *et al.*, 1994]:

$$\Delta H = 2\sqrt{2r_t\Delta Z}, \quad (4.32)$$

where r_t is the radius of the tangent point. Using Eq. (4.32), we see that the vertical resolutions of 1.4 km and 0.5 km correspond to horizontal resolutions of ~ 270 km and ~ 160 km, respectively. Further improvement of the vertical resolution can be reached by applying a diffraction correction based on the synthetic aperture principle [Gorbunov and Gurvich, 1998]. A detailed treatment of the resolution associated with radio occultation measurements, including effects of spherical asymmetry, is given by Ahmad and Tyler°[1998].

4.3 Derivation of atmospheric parameters

4.3.1 Density

After correcting for ionospheric effects, only the dry and moist atmospheric contributions to refractivity remain. In regions where the atmosphere is drier than a specific humidity of about 10^{-4} kg/kg [cf. Kursinski, 1997], the moist term can be neglected and refractivity can be expressed as (see chapter 2):

$$N = k_1 \frac{P}{T}, \quad \text{with } k_1 = 77.60 \text{ K/hPa}. \quad (4.33)$$

Using the ideal gas law (and ignoring the small effects of nonideal gas behavior within the first scale height above surface), the density of dry air (ρ_d [kg/m³]) can thus be written as function of refractivity [cf., e.g., Kirchengast and Ladreiter, 1996 and Kursinski *et al.*, 1996]:

$$\rho_d(z) = \frac{m_d}{R^* k_1} N(z) = b_1 N(z), \quad (4.34)$$

where the new constant b_1 is given by:

$$b_1 = \frac{m_d}{R^* k_1} = 4.4892 \cdot 10^{-3} \text{ kg/m}^3. \quad (4.35)$$

4.3.2 Pressure

In dry air below the homopause (about 100 km), m_d is a constant and the pressure p ($= p_d$) can be obtained from density by integrating the hydrostatic equation:

$$\frac{\partial p}{\partial z} = -\rho_d(z)g(z), \quad (4.36)$$

where g is the acceleration of gravity. The integration is started at a height level high in the atmosphere where $p \cong 0$ is reasonably assumed (practically ~ 120 km is a good choice and some climatological initial pressure is guessed there). Initial pressure errors decrease exponentially with scale height as the integration moves deeper into the atmosphere. $g(z)$ can be accurately specified by a standard gravity formula (e.g., Eq. (3.20)). The pressure p [hPa] as a function of height is then given by:

$$p(z) = \int_z^{\infty} g(z')\rho_d(z') dz' = b_2 \int_z^{\infty} g(z')N(z') dz', \quad (4.37)$$

with

$$b_2 = 4.4892 \cdot 10^{-5} \text{ hPa s}^2/\text{m}^2. \quad (4.38)$$

4.3.3 Temperature

The temperature profile can be derived using the ideal gas law again:

$$T(z) = \frac{m_d p(z)}{R^* \rho_d(z)} = \frac{m_d b_2}{R^* b_1 N(z)} \int_z^{\infty} g(z')N(z') dz' = \frac{b_3}{N(z)} \int_z^{\infty} g(z')N(z') dz', \quad (4.39)$$

where b_3 is defined as:

$$b_3 \equiv \frac{m_d}{R^*} \frac{b_2}{b_1} = 3.4836 \cdot 10^{-3} \text{ K s}^2/\text{m}^2. \quad (4.40)$$

The presence of significant amounts of tropospheric water vapor complicates the interpretation of refractivity. Near the surface in the tropics, water vapor can contribute up to 30% to net refractivity and can dominate the vertical refractivity gradients and bending [cf. Kursinski *et al.*, 1997]. Neglecting the contribution of water vapor results in retrieved temperatures which are too low. This systematic error is also known as "cold bias" and can reach several tens of Kelvin at low altitudes [cf. Rocken *et al.*, 1997].

4.3.4 Water vapor

If independent information on the height distribution of pressure and temperature is available from observations or meteorological analyses, the water vapor density profile $\rho_w(z)$ can be derived using the Smith-Weintraub representation of the refractivity formula (see chapter 2):

$$\rho_w(z) = \frac{N_w(z) m_w}{\left(k_2 - k_1 + \frac{k_3}{T(z)}\right) R^*} = \frac{N_w(z) m_w}{\tilde{k}_4 R^*}, \quad (4.41)$$

where N_w is the part of the refractivity which is attributable to water vapor (see chapter 2 for details). Alternatively we can write in terms of water vapor partial pressure:

$$e(z) = \frac{T^2(z)N(z) - k_1 p(z)T(z)}{k_4}. \quad (4.42)$$

Equation (4.42) can be used to determine the sensitivity of water vapor partial pressure to errors in pressure, temperature, and refractivity [cf. Kursinski *et al.*, 1995]. Under the assumption, that the errors in T , N , and p (σ_T , σ_N , and σ_p , respectively) are uncorrelated, propagating these errors through (4.42) by evaluation of the total differential yields for σ_e , the error in the water vapor partial pressure:

$$\sigma_e = \left[\left(\frac{2N}{k_4} T - \frac{k_1}{k_4} p \right)^2 \sigma_T^2 + \frac{T^4}{k_4^2} \sigma_N^2 + \frac{k_1^2 T^2}{k_4^2} \sigma_p^2 \right]^{1/2}. \quad (4.43)$$

If we assume that N and p are, relatively seen, free of errors and insert typical values for N , p , and T (288, 1013.25 hPa, and 273.15 K, respectively), we see that:

$$\sigma_e \cong \left(\frac{2N}{k_4} T - \frac{k_1}{k_4} p \right) \sigma_T = 0.21 \sigma_T . \quad (4.44)$$

If we want to determine the vapor pressure with an accuracy of < 0.5 hPa, the temperature should be known with an accuracy of < 2.4 K. The best humidity data can therefore be expected at low latitudes, where temperature variations with time and location are small compared with variations of the vapor pressure. Kursinski *et al.* [1995] estimated that individual profile accuracies of better than 5% in the boundary layer and 20% up to 6 or 7 km altitude can be achieved at low latitudes. At 45° latitude, the altitude of 20% accuracy is expected to vary from about 5 km in summer to 2-3 km in winter.

4.4 The GPS/MET experiment

The GPS/MET experiment, led by the University Corporation for Atmospheric Research (UCAR), Boulder, U.S.A., was the first practical demonstration of GPS based radio occultation in the Earth's atmosphere [Ware *et al.*, 1996].

4.4.1 The LEO spacecraft and the GPS receiver

The GPS/MET instrument consists of a high precision TurboRogue receiver aboard the MicroLab-1 satellite. The small research satellite was launched in April 1995 and orbits the Earth at an altitude of about 730 km with a period of about 100 min in a near circular orbit with an inclination of about 70° . The dual-frequency GPS receiver was built by the Jet Propulsion Laboratory (JPL) and weighs less than 10 kg. Until March 1997, it measured excess phase and amplitudes during the setting of the GPS-to-LEO link through the Earth's atmosphere [see Rocken *et al.*, 1997].

4.4.2 Occultation events

For a LEO satellite at the indicated altitude, approximately 250 setting occultations occur each day (cf. chapter 5). However, because of gaps in the ground tracking network and memory limitations on board MicroLab-1, only about 150 soundings per day were performed under optimal conditions. Roughly 70 000 occultations have been collected and archived according to Rocken *et al.* [1997].

Up to eight GPS satellites were tracked simultaneously, providing carrier phase observations up to fifty times per second and pseudo-range (navigation) measurements up to once per second. During four "prime times", when the Anti-Spoofing encryption of the GPS signal was turned off, both GPS carrier frequencies could be sampled with the same performance. Double Differencing (incorporating co-visible ground stations) was applied to eliminate clock errors and the influence of Selective Availability (cf. chapter 3).

Each occultation event had a typical duration of about 1.5 min, spanning a height range from about 100 km down into the troposphere. Due to tracking problems, GPS/MET soundings frequently failed to penetrate the lowest 5 km of the troposphere in the presence of significant water vapor. The tracking problems result from the relatively low gain antenna used for the GPS/MET proof-of-concept instrument. Future instruments will use higher gain directional antennas, providing measurements down to lower altitudes [cf., e.g., Rocken *et al.*, 1997].

4.4.3 Accuracy of the derived atmospheric profiles

Steiner *et al.* [1999] validated GPS/MET-derived refractivity and temperature profiles through comparison with co-located profiles from European Center for Medium-range Weather Forecasts (ECMWF) analyses. The mean refractivity differences were found to be less than 0.25% between 8 and 22 km, and the standard deviations less than 0.8%. Mean temperature differences were found to be less than 0.5 K between 10 and 22 km and the standard deviation is less than 1.5 K.

Rocken *et al.* [1997] demonstrated by comparison of several thousands of GPS/MET data profiles with the best correlative data sets (radiosondes, etc.), that a ~ 1 K agreement in the retrieved temperature can be reached between 1 and 40 km. The authors showed furthermore, that National Centers for Environmental Prediction (NCEP) and GPS/MET water vapor partial pressure agreed to better than 0.5 hPa down to an altitude of 1 km.

5 SPACE-TIME SAMPLING CONDITIONS FOR ATMOSPHERIC IMAGING

5.1 Introduction

5.1.1 Definition of atmospheric imaging

Atmospheric imaging, as understood here, means the reconstruction of a two- or more-dimensional field of an atmospheric variable either based on remote atmospheric sounding measurements alone ("true" imaging) or, in the variant of *data assimilation*, invoking atmospheric model assistance ("model-assisted" imaging). In chapter 4 we have demonstrated that radio occultation, as a standalone technique, delivers accurate profiles of atmospheric bending angles, refractivity, temperature, and humidity, with high vertical resolution.

Imaging becomes feasible, when occultation data are combined with other sounding data (e.g., ground-based GPS measurements as discussed in chapter 3) and/or model information.

After formulating space-time and co-location requirements for "model-assisted" and "true" imaging, we will provide quantitative information on imaging feasibility based on realistic atmospheric sounding geometry simulations which have been performed for different imaging scenarios [for a detailed description, see Foelsche and Kirchengast, 1997].

5.1.2 Simulation of occultation events

Occultation observations can be simulated using the End-to-end GNSS (Global Navigation Satellite System) Occultation Performance Simulator (EGOPS) with appropriate satellite orbit element files [see Kirchengast, 1996 and 1997 for details on the EGOPS tool].

As a representative example for the individual geometrical (space-time) behavior of realistic occultation events, Fig. 5.1 shows in the left-hand panels the slowing of the signal due to the neutral atmosphere (solid lines) for three selected occultation events at low, middle, and high latitudes.

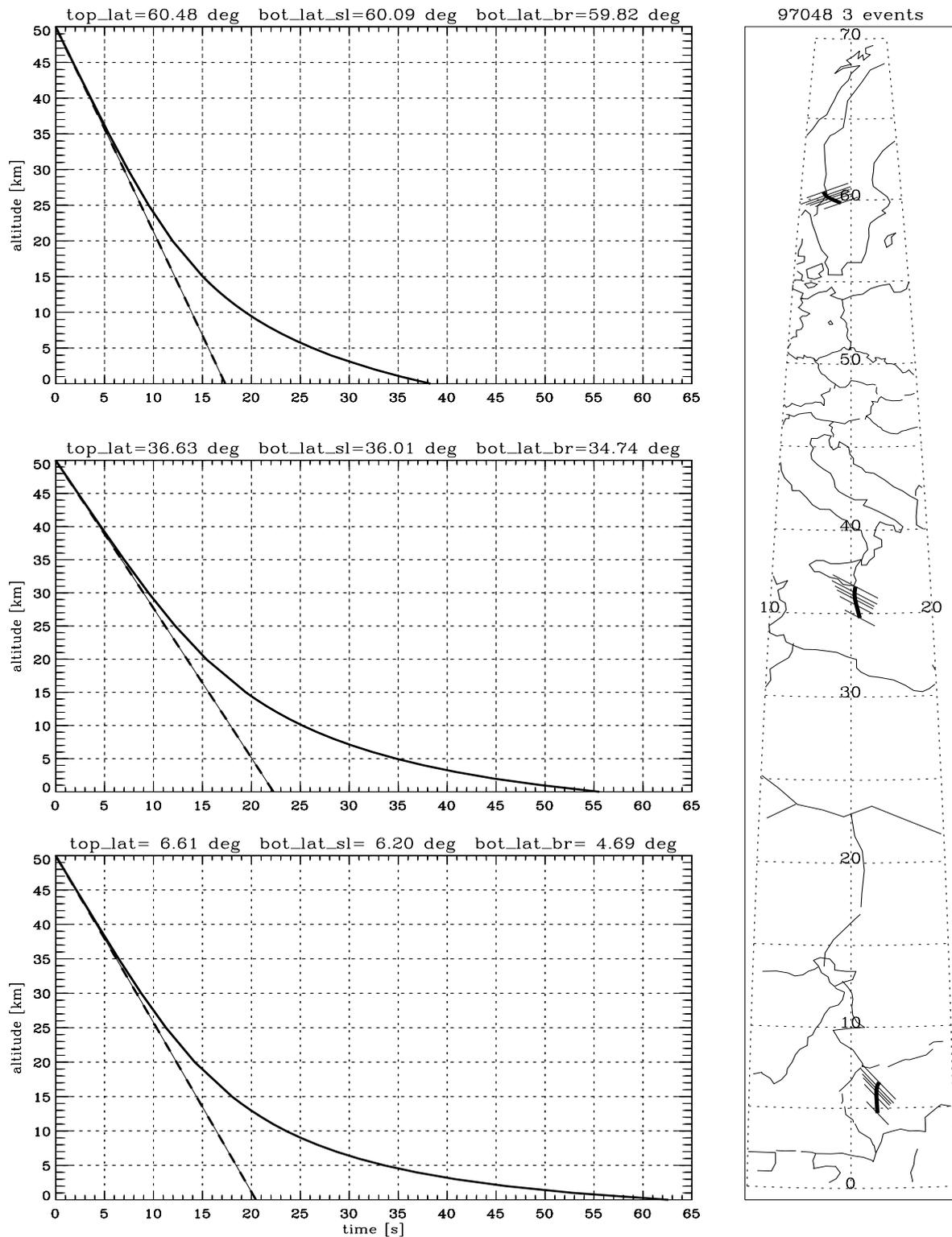


Fig. 5.1. Left panels: Height of ray perigee vs. time for three selected occultation events, straight rays (dashed line) and bended rays (solid line). Right panel: location of the events, indicated by the following symbol: ground projection of the tangent point trajectory (heavy line), ray-segments at 1 km, 5 km, 10 km, 15 km, 25 km, and 50 km (light lines), with lengths of ± 150 km (50 km level), decreasing to ± 100 km (1km level).

Dashed lines indicate ray propagation without atmospheric bending. In the high latitude case, signal slowing is primarily due to the dry atmosphere, whereas the other two panels show the increasing influence of atmospheric water vapor with decreasing latitude. The right-hand panel shows the location of the occultation events, the ground projection of the tangent point trajectory (heavy line), and ray-segments at 1 km, 5 km, 10 km, 15 km, 25 km, and 50 km (light lines), with lengths of ± 150 km (50 km level), decreasing to ± 100 km (1 km level), corresponding roughly to the horizontal resolution of the radio occultation measurements at these altitudes. In unfavorable cases, the length of the tangent point trajectory projection (smearing length) can reach several hundred kilometers, as in the case of the occultation event south of Sicily, Italy.

The following simulations on space-time sampling are based on the operational system of 48 GNSS satellites: 24 GPS (Global Positioning System) and 24 GLONASS (GLObal Navigation Satellite System). We always considered bended rays. August 1, 1996, was chosen as reference day (the choice was for convenience only as satellite orbit elements were available for this date; the absolute date does not matter for these simulations).

5.2 Observational requirements

Space-time sampling requirements for atmospheric imaging are considered for two different scenarios:

- "*Model-assisted imaging*", i.e., imaging by assimilation of GNSS occultation data into global or regional atmospheric models (see Table 5.1 below).
- "*True imaging*" by directly combining GNSS occultation data with ground-based GNSS sounding data for two- or more-dimensional field reconstruction (see Table 5.2 below). This second approach is the one we then focus in chapter 6 in terms of two-dimensional field reconstruction by tomographic methodology.

5.2.1 Requirements for model-assisted imaging

As the vertical resolution of occultation data is known to be good (about 0.5-1.5 km by geometric optics concepts) and readily sampled properly by occultation instruments, which typically use 10 - 50 Hz sampling rates for vertical scanning velocities ≤ 3 km/s, we will focus below on the investigation of horizontal sampling and time sampling requirements.

	Global Assimilation		Regional Assimilation	
	Ideal data	Outstanding data	Ideal data	Outstanding data
Horizontal domain	Global	Global	Regional	Regional
Horizontal sampling	< 300 km	< 1000 km	< 50 km	< 300 km
Time sampling	1 - 6 hours	< 12 hours	1 - 6 hours	< 12 hours
Vertical domain	0 - 80 km	0 - 80 km	0 - 80 km	0 - 80 km
Vertical sampling	0.1 - 0.5 km	< 0.7 km	0.1 - 0.5 km	< 0.7 km

Table 5.1: Requirements for "model assisted" imaging [Sources: ESA, 1996a; GRAS-SAG, 1998; D. Offiler, priv. communications, 1997; G. Kirchengast, priv. communications, 1998].

"Regional" in Table 5.1 is understood to be the typical region covered by LAMs (Local Area Models). In this study the European Area spanning $30^\circ\text{N} < \text{latitude} < 72^\circ\text{N}$, $20^\circ\text{W} < \text{longitude} < 37^\circ\text{E}$ is considered representative for such a regional domain.

"Ideal data" means that state-of-the-art requirements (as, e.g., set out by the World Meteorological Organization) are entirely fulfilled. "Outstanding data" means that the observing system would still outperform traditional data sources, like radiosondes.

Vertical domain requirements refer to bending angle data, temperature soundings are needed in the 0-50 km domain, and humidity profiles in the 0-10 km (with some margin the 0-15 km) domain.

5.2.2 Requirements for "true" imaging

"True" imaging requires local chains or networks of ground-based GNSS receivers and the occurrence of occultation events in the area of interest. It is practically only feasible and useful for atmospheric humidity (an extensive variable, while refractivity and temperature are intensive variables [see, e.g., Peixoto and Oort, 1992]). The vertical domain is therefore limited for the "true" imaging geometry simulations to 0-15 km.

	Ideal data	Outstanding data
Ground station spacing	10 - 50 km	≤ 100 km
Time co-location	≤ 30 min	≤ 1 hour
Occultation plane-station distance	$< \pm 25$ km	$< \pm 50$ km
Occultation plane obliquity	$< 45^\circ$	$< 75^\circ$
Vertical domain	0 - 15 km	0 - 15 km
Vertical sampling	< 0.25 km	≤ 0.5 km

Table 5.2: Requirements for "true" imaging [Sources: Høeg *et al.*, 1995; Leitinger *et al.*, 1997; A. Rius, *priv. communications*, 1997; G. Kirchengast *priv. communications*, 1998]

Again the requirement on vertical sampling (and domain) can be fulfilled with no problems by occultation receivers and we thus focus below on assessing the other requirements of Table 5.2. We note just with respect to the vertical sampling requirements that indeed a vertical resolution of the order of 100 m can be intrinsic to the water vapor data in the troposphere when using advanced diffraction correction techniques [for details see, e.g., Gorbunov *et al.*, 1996b and Karayel and Hinson, 1997].

5.3 LEO satellite constellations

Four different scenarios (1, 2, 8, and 20 satellites) are considered which are introduced below. Orbital elements for the Low Earth Orbit (LEO) satellites are listed in Table 5.3 together with an example file illustrating the NORAD/NASA TLE (two-line element set) format [cf. Kelso, 1988, additional information can be found on the "Celes Trak" homepage: <http://celestrak.com/NORAD/elements>].

(1) Single satellite constellation: In this case we considered a receiver onboard the METOP (METeorological OPerational) platform. This weather satellite shall be launched in 2003 and will operate in a nearly circular sun-synchronous orbit (eccentricity 0.001, inclination 98.7° , mean altitude 826 km) with a mean argument of perigee of 90° (to minimize orbit perturbations due to earth oblateness). METOP will be equipped with a GRAS (GPS/GLONASS Receiver for Atmospheric Sounding) instrument which will collect occultations. Right ascension of the ascending node and mean anomaly are chosen in a way that METOP reaches the descending node at nominal 09:00 local solar

time [cf. Potti and Carmona, 1995]. The azimuthal antenna field-of-view was assumed to be $\pm 45^\circ$ (for all GNSS receiving antennas for all satellite scenarios).

- (2) Tandem constellation: Atmospheric Chemistry Explorer (ACE) and METOP. In order to exploit the synergy with the METOP instruments, ACE (if realized within ESA'S Earth Explorer programme) may be operating in the same orbit as METOP but separated by about 27.7° , so that ACE's limblooking instruments can sound roughly the same atmospheric region as METOP's downlooking instruments [cf. ESA, 1996b].
- (3) Eight satellite constellation: 6 AMICO-satellites (Atmospheric profiling Microsat Constellation), ACE and METOP. The AMICO satellites are evenly distributed (in-orbit separation 120°) in two orbits with 70° inclination (node separation 90°). Eccentricity and argument of perigee are the same as for METOP, the mean altitude is slightly lowered to 800 km.
- (4) Twenty satellite constellation: 12 Super-AMICOs (SAMICO), 6 AMICOs, ACE and METOP. The SAMICO satellites are evenly distributed (in-orbit separation 60°) in two orbits with 80° inclination. Eccentricity, argument of perigee and mean altitude are the same as for AMICO, but the SAMICO orbit positions are shifted against the AMICO positions to evenly cover all mean-anomaly angles (see Table 5.3).

5.4 Distribution of occultation events

Figure 5.2 to Figure 5.5 display the global and regional distribution of occultation events during a 24 hour interval for several cases based on LEO constellations described above. We used the equal-area Hammer-Aitoff projection for all maps to show the area coverage in a proper way. The Figures are presented in order to illustrate and discuss the compliance with the requirements for model-assisted imaging, i.e., global and regional assimilation (Table 5.1). The considered region ($30^\circ\text{N} < \text{latitude} < 72^\circ\text{N}$, $20^\circ\text{W} < \text{longitude} < 37^\circ\text{E}$) is hereinafter understood as "Europe". August 1, 1996 was chosen as reference day, and the corresponding GPS and GLONASS ephemeris data (also available in NORAD/NASA TLE format) were used for the simulation.

Satellite	i [°]	ε [1]	Ω [°]	ω [°]	T [min]
METOP	98.7	0.00116	264.4	294.1	101.3
ACE	98.7	0.00116	264.4	266.4	101.3
AMICO 1	70.0	0.00116	204.4	294.1	100.8
AMICO 2	70.0	0.00116	204.4	54.1	100.8
AMICO 3	70.0	0.00116	204.4	174.1	100.8
AMICO 4	70.0	0.00116	294.4	1354.1	100.8
AMICO 5	70.0	0.00116	294.4	114.1	100.8
AMICO 6	70.0	0.00116	294.4	234.1	100.8
SAMICO 1	80.0	0.00116	234.4	309.1	100.8
SAMICO 2	80.0	0.00116	234.4	9.1	100.8
SAMICO 3	80.0	0.00116	234.4	69.1	100.8
SAMICO 4	80.0	0.00116	234.4	129.1	100.8
SAMICO 5	80.0	0.00116	234.4	189.1	100.8
SAMICO 6	80.0	0.00116	234.4	249.1	100.8
SAMICO 7	80.0	0.00116	324.4	324.1	100.8
SAMICO 8	80.0	0.00116	324.4	24.1	100.8
SAMICO 9	80.0	0.00116	324.4	84.1	100.8
SAMICO 10	80.0	0.00116	324.4	144.1	100.8
SAMICO 11	80.0	0.00116	324.4	204.1	100.8
SAMICO 12	80.0	0.00116	324.4	264.5	100.8

Table 5.3: Orbital elements of the LEO satellites, i is the inclination, ε the numerical eccentricity, Ω the right ascension of the ascending node, ω the argument of Perigee at a given instant, and T is the revolution period. A part of the corresponding TLE file is shown below, the ephemeris data are contained for each satellite in the respective second line.

```

METOP-1                                M1
 1   1   96213USR 96213.50000000 .00000000 00000-0 00000-0 0 0010
 2   1   98.7047 264.3857 0011650 90.0000 294.1230 14.21635855 0
ACE                                       E1
 1   1   96213USR 96213.50000000 .00000000 00000-0 00000-0 0 0010
 2   1   98.7047 264.3857 0011650 90.0000 266.4230 14.21635855 0

```

5.4.1 Single satellite constellation

A GPS receiver on board of a single LEO satellite with restriction to setting occultations (as in the GPS/MET experiment) provides up to 250 globally distributed occultations per day (Fig. 5.2). The additional use of GLONASS satellites and rising occultations (Fig. 5.3) leads to more than 1000 occultations per day. This constellation (or a four-satellite constellation with restriction to GPS and setting occultations) fulfills the "Outstanding data"-requirements for imaging by assimilation of occultation data into global models. Note, by the way, that this shows the big advantage of exploiting the full GNSS system together with using both fore- and off-looking receiving antennas: A single satellite already provides outstanding data for global assimilation. The actual METOP/GRAS instrument will, though, according to current plans, exploit only GPS setting and rising occultations leading to about 500 occultations per day, a case in fact "between" Fig. 5.2 and Fig. 5.3.

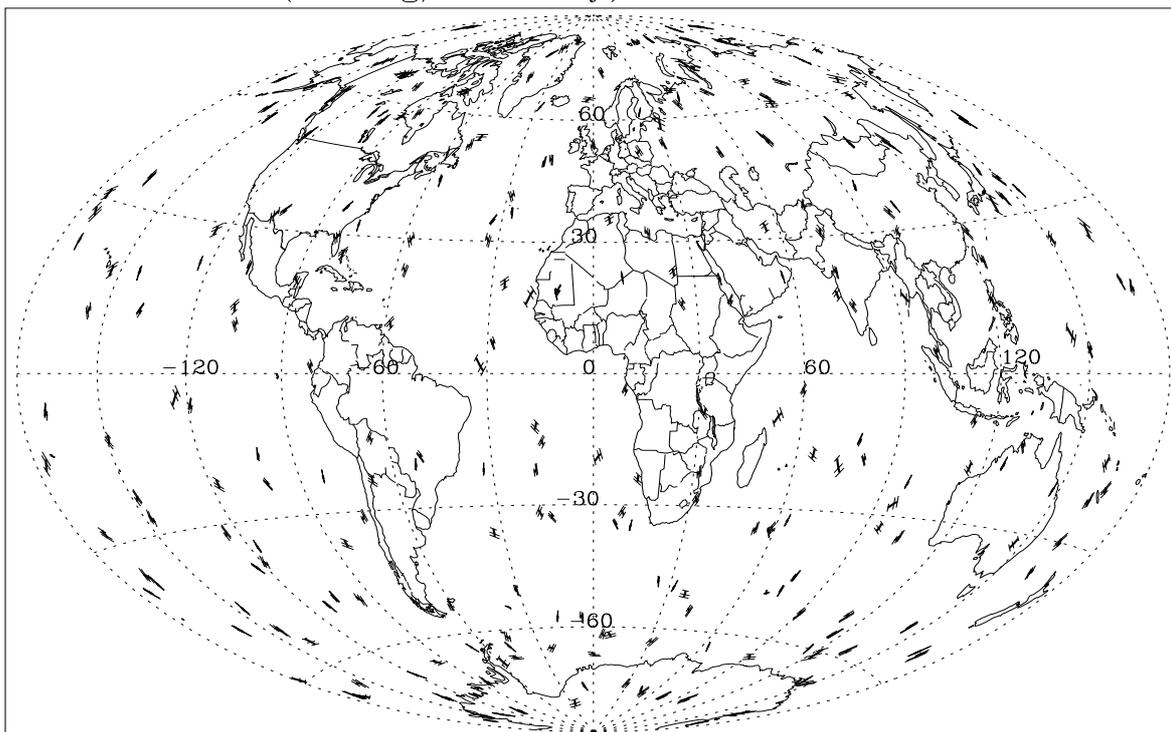
5.4.2 Tandem constellation

The tandem constellation (METOP and ACE, Fig. 5.4) offers pairs of occultation events with small spatial and temporal separation, clearly visible in the bottom panel of Fig. 5.4. The total number of profiles already exceeds that provided by the global radiosonde network. The occultation events are furthermore uniformly distributed, in contrast to the radiosonde measurements, which are sparse over the oceans and in the Southern Hemisphere. The space-time clustering of events may be useful for "tandem processing" algorithms which may intrinsically account for horizontal atmospheric gradients.

5.4.3 Eight satellite constellation

Eight satellites equipped with GNSS receivers provide about 8000 occultation events per day (approximately 1000 per satellite), if the receivers are capable to track setting and rising occultations, and the entire GNSS can be exploited. We do not illustrate this case in detail here, a map such as in Figures 5.2 - 5.5 can be found in Foelsche and Kirchengast [1997].

METOP (setting/GPS only) 24 hours 262 events



METOP (setting/GPS only) 24 hours 10 events

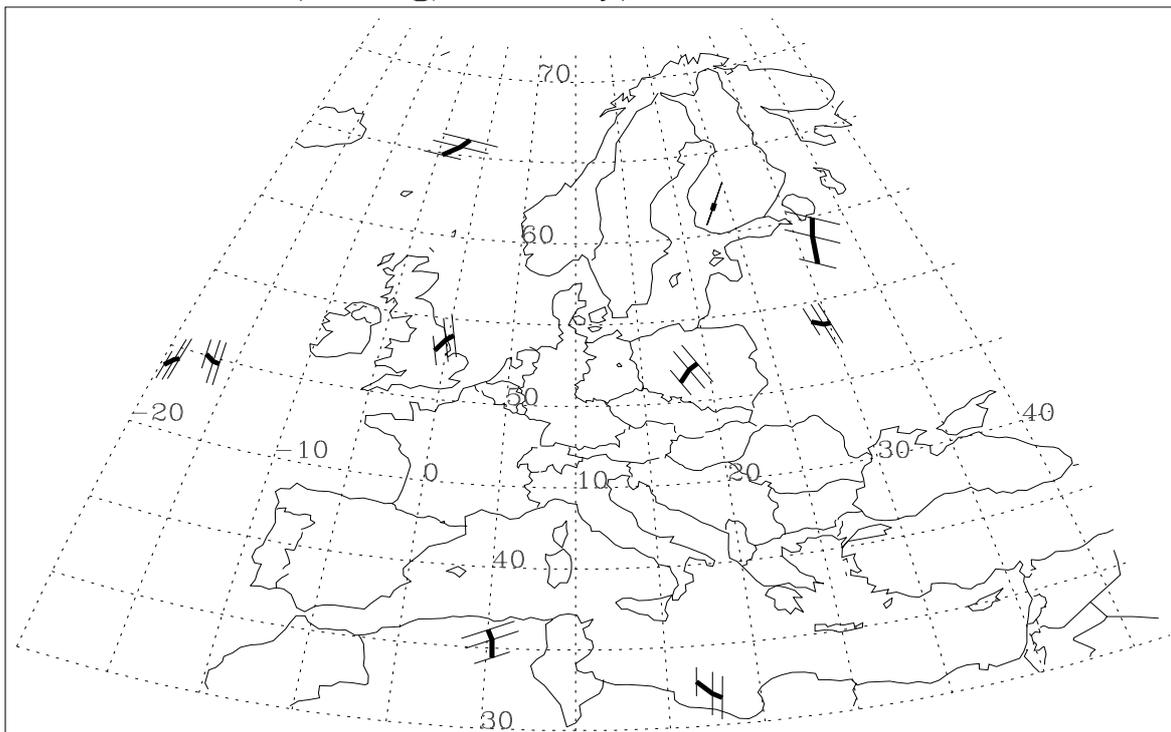
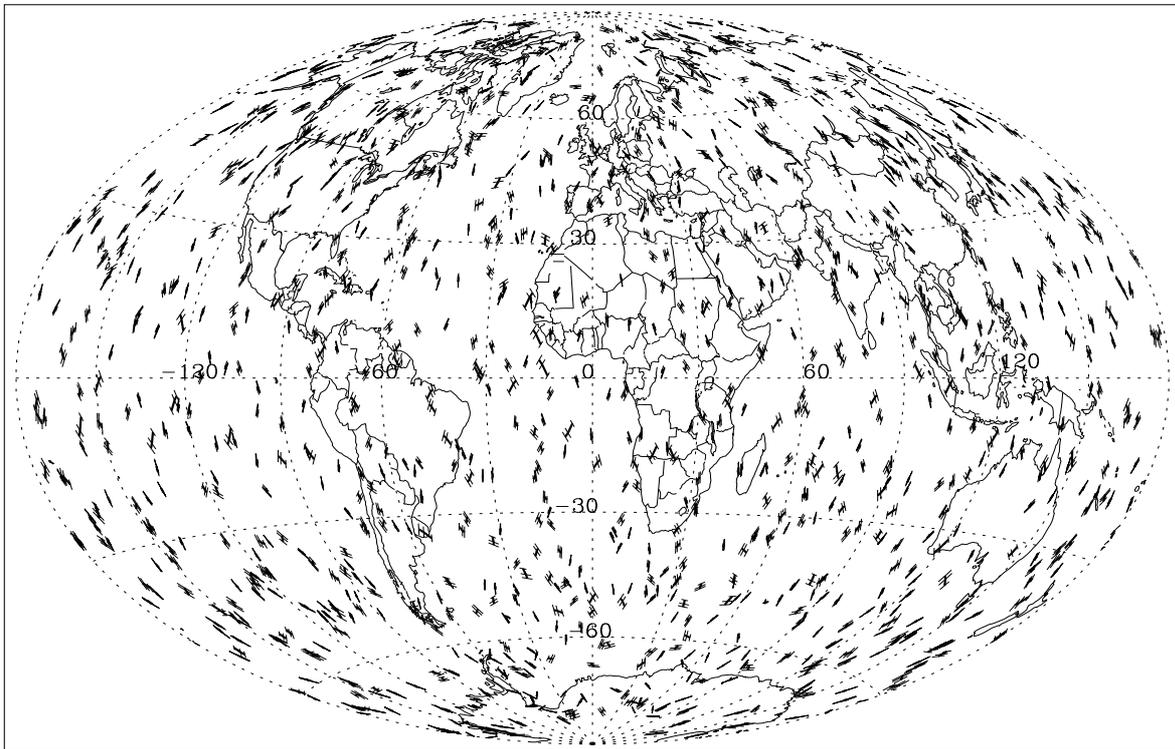


Fig. 5.2: Simulation of the geographical distribution of occultation events during 24 hours (August 1, 1996) for the one LEO constellation (METOP), only setting occultations (bended rays) and GPS satellites are considered. Global map (top panel), Europe (bottom panel), symbols as in Fig. 5.1 but only three height levels (1km, 15km, 50km).

METOP (total) 24 hours 1041 events



METOP (total) 24 hours 43 events

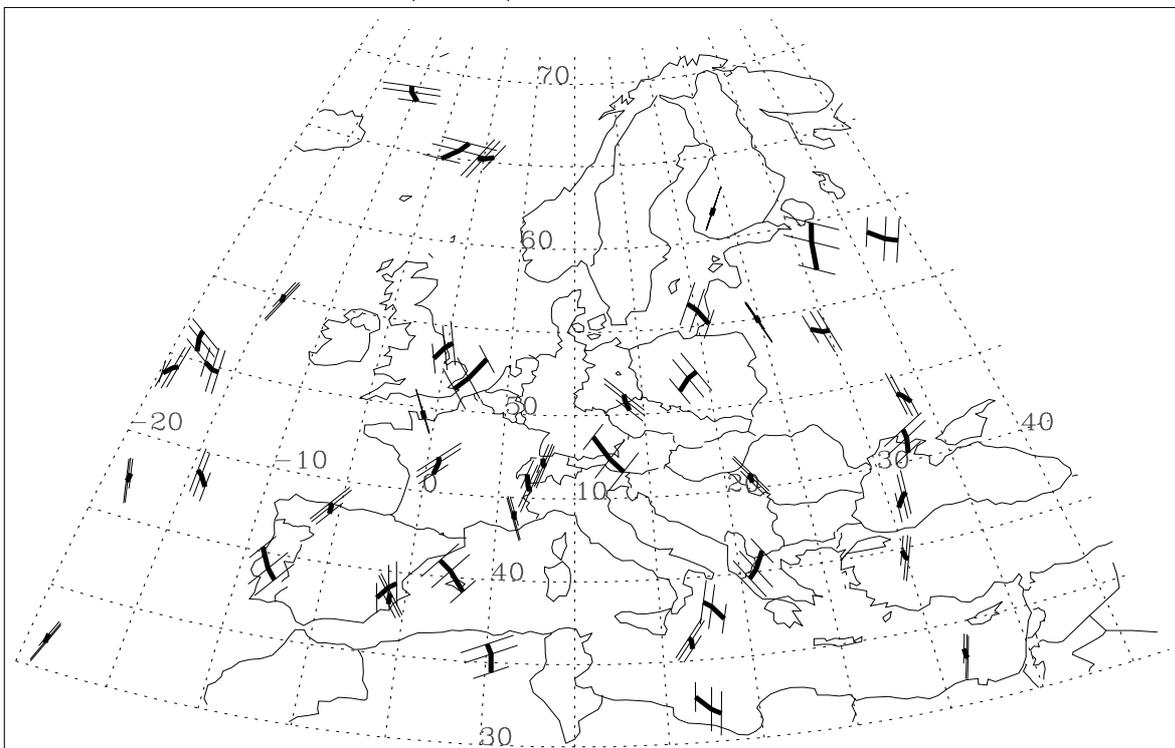


Fig. 5.3: Simulation of the geographical distribution of occultation events during 24 hours for the one LEO constellation (METOP), GPS and GLONASS, setting and rising occultations (indicated as "total"). Same maps and symbols as in Fig. 5.2.

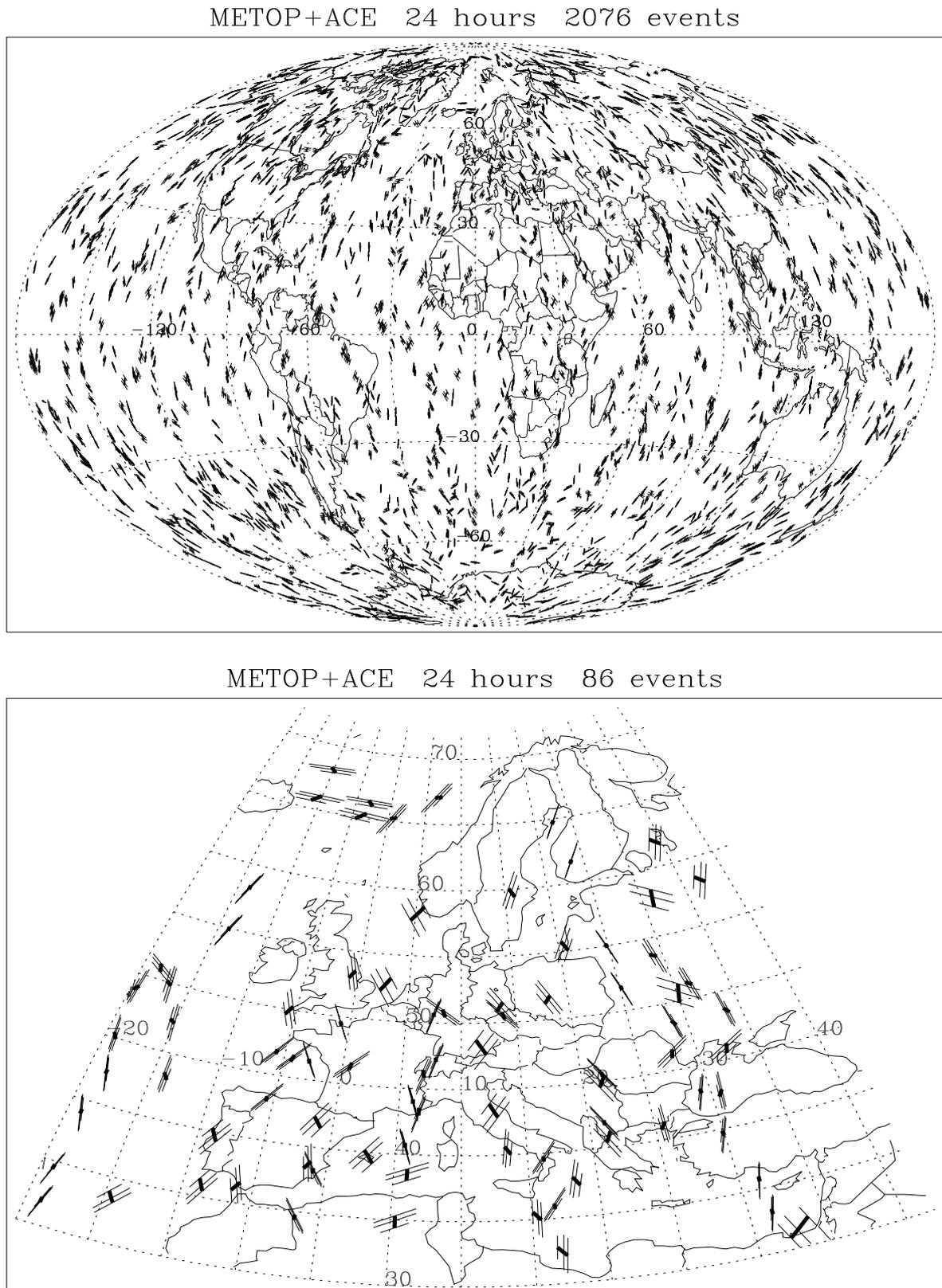


Fig. 5.4: Simulation of the geographical distribution of occultation events during 24 hours for the two LEO constellation (METOP and ACE), GPS and GLONASS satellites, setting and rising occultations. Same maps and symbols as in Fig. 5.2.

5.4.4 Twenty satellite constellation

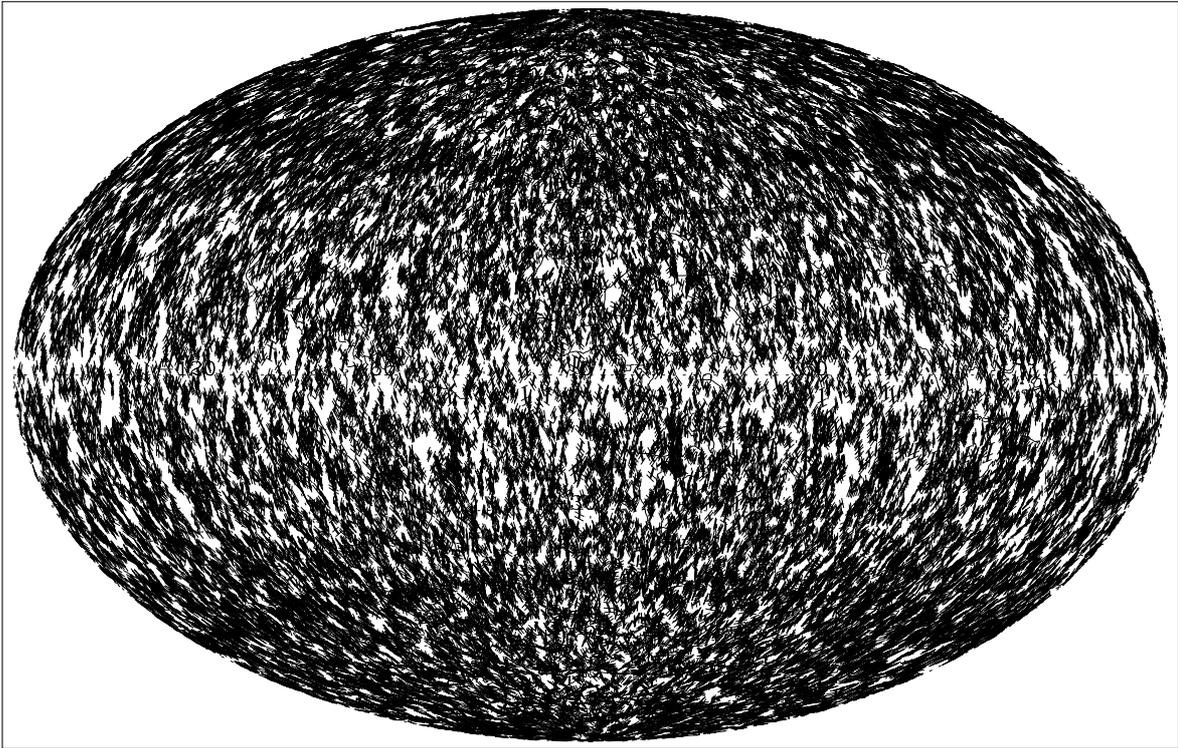
Considering a prospective future scenario of 20 LEOs (Fig. 5.5), even regions in the size of an European Area can be covered with almost 1000 events per day, sufficient to fulfill the "Outstanding data"-requirements for regional assimilation. Beyond that, this constellation marginally fulfills even the "Ideal data"-requirements for global assimilation. Approximately 23 000 occultations per day (23 LEO satellites) would be needed to rigorously achieve a horizontal resolution of < 300 km in combination with a temporal resolution of < 6 hours. ("Ideal data" for regional assimilation would require the currently unrealistic number of about 800 LEOs; we not on this matter that one can also argue that a number of more than about 50-100 LEOs begins to provide increasingly limited add-on value [see, e.g., Ramsauer and Kirchengast, 1999]).

5.4.5 Temporal distribution of occultation events

Figure 5.6 shows the number of occultation events over Europe during 24 one-hour intervals for three different LEO constellations. In the single satellite case (top panel), two intervals with eight occultations per hour occur, when the suborbital satellite track crosses the European domain. The middle panel (6 AMICO satellites) and the bottom panel (12 SAMICO satellites), both cases without METOP and ACE, indicate that satellites evenly distributed in space do not guarantee occultation events evenly distributed in time in a selected geographical area.

This situation is caused by "resonance" effects, as the GPS satellites have a 12 hour orbital period and the LEO satellites traverse a selected region twice per day. Long-time investigation over several months shows as illustrated in Fig. 5.7 for AMICO, that the pronounced minimum in the AMICO histogram is not stationary, three months later, it will occur between 16:00 and 18:00 UT, as it moves with the regression of the nodes of the satellite orbits. Note that this leads to Local-Time-Season interdependence for regional coverage (see Fig. 5.7).

METOP+ACE+AMICO+SAMICO 24 hours 20201 events



METOP+ACE+AMICO+SAMICO 24 hours 908 events

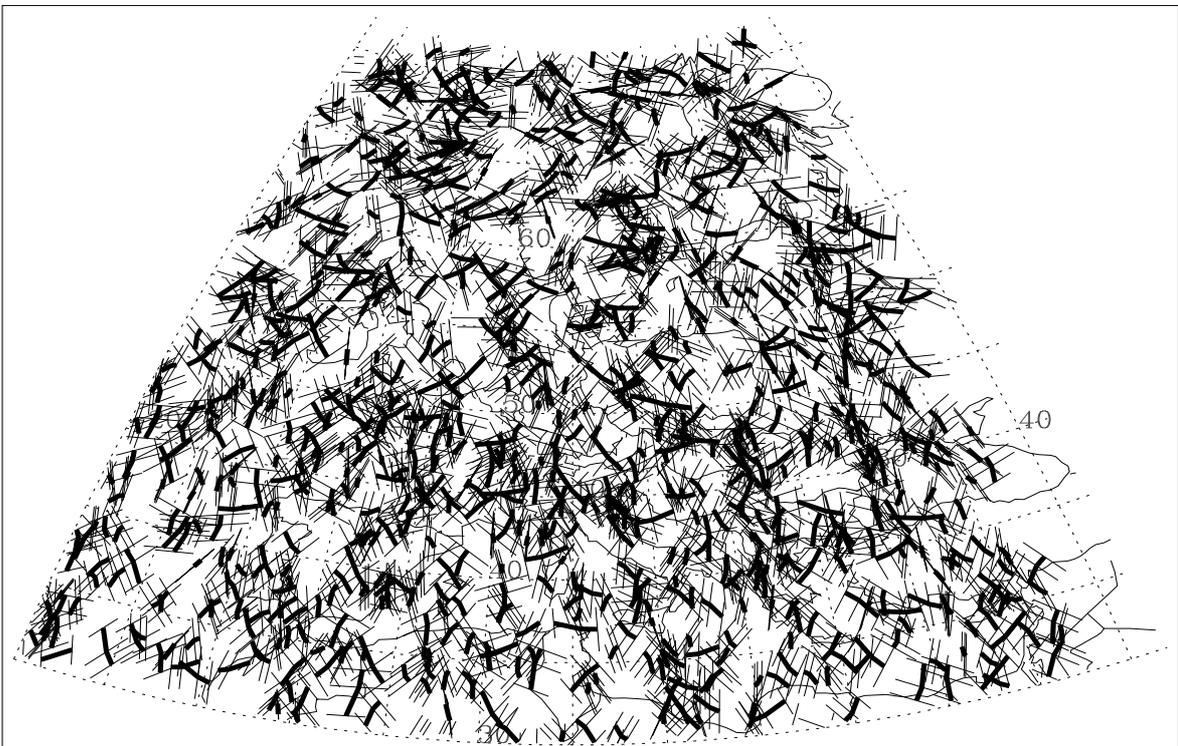


Fig. 5.5: Simulation of the geographical distribution of occultation events during 24 hours for 20 LEOs (METOP, ACE, 6 AMICOs and 12 Super-AMICO satellites, GPS and GLONASS, setting and rising occultations. Same maps and symbols as in Fig. 5.2.

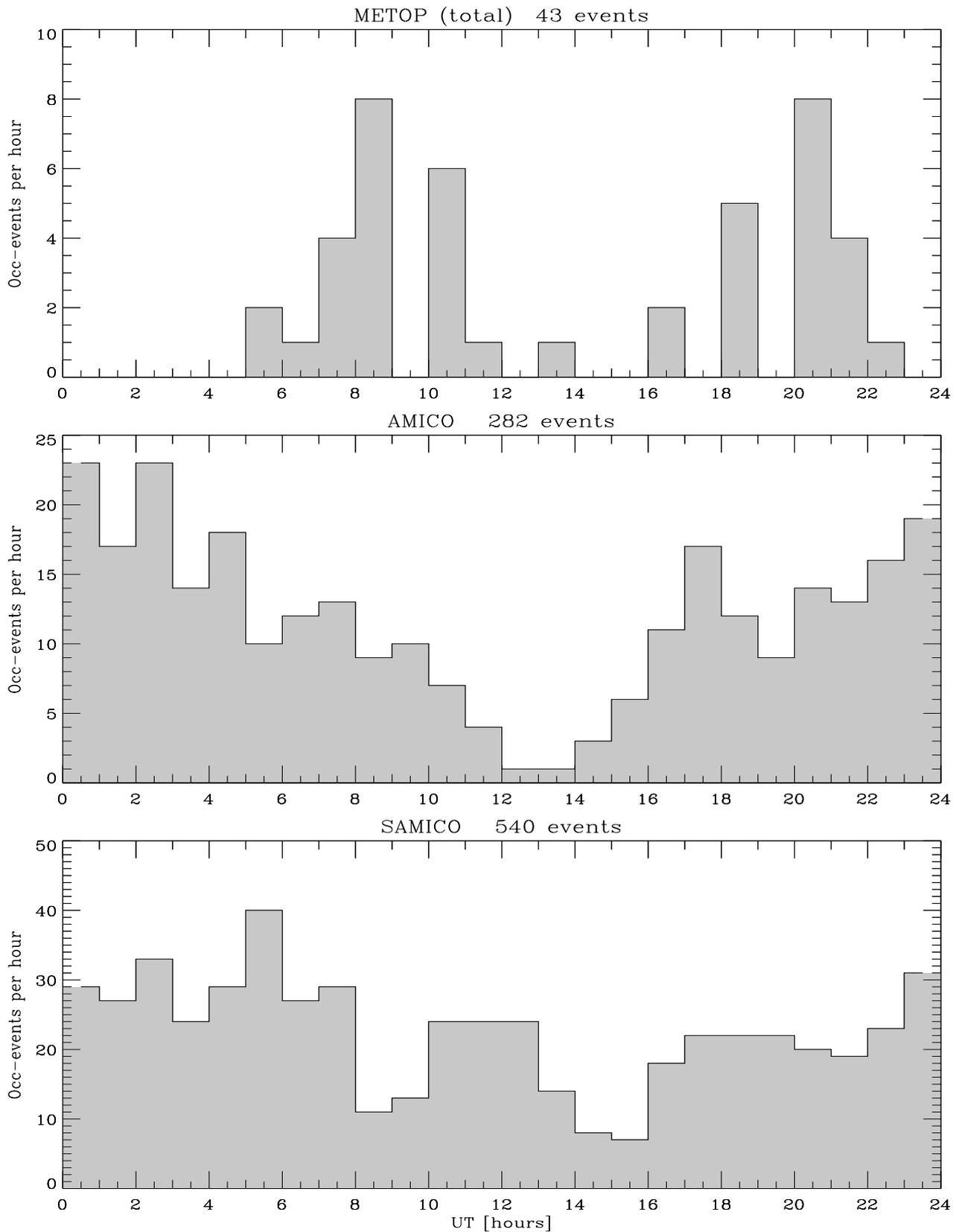


Fig. 5.6: Number of occultation events in Europe during 24 one-hour intervals; (top panel) one LEO (METOP, as in Fig. 5.3, bottom panel), (middle panel) 6 LEOs (AMICO constellation), (bottom panel) 12 LEOs (Super-AMICO constellation).

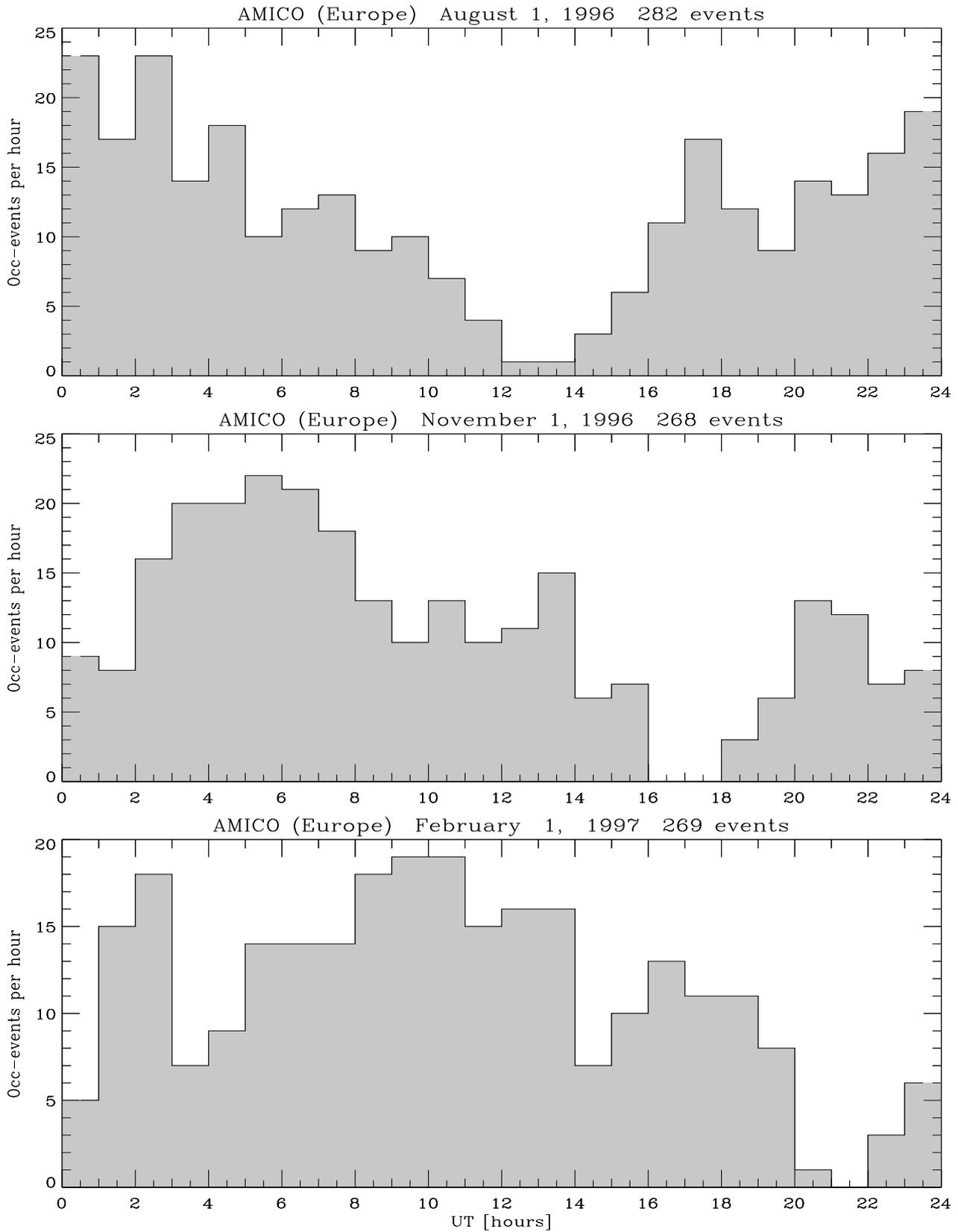


Fig. 5.7: Number of occultation events in Europe during 24 one-hour intervals for the AMICO constellation (as in Fig. 5.6, middle panel) at the begin of the simulation (top panel), three months later (middle panel), and half a year later (bottom panel).

5.4.6 Spatial distribution during short time-intervals

Figure 5.8 and Figure 5.9 display the geographical distribution of occultation events over Europe for the single satellite and the twenty satellite constellation, respectively, during a selected 30-min (top panels) and a one-hour interval (bottom panels) in order to set the scene with respect to the observational requirements for "true" imaging.

The selected 30-min interval, 08:30-09:00 Universal Time (UT) on the reference day, is favorable for the single satellite constellation (8 events per hour for METOP, cf. top panel of Fig. 5.6), but particularly unfavorable for the SAMICO constellation (cf. Fig. 5.6, bottom panel). The one-hour interval lasts from 08:00 to 09:00 UT. We note that there are several one-hour intervals with more than 50 occultation events over Europe for the twenty satellite constellation. The occultation events are indicated with the same symbols as in Fig. 5.1, but only 4 height levels (1km, 5km, 10km, and 15km) are shown, corresponding to the vertical domain of interest in the case of "true" imaging.

Even when allowing for the 1 hour interval, all eight events in the single satellite case (and all 14 events in the tandem satellite case, not shown as extra map) occur during the 30-min interval from 8:30-9:00 UT (cf. Fig. 5.8). The eight satellite constellation provided 19 occultation events during the selected 30 min interval and 23 events during the 1 hour interval. The simulation results for 20 LEOs and the selected time intervals are as displayed in Fig. 5.9. Generally, with 8 and 20 LEOs average values of 15 and 38 occultation events per hour, respectively, can be expected over Europe.

5.5 Combination with ground-based measurements

We will focus here on "true" imaging by combination of occultation data with ground-based GNSS measurements. Another interesting approach were to combine occultation data with spaceborne downlooking sounder data (passive IR/Microwave radiance). The high vertical but moderate horizontal resolution of the occultation data (cf. chapter 4) combined with the good horizontal but comparatively poor vertical resolution of a humidity sounder like AMSU-B (Advanced Microwave Sounding Unit) aboard METOP has the potential of providing frequently accurate images of the tropospheric moisture distribution. Requirements for this imaging technique are considered in Foelsche and Kirchengast [1997].

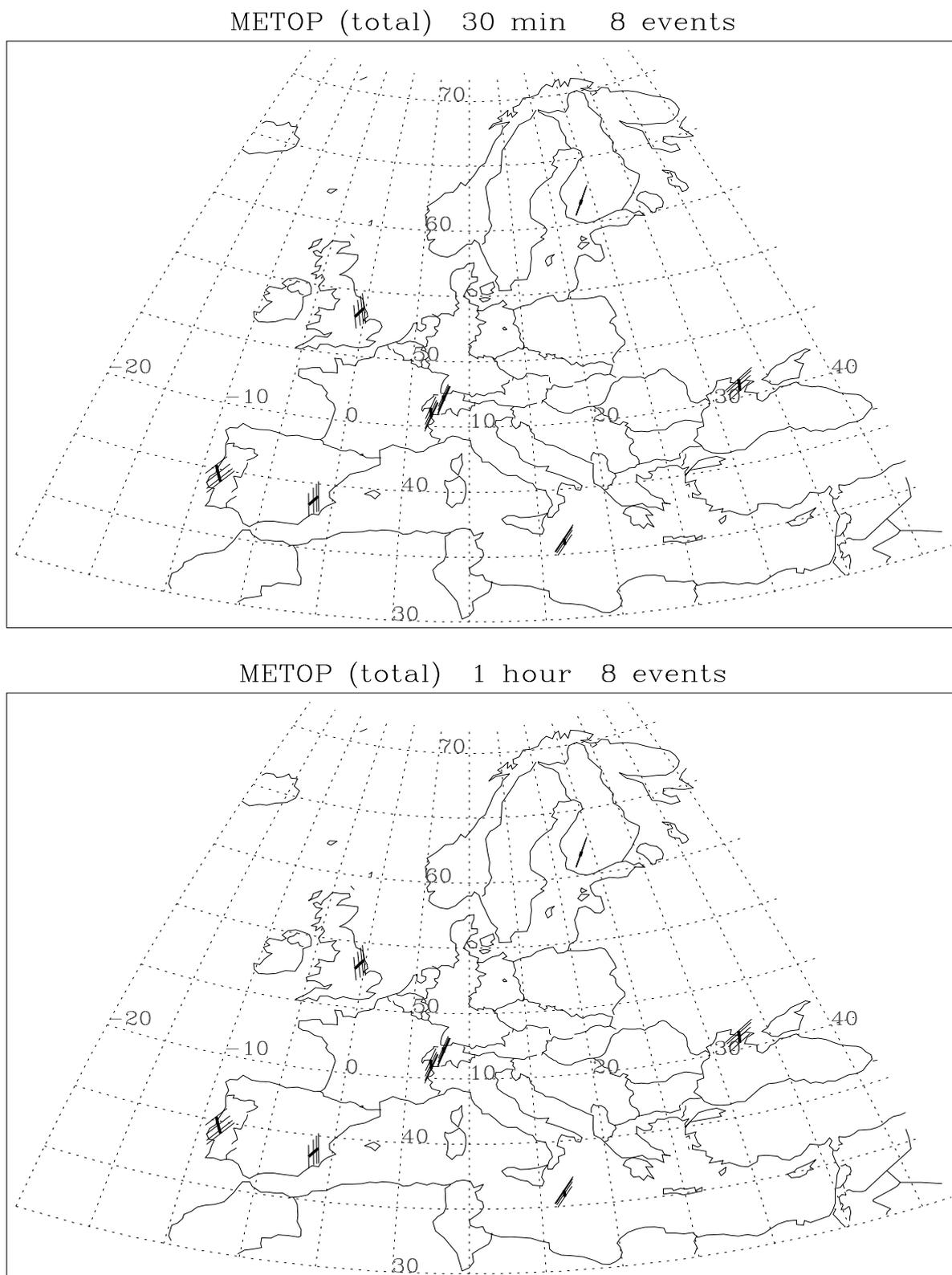


Fig. 5.8: Geographical distribution of occultation events in Europe during a 30-min interval (8:30 - 9:00 UT, top panel) and a one-hour interval (8:00 - 9:00 UT, bottom panel) on the reference day for the one LEO constellation (METOP). Symbols for indicating the events as in Fig. 5.1, but only 4 height levels (1km, 5km, 10km, and 15km).

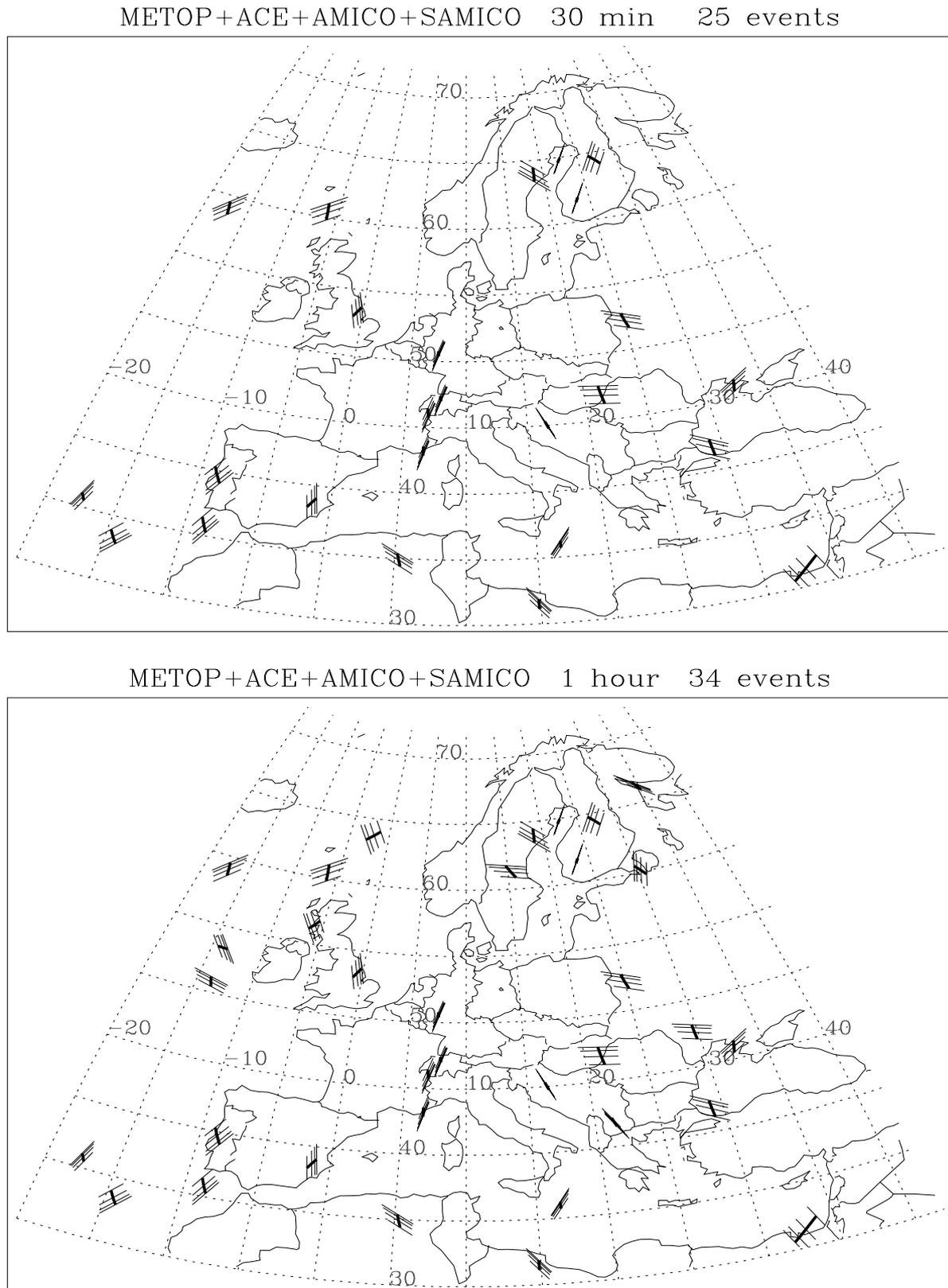


Fig. 5.9: Geographical distribution of occultation events in Europe during a 30-min (top panel) and a one-hour interval (bottom panel) for 20 LEOs (METOP, ACE, 6 AMICOs, and 12 SAMICOs, as in Fig. 5.5). Same intervals and symbols as in Fig. 5.8.

5.5.1 Simulation of ground-based measurements

In order to illustrate and discuss the compliance with the requirements for "true" imaging, Fig. 5.10 shows ground projections of GNSS ray paths (assumed to be straight lines) between receiver and 15 km level for 31 (existing) European IGS stations (International GPS Service for Geodynamics) and three synthetic chains during the selected 30-min and one-hour intervals (rays indicated at 10 min steps). Geographic coordinates of the IGS stations can be found on the IGS website: <http://igs.cb.jpl.nasa.gov/network.html>.

One of the synthetic, meridionally oriented local chains is located in France and consists of 20 stations with $1/4^\circ$ spacing. The two others are located in Finland (containing the IGS station Kirkkonummi), each of them consists of 10 stations with 50 km spacing). These synthetic chains shall represent dense local-to-regional receiver networks currently built up at various places worldwide in particular in Japan, U.S.A., and Europe. It is such dense network which can be the basis for "true" imaging.

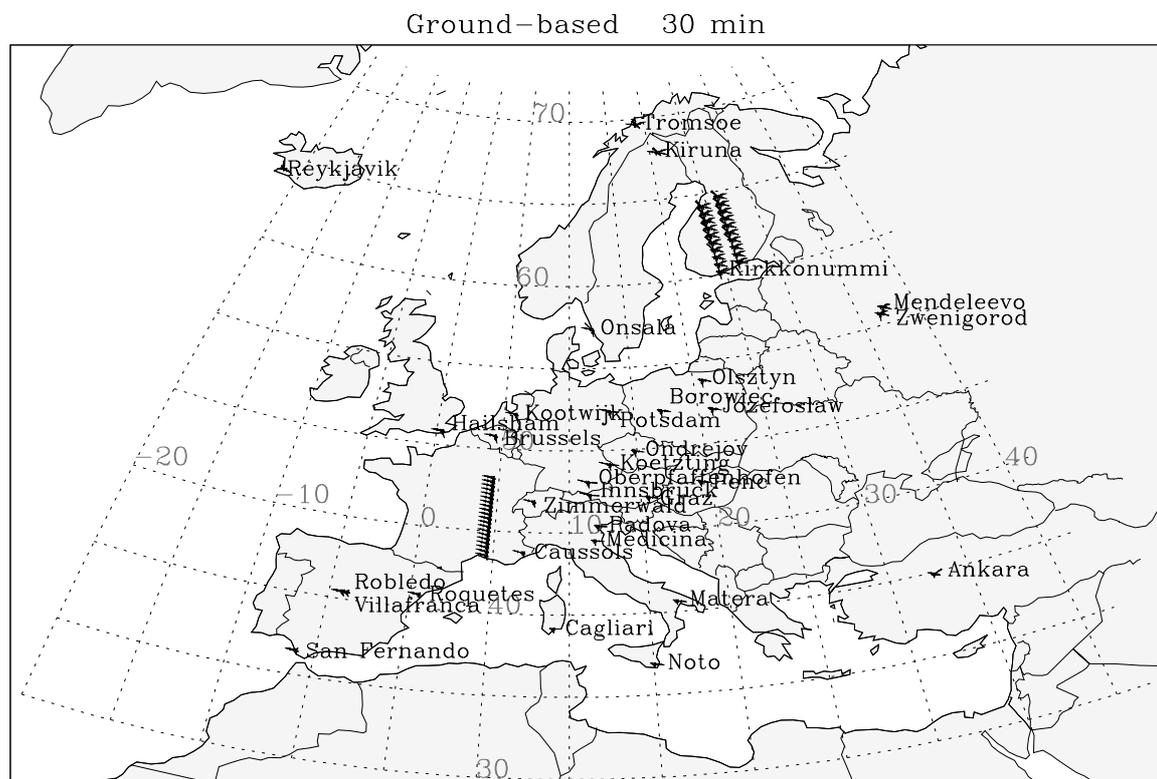


Fig. 5.10: Ground-based GPS, 31 (existing) IGS stations and three (synthetic) local chains (one in France, two in Finland) during the selected 30-min interval (8:30 - 9:00 UT). Straight rays (10 min steps) between station and the 15 km-level are shown as light lines.

Fig. 5.11 presents the combinations of occultation events for 20 LEOs (Fig. 5.9) and ground-based GPS stations (Fig. 5.10) during the selected 30-min interval. Having water vapor imaging in mind, occultations and ground-based-rays are displayed for the 1-15 km domain.

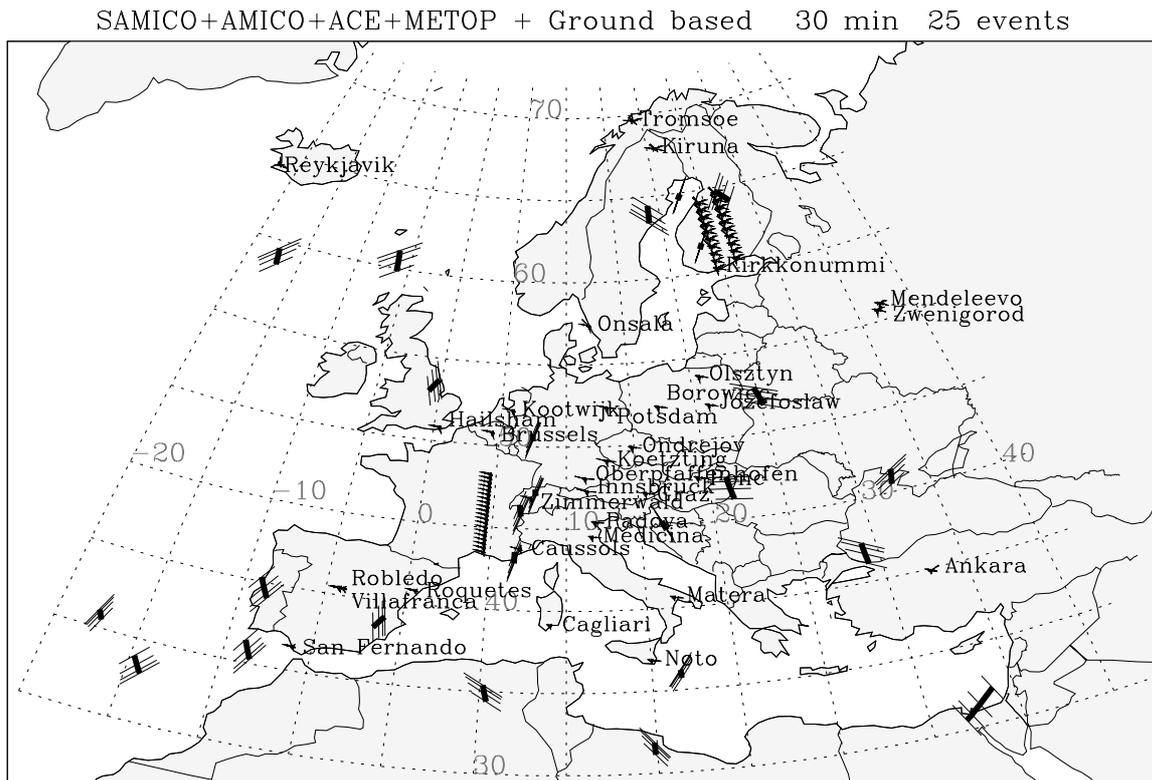


Fig. 5.11: Map showing combination of ground-based GPS (as in Fig. 5.10) and occultation events (as in Fig. 5.9) during the selected 30-min interval (8:30 - 9:00 UT).

One can conclude from Fig. 5.11 that occultations, even in case of a multi-satellite constellation, can be exploited in a "true" imaging sense only intermittently in time when co-located with an appropriate regional network.

In southern Finland, for example, where the synthetic chains representing typical regional GPS receiver networks have been placed, occultations and ground-based GPS rays marginally fulfill the requirements for "true" imaging. Figure 5.12 (for METOP only) and Fig. 5.13 (for 20 LEOs) zoom into this geographic region (a map spanning the area $58^{\circ}\text{N} < \text{latitude} < 67^{\circ}\text{N}$, $21^{\circ}\text{E} < \text{longitude} < 32^{\circ}\text{E}$ is shown) in order to depict the co-location situation more clearly.

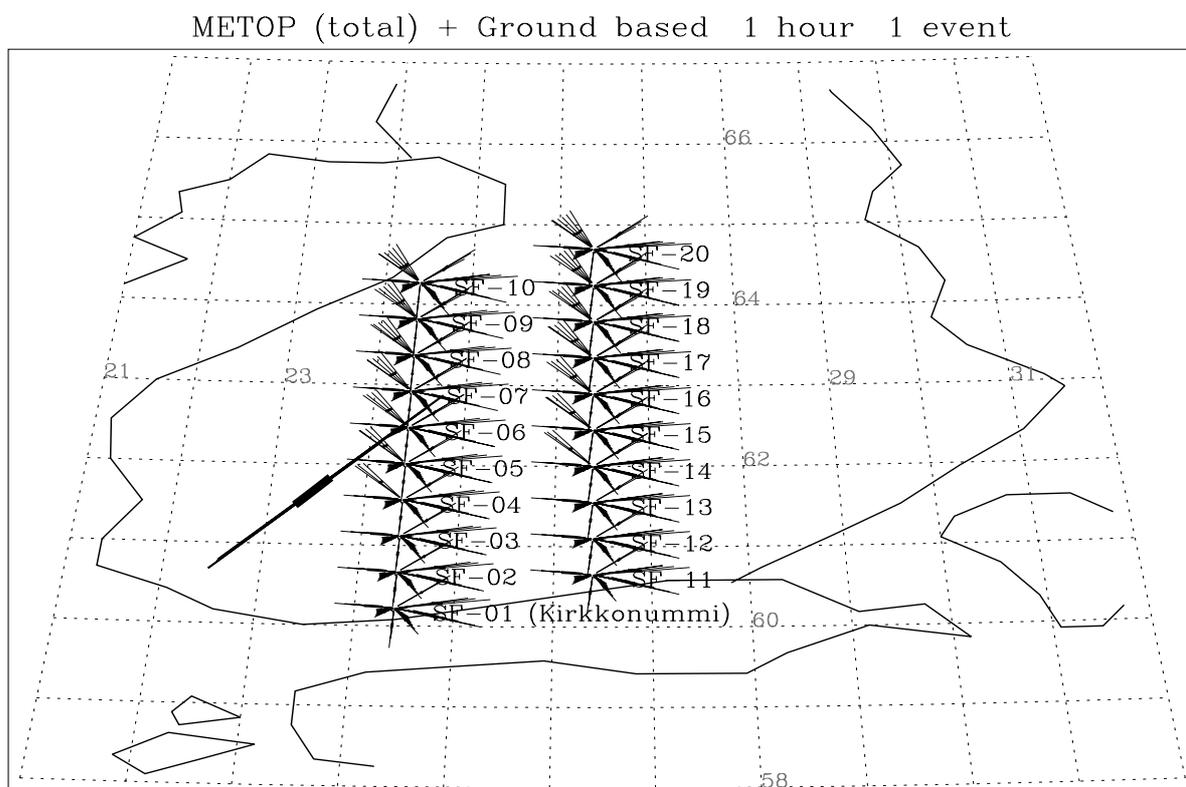
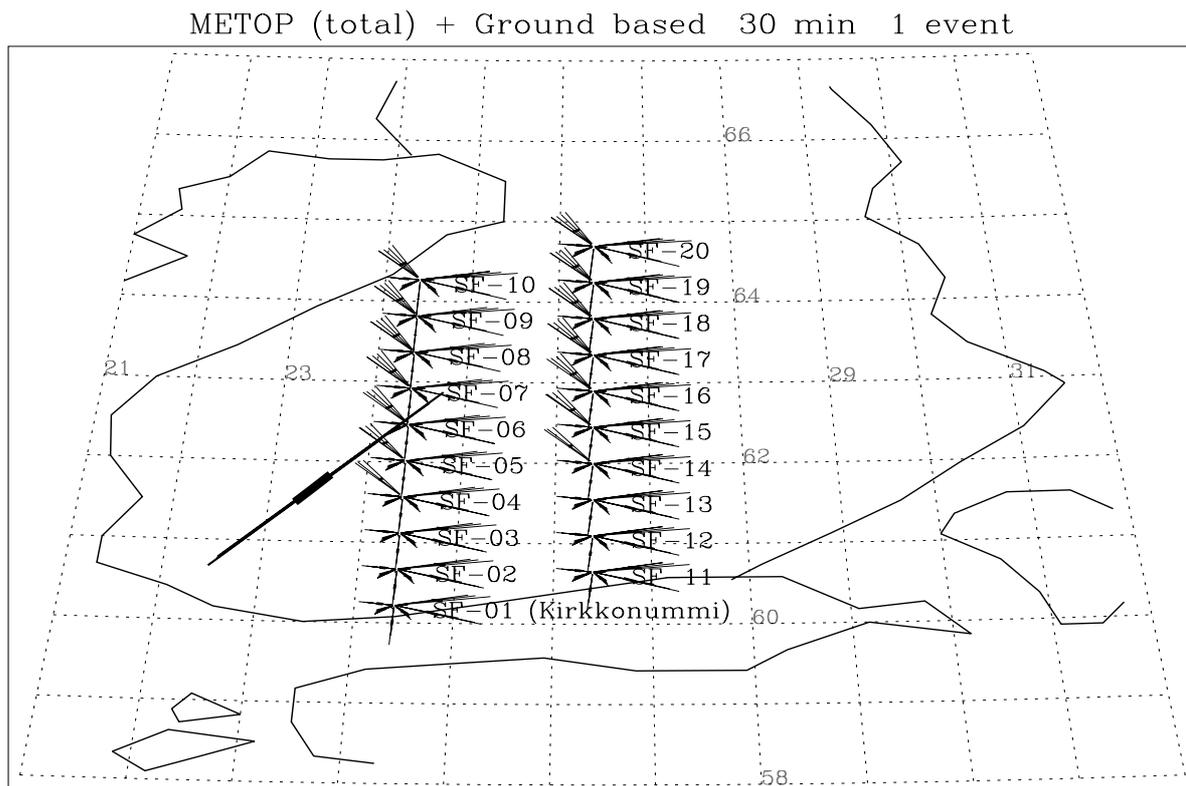
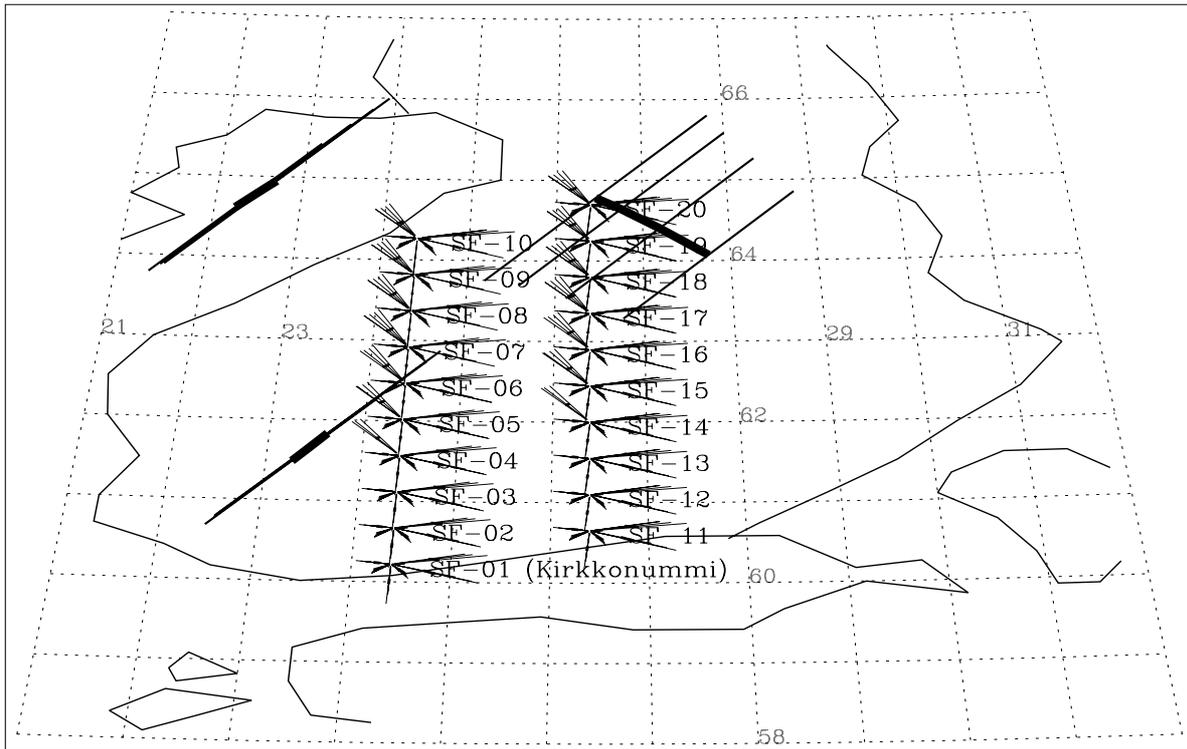


Fig. 5.12: Map showing combination of occultation events (one LEO) and ground based GPS (IGS station Kirkkonummi and two synthetic local chains) in southern Finland ($58^{\circ}\text{N} < \text{latitude} < 67^{\circ}\text{N}$, $21^{\circ}\text{E} < \text{longitude} < 32^{\circ}\text{E}$); same time intervals as in Fig. 5.8.

SAMICO+AMICO+ACE+METOP + Ground based 30 min 3 events



SAMICO+AMICO+ACE+METOP + Ground based 1 hour 4 events

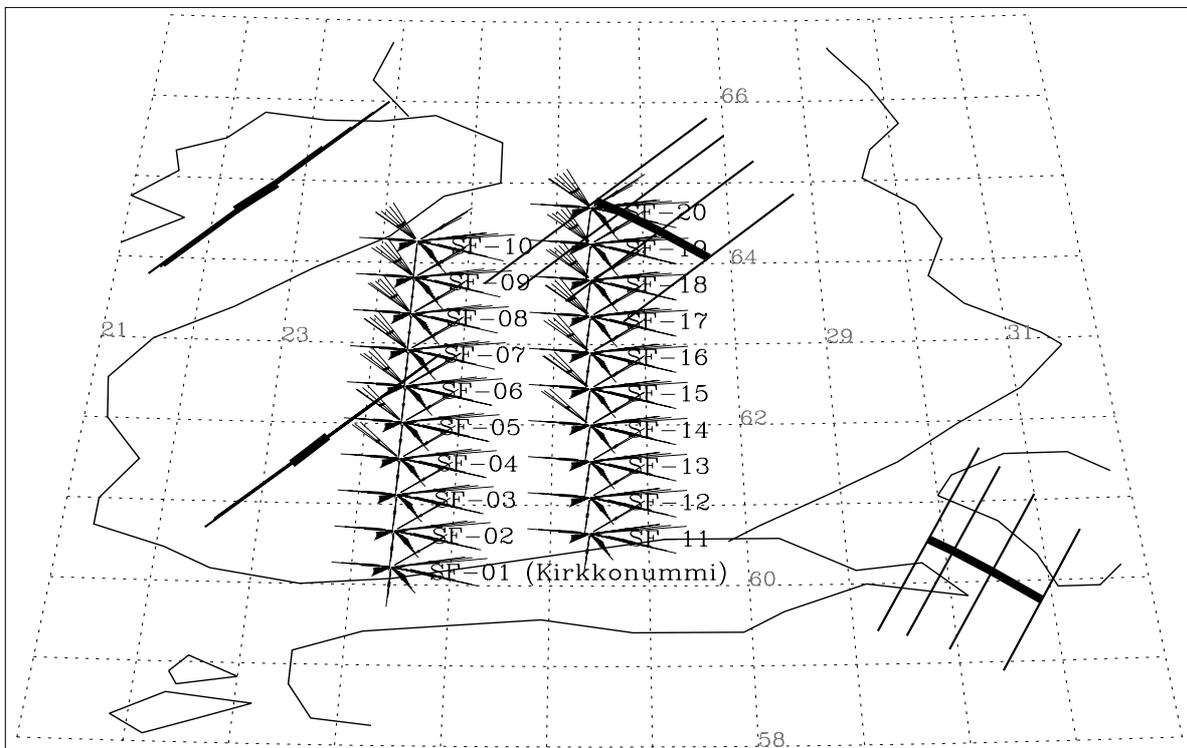


Fig. 5.13: Map showing combination of occultation events (20 LEOs) and ground-based GPS (IGS station Kirkkonummi and two synthetic local chains) in southern Finland ($58^{\circ}\text{N} < \text{latitude} < 67^{\circ}\text{N}$, $21^{\circ}\text{E} < \text{longitude} < 32^{\circ}\text{E}$); same time intervals as in Fig. 5.8.

We can conclude that "true" (water vapor) imaging combining ground-based GNSS with spaceborne GNSS can be expected to be possible intermittently in time when individual occultations are co-located with an appropriate regional network. The "true" imaging requirements, asking for sufficient space-time co-location, call for a constellation of satellites at least at the 10 LEO level in order to be frequently fulfilled for given regional networks as we illustrated by comparing the 20-LEO simulation results with the one-LEO results.

In practice it is clear that in case of interest in a specific atmospheric field such as water vapor, the method of choice will be, for most applications such as routine meteorological and climate analysis, not "true" imaging (which will be most interesting in relation to atmospheric process studies) but model-assisted imaging, i.e., data assimilation. The reason is that the multidimensional information contained in combination in the quite different spaceborne and ground-based data types addressed in this study is certainly best exploitable by the assimilation approach.

In the next chapter we will endeavor, though, to investigate "true" water vapor imaging methodology in some depth and, as will be seen, we find an attractive setup for two-dimensional imaging which will allow to optimally combine GNSS occultation profiles with ground-based GNSS sounding data. Moreover, the scheme developed allows also any other vertical water vapor profile (e.g., radiosonde profiles) to be seamlessly combined with ground-based GNSS sounding data.

6 TROPOSPHERIC WATER VAPOR IMAGING

Multi-satellite constellations designed for analyses of the Earth's atmosphere using the radio occultation technique shall emerge in the near future. Within the COSMIC project, for example, the simultaneous launch of eight LEO satellites, each equipped with a GPS/MET-like occultation sensor, is planned for June 2001 (detailed information can be found on the COSMIC homepage, <http://cosmic.gpsmet.ucar.edu>). In chapter 5 we have demonstrated, that "true" water vapor imaging by combination of spaceborne and ground-based GNSS measurements becomes in this case feasible, at least intermittently in time.

We have developed an imaging technique for two-dimensional "true" water vapor imaging based on a ground receiver network and co-located occultation events. As an alternative to occultation profiles also vertical water vapor profiles from other sources like radiosondes may be utilized by our method.

In this chapter, we will present the technique and selected retrieval results, based on simulated sounding data.

6.1 Forward model - geometry

In order to demonstrate the feasibility of "true" water vapor imaging, we selected a somewhat idealized but not unrealistic geometry. We considered a two-dimensional planar cut through the neutral atmosphere, the reconstruction plane, which can be thought to be defined by the occultation event.

6.1.1 Reconstruction plane and pixel geometry

The domain of the reconstruction plane, as used for the forward model, is divided into 45 000 picture elements (pixels). The height interval between 0 km and 10 km is represented by 100 levels with 100 m height, the interval between 10 km and 50 km by 200 levels with a respective height of 200 m. In the horizontal domain we divided the plane into 150 equidistant latitudinal sectors with a width of 0.05° (corresponding to 5.6 km at ground level), the sector thus spans a total width of 7.5° (~ 800 km).

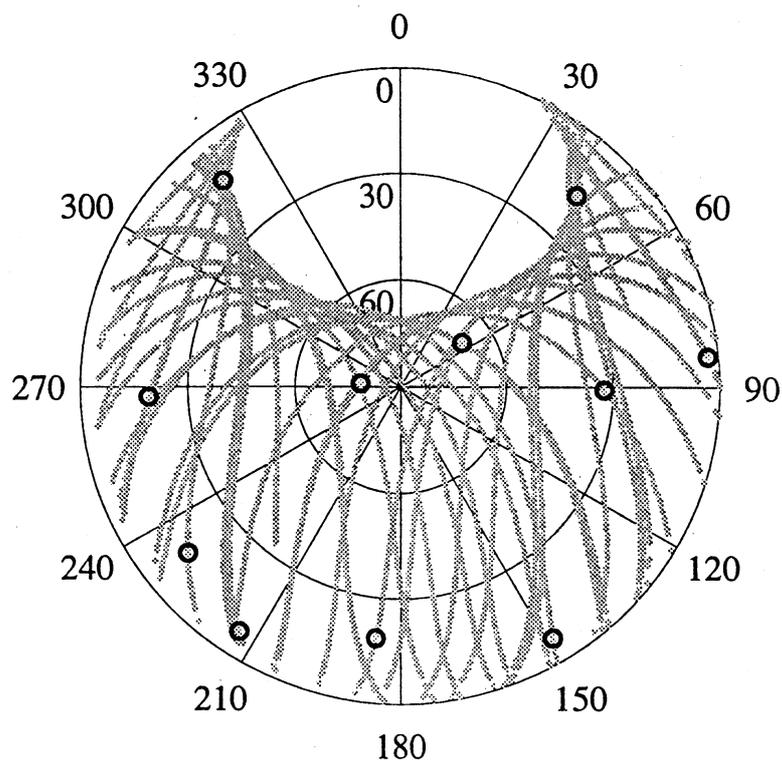
In general, the ground receiver and the satellite positions will laterally deviate from the fixed reconstruction plane. This problem can be overcome by introducing "three

dimensional pixels" with a certain extent perpendicular to the occultation plane [cf. Leitinger *et al.*, 1997], and/or by performing, in pre-processing, a projection of rays into this plane.

6.1.2 Satellites and ground stations

For our simulations we assumed a vertical occultation event, co-located with a meridionally oriented chain of 10 equidistant ground-based receivers. A considerable number of occultation events can be approximated by a vertical plane (see, e.g., Fig. 5.13). The ground stations are assumed to be separated by 0.25° (27.8 km), comparable to existing networks. In practice, one would use a subset of stations of a given receiver network, that is best co-located with the occultation plane.

As noted in section 3.1.1, 5 to 12 GPS satellites are in view at a given time for a ground based receiver. If signals from GLONASS satellites can be used in addition, the number nominally doubles.



*Fig. 6.1: GPS satellite elevation and azimuth tracks (gray curves) observed at Table Mountain, Colorado, during 24 hours. Circles show the satellite positions at a given point in time [from Ware *et al.*, 1997].*

We assumed 3 GNSS satellites to be approximately coplanar with the occultation plane. One satellite is located near the zenith, the two others near the horizon. These assumptions are not unrealistic even for the pure GPS case, as can be seen from the typical elevation-azimuth map depicted in Fig. 6.1. The time interval was restricted to 30 minutes, in order to fulfill the "ideal data" requirements for "true" imaging (cf. section 5.2.2). During the selected time interval, the atmospheric water vapor field can be expected to remain fairly constant. Within a 30 minute period, a GPS (or GLONASS) satellite moves approximately 15° across the sky.

Assuming a signal integration time of about 4 minutes per transmitter-to-receiver ray [cf. Rocken *et al.*, 1993], we considered 8 rays per satellite, leading to a total number of 240 rays for 10 stations each receiving 3 satellites.

An illustration of the geometry of the experiment is given below in section 6.5 where we discuss the water vapor retrieval setup (Fig. 6.4). The forward model (FOMOD) geometry is essentially the same as the retrieval geometry discussed there, but we used FOMOD pixels with a finer resolution and a FOMOD domain in the reconstruction plane extending further in height (0 km - 50 km for FOMOD instead of 0 km - 6/10 km for retrievals) and latitude ($\pm 3.75^\circ$ for FOMOD instead of $\pm 1.25^\circ$ for retrievals). This was just done for the sake of numerical accuracy in the along-ray integration.

We selected an elevation cutoff angle of 7° , comparable to cutoff angles used in field experiments in order to minimize multi-path effects. At elevations above 7° , ray bending can be neglected and the rays can be considered as straight lines, simplifying the forward model considerably [see, e.g., Elgered, 1993].

When performing the forward modeling, i.e., the generation of quasi-realistic measurement data we employed two different approaches, a simple one and a more realistic one, as described in section 6.2 and 6.3, respectively, below.

6.2 Forward model - synthetic refractivity fields

In order to basically understand and check the performance of the retrieval algorithm we investigated several scenarios based on simple synthetic water vapor density fields which we constructed analytically. The synthetic model fields are displayed and described in section 6.6 and discussed in comparison with the retrieved fields.

In these analytical water vapor density scenarios we focused on learning about characteristic features of the retrieval algorithm and not on a quasi-realistic representation of the measurement procedure. Forward modeling in these cases thus just meant that slant integrated water vapor (SIWV) along the different ray-paths through the analytical density fields was computed and, in most cases, randomly perturbed by typical rms errors in SIWV. These synthetic SIWV data were then used as direct for the retrieval algorithm.

6.3 Forward model - ECMWF-derived refractivity fields

In order to have a highly realistic forward computation of slant integrated water vapor (SIWV), we followed exactly the full procedure from excess phase paths to SIWV as described in chapter 3. These FOMOD simulations and the subsequent inversion to water vapor images, aimed at demonstrating the feasibility of tomographic water vapor imaging in a realistic manner.

6.3.1 From ECMWF profiles to refractivity fields

European Centre for Medium-range Weather Forecast (ECMWF) analysis fields for October 20, 1995, were used to construct realistic refractivity fields (arbitrarily selected day). Examples of water vapor density fields are shown in section 6.7 and discussed in comparison with the retrieved fields. The ECMWF data were cubic interpolated to fit the FOMOD pixel geometry. In the case of temperature and specific humidity, values at the center of the picture elements were taken as representative for the pixel area.

In the case of total atmospheric pressure we assumed, In order to minimize the discretization error, exponential decrease with height within each pixel, with a scale height corresponding to the temperature at the center of the pixel (H). The mean pressure within the pixel (indicated with an overbar) is therefore given by:

$$\bar{p} = p \frac{H}{\Delta z} \left[1 - \exp\left(-\frac{\Delta z}{H}\right) \right], \quad (6.1)$$

where Δz is the pixel height and p is the pressure at the lower edge of the pixel (the surface pressure in the case of the lowest pixel).

The ECMWF profiles extend only to a height of about 30 km. Neglecting the influence of the atmosphere above 30 km introduces an unacceptable error into the forward model. To overcome this problem, we used MSISE-90 climatological data [cf. Hedin, 1987 and 1991]

for the height interval between 30 km and 50 km, performing a smoothing transition between the two different profiles.

The water vapor partial pressure was computed using Eq. (1.2), the inverse compressibility factors of dry air and water vapor with Eq. (2.90) and Eq. (2.91), respectively. The refractivity was computed using the Thayer equation (2.89) with the constants proposed by Bevis *et al.* [1994] (see section 2.5.3).

6.3.2 Slant neutral delays

We considered only the influence of the neutral atmosphere, making the reasonable assumption that the ionospheric delay can be virtually removed (to 1 mm accuracy) using appropriate correction techniques (cf. section 3.1.4). As we assume straight line propagation (ray bending can be ignored at elevations above 7°, the selected elevation cutoff angle) and constant refractivity within the picture elements, the excess path delay (Eq. 3.3) for a particular ray can be written as:

$$\Delta L = \int_S (n - 1) ds = 10^{-6} \sum_i N_i s_i, \quad (6.2)$$

where s_i is the ray path length in pixel i and N_i is the corresponding refractivity.

The forward model is therefore a system of linear equations, which can be written as:

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \boldsymbol{\varepsilon}, \quad (6.3)$$

where \mathbf{y} is the *measurement vector*, containing the (simulated) excess path delay measurements along different ray paths, \mathbf{A} is the matrix of ray path lengths within the pixels, \mathbf{x} is the *state vector*, describing the state of the system and containing the pixel refractivities, and $\boldsymbol{\varepsilon}$ is the measurement error vector. A random Gaussian noise with a standard deviation of 0.4% (corresponding to 1.0 mm in zenith direction) was added to the modeled excess path delays in order to simulate measurement errors (e.g., residual ionospheric error).

In such a case, the *design matrix* \mathbf{A} depends only on the geometry of the experiment and not on the state of the atmosphere. For the selected geometry, the system is over-determined but close to ill-conditioned.

6.3.3 From surface pressure to slant hydrostatic delay

Taking the surface pressure values of the ECMWF data sets, zenith hydrostatic delays at the 10 ground stations were computed using Eq. (3.25). For the surface pressure measurements we assumed and implemented an rms error of 0.2 hPa. In order to obtain slant hydrostatic delays along the ray paths to the different GPS satellite positions, we had to use an appropriate mapping function (see section 3.3).

Comparing derived and "true" slant water vapor contents, we obtained the best results, using the Niell mapping function (cf. section 3.3.2) at low latitudes (up to 40°), and the geometric mapping function (cf. section 3.3.3) poleward of 40° .

6.3.4 From slant wet delay to slant integrated water vapor

The slant wet delay is then simply the difference of the measured excess path delay and the zenith hydrostatic delay mapped into the direction of the GPS satellite (cf. Eq. 3.41). An example, located in Florida, is shown in Fig. 6.2 below. Slant hydrostatic delay was in this case obtained using the Niell mapping function.

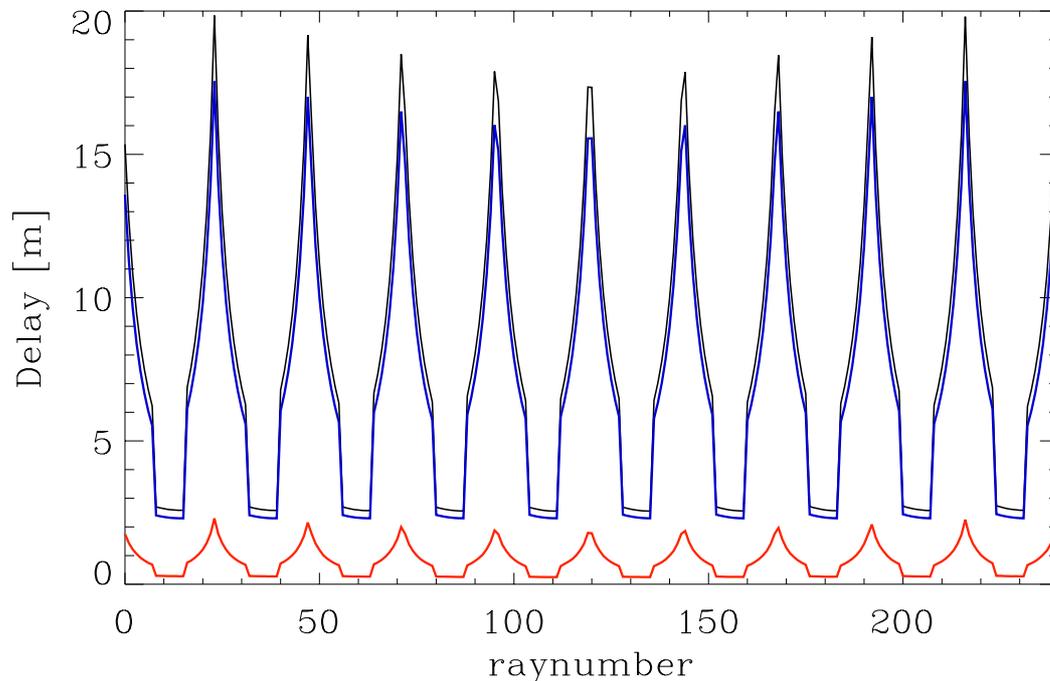


Fig. 6.2: Excess path delay (black), slant hydrostatic delay (blue) and wet delay (red) for a refractivity field located in Florida (centered at Palm Beach, 27°N , 80°W), using the experimental setup described in the text. The leftmost 24 ray-numbers correspond to ray-paths between the different satellite positions and the first ground station, the second 24 ray-numbers to the second ground station, etc.

As the wet delay is obtained by subtraction of two quantities of comparable magnitude (see Fig. 6.2), the forward modeling should be performed as precise as possible.

The slant integrated water vapor (SIWV) was derived, finally, by multiplication of the slant wet delay with the factor Π (Eq. 3.47). The mean weighted temperature T_m was estimated, using the linear regression, Eq. (3.55), proposed by Bevis *et al.* [1992], and the surface temperature values contained in the ECMWF data sets. No surface temperature measurement errors were assumed, as they are negligible compared with the error introduced by Eq. (3.55).

The derived SIWV values were compared with the corresponding "true" SIWV values, obtained for comparison purposes by direct integration of the ECMWF water vapor densities along the ray-paths. An example is shown in Fig. 6.3 below. The largest differences correspond to low elevation angles, reflecting the influence of insufficient mapping of the hydrostatic delay. We note that the cosecant mapping function (Eq. (3.30)) would lead in this case to an underestimation of the SIWV of more than 50% at low elevation angles.

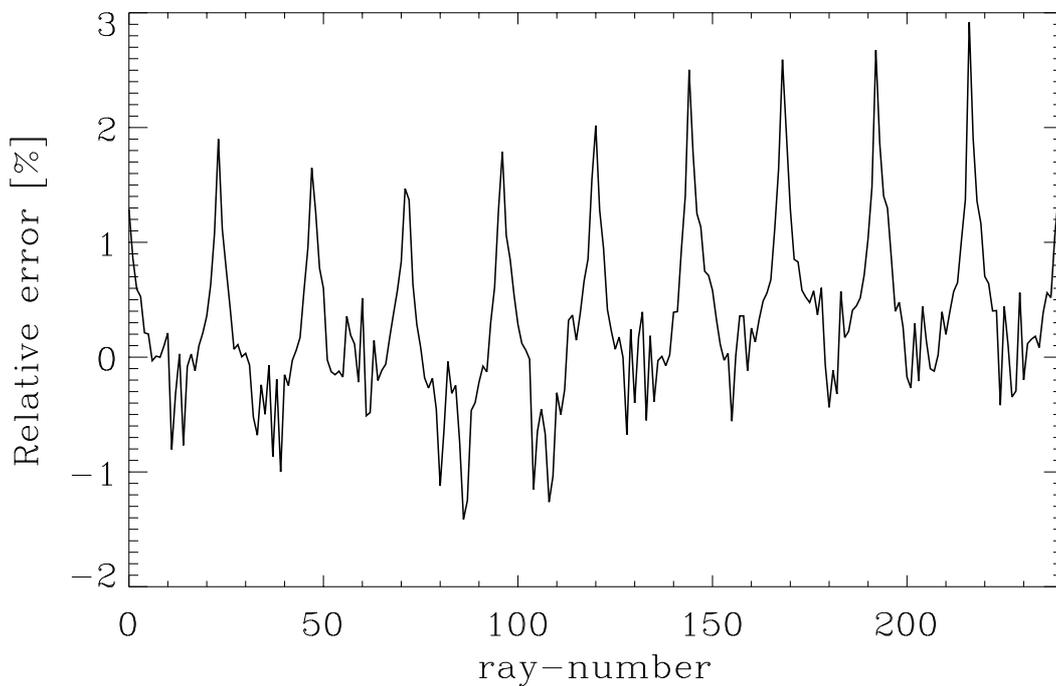


Fig. 6.3: Relative integrated water vapor error [%] for an ECMWF refractivity field located in Florida (the same field as used for Fig. 6.2).

6.4 Discrete inverse theory

In the following section we will give a short review of aspects of inverse theory relevant to our imaging problem. We will confine ourselves to discrete inverse theory and to linear problems as our problem is discrete and linear (i.e. the forward model is a linear function of the state). For details see, e.g., Menke [1989], Storch and Zwiers [1997], and Rodgers [1998].

6.4.1 Forward model

After the derivation of SIWV we can rewrite the linear forward model in the same form as Eq. (6.3) but with slightly different meaning:

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \boldsymbol{\varepsilon}. \quad (6.4)$$

The *measurement vector* \mathbf{y} contains now the forward modeled SIWV measurements along the different ray paths, \mathbf{A} is the matrix of ray path lengths within the retrieval pixels, and the *state vector* \mathbf{x} contains the water vapor densities within the picture elements. The *design matrix* \mathbf{A} depends only on the geometry of the experiment, as long as ray bending can be neglected (ray bending depends on the state \mathbf{x} and makes to forward model non-linear).

Measurements are always discrete quantities, where the true continuous state function is replaced by a representation in terms of a finite number of parameters (in our problem the water vapor densities within the picture elements).

Within the framework of discrete inverse theory, the forward and the inversion problem can be considered as an algebraic mapping between vector spaces. In the forward model, the linear operator \mathbf{A} (called the *design matrix*), representing the act of measuring, describes the mapping from the state space into the measurement space (apart from the measurement error $\boldsymbol{\varepsilon}$), it is a $m \times n$ matrix corresponding to m equations (measurements) and n unknowns (elements of the state vector). Each row of \mathbf{A} can be considered as a vector \mathbf{a}_i in the state space. The Design matrix is singular, if $m \neq n$ or if two or more \mathbf{a}_i vectors are linearly dependent.

6.4.2 Range and nullspace of the Design matrix

For singular matrices, the concepts of *nullspace* and *range* are important. In the case of a singular matrix \mathbf{A} , there is some subspace of \mathbf{x} , called the *nullspace*, that is mapped to zero. The dimension of the nullspace is called the *nullity* of \mathbf{A} . The subspace of \mathbf{x} that can be "reached" by inversion of measurements is called the *range* of \mathbf{A} . The dimension of the range, p , is called the *rank* of \mathbf{A} . The sum of nullity and rank is therefore the dimension of the state vector, n (if $m > n$, as in our problem). The matrix \mathbf{A} is *rank deficient* if $p < n$. The \mathbf{a}_i vectors span in this case only a p -dimensional subspace of \mathbf{x} . Only those components of the state vector, which lie in this subspace will contribute to the measurement vector, all other components (orthogonal to this space) are immeasurable.

6.4.3 The inverse problem

In order to solve the inverse problem we have to find an operator for the inverse mapping from the measurement space into the state space, the generalized inverse of \mathbf{A} , denoted \mathbf{A}^{-g} . The inverse problem can then be written (with the retrieved state vector \mathbf{x}_{retr}) as:

$$\mathbf{x}_{\text{retr}} = \mathbf{A}^{-g} \mathbf{y} . \quad (6.5)$$

A unique solution exists only if $p = n = m$ (i.e., the problem is even-determined), the generalized inverse is in this case simply the inverse of the Design matrix, \mathbf{A}^{-1} .

The system is under-determined if there are less measurements than unknowns ($m < n$). In this case the solution is non-unique (i.e., the number of solutions is infinite) since there are components of the state which cannot be determined.

6.4.4 Least squares solution

The problem is over-determined if there are more measurements than unknowns ($m > n$) and the matrix is of full rank ($p = n$). The solution is also non-unique as the system is inconsistent, but it is possible to define a best approximate solution determined by the minimum misfit principle. The generalized solution is the one that minimizes the norm $\|\mathbf{Ax} - \mathbf{y}\|$, which becomes the well-known least squares solution if the L_2 -norm is used. The generalized solution is in this case given by:

$$\begin{aligned} \mathbf{A}^{-g} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T , \quad \text{yielding} \\ \mathbf{x}_{\text{retr}} &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} . \end{aligned} \quad (6.6)$$

6.4.5 Truncated SVD

A problem can also be simultaneously over- and under-determined (i.e. mixed-determined). In this case there are more measurements than unknown ($m > n$) but the matrix \mathbf{A} is rank-deficient ($p < n$). A generalized solution can be found in this case by performing a type of eigenvalue decomposition of the design matrix, called *Singular Value Decomposition* (SVD) [for details see Press *et al.*, 1992]. Any ($m \times n$) matrix (with $m \geq n$) can be written as the product of an ($m \times n$) column-orthogonal matrix \mathbf{U} , an ($n \times n$) diagonal matrix $\mathbf{\Lambda}$ with positive or zero elements (the *singular values*, λ_i), and the transpose of an ($n \times n$) orthogonal matrix \mathbf{V} :

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \quad \text{with} \\ \mathbf{U}^T \mathbf{U} &= \mathbf{V}^T \mathbf{V} = \mathbf{I}, \end{aligned} \quad (6.7)$$

where \mathbf{I} is the identity matrix. The singular values equal the roots of the eigenvalues of $\mathbf{A}^T \mathbf{A}$ if ($m \geq n$) or of $\mathbf{A} \mathbf{A}^T$ if ($m \leq n$). When \mathbf{A} is of full rank ($p = \min(m,n)$), its singular values are all > 0 and $\mathbf{\Lambda}$ can be inverted yielding:

$$\mathbf{A}^{-g} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{U}^T, \quad (6.8)$$

where the elements of $\mathbf{\Lambda}^{-1}$ are simply the reciprocals of the singular values.

If \mathbf{A} is rank deficient, there are singular values = 0, introducing the problem of infinity ($1/0$) and $\mathbf{\Lambda}^{-1}$ does not exist. The problem can be solved by replacing those values $1/\lambda_i$ which are infinite with $1/\lambda_i = 0$ establishing a truncated matrix $\mathbf{\Lambda}^{-g}$.

The generalized solution of the mixed-determined problem, known as *truncated SVD*, can therefore be written as:

$$\mathbf{x}_{\text{retr}} = \mathbf{A}^{-g} \mathbf{y} = \mathbf{V} \mathbf{\Lambda}^{-g} \mathbf{U}^T \mathbf{y}. \quad (6.9)$$

The *condition number* of a matrix is defined as the ratio of the largest λ_i to the smallest λ_i . A matrix is singular if its condition number is infinite, and it is *ill-conditioned* if its condition number is so large, that its reciprocal reaches the computer's numerical precision [cf. Press *et al.*, 1992].

In an ill-posed problem it is thus appropriate to use a truncated SVD solution even for $\lambda_i \approx 0$ (typically if $\lambda_{\text{max}}/\lambda_i > \sim 10^3$), as the very small singular values lead to an intolerably high error amplification when the inverse mapping, Eq. (6.8), is applied to the measurement vector.

For the selected geometry, the imaging problem is over-determined but close to ill-conditioned (the condition number is $\sim 10^{11}$!). Employing only the 120 largest singular values yields a more reasonable "residual" condition number of 1726.

6.4.6 Optimal estimation

In the case of an ill-conditioned problem, it is generally unavoidable to incorporate additional or *a priori* information. The *Bayesian approach* provides a consistent theoretical framework to incorporate *a priori* information (\mathbf{x}_{ap}) in a transparent and explicit manner. A detailed derivation of the following formula, known as *optimal estimation*, can be found in Rodgers [1998]. In the case of a linear forward model with the assumption of Gaussian probability distribution functions for the measurement and *a priori* errors, the retrieved state vector is given by:

$$\mathbf{x}_{\text{retr}} = \mathbf{x}_{\text{ap}} + (\mathbf{A}^{-\text{T}}\mathbf{S}_{\varepsilon}^{-1}\mathbf{A} + \mathbf{S}_{\text{ap}}^{-1})^{-1}\mathbf{A}^{\text{T}}\mathbf{S}_{\varepsilon}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_{\text{ap}}), \quad (6.10)$$

where $\mathbf{S}_{\varepsilon}^{-1}$ is the measurement error covariance matrix (a diagonal matrix in our case of independently sampled SIWV measurements) and $\mathbf{S}_{\text{ap}}^{-1}$ is the *a priori* covariance matrix.

The covariance matrix of the retrieved state vector is given by:

$$\mathbf{S}_{\text{retr}} = (\mathbf{A}^{-\text{T}}\mathbf{S}_{\varepsilon}^{-1}\mathbf{A} + \mathbf{S}_{\text{ap}}^{-1})^{-1}. \quad (6.11)$$

6.5 Retrieval geometry

With the exception of the size and the number of the picture elements, the retrieval geometry is exactly the same as the FOMOD geometry described in section 6.1.

6.5.1 Reconstruction plane and pixel geometry

The domain of the reconstruction plane for the simple synthetic SIWV data, is divided into 200 picture elements (20×10). The height interval between 0 km and 10 km is represented by 20 levels with a respective height of 500 m. In the horizontal domain we divided the plane into 10 equidistant latitudinal sectors with a width of 0.25° (corresponding to 27.8 km at ground level), symmetric with respect to the ground stations.

In the case of realistic ECMWF-derived data, the height of the retrieval domain within the reconstruction plane was decreased to 6 km with 12 height levels (500 m each). The pixel number is therefore decreased to 120 (12×10).

Figure 6.4 illustrates the tomographic imaging setup for the case of the ECMWF data imaging. Note that the chosen aspect ratio ($\sim 1 : 22$) is very much stretching the vertical against the horizontal scale; the most slant rays indicated are at 7° elevation angle (despite they still look almost vertical with this aspect ratio).

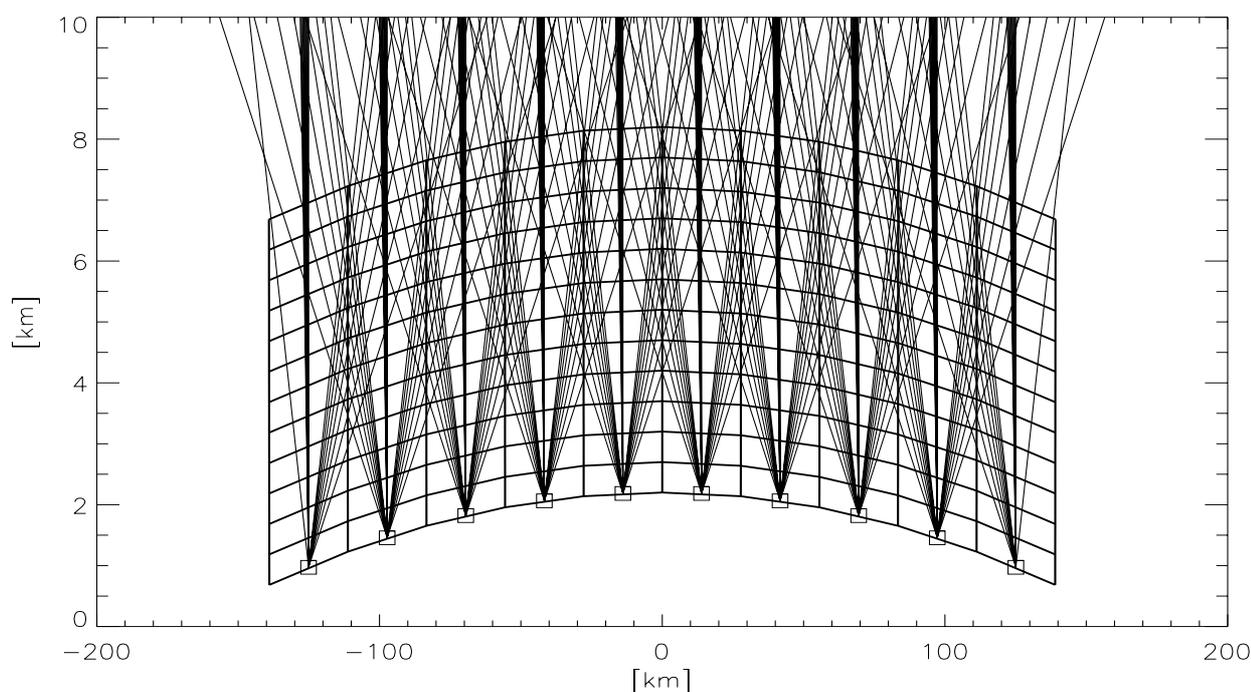


Fig. 6.4: Pixel geometry and rays from 24 satellite positions to 10 ground stations (indicated by squares). A total of 226 rays is shown, as only rays, that do not leave the reconstruction field sideward were used for the retrieval.

6.6 Retrieval results - synthetic data

As described in section 6.2, for basically investigating the performance of our imaging algorithm, we used directly SIWV along the different ray-paths as input for the retrieval algorithm. The *a priori* density field used in this case was simply the mean density profile extended over the entire retrieval domain. The latitude range used for the following figures was arbitrarily chosen centered at 45° .

6.6.1 Exponential atmosphere with horizontal gradient

In this simple but not unrealistic baseline scenario we assumed an exponential decrease of water vapor partial pressure with height and a linear increase of surface water vapor with decreasing latitude (see Fig. 6.5 below). We selected a typical climatological water vapor scale height of 2 km (see section 3.4 and Eq. (3.59)). No SIWV rms error was applied, i.e., perfect data were assumed.

The linear latitudinal gradient was chosen in a way that the surface water vapor partial pressure increased within the retrieval domain from 15 hPa to 25 hPa (corresponding to water vapor densities from 10.9 g/m^3 to 17.9 g/m^3).

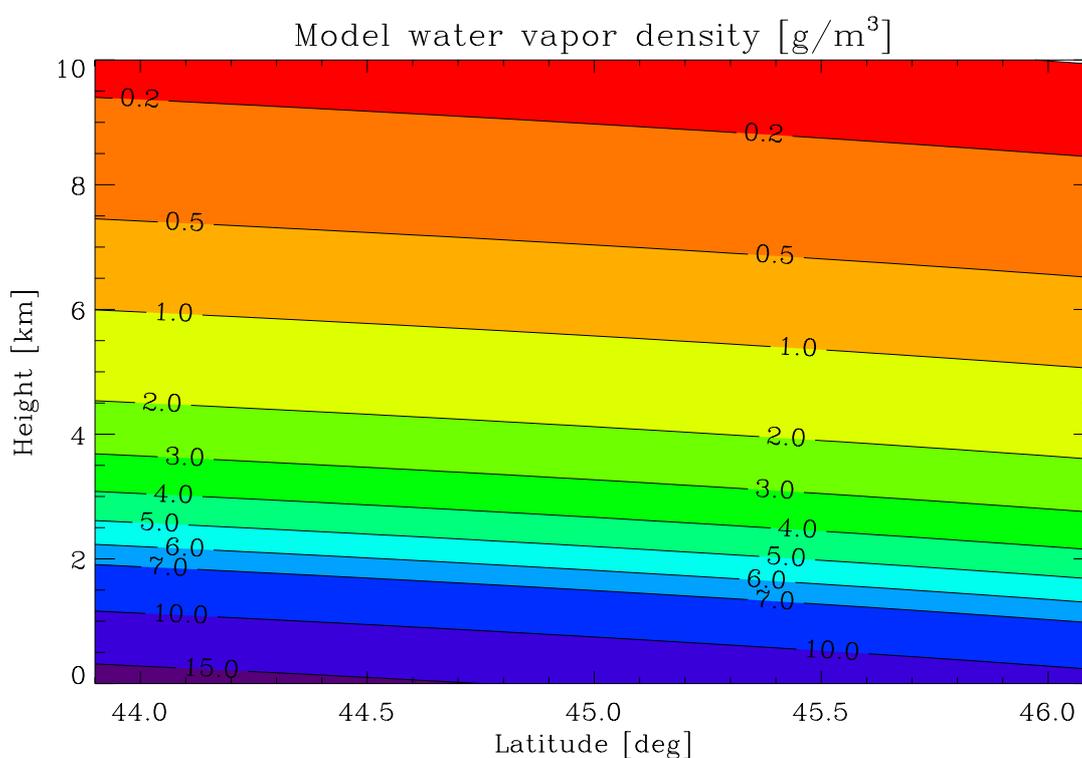


Fig. 6.5: Synthetic exponential atmosphere with linear horizontal gradient. The exponential decrease with height is described by a climatological (constant) water vapor scale height of 2 km.

The relative difference between the model field and the corresponding *a priori* field (average over the horizontal domain) is shown in Fig. 6.6 below, it exceeds 10% in more than half of the domain.

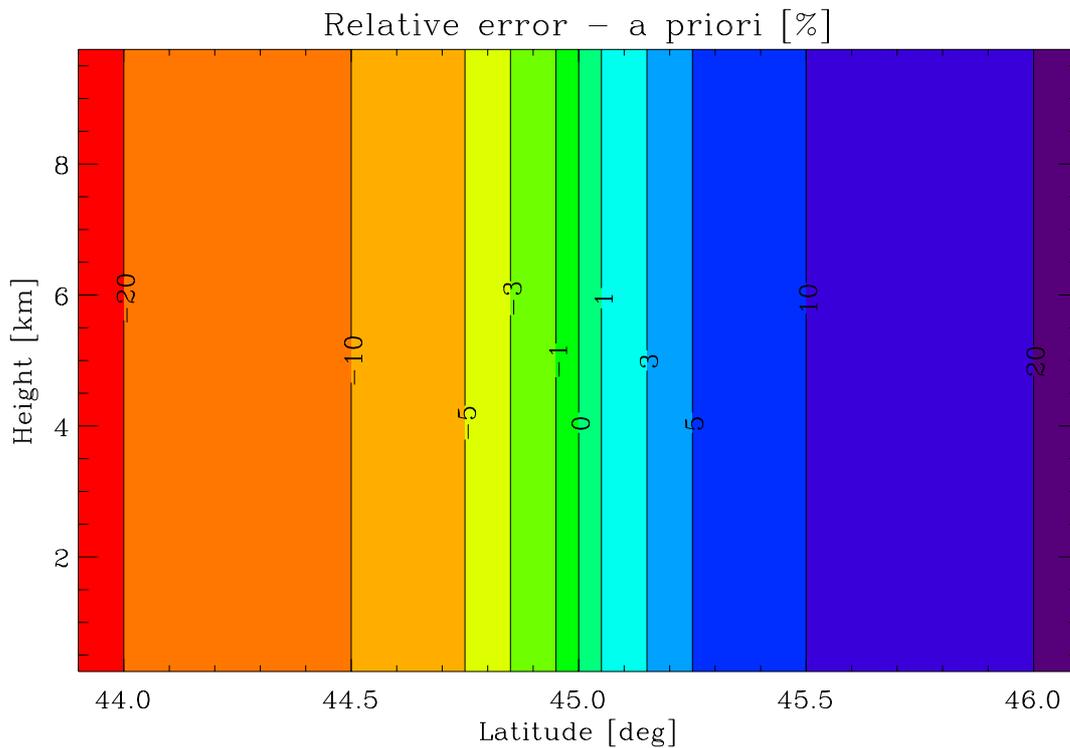


Fig. 6.6: Relative difference between the model field and the corresponding a priori field for the synthetic exponential atmosphere with linear horizontal gradient.

Given these perfect measurements, it is possible to obtain in this case a truncated SVD solution which at least somehow resembles the original water vapor density field (see subsection 6.4.5). The singular values of the *Design matrix* \mathbf{A} are plotted in Fig. 6.7. The SVD result, depicted in Fig. 6.8, obtained by using the 120 largest (of 200 possible) singular values, shows an almost linear decrease with height, yielding surface values which are far too low (about 50%) and negative densities at heights above 8 km.

The solution is furthermore extremely sensitive to the employed number of singular values, with only one more or less, the performance already decreases dramatically. The relative errors corresponding to the SVD result (Fig. 6.8) are displayed in (Fig. 6.9).

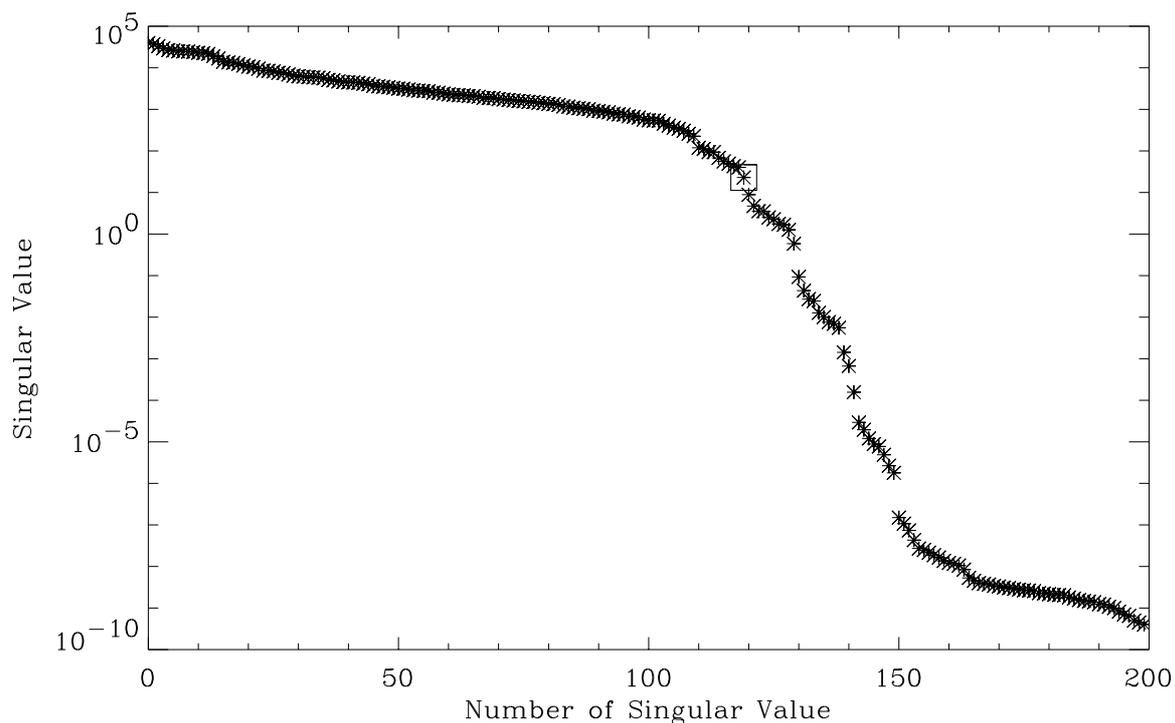


Fig. 6.7: Logarithmic plot of the singular values of the design matrix A (in decreasing order). The smallest singular value used for the truncated SVD is indicated with a square.

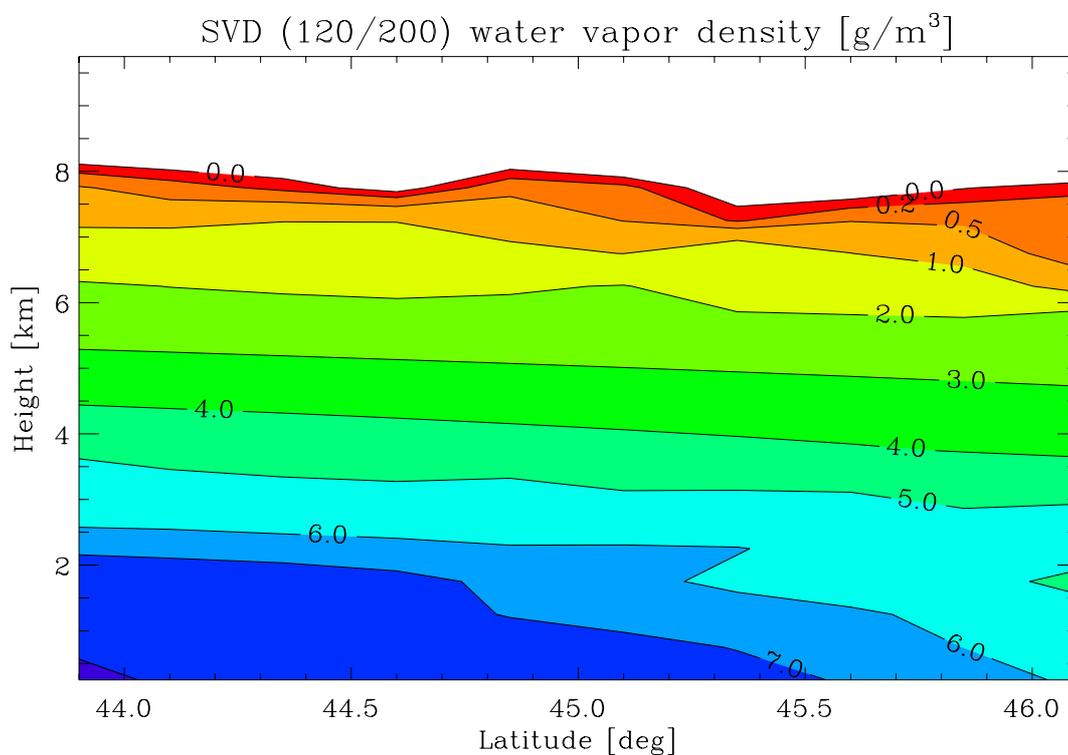


Fig. 6.8: Truncated SVD solution for the scenario described above, using 120 (of 200 possible) singular values. The white area corresponds to negative water vapor densities.

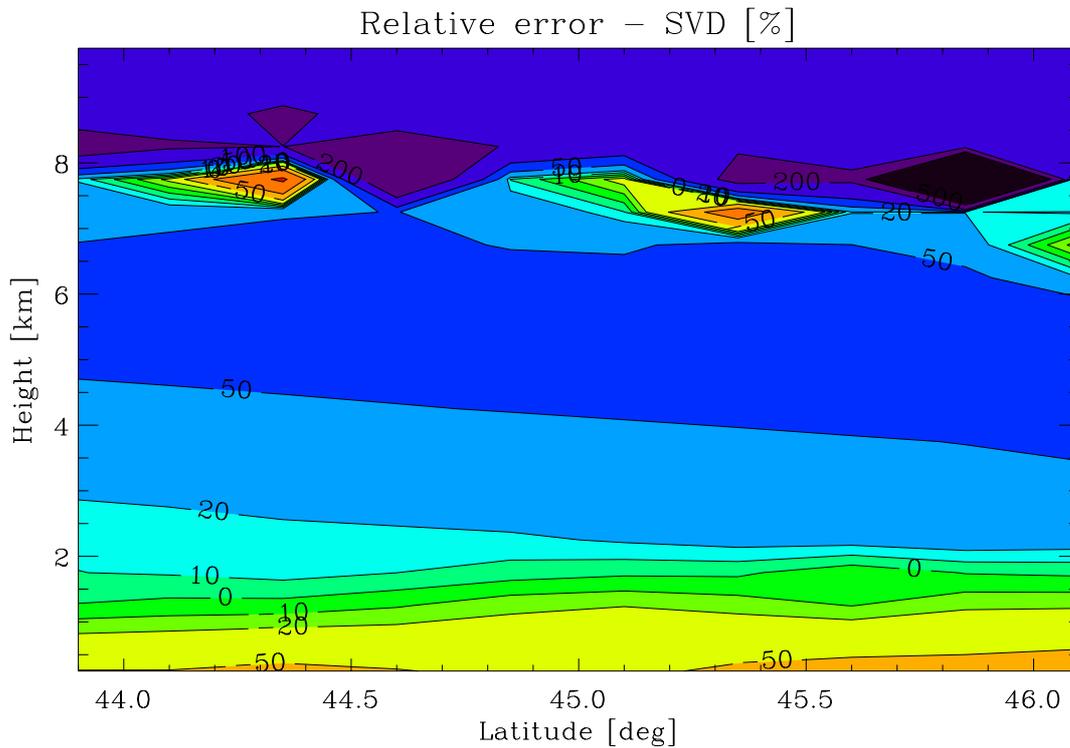


Fig. 6.9: Relative difference between the truncated SVD solution shown in Fig. 6.8 and the original water vapor density field (Fig. 6.5).

Combining the SIWV data with the *a priori* field via optimal estimation (Eq. (6.10)) yields, on the other hand the satisfying result shown in Fig. 6.10. The relative error (shown in Fig. 6.11) is smaller than 10% almost in the entire area of the retrieval domain.

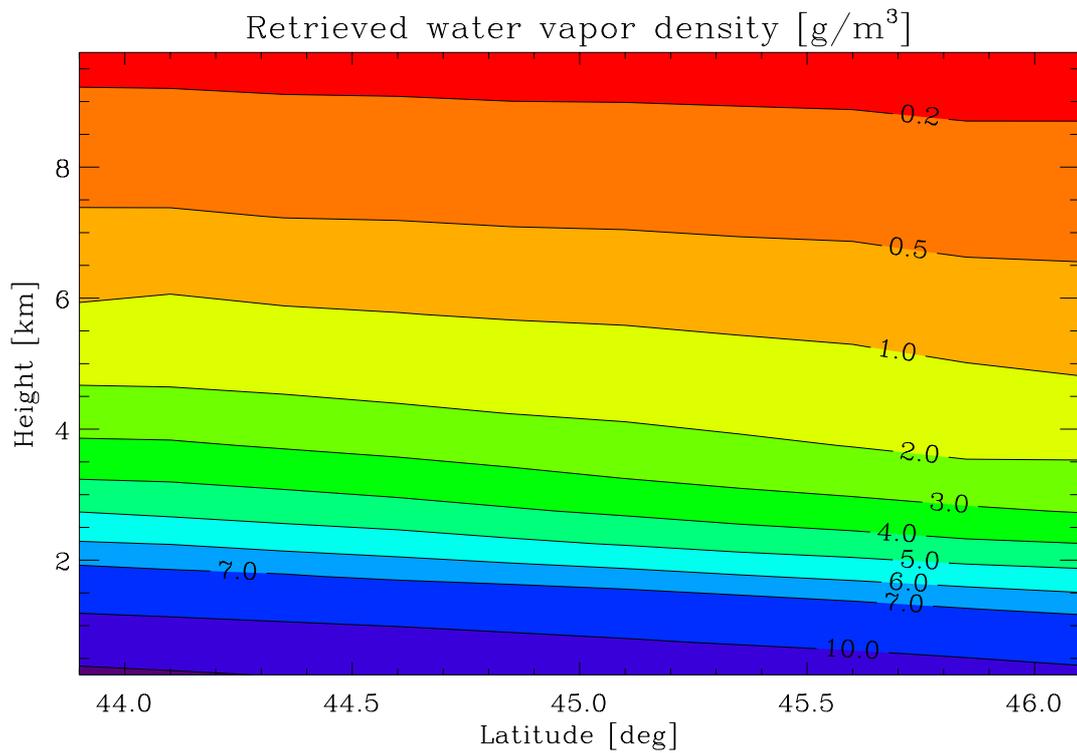


Fig. 6.10: Retrieval of the synthetic exponential atmosphere with linear horizontal gradient, obtained with optimal estimation.

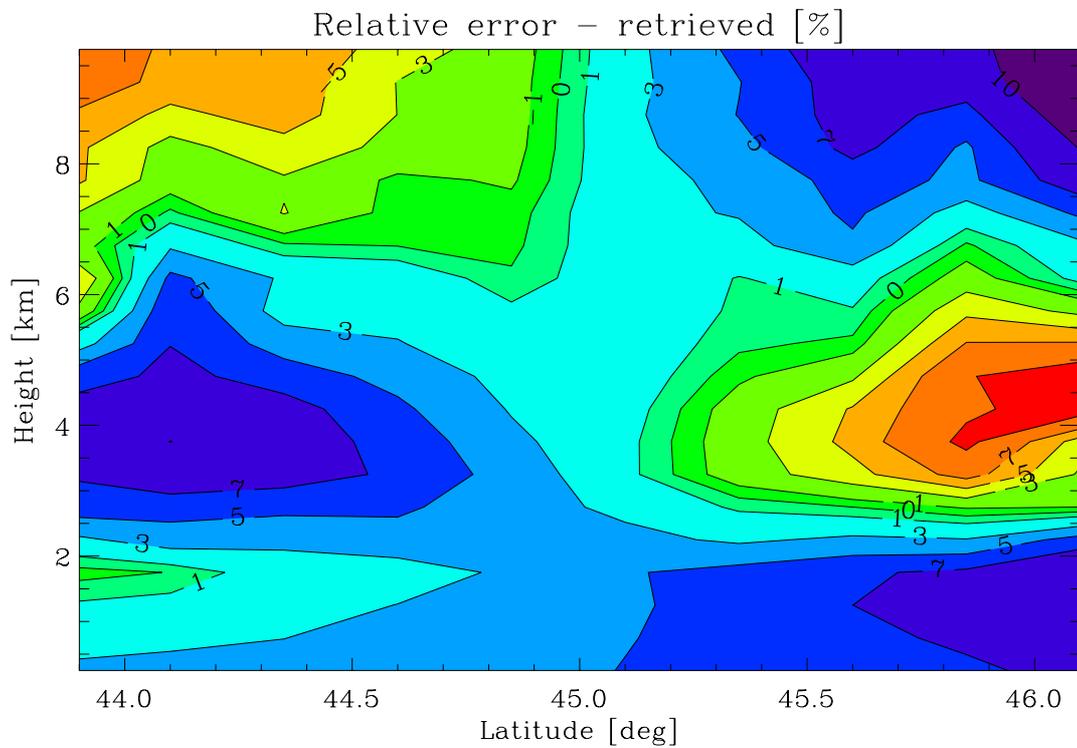


Fig. 6.11: Relative difference between the optimal estimation retrieval (Fig. 6.10) and the original water vapor density field (Fig. 6.5).

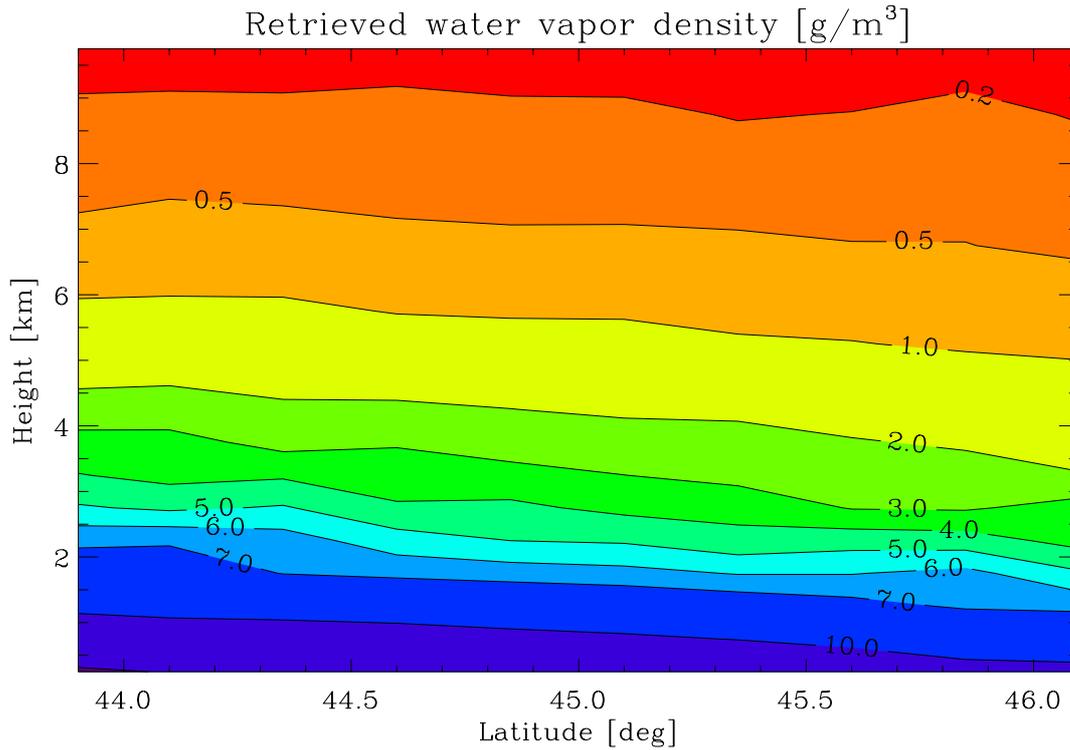


Fig. 6.12: Optimal estimation retrieval of the synthetic exponential atmosphere with linear horizontal gradient, assuming an rms IWV error of $1.5 \text{ kg}/\text{m}^2$.

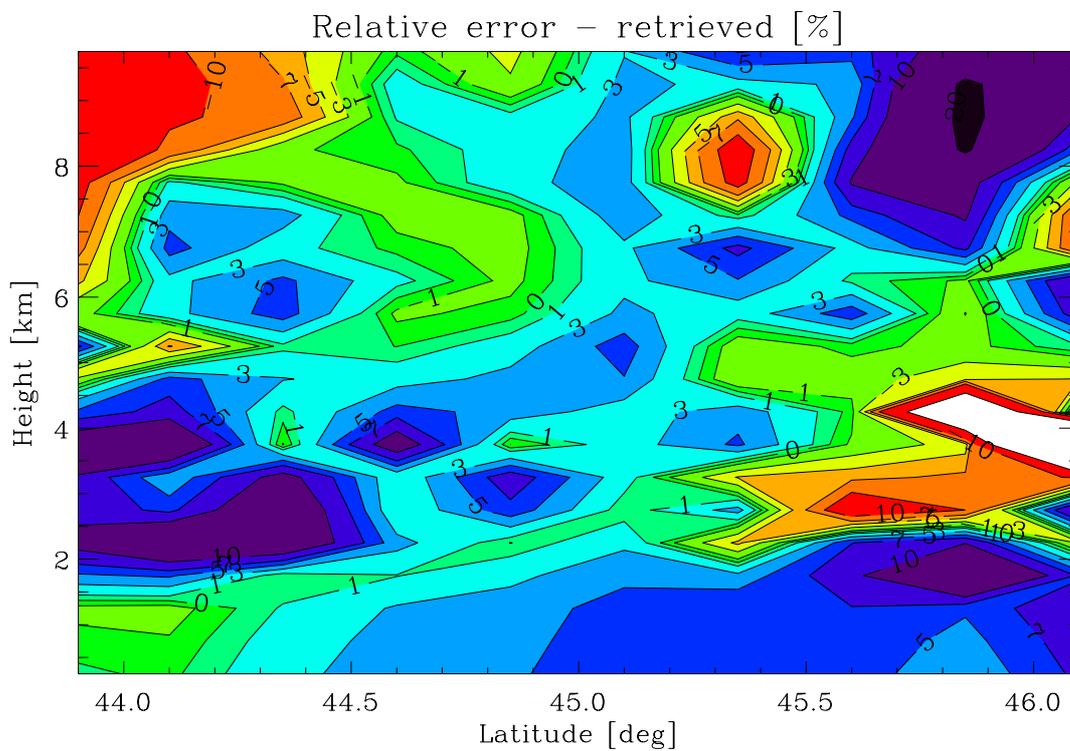


Fig. 6.13: Relative difference between the optimal estimation retrieval (Fig. 6.12) and the original water vapor density field (Fig. 6.5).

Adding a realistic rms error of 1.5 kg/m^2 [cf., e.g., Ware *et al.*, 1997] to the SIWV "measurements" yields the (only slightly degraded) optimal estimation retrieval shown in Fig. 6.12 and the corresponding relative error shown in Fig. 6.13. On the contrary such realistic error completely destroys the truncated SVD solution as is displayed by Fig. 6.14.

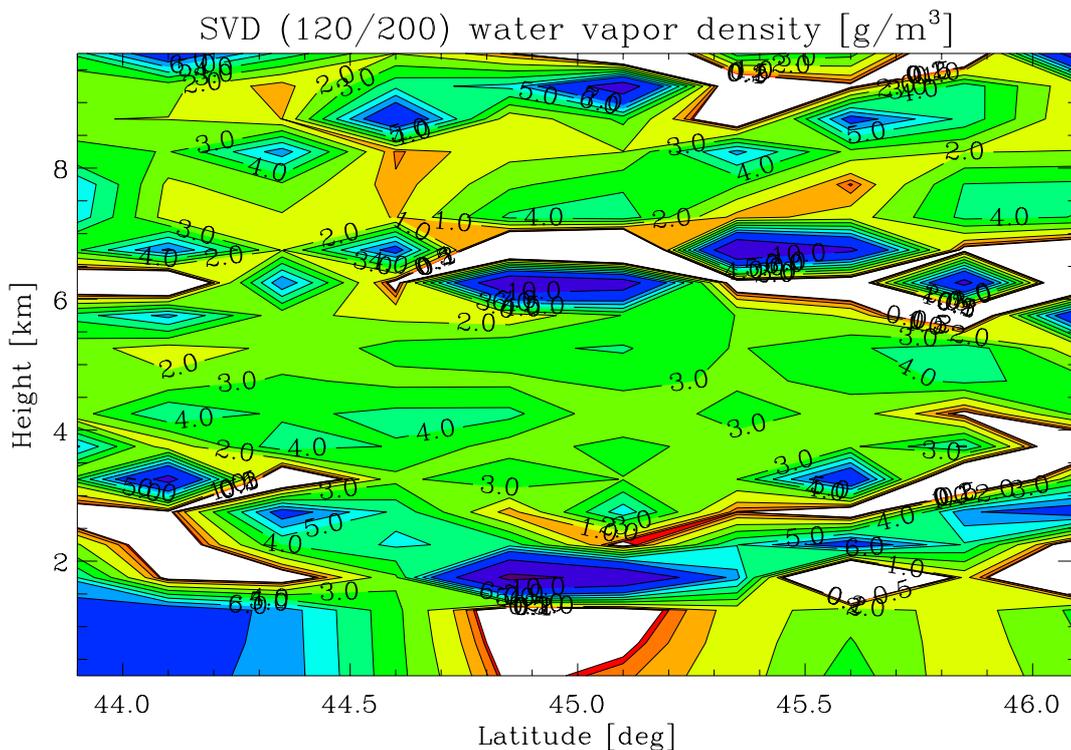


Fig. 6.14: Truncated SVD "solution" using 120 (of 200 possible) singular values and assuming an rms IWV error of 1.5 kg/m^2 (cf. Fig. 6.8 based on perfect measurements).

As very clearly shown by this simple example case, the information on the vertical water vapor structure, brought into the optimal estimation imaging seamlessly as prior information, dramatically improves the conditioning of the imaging problem and is indeed a strict requirement for rendering the reconstruction feasible.

6.6.2 Isolated Gaussian blob

In this scenario we assumed an analytical model water vapor density in the form a bivariate normal distribution (a so-called "Gaussian blob") with a peak water vapor density of 10 g/m^3 in a height of 3 km, a vertical half width (standard deviation) of 1 km, and a horizontal half width of 0.15° . We added again a realistic rms error of 1.5 kg/m^2 to the simulated SIWV measurements. The model water vapor density field is displayed in Fig.

6.15. This case serves to test and indicate the ability of the reconstruction algorithm to resolve small-scale structures in the water vapor field.

The corresponding retrieval result, employing optimal estimation, is shown in Fig. 6.16. The overall structure of the Gaussian blob is quite well resolved, but slightly broadened in latitude. The small-amplitude structures sideward of the Gaussian are introduced by the *a priori field* (again the average profile over the horizontal domain), which is in this case a "dam-like" structure with a peak water vapor density of less than 3 g/m^2 .

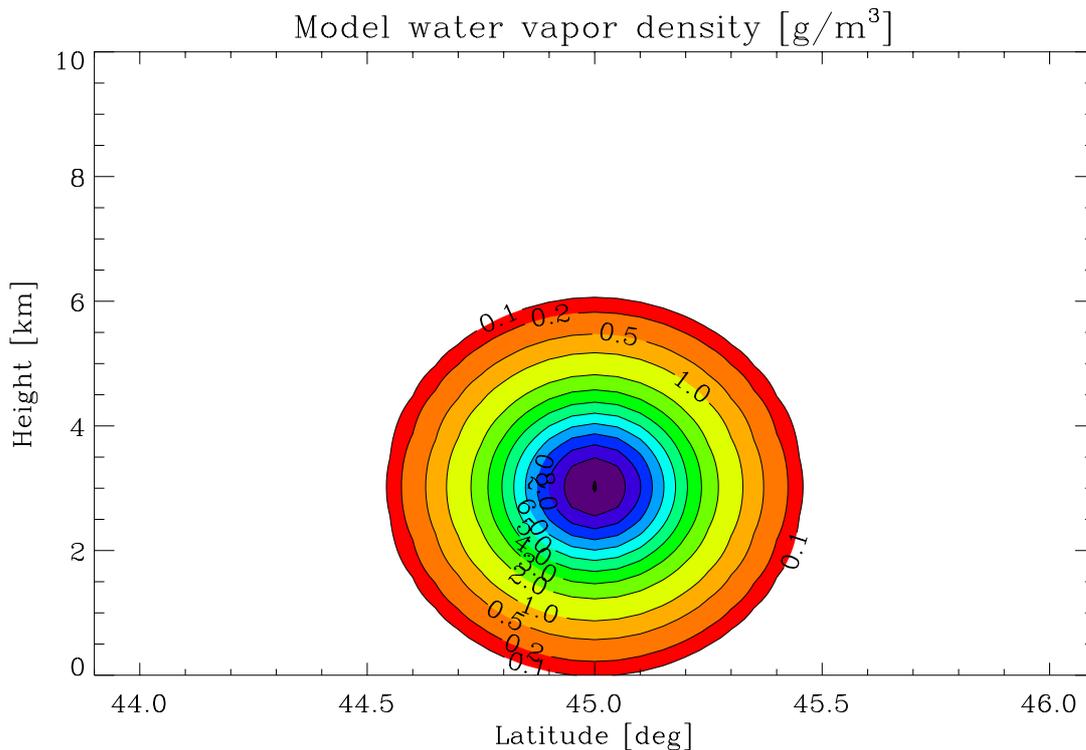


Fig. 6.15: Model water vapor density for an isolated Gauss peak with a vertical half width of 1 km, and a horizontal half width of 0.15° .

The difference between model and result is in this case displayed in absolute values (Fig. 6.17), as the quasi-zero density values outside of the Gaussian inhibit the display of relative errors in the retrieved density field. Obviously, our method allows reasonable well to recover small scale structures.

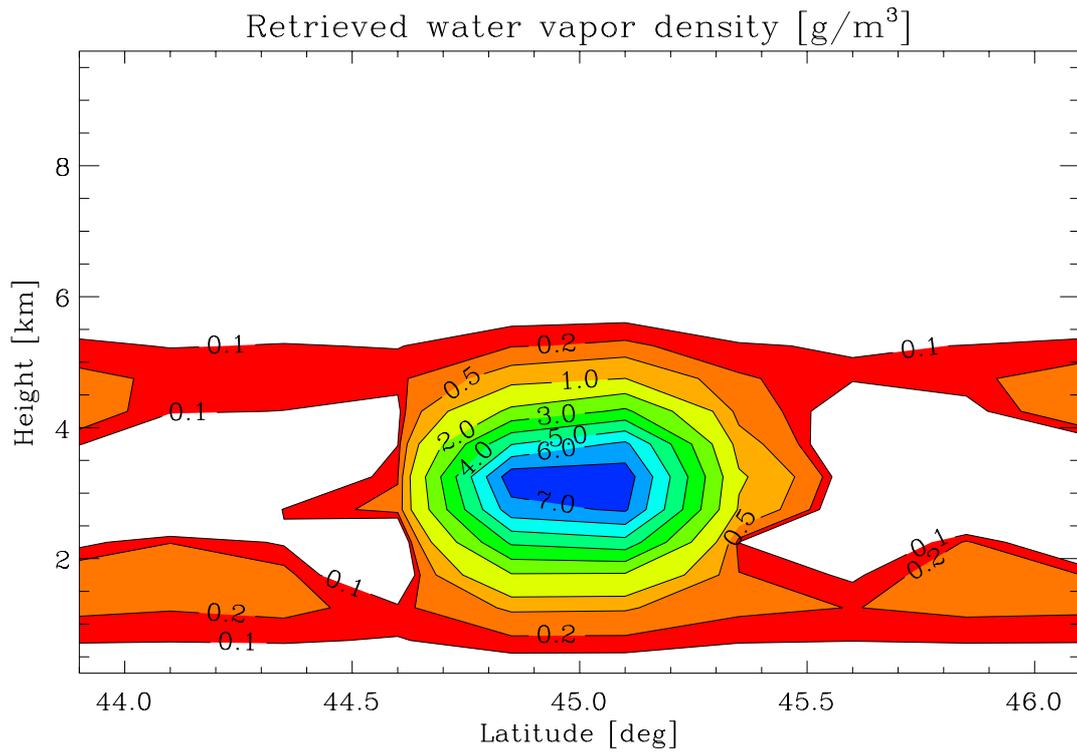


Fig. 6.16: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.15, assuming an rms IWV error of 1.5 kg/m².

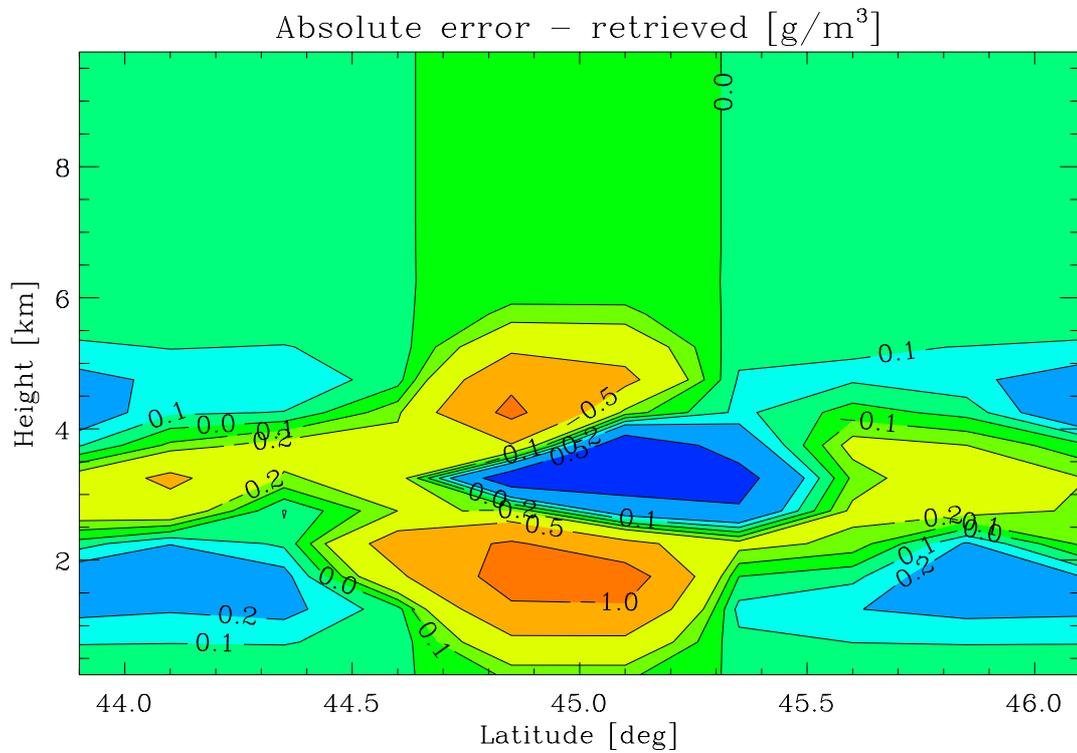


Fig. 6.17: Absolute difference between model (Fig. 6.15) and retrieved water vapor density (Fig. 6.16) for the scenario of an isolated Gauss peak.

6.6.3 Centered Gaussian blob with exponential background

In this scenario a Gaussian blob (the same as in sub-section 6.6.2) was embedded into an exponentially decreasing background density field. The background density field is determined by a constant surface density value of 20 g/m^3 and by a climatological (constant) water vapor scale height of 2 km, describing the exponential decrease with height. The blob was placed here in the middle of the retrieval domain shown in Fig. 6.18; in the next sub-section the results are shown when placing the blob near the of the domain. The simulated SIWV measurements were again disturbed with an rms error of 1.5 kg/m^2 in order to simulate (conservatively assumed) measurement errors.

The optimal estimation retrieval result for the centered Gaussian is displayed in Fig. 6.19, the relative error in Fig. 6.20, respectively.

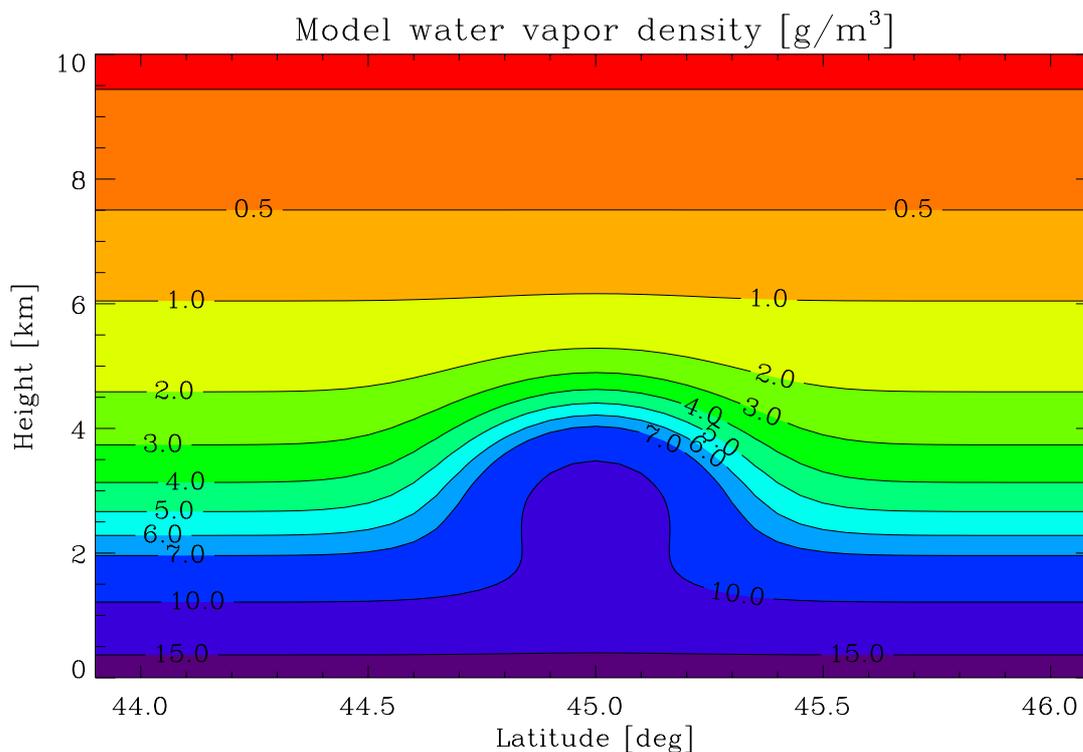


Fig. 6.18: Exponentially decreasing background water vapor density field with centered Gaussian blob.

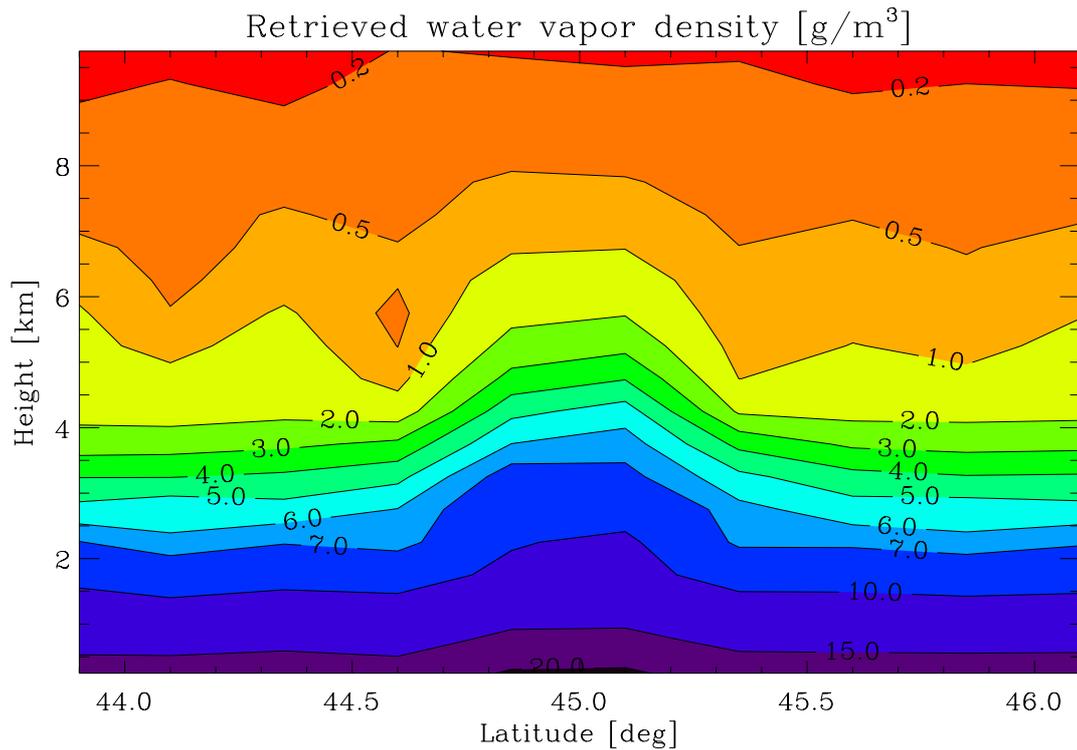


Fig. 6.19: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.18, assuming an rms SIWV error of $1.5 \text{ kg}/\text{m}^2$.

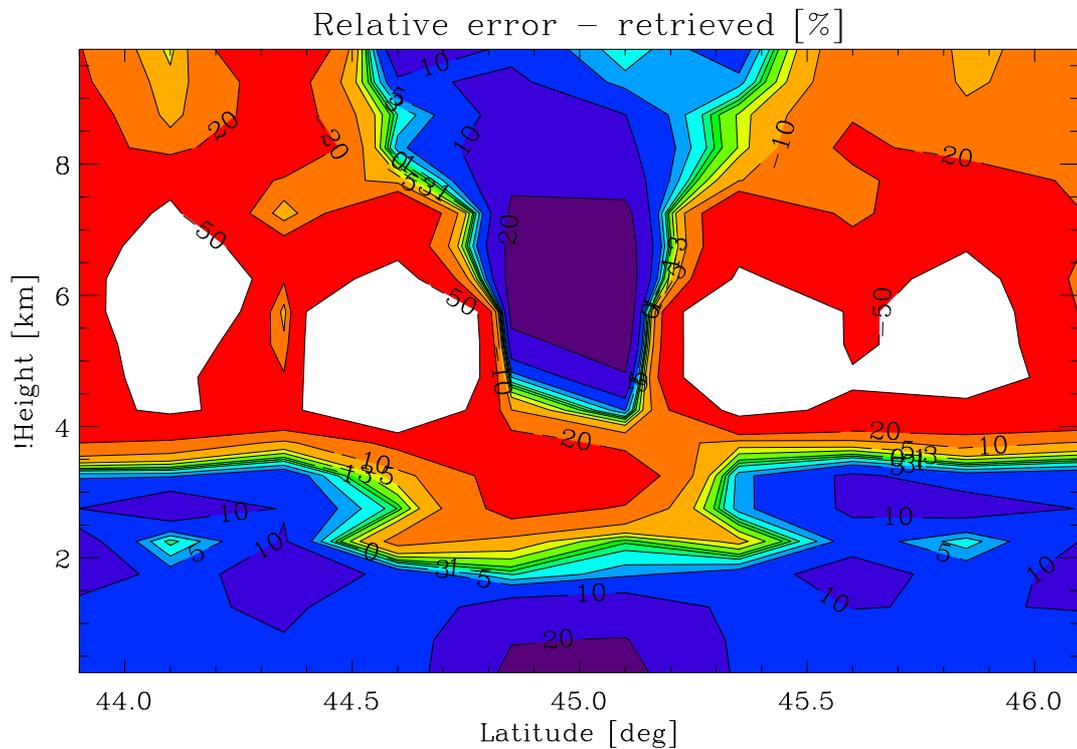


Fig. 6.20: Relative difference between model (Fig. 6.18) and retrieved water vapor density (Fig. 6.19) for the scenario of a Gaussian blob within an exponentially decreasing background density field.

The scenario described above represents a rather difficult problem for the retrieval algorithm, the overall structure of the Gaussian blob is nevertheless reasonably resolved. The pronounced overestimation of the water vapor density below and above the Gaussian blob is compensated by regions with underestimated water vapor densities, visible as white areas in Fig. 6.20, corresponding to relative errors of more than 50%. Thus such strong small scale structures can still be semi-quantitatively imaged though in the already fairly dry middle troposphere errors exceeding 50% are observed.

6.6.4 Offset Gaussian with exponential background

The background density field in this scenario was exactly the same as in sub section 6.6.3, but the Gaussian blob, though also the same in shape and magnitude, was now placed near the "northern" edge of the retrieval domain. The corresponding model water vapor density field is shown in Fig. 6.21. In this case it should be taken into account that only a few rays penetrate a small structure near the edge of the retrieval domain. Considering this fact, the optimal estimation retrieval result (shown in Fig. 6.22) for this rather challenging synthetic case is still reasonable.

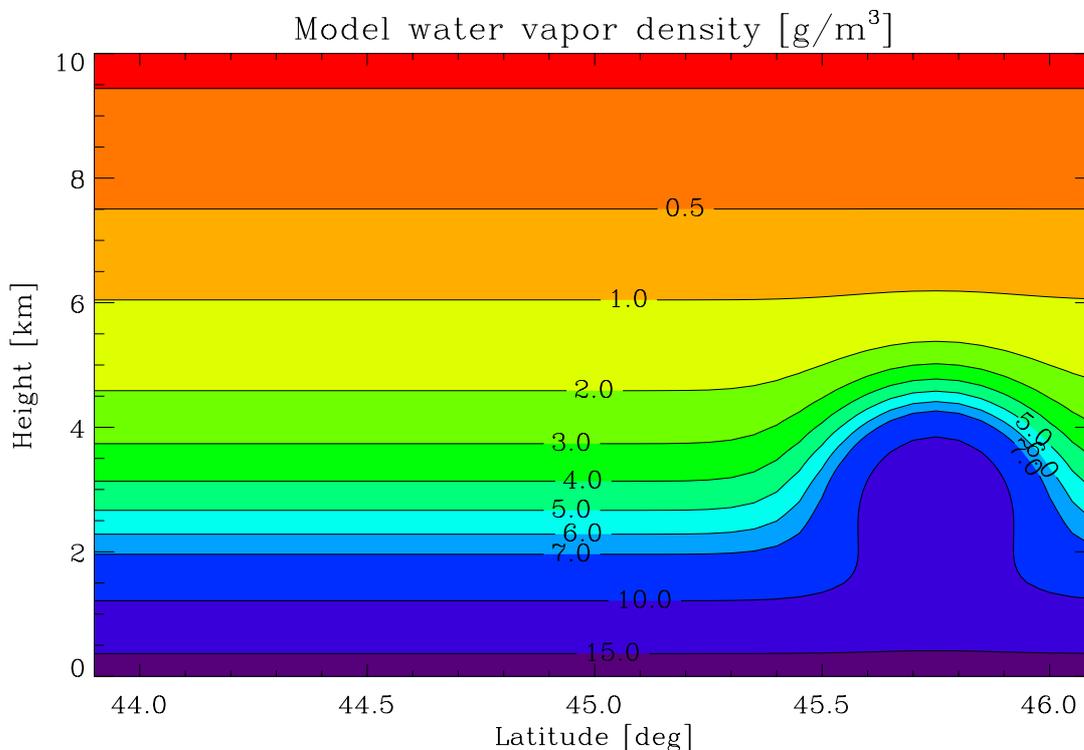


Fig. 6.21: Exponentially decreasing background water vapor density field with a Gaussian blob near the edge of the retrieval domain.

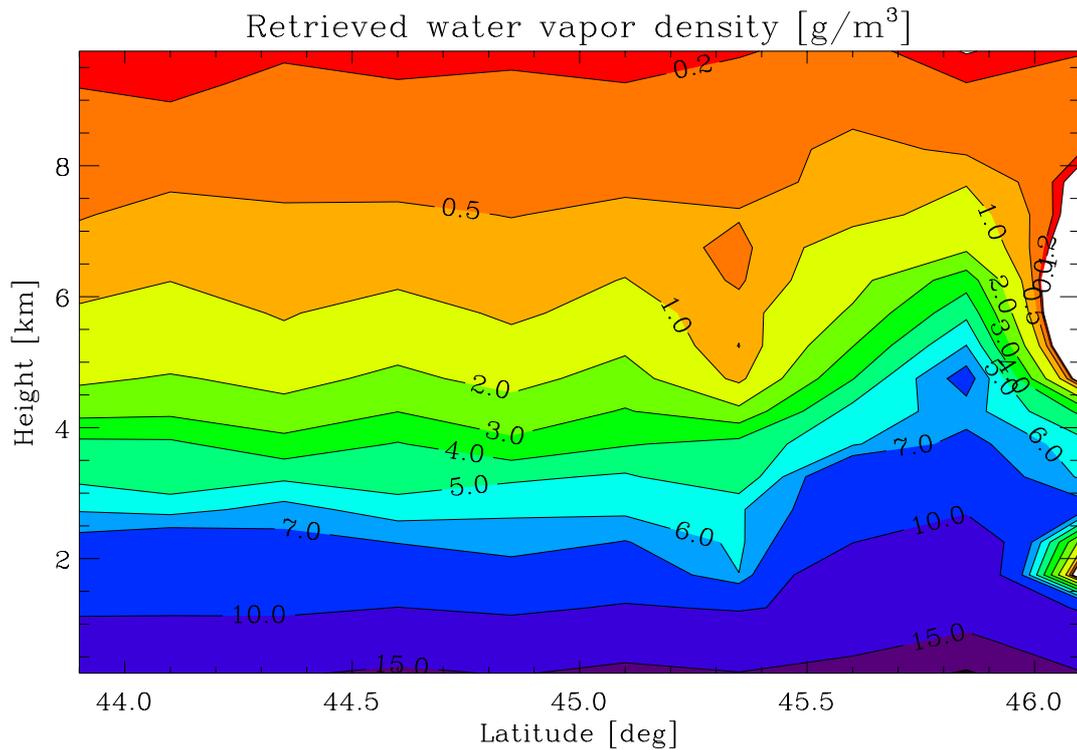


Fig. 6.22: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.22, assuming an rms SIWV error of $1.5 \text{ kg}/\text{m}^2$.

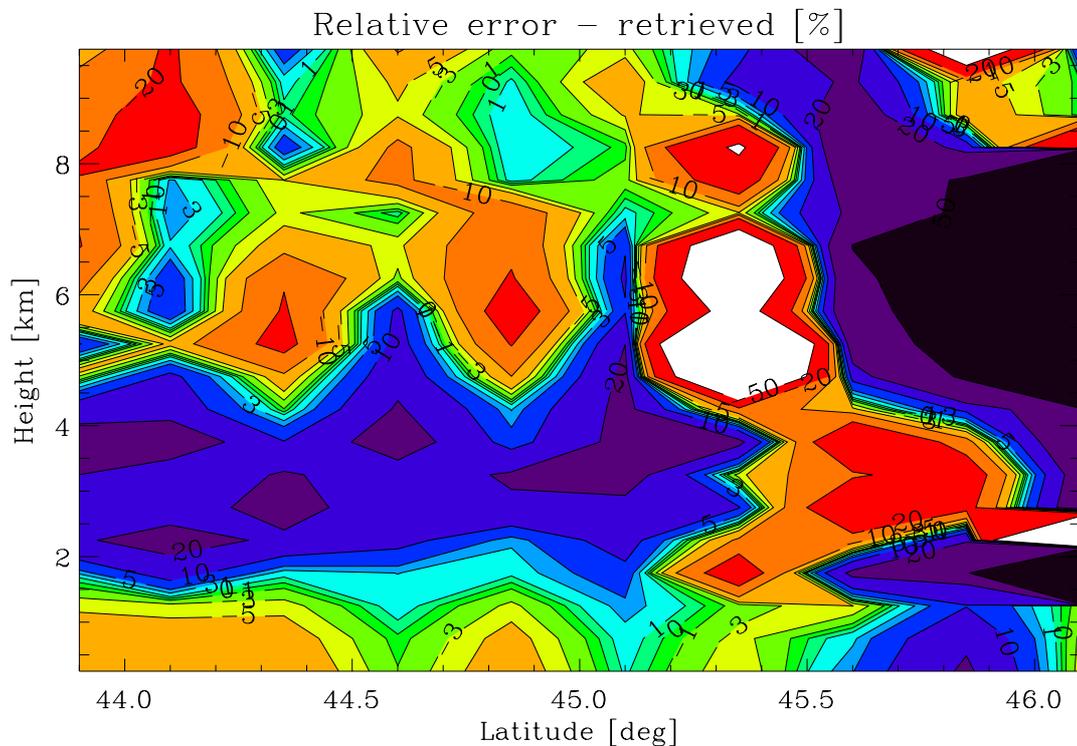


Fig. 6.23: Relative difference between model (Fig. 6.21) and retrieved water vapor density (Fig. 6.22) for the scenario of a Gaussian blob within an exponentially decreasing background density field near the edge of the retrieval domain.

The relative error shown in Fig. 6.23, indicates that serious qualitative reconstruction difficulties exist in the northern part of the domain but that most of the domain is still reconstructed with errors below about 20%.

6.6.5 Strong sub-scale Gaussian blob

In the last simple analytical scenario, a very narrow and strong Gaussian blob with a vertical half width of 0.2 km, and a latitudinal half width of 0.05° (i.e., with a blob dimension smaller than a single picture element) within an exponentially decreasing background was used to investigate the response of the reconstruction system to strong sub-scale perturbations [cf. Fougere, 1995, who performed similar tests when dealing with ionospheric imaging].

The model water vapor density field, shown in Fig. 6.24, is essentially the same as in Fig. 6.18, with the exception that the size of the Gaussian blob is greatly reduced while keeping the same magnitude.

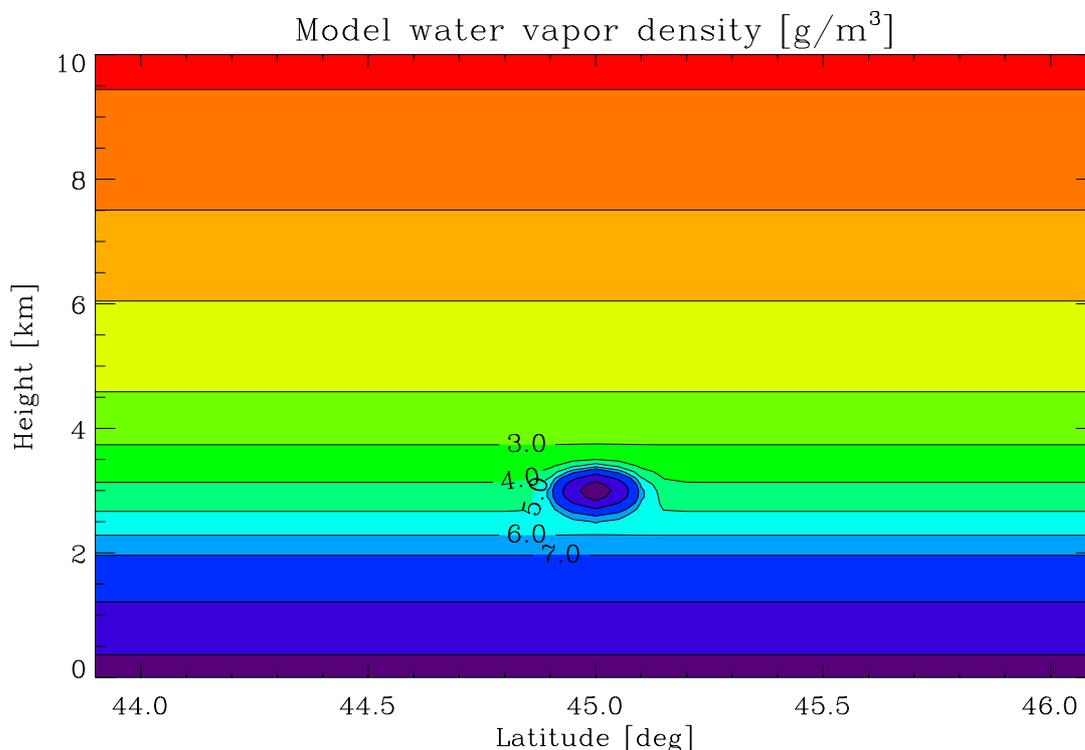


Fig. 6.24: Exponentially decreasing background water vapor density field with a centered strong sub-scale Gaussian blob.

The optimal estimation retrieval result, displayed in Fig. 6.25, demonstrates that any localization information on the sub-scale Gaussian blob is lost. The characteristic wave-like response features well visible in Fig. 6.25 can also be detected, though they are less pronounced, in Fig. 6.19.

The relative error is now, not surprisingly, large above about 4 km since the few rays probing the sub-scale blob spread out significant absolute density errors into the middle and upper troposphere. On the other hand the moist lower troposphere below about 4 km shows still fairly reasonable results.

We should also note, finally, that a sub-scale structure as strong as the example used here for illustration will be very rare in the real atmosphere.

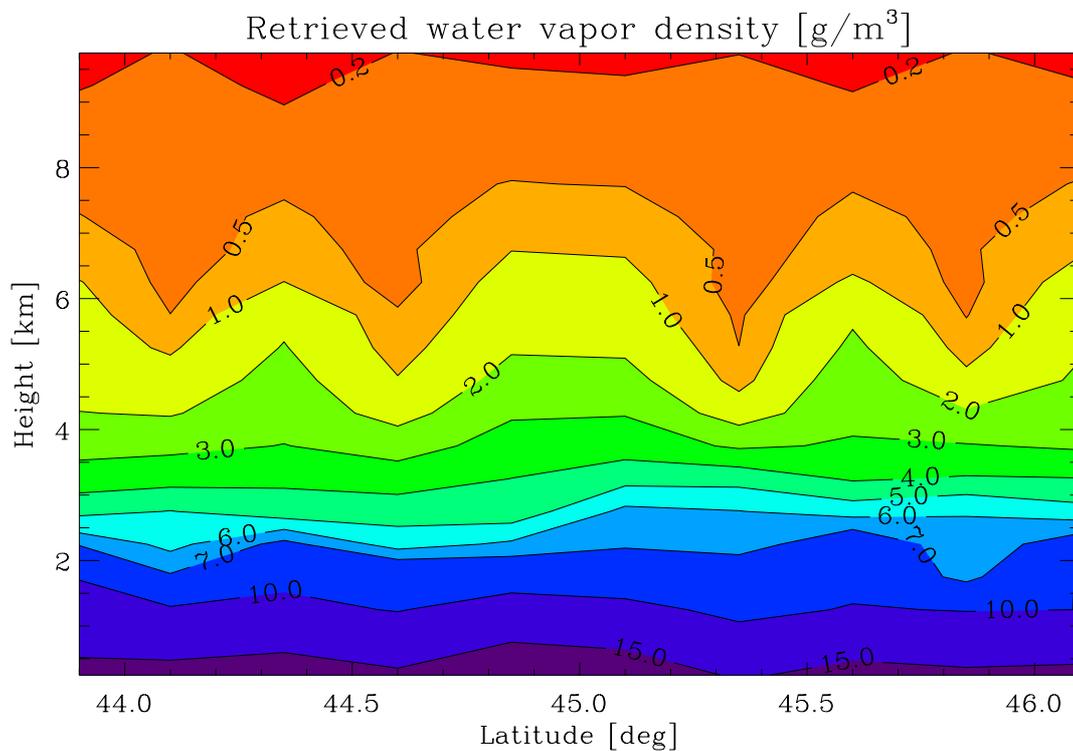


Fig. 6.25: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.24, assuming an rms SIWV error of $1.5 \text{ kg}/\text{m}^2$.

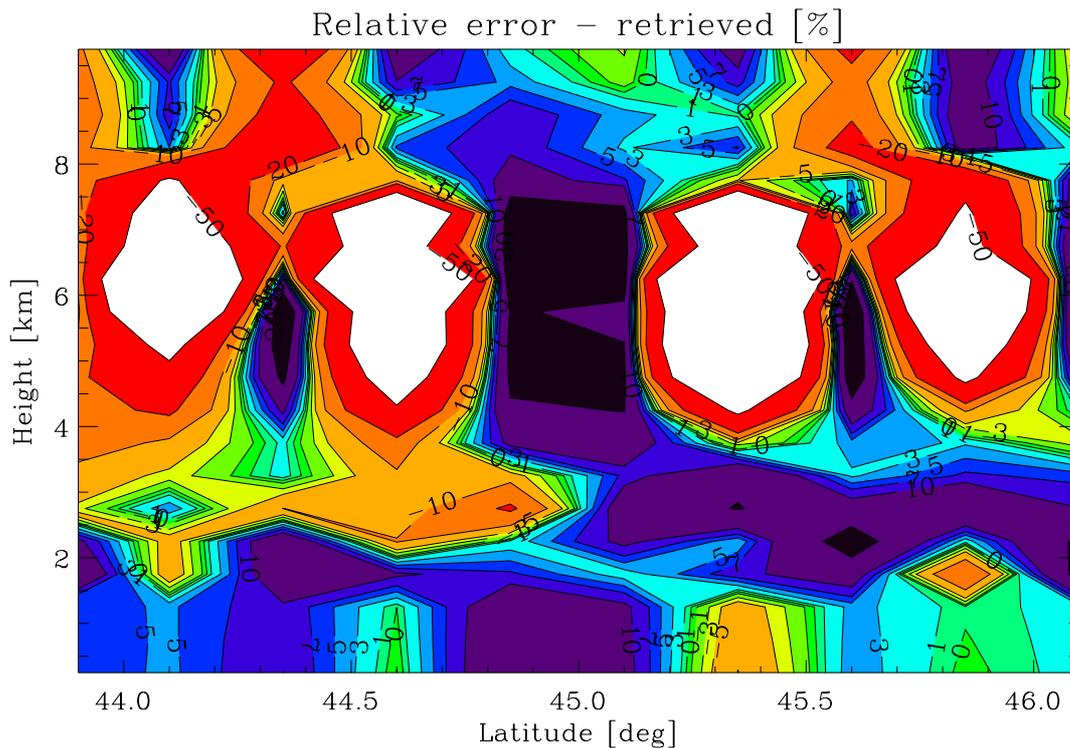


Fig. 6.26: Relative difference between model (Fig. 6.24) and retrieved water vapor density (Fig. 6.25) for the scenario of a strong sub-scale Gaussian blob in an exponentially decreasing background density field.

6.7 Retrieval results - ECMWF-derived data

The results of sub-section 6.6.1 clearly demonstrated that it is generally not possible to obtain any useful (two-dimensional) information about the water vapor density field with ground-based sounding data alone (see Fig. 6.14). It is therefore indispensable to incorporate additional information about the height distribution of water vapor.

6.7.1 Realistic a priori information

We assumed that a radio occultation measurement co-located with the ground stations delivers an accurate mean refractivity profile which is representative for the whole retrieval domain, as the horizontal range of the domain corresponds roughly to the horizontal resolution of the radio occultation technique (cf. sub-section 4.2.9). Thus we mimicked occultation derived refractivity profiles by computing the horizontally averaged profile over the domain.

Pressure and temperature data from the latest available ECMWF analysis field (3 hours before the field under study) were then used to compute a mean dry refractivity profile via Eq. (2.65). The water vapor density profile was then obtained via Eq. 4.41. We note that using the latest ECMWF analysis to derive pressure and temperature profile information represents a rather conservative approach, in practice one would use a short-range forecast. As the vertically integrated water vapor overlying a specific station is accurately delivered by the ground-based GPS measurements, the 10-station-average value was used to adjust (shrink or blow up) the water vapor density profile, derived via Eq. (4.41), in order to match this average value. In this way some potential bias in the density profile is effectively mitigated.

Furthermore, as the occultation measurements can generally not be expected to provide data within the first few hundred meters, an adjustment procedure similar to the one described above will have to be frequently employed together with short-range extrapolation in order to extend the water vapor density profile reasonably down to the ground. Alternatively no adjustment at all can be used in this case since it is not a necessary condition in our methodology that the *a priori* field extends over all pixels of the reconstruction domain.

The resulting water vapor density profile was then assumed to be representative for the entire retrieval domain (assuming it to be distributed spherically symmetric), furnishing the *a priori* field. An inherent advantage of the radio occultation technique is to deliver a mean profile, which is representative over the entire retrieval domain. In principle, any vertical water vapor density profile (e.g., a co-located radiosonde profile) can be used as an *a priori* as well.

The *a priori* covariance matrix, \mathbf{S}_{ap} , was constructed with the following reasonable assumptions: We considered the surface *a priori* water vapor density to be accurate to 10% at center, gradually increasing to 30% at the edges of the retrieval domain. At the 6 km level, the respective values were assumed to be 65% (center) and 200% (edge). The decrease of accuracy with height was assumed to be linear between the values and off-diagonal elements (covariances) were assumed negligible. These assumptions aim at well representing the error of occultation derived profiles over the domain. The reconstruction algorithm is only weakly sensitive to the exact variances assumed.

The measurement error covariance matrix, \mathbf{S}_e , was constructed with the assumption of an rms measurement error of 1.5 kg/m^2 uncorrelated between rays (off-diagonal zero), corresponding to a $\sim 1 \text{ cm}$ error in the wet delay.

6.7.2 Florida case

Retrieval results for an ECMWF-derived a refractivity field located in Florida (centered near Palm Beach, at 27°N , 80°W) are presented as a representative example for low latitudes with a rather moist troposphere. The model water vapor density field is displayed in Fig. 6.27 and the retrieval result, obtained by optimal estimation, in Fig. 6.28.

A pronounced secondary maximum at a height of $\sim 3 \text{ km}$ (near the northern edge of the retrieval domain, see Fig. 6.27), associated with the boundary of the tradewind inversion, is clearly visible and well resolved in the retrieved water vapor density field (see Fig. 6.28). The a priori profile is plotted in Fig. 6.29. The relative error of the retrieval is shown in Fig. 6.30. Even though this shows considerable structure the retrieval accuracy is better than 10% in most of the domain up to about 5 km height.

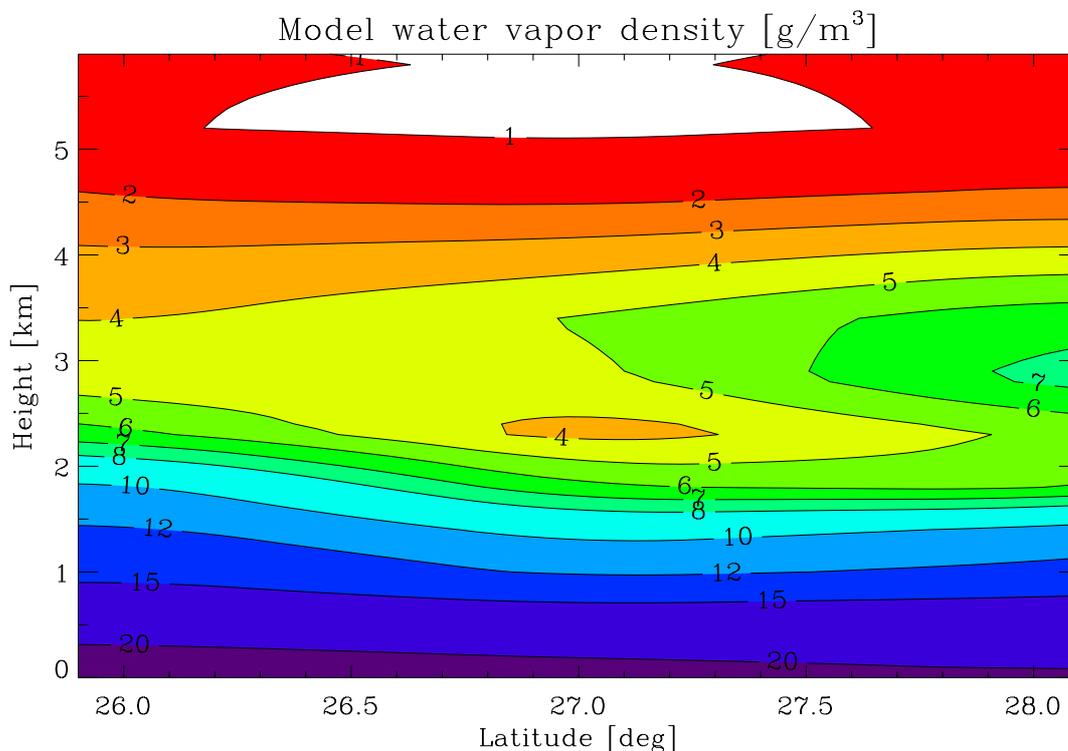


Fig. 6.27: Model water vapor density, derived from ECMWF analysis data over Florida.

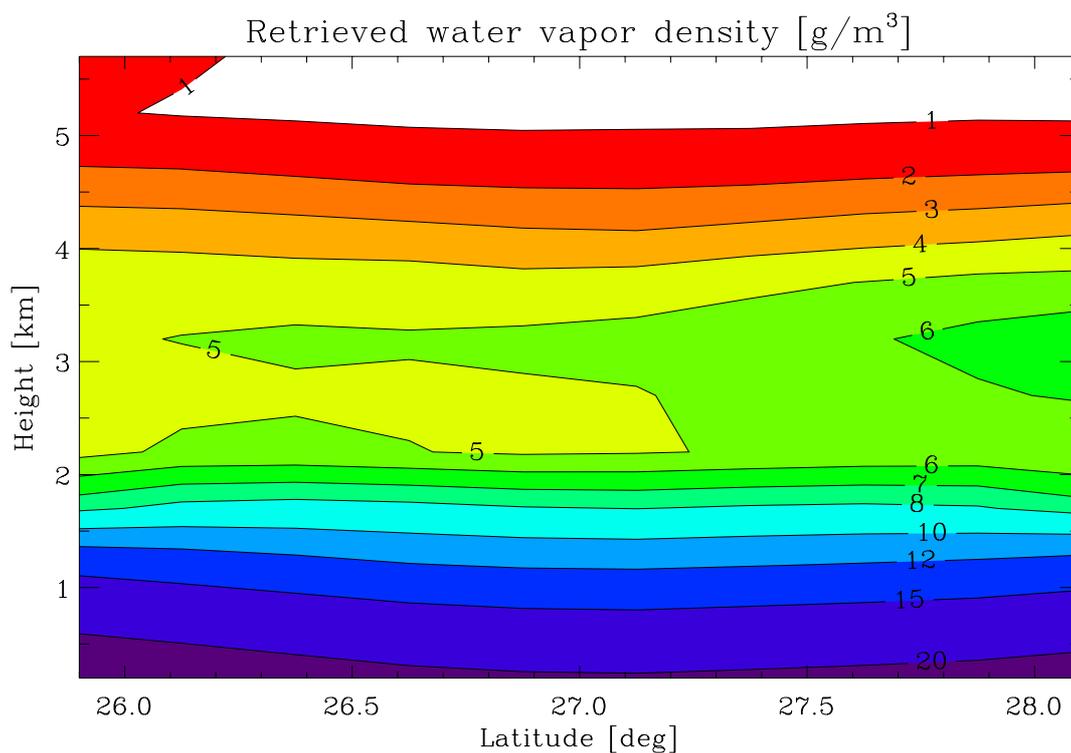


Fig. 6.28: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.27 (Florida case).

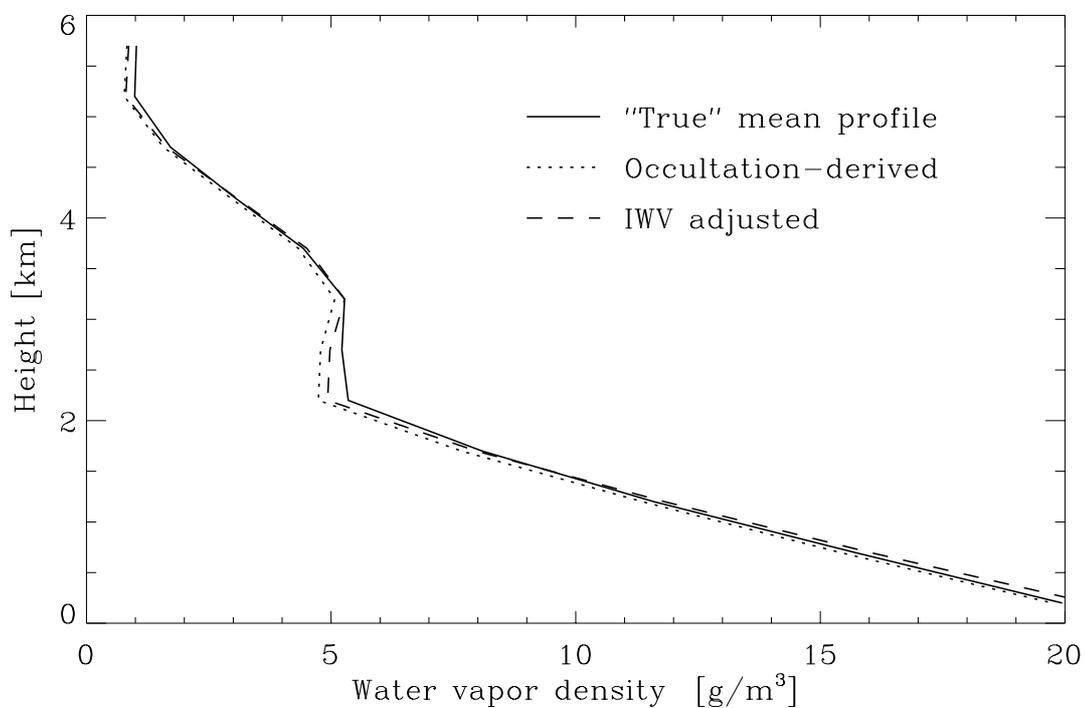


Fig. 6.29: Averaged profile of the water vapor density field shown in Fig. 6.27 (solid line), density profile obtained by mimicked radio occultation (dotted), and occultation profile with IWV adjustment (dashed), the latter used as a priori profile.

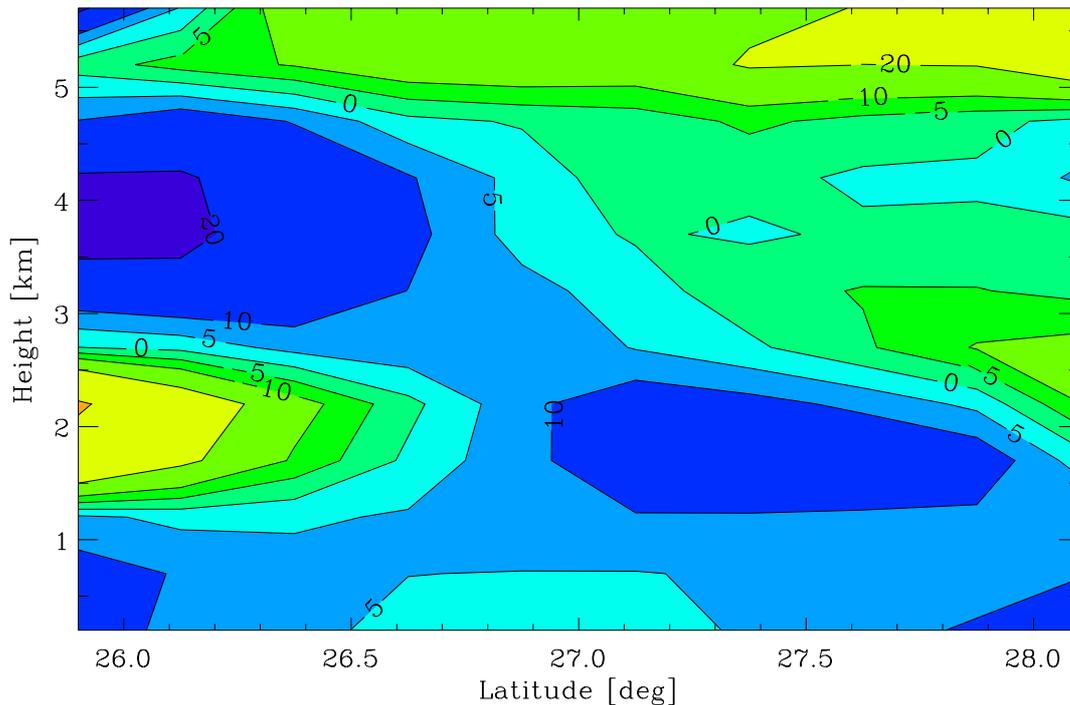


Fig. 6.30: Relative difference between model (Fig. 6.27) and retrieved water vapor density (Fig. 6.28) - Florida case.

6.7.3 Finland case

Retrieval results for an ECMWF-derived a refractivity field located in Finland (centered at 62°N, 26°E) are shown as representative example for high latitudes. The slant integrated water vapor was in this case derived using the geometric mapping function (see sub-section 3.3.3). The overall structure of the model water vapor density field (shown in Fig. 6.31) is similar to the synthetic field in sub-section 6.6.1, but the absolute humidity and the horizontal gradient are much lower. The maximum water vapor density is 7.1 g/m³, in the synthetic atmosphere it was assumed to be 17.9 g/m³ (see Fig. 6.5).

The mimicked occultation profile underestimates the water vapor density within the entire height range, the IWV adjustment improved the situation considerably (see Fig. 6.33). A residual underestimation in the upper part leads to very low water vapor densities above 5 km in the result which show values < 0.2 g/m³ above about 4.5 km (see Fig. 6.32). The corresponding relative error is therefore huge (see Fig. 6.34), even though the absolute error is only about 0.2 g/m³. Below 4 km the results are fairly reasonable and errors below ~ 20% are achieved over most of the domain.

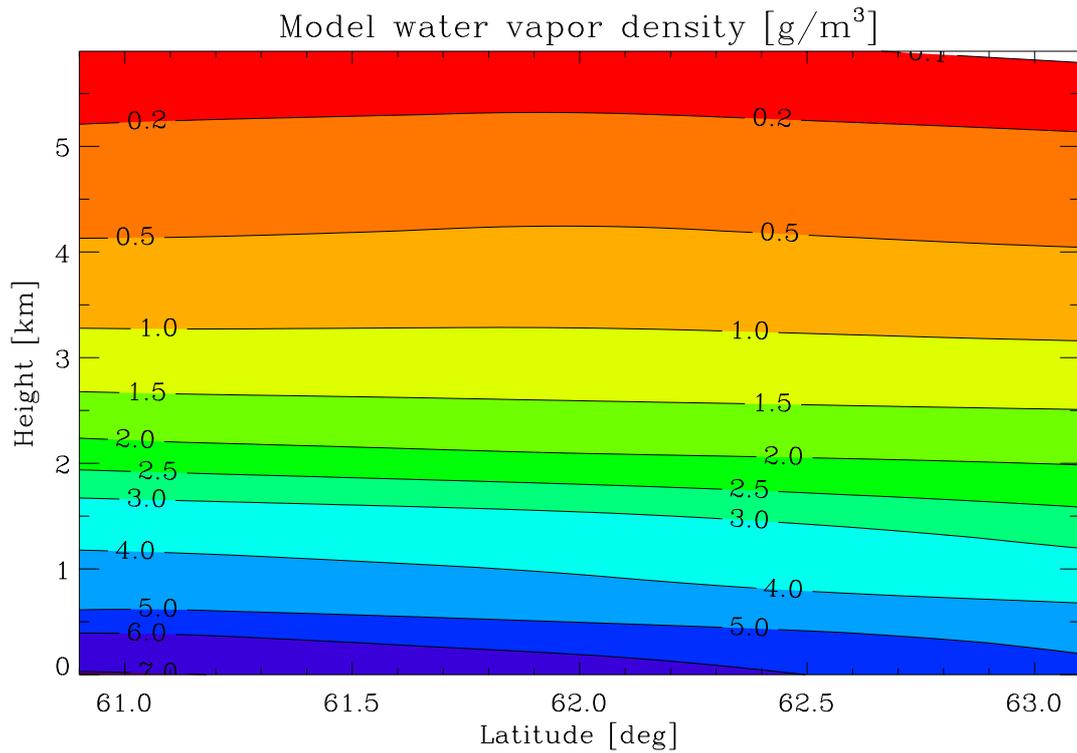


Fig. 6.31: Model water vapor density, derived from ECMWF analysis data over Finland.

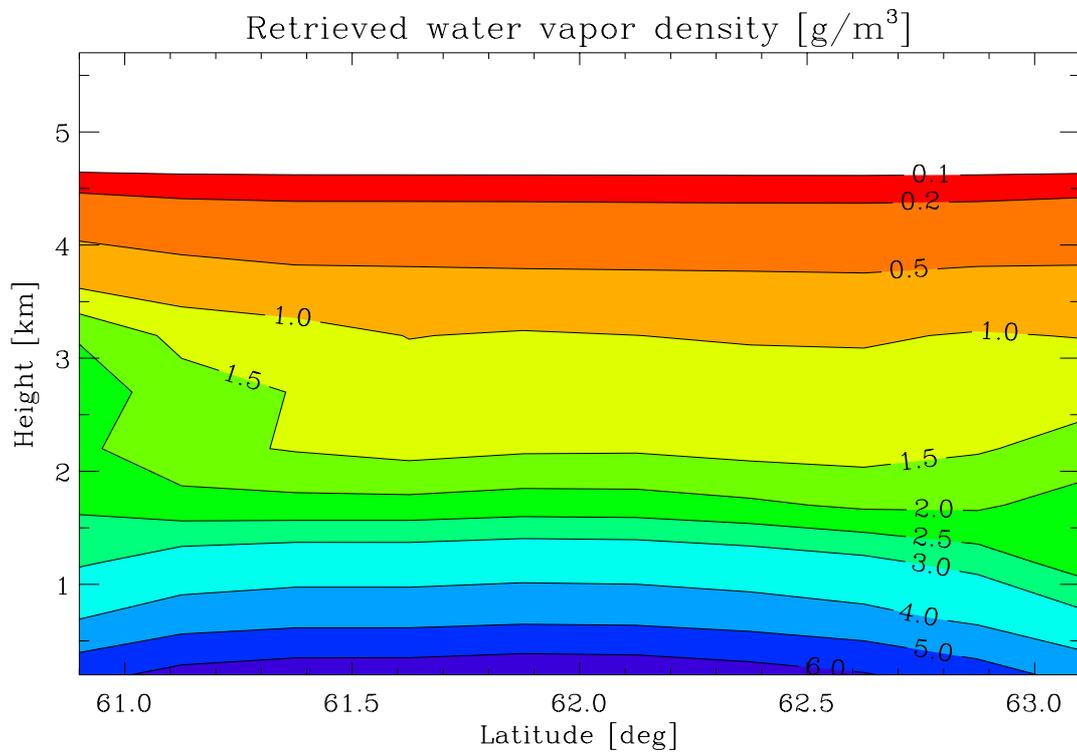


Fig. 6.32: Optimal estimation retrieval of the water vapor density field shown in Fig. 6.31, (Finland case).

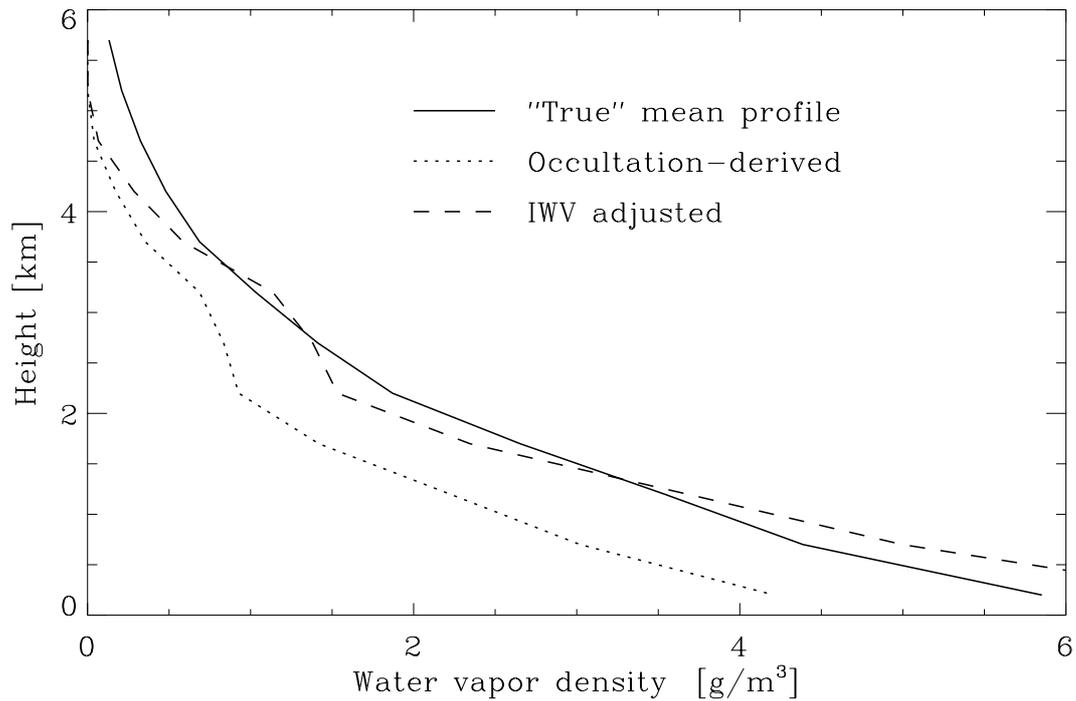


Fig. 6.33: Averaged profile of the water vapor density field shown in Fig. 6.31 (solid line), density profile obtained by mimicked radio occultation (dotted), and occultation profile with IWV adjustment (dashed), the latter used as a priori profile.

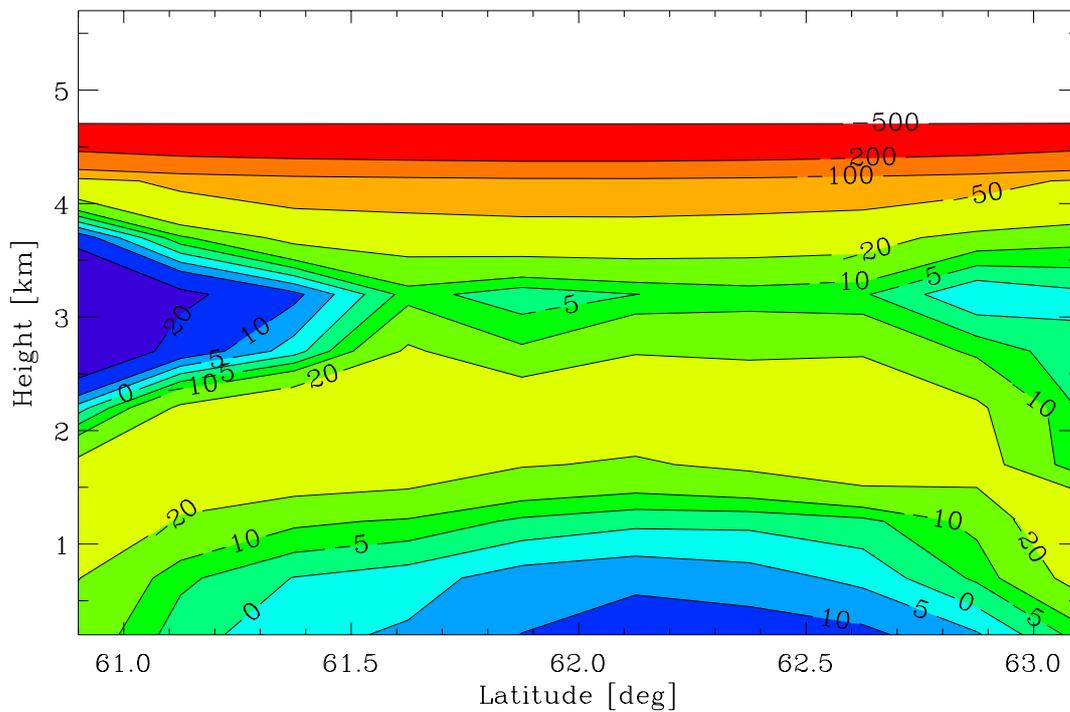


Fig. 6.34: Relative difference between model (Fig. 6.31) and retrieved water vapor density (Fig. 6.32) - Finland case.

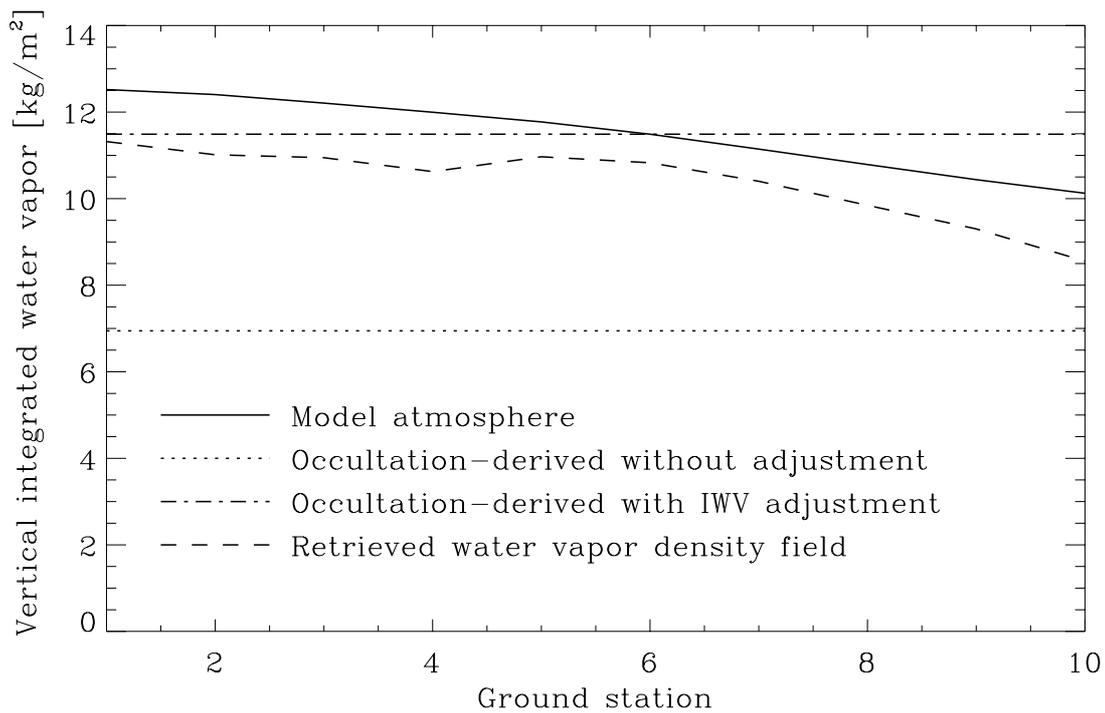


Fig. 6.35: Vertical integrated water vapor over the ground-stations derived from the model atmosphere (solid line), from the mimicked radio occultation without adjustment (dotted), with IWV adjustment (dashed-dotted), and from the retrieval result (dashed).

For assessing how well the vertically integrated water vapor (IWV) and its horizontal structure are captured by the reconstructed image, it is instructive to look also into this quantity. We depicted one such example in Fig. 6.35 for the Finland case: The "true", the *a priori*, and the retrieved vertical integrated water vapor values overlaying the 10 ground-stations are shown. Though the Finland case with its fairly low densities is a case with comparatively degraded performance relative to low latitude images the IWV is obviously captured well. Nevertheless, if the interest were in IWV only, one should better derive the IWV directly from the slant path delays.

SUMMARY AND CONCLUSIONS

This work was concerned with the development of a water vapor imaging technique, based on the combination of ground-based and spaceborne GNSS sounding data. First we motivated why an improved knowledge on the distribution of water vapor in the atmosphere and a deeper understanding of the variations in space and time are crucial, e.g., for the assessment of the impact of anthropogenic climate change.

The atmospheric refractivity is a parameter of fundamental importance for the remote sensing techniques used within this work. We explored and described thus thoroughly how the macroscopic refractivity relates to microscopic properties (i.e., the atomic or molecular polarizability and the permanent dipole moment, in case of a polar molecule like the water molecule) and obtained theoretical results that are in excellent agreement with published measurement values [cf., e.g., Thayer, 1974].

Ground-based GPS meteorology represents a new powerful tool which can be used to derive the vertical integrated water vapor overlaying a receiver station with an accuracy of about 1 kg/m^2 [cf., e.g., Elgered *et al.*, 1997], which is equivalent to 1 mm accuracy in precipitable water, the height of an equivalent column of liquid water. Slant integrated water vapor along ray paths with arbitrary elevation angles can be derived with almost the same accuracy [Ware *et al.*, 1997]. We introduced the theoretical background and the practical concept of the methodology of ground-base GPS meteorology and focused on the dominant error sources.

An important feature of the retrieval of integrated water vapor is the mapping function, that relates the zenith hydrostatic delay (derived from pressure measurements at the ground receiver) to the hydrostatic delay in the direction of the GPS satellite. We developed a new mapping function, based on geometrical considerations. The mapping function was tested using simulated data. Poleward of about 40° latitude, the obtained results were better than the ones obtained by employing widely used mapping functions.

The radio occultation technique is based on GNSS measurements with a receiver onboard a low Earth orbit (LEO) satellite. As a standalone technique, it delivers accurate near-vertical profiles of atmospheric refractivity, which can be converted into water vapor profiles,

given independent information on the height distribution of temperature. We gave an overview of the theoretical background and the methodology of the radio occultation technique as we propose to use such profiling information in combination with the ground-based data.

Tomographic imaging becomes feasible when ground-based measurements are combined with spaceborne measurements, which requires co-location of ground receivers and occultation events. After formulating space-time sampling requirements for atmospheric imaging, we investigated how well they are fulfilled for different LEO constellations.

We developed a technique for tropospheric water vapor tomographic imaging, where the ground-based measurements are combined with the occultation profile employing optimal estimation.

The retrieval algorithm was tested with the aid of simulated data in different scenarios. We can conclude that the presented retrieval algorithm is capable to reproduce realistic atmospheric features, like secondary water vapor maxima near the top of the tradewind inversion.

In areas with low absolute humidities, the retrieval result is significantly affected by the accuracy of the required a priori temperature and pressure profiles. A procedure like the IWV adjustment (presented in sub-section 6.7.1) should be employed if the used a priori temperature and pressure profile is expected to be inaccurate. But even in areas with low absolute humidities, like in the Finland example presented in the work, useful two-dimensional information can be obtained with the optimal estimation approach.

In areas with high absolute humidities, variations of the water vapor density are generally less pronounced, the occultation-derived profile is less sensitive to errors in the a priori temperature and pressure profile, and the retrieval results are generally of good quality. We are confident that the proposed methodology will find its way for fruitful application to genuine data and thus contribute to the provision of much needed information on regional and global water vapor distribution.

REFERENCES

- Ahmad, B., and G.L. Tyler,** The two-dimensional resolution kernel associated with retrieval of ionospheric and atmospheric refractivity profiles by Abelian inversion of radio occultation data, *Radio Sci.*, **33**, 129-142, 1998.
- Alber, C., R. Ware, C. Rocken, and F. Solheim,** GPS surveying with 1 mm precision using corrections for atmospheric slant path delay, *Geophys. Res. Lett.*, **24**, 1859-1862.
- Askne, J., and H. Nordius,** Estimation of tropospheric delay for microwaves from surface weather data, *Radio Sci.*, **22**, 379-386, 1987.
- Atkins, P.W.,** *Physikalische Chemie* (in German), VCH Verlagsgesellschaft, Weinheim, Germany, 1990.
- Bevis, M., S. Businger, T.A. Herring, C. Rocken, R.A. Anthes, and R.H. Ware,** GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System, *J. Geophys. Res.*, **97**, 15 787-15 801, 1992.
- Bevis, M., S. Businger, S. Chiswell, T.H. Herring, R.A. Anthes, C. Rocken, and R.H. Ware,** GPS meteorology: Mapping zenith wet delays onto precipitable water, *J. Appl. Met.*, **33**, 379-386, 1994.
- Bony, S., and J.-P. Duvel,** Influence of the vertical structure of the atmosphere on the seasonal variation of precipitable water and greenhouse effect, *J. Geophys. Res.*, **99**, 12 963-12 980, 1994.
- Born, M., and E. Wolf,** *Principles of Optics*, Pergamon Press, Oxford, UK, 1993.
- Broecker, W.S.,** Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO₂ upset the current balance? *Science*, **278**, 1582-1588, 1997.
- Bronstein, I.N., and K.A. Semendjajew,** *Taschenbuch der Mathematik*, Teubner Verlag, Leipzig, Germany, 1989.
- Brunner, F.K., and M. Gu,** An improved model for the dual frequency ionospheric correction of GPS observations, *Manuscripta Geodaetica*, **16**, 205-214, 1991.
- Bucci Mockler, S.,** Water vapor in the climate system, *AGU special report*, 1995.
- Bürki, B., H. Hirter, M. Cocard, and H.G. Kahle,** Mikrowellen-Wasserdampf-Radiometrie und deren Anwendung in der Geodäsie (in German), Bericht 234 des

- Instituts für Geodäsie und Photogrammetrie der Eidgenössischen Technischen Hochschule, Zürich, 1994.
- Businger, S., S.R. Chiswell, M. Bevis, J. Duan, R.A. Anthes, C. Rocken, R.H. Ware, M. Exner, T. van Hove, and F.S. Solheim**, The promise of GPS in atmospheric monitoring, *Bull. of the American Meteorol. Soc.*, **77**, 5-18, 1996.
- Chahine, M.T.**, The hydrological cycle and its influence on climate, *Nature*, **359**, 373-380, 1992.
- Coakly, J.A.**, Cloud climatology, in *Encyclopedia of climate and weather*, edited by S.H. Schneider, Oxford University Press, NY, 156-162, 1996.
- CRC press**, *Handbook of Chemistry and Physics*, 74th edition, edited by D.R. Lide, CRC press, Boca Raton, Florida, 1993.
- Davis, J.L., T.A. Herring, I.I. Shapiro, A.E.E. Rogers, and G. Elgered**, Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimation of baseline length, *Radio. Sci.*, **20**, 1 593-1 607, 1985.
- Davis, J.L., G. Elgered, A.E. Niell, and C.E. Kuehn**, Ground-based measurements of gradients in the “wet” radio refractivity of air, *Radio Sci.*, **28**, 1 003-1 018, 1993.
- Del Genio, A.D., A.A. Lacis, and R.A. Ruedy**, Simulations of the effect of a warmer climate on atmospheric humidity, *Nature*, **351**, 382-384, 1991.
- Del Genio, A.D., W. Kovari, and M.-S. Yao**, Climatic implications of the seasonal variation of upper tropospheric water vapor, *Geophys. Res. Lett.*, **21**, 2 701-2 704, 1994.
- Dixon, T.H.**, An introduction to the Global Positioning System and some geological applications, *Rev. Geophys.*, **29**, 249-276, 1991.
- Doviak, R.J., and D.S. Zrnic**, *Doppler radar and weather observations*, Academic Press, London, UK, 1984.
- Duan, J., M. Bevis, P. Fang, Y. Bock, S. Chiswell, S. Businger, C. Rocken, F. Solheim, T. van Hove, R. Ware, S. McClusky, T.A. Herring, and R.W. King**, GPS meteorology: Direct estimation of the absolute value of precipitable water, *J. Appl. Met.*, **35**, 830-838, 1996.
- Elgered, G.**, Tropospheric radio path delay from ground-based microwave radiometry, in *Atmospheric remote sensing by microwave radiometry*, edited by M.A. Janssen, Wiley & Sons, NY, 215-258, 1993.

- Elgered, G., J.M Johansson, B.O. Rönnäng, and J.L. Davis**, Measuring regional atmospheric water vapor using the Swedish permanent GPS network, *Geophys. Res. Lett.*, **24**, 2 663-2 666, 1997.
- Emardson, T.R., G. Elgered, and M. Johansson**, Three months of continuous monitoring of atmospheric water vapor with a network of Global Positioning System receivers, *J. Geophys. Res.*, **103**, 1 807-1 820, 1998.
- ESA**, Atmospheric profiling mission (Rep. for assessment on a candidate Earth Explorer mission), ESA SP-1196(7), ESA/ESTEC, Noordwijk, NL, 1996a.
- ESA**, Atmospheric chemistry mission (Rep. for assessment on a candidate Earth explorer mission), ESA SP-1196(6), ESA/ESTEC, Noordwijk, NL, 1996b.
- Eshleman, V.R.**, The radio occultation method for the study of planetary atmospheres, *Planet. Space Sci.*, **21**, 1 521-1 531, 1973.
- Fehmers, G.**, Tomography of the Ionosphere, Ph.D. thesis, Eindhoven University of Technology, The Netherlands, 1996.
- Foelsche, U., and G. Kirchengast**, Space-time resolution conditions for atmospheric imaging involving GNSS occultation: A quantitative study, Techn. Report for ESA/ESTEC-No.3/1997, 38pp., Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1997.
- Fougere, P.F.**, Ionospheric radio tomography using maximum entropy: 1. Theory and simulations, *Radio Sci.*, **30**, 429-444, 1995.
- Fjeldbo, G., and V.R. Eshleman**, The bistatic radar-occultation method for the study of planetary atmospheres, *J. Geophys. Res.*, **70**, 3 217-3 225, 1965.
- Fjeldbo, G., and V.R. Eshleman**, The atmosphere of Mars analyzed by integral inversion of the Mariner IV occultation data, *Planet. Space Sci.*, **16**, 1 035-1 059, 1968.
- Fjeldbo, G., V.R. Eshleman, O.K. Garriott, and F.L Smith III**, The two-frequency bistatic radar-occultation method for the study of planetary ionospheres, *J. Geophys. Res.*, **70**, 3 701-3 710, 1965.
- Fjeldbo, G., A.J. Kliore, and V.R. Eshleman**, The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments, *Astron. J.*, **76**, 123-140, 1971.
- Fjeldbo, G., A.J. Kliore, D. Sweetnam, P. Esposito, B. Seidel, and T. Howard**, Occultation of Mariner 10 by Mercury, *Icarus*, **29**, 439-444, 1976.

- Fjeldbo, G., D. Sweetnam, J. Brenkle, E. Christensen, D. Farless, J. Mehta, B. Seidel, W. Michael, A. Wallio, and M. Grossi**, Viking radio occultation measurements of the Martian atmosphere and topography: Primary mission coverage, *J. Geophys. Res.*, **82**, 4 317-4 324, 1977.
- Fließbach, T.**, *Elektrodynamik, Lehrbuch zur Theoretischen Physik II* (in German), Spektrum Lehrbuch, Spektrum Akademischer Verlag, Heidelberg-Berlin-Oxford, 1997.
- Gaffen, D.J., W.P. Elliot, and A. Robock**, Relationships between tropospheric water vapor and surface temperature as observed by radiosondes, *Geophys. Res. Lett.*, **19**, 1 839-1 842, 1992a.
- Gaffen, D.J., A. Robock, and W.P. Elliot**, Annual cycles of tropospheric water vapor, *J. Geophys. Res.*, **97**, 18 185-18 193, 1992b.
- Gorbunov, M.E.**, Solution of inverse problems of remote atmospheric refractometry on limb paths, *Atmos. and Oceanic Physics*, **26**, 86-91, 1990.
- Gorbunov, M.E.**, Three-dimensional satellite refractive tomography of the atmosphere: Numerical simulation, *Radio Sci.*, **31**, 95-104, 1996.
- Gorbunov, M.E., and S.V. Sokolovskiy**, Remote sensing of refractivity from space for global observations of atmospheric parameters, Report No. 119, Max-Planck Institute for Meteorology, Hamburg, Germany, 1993.
- Gorbunov, M.E., and A.S. Gurvich**, Microlab-1 experiment: Multipath effects in the lower troposphere, *J. Geophys. Res.*, **103**, 13 819-13 826, 1998.
- Gorbunov, M.E., S.V. Sokolovskiy, and L. Bengtsson**, Space refractive tomography of the atmosphere: Modelling of direct and inverse problems, Rep. No. 210, Max-Planck Institute for Meteorology, Hamburg, Germany, 1996a.
- Gorbunov, M.E., A.S. Gurvich, and L. Bengtsson**, Advanced algorithms of inversion of GPS/MET satellite data and their application to reconstruction of temperature and humidity, Rep. No. 211, 40pp., Max-Planck Institute for Meteorology, Hamburg, Germany, 1996b.
- GRAS-SAG**, The GRAS instrument on METOP, Report of the ESA/EUMETSAT GNSS Receiver for Atmospheric Sounding Science Advisory Group, ESA/ESTEC, Noordwijk, NL, 1998.
- Gurvich, A.S., and T.G. Krasil'nikova**, Navigation satellites for radio sensing of the Earth's atmosphere, *Sov. J. Remote Sensing*, **7(6)**, 1 124-1 131, 1990.

- Hafner, J.**, Refraktivitätsfeld der unteren Atmosphäre: Physikalische Grundlagen und ein globaler Überblick über den Radiofrequenzbereich (in German), M.Sc. thesis, Inst. for Meteorology and Geophysics, Univ. of Graz, Austria, 1997.
- Hajj, G.A., R. Ibañez-Meier, E.R. Kursinski, and L.J. Romans**, Imaging the ionosphere with the Global Positioning System, *I. J. Imaging Systems and Technology*, **5**, 174-184, 1994a.
- Hajj, G.A., E.R. Kursinski, W.I. Bertiger, L.J. Romans, and K.R. Hardy**, Assessment of GPS occultations for atmospheric profiling, in *Proceedings of the 7th Conference on Satellite Meteorology and Oceanography*, Monterey, CA, American Meteorological Society, pp J7-J10, 1994b.
- Hajj, G.A., E.R. Kursinski, W.I. Bertiger, S.S. Leroy, T.K. Meehan, L.J. Romans, and J.T. Schofield**, Initial results of GPS-LEO occultation measurements of Earth's atmosphere obtained with the GPS-MET experiment, in *Proceedings of the IAG Symposium G1 „GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications*, IUGG XXI General Assembly, Boulder, CO, Jul. 2-14, 1995.
- Handel, M.D.**, Water vapor, in *Encyclopedia of climate and weather*, edited by S.H. Schneider, Oxford University Press, NY, 823-829, 1996.
- Hanel, R.A., B.J. Conrath, D.E. Jennings, and R.E. Samuelson**, *Exploration of the solar system by infrared remote sensing*, Cambridge planetary science series, **7**, Cambridge university press, Cambridge, 1992.
- Hardy, W.E., and J.L. VanHove**, Microwave Humidity Sounder (MHS): A new generation of instrument, Proceedings of the 45th congress of the International Astronautical Federation, October 9-14, 1994, Jerusalem/Israel, 1995.
- Hardy, K.R., D.P. Hinson, G.L. Tyler, and E.R. Kursinski**, Atmospheric profiles from active space-based radio measurements, *Preprint of the 6th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Boston, MA, 420-423, 1992.
- Hardy, K.R., G.A. Hajj, and E.R. Kursinski**, Accuracies of atmospheric profiles obtained from GPS occultations, *I. J. of Satellite Communications*, **12**, 463-473, 1994.
- Hedin, A.E.**, MSIS-86 thermosphere model, *J. Geophys. Res.*, **92**, 4 649, 1987.
- Hedin, A.E.**, Extension of the MSIS thermosphere model into the middle and lower atmosphere, *J. Geophys. Res.*, **96**, 1159-1172, 1991.

- Hinson, D.P., and J.A. Magalhaes**, Equatorial waves in the stratosphere of Uranus, *Icarus*, **94**, 64-91, 1991.
- Hocke, K.**, Inversion of GPS meteorology data, *Ann. Geophys.*, **14**, 443-450, 1997.
- Hocke, K., G. Kirchengast, and A. Steiner A.**, Ionospheric correction and inversion of GNSS occultation data: Problems and solutions, Techn. Report for ESA/ESTEC-No.2/1997, 35pp., Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1997.
- Høeg, P., A. Hauchecorne, G. Kirchengast, S. Syndergaard, B. Belloul, R. Leitinger, and W. Rothleitner**, Derivation of atmospheric properties using a radio occultation technique, DMI Scientific Report for ESA/ESTEC, No. 4/1995, 208pp., Danish Met. Inst., Copenhagen, Denmark, 1995.
- Hofman-Wellenhof, B., H. Lichtenegger, and J. Collins J.**, *GPS theory and practice*, Springer-Verlag, Vienna, Austria, 1994a.
- Hofman-Wellenhof, B., G. Kienast, and H. Lichtenegger H.**, *GPS in der Praxis*, Springer-Verlag, Vienna, Austria, 1994b.
- Houghton, J.T., L.G. Meira Filho, B.A. Challander, N. Harris, A. Kattenberg, and K. Maskell**, *Climate Change, The science of climate change*, Cambridge Univ. Press, 1996.
- IGS**, International GPS Service for Geodynamics, 1994 Annual Report, JPL, Pasadena/CA, 1995.
- Jackson, J.D.**, *Klassische Elektrodynamik* (in German), de Gruyter, Berlin, Germany, 1985.
- Janssen, M.A. (Ed.)**, *Atmospheric remote sensing by microwave radiometry*, Wiley & Sons, NY, 1993
- Jeske, H.**, Meteorological optics and radiometeorology, in *Landolt-Börnstein, Numerical data and functional relationships in science and technology, Group V, Vol. 4, Meteorology, Subv. b, Physical and Chemical Properties of the Air*, edited by G. Fischer, Springer, Berlin, Heidelberg, 187-348, 1988.
- Jones, D.S.**, *Acoustic and electromagnetic waves*, Clarendon press, Oxford, 1986.
- Karayel, E.T., and D.P. Hinson**, Sub-Fresnel-scale vertical resolution in atmospheric profiles from radio occultation, *Radio Sci.*, **32**, 411-423, 1997.
- Kelly K.K., A.F. Tuck, and T. Davies**, Wintertime asymmetry of upper tropospheric water between the Northern and Southern Hemispheres, *Nature*, **353**, 244-247, 1991.

- Kelso, T.**, Models for Propagation of NORAD Element Sets (F.R. Hoots and R.L. Roehlich); Package by T.S. Kelso, Spacetrack Report No. 4, Dep. of Commerce/Nat'l Tech. Information Service, Springfield/Va, 1980/1988.
- Kirchengast, G.**, End-to-end GNSS Occultation Performance Simulator (EGOPS) functionality definition, Techn. Report for ESA/ESTEC-No.1/1996, 25pp. Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1996.
- Kirchengast, G.**, End-to-end GNSS Occultation Performance Simulator overview and exemplary applications, Wissenschaftl. Ber. No. 2/1998, 142pp., Inst. Meteorol. Geophys., Univ. Graz, Austria, 1998.
- Kirchengast, G., and H.P. Ladreiter**, The potential of the radio-occultation technique based on GPS/GLONASS-signals for determining fundamental atmospheric parameters (in German), *Kleinheubacher Berichte*, **39**, 677-686, 1996.
- Kirchengast, G., J. Ramsauer, J. Skakkebaek, and K. Schultz**, Atmospheric Sounding with GNSS Occultation, User Requirements/Software Requirements Document, CRI /ATMSOU/URD-SRD, 1997.
- Kliore, A.J., G. Fjeldbo, B.L. Seidel, D.N. Sweetnam, T.T. Sesplaukis, P.M. Woiceshyn, and S.I. Rasool**, The atmosphere of Io from Pioneer 10 radio occultation measurements, *Icarus*, **24**, 407-410, 1975.
- Kuo, Y.-H., X. Zou, and Y.-R. Guo**, Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model, *Mon. Wea. Rev.*, **124**, 122-147, 1996.
- Kursinski, E.R.**, The GPS radio occultation concept: Theoretical performance and initial results, Ph.D. thesis, Caltech, Ca, 1997.
- Kursinski, E.R., G.A. Hajj, K.R. Hardy, L.J. Romans and L.T. Schofield**, Observing tropospheric water vapor by radio occultation using the Global Positioning System, *Geophys. Res. Lett.*, **22**, 2 365-2 368, 1995.
- Kursinski, E.R., G.A. Hajj, W.I. Bertiger, S.S. Leroy, T.K. Meehan, L.J. Romans, J.T. Schofield, D.J. McCleese, W.G. Melbourne, C.L. Thornton, T.P. Yunck, J.R. Eyre, and R.N. Nagatani**, Initial results of radio occultation observations of Earth's atmosphere using the Global Positioning System, *Science*, **271**, 1107-1110, 1996.
- Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, and K.R. Hardy**, Observing Earth's atmosphere with radio occultation measurements using GPS, *J. Geophys. Res.*, **102**, 23 429-23 465, 1997.

- Laube, M., and H. Höller**, Thermodynamics of dry air, water, and water vapor, in *Landolt-Börnstein, Numerical data and functional relationships in science and technology, Group V, Vol. 4, Meteorology, Subv. b, Physical and Chemical Properties of the Air*, edited by G. Fischer, Springer, Berlin, Heidelberg, 28-41, 1988.
- Legates D.R.**, Precipitation, in *Encyclopedia of climate and weather*, edited by S.H. Schneider, Oxford University Press, NY, 608-612, 1996.
- Leitinger, R.**, Tomography, in *Modern Ionospheric Science*, edited by H. Kohl, R. Rüster, and K. Schlegel, EGS, Katlenburg-Lindau, 346-370, 1996.
- Leitinger, R., H.P. Ladreiter, and G. Kirchengast**, Ionosphere tomography with data from satellite reception of Global Navigation Satellite System signals and ground reception of Navy Navigation Satellite System signals, *Radio Sci.*, **32**, 1 657-1 669, 1997a.
- Leitinger, R., H.P. Ladreiter, and G. Kirchengast**, Ionosphere tomography combining data from spaceborne GNSS and ground-based NNSS reception, Techn. Report for ESA/ESTEC-No.1/1997, 25pp., Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1997b.
- Leitinger, R., H.P. Ladreiter, and G. Kirchengast**, Ionosphere tomography involving real spaceborne GNSS (GPS/MET) and ground-based NNSS Data, Techn. Report for ESA/ESTEC-No.4/1997, Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1997c.
- Liljequist, G.H. and K. Cihak**, *Allgemeine Meteorologie*, 3. Auflage, Vieweg & Sohn, Braunschweig-Wiesbaden, 1984.
- Lindal, G.F.**, The atmosphere of Neptune: An analysis of radio occultation data acquired with Voyager 2, *Astron. J.*, **103**, 967-982, 1992.
- Lindal, G.F., H.B. Hotz, D.N. Sweetnam, Z. Shippony, J.P. Brenkle, F.V. Hartsell, and R.T. Spear**, Viking radio occultation measurements of the atmosphere and topography of Mars: Data acquired during 1 Martian year of tracking, *J. Geophys. Res.*, **84**, 8 443-8 456, 1979.
- Lindal, G.F., G.E. Wood, G.S. Levy, J.D. Anderson, D.N. Sweetnam, H.B. Hotz, B.J. Buckles, D.P. Holmes, P.E. Doms, V.R. Eshleman, G.L. Tyler, and T.A. Croft**, The atmosphere of Jupiter: An analysis of the Voyager radio occultation measurements, *J. Geophys. Res.*, **86**, 8 721-8 727, 1981.

- Lindal, G.F., G.E. Wood, H.B. Hotz, D.N. Sweetnam, V.R. Eshleman, and G.L. Tyler,** The atmosphere of Titan: An analysis of the Voyager 1 radio occultation measurements, *Icarus*, **53**, 348-363, 1983.
- Lindal, G.F., D.N. Sweetnam, and V.R. Eshleman,** The atmosphere of Saturn: An analysis of the Voyager radio occultation measurements, *Astron. J.*, **90**, 1 136-1 146, 1985.
- Lindal, G.F., J.R. Lyons, D.N. Sweetnam, V.R. Eshleman, D.P. Hinson, and G.L. Tyler,** The atmosphere of Uranus: Results of radio occultation measurements with Voyager 2, *J. Geophys. Res.*, **92**, 14 987-15 001, 1987.
- Lindal, G.F., J.R. Lyons, D.N. Sweetnam, V.R. Eshleman, D.P. Hinson, and G.L. Tyler,** The atmosphere of Neptun: Results of radio occultation measurements with the Voyager 2 spacecraft, *Geophys. Res. Lett.*, **17**, 1733-1736, 1990.
- Linfield, R.P., S.J. Keihm, L.P. Teitelbaum, S.J. Walter, M.J. Mahony, R.N. Treuhaft, and L.J. Skjerve,** A test of water vapor radiometer-based troposphere calibration using very long baseline interferometry observations on a 21-km baseline, *Radio Sci.*, **31**, 129-146, 1996.
- Liou, K.N.,** *An introduction to atmospheric radiation*, International Geophysics series, **26**, Academic press, San Diego, 1980.
- Lusignan, B., G. Modrell, A. Morrison, J. Pomalaza, and S.G. Ungar,** Sensing the Earth's atmosphere with occultation satellites, *Proc. Of the IEEE*, **4**, 458-467, 1969.
- Melbourne, W.G., E.S. Davis, C.B. Duncan, G.A. Hajj, K.R. Hardy, E.R. Kursinski, T.K. Meehan, L.E. Young, and T.P. Yunck T.P.,** The application of spaceborne GPS to atmospheric limb sounding and global change monitoring, *JPL Publication 94-18*, 147pp., Jet Propulsion Lab., Pasadena, CA, 1994.
- Menke, W.,** *Geophysical data analysis: Discrete inverse theory*, International Geophysics Series, Vol. 45, Academic Press, San Diego, CA, 1989.
- Moore, W.J., and D.O. Hummel,** *Physikalische Chemie* (in German), de Gruyter, Berlin, 1973.
- Mortensen, M.D., and P. Høeg,** Inversion of GPS occultation measurements using Fresnel diffraction theory, *Geophys. Res. Lett.*, **23**, 2 441-2 444, 1998.
- Niell, A.E.,** Global mapping functions for the atmosphere delay at radio wavelengths, *J. Geophys. Res.*, **101**, 3 227-3 246, 1996.

- Peixoto, J.P., and A.H. Oort**, *Physics of climate*, American Inst. of Physics, NY, 1992.
- Phinney, R.A., and D.L. Anderson D.L.**, On the radio occultation method for studying planetary atmospheres, *J. Geophys. Res.*, **73**, 1 819-1 827, 1968.
- Potti, J., and J.C. Carmona**, Study of autonomous orbit and attitude determination techniques for Low-Earth-Orbit observation systems, *Techn. Report No. 4 on Visibility Analysis for Atmospheric Profiling by Radio Occultation*, ESTEC techn. Report, 1995.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery**, *Numerical recipes in FORTRAN: the art of scientific computing*, Cambridge Univ. Press, Cambridge, UK, 1992.
- Purcell, E.M.**, *Elektrizität und Magnetismus*, Berkely Physik Kurs, **2**, Vieweg + Sohn, Braunschweig, 1984.
- Rahmstorf, S.**, Rapid climate transitions in a coupled ocean-atmosphere model, *Nature* **372**, 1994, 82-85, 1994.
- Ramsauer, J., and G. Kirchengast**, Impact of an Atmospheric Profiling Mission on NWP: Review of Sampling Requirements and and Mission Scenario Study, Techn. Report for ESA/ESTEC No. 1/1999, 60pp., Inst. for Meteorol. Geophys., Univ. of Graz, Austria, 1999.
- Raschke, E. (Ed.)**, *Radiation and water in the climate system, remote measurements*, NATO ASI Series I, Springer-Verlag, Berlin-Heidelberg-New York, 1996.
- Rocken, C., R. Ware, T. van Hove, F. Solheim, C. Alber, J. Johnson, and M. Bevis**, Sensing atmospheric water vapor with the Global Positioning System, *Geophys. Res. Lett.*, **20**, 2 631-2 634, 1993.
- Rocken, C., T. van Hove, J. Johnson, F. Solheim, R. Ware R., M. Bevis, S. Chiswell, and S. Businger**, GPS/STORM - GPS sensing of atmospheric water vapor for meteorology, *J. of Atmosph. and Oceanic Technology*, **12**, 468-478, 1995.
- Rocken, C., T. van Hove, R. Ware**, Near real-time GPS sensing of atmospheric water vapor, *Geophys. Res. Lett.*, **24**, 3221-3224, 1997.
- Rodgers, C.D.**, *Inverse methods for atmospheric sounding: Theory and practise*, (to be published, available on-line from <http://www.atm.ox.ac.uk/user/rodgers/>), cited 1998.
- Ross, R.J., and S. Rosenfeld**, Estimating mean weighted temperature of the atmosphere for Global Positioning System applications, *J. Geophys. Res.*, **102**, 21 719-21 21730, 1997.

- Rossow, W.B.**, Remote sensing of atmospheric water vapor, in *Radiation and water in the climate system, remote measurements*, NATO ASI Series I, Springer-Verlag, edited by E. Raschke, Berlin-Heidelberg-New York, 175-191, 1996.
- Rothacher, M.**,: Orbits of satellite systems in space geodesy, *Geodätisch-geophysikalische Arbeiten in der Schweiz*, **46**, 1992.
- Saastamoinen, J.**, Introduction to practical computation of astronomical refraction, *Bulletin Géodésique*, **106**, 383-397, 1972.
- Saastamoinen, J.**, Contributions to the theory of atmospheric refraction, *Bulletin Géodésique*, **107**, 13-34, 1973.
- Salby, M.L.**, *Fundamentals of Atmospheric Physics*, Academic Press, San Diego, CA, 1996.
- Schneider, S.H.**, *Encyclopedia of climate and weather*, Oxford University Press, New York-Oxford, 1996.
- Smith, E.K., and S. Weintraub**, The constants in the equation for atmospheric refractive index at radio frequencies, *Proceedings of the I.R.E.*, **41**, 1 035-1 037, 1953.
- Steiner, A.K.**, High resolution sounding of key climate variables using the radio occultation technique (Ph.D. thesis), Wissenschaftl. Ber. No. **3/1998**, 170pp., Inst. for Meteorol. and Geophys., Univ. of Graz, Austria, 1998.
- Steiner, A.K., G. Kirchengast, and H.P. Ladreiter**, Inversion, error analysis, and validation of GPS/MET occultation data, *Ann. Geophys.*, **17**, 122-138, 1999.
- Storch, H.v., and F.W. Zwiers**, Statistical analysis in climate research, Cambridge Univ. Press (in press), 528pp., 1999.
- Syndergaard, S.**, Modeling the impact of the Earth's oblateness on the retrieval of temperature and pressure profiles from limb sounding, *J. Atmos. Terr. Phys.*, **60**, 171-180, 1998.
- Thayer, G.D.**, An improved equation for the radio refractive index of air, *Radio Sci.*, **9**, 803-807, 1974.
- Tralli, D.M., and S.M. Lichten**, Stochastic estimation of tropospheric path delays in Global Positioning System geodetic measurements, *Bulletin Géodésique*, **64**, 127-159, 1990.

- Tregoning, P., R. Boers, D. O'Brien, and M. Handy,** Accuracy of absolute precipitable water vapor estimates from GPS observations, *J. Geophys. Res.*, **103**, 28 701-28 710, 1998.
- Treuhaff, R.N., and G.E. Lanyi,** The effect of the dynamic wet troposphere on radio interferometric measurements, *Radio Sci.*, **22**, 251-265, 1987.
- Volkov, A.A., G.M. Grechko, A.S. Gurvich, V. Kann, and S.V. Sokolovskiy,** Recovery of the vertical temperature profile of the atmosphere from refraction measurements made from the Salyut-7 station, *Phys. of Atmos. and Oceans*, **23**, 914-915, 1987.
- Vorob'ev, V.V., and T.G. Krasil'nikova,** Estimation of the accuracy of the atmospheric refractive index recovery from Doppler shift measurements at frequencies used in the NAVSTAR system, *Phys. of Atmos. and Oceans*, **29**, 602-609, 1994.
- Ware, R., C. Rocken, F. Solheim, T. van Hove, C. Alber, and J. Johnson J.,** Pointed water vapor radiometer corrections for accurate Global Positioning System surveying, *Geophys. Res. Lett.*, **20**, 2 635-2 638, 1993.
- Ware, R., M. Exner, D. Feng, M.E. Gorbunov, K. R. Hardy, B. Herman, Y.-H. Kuo, T.K. Meehan, W.G. Melbourne, C. Rocke, W. Schreiner, S.V. Sokolovskiy, F. Solheim, X. Zou, R. Anthes, S. Businger, and K. Trenberth,** GPS Sounding of the atmosphere from Low Earth Orbit: Preliminary results, *Bull. Am. Meteorol. Soc.*, **77**, 19-40, 1996.
- Ware, R., C. Alber, C. Rocken, and F. Solheim,** Sensing integrated water vapor along GPS ray paths, *Geophys. Res. Lett.*, **24**, 417-420, 1997.
- Webster, P.J.,** The role of hydrological processes in ocean-atmosphere interactions, *Rev. Geophys.*, **32**, 427-476, 1994.
- Yuan, L.L., R.A. Anthes, R.H. Ware, C. Rocken, W.D. Bonner, M.G. Bevis, and S. Businger,** Sensing climate change using the Global Positioning System, *J. Geophys. Res.*, **98**, 14 925-14 937, 1993.
- Yunck, T.P., G.F. Lindal, and C.H. Liu,** The role of GPS in precise Earth observation, paper presented at the Position, Location, and Navigation Symposium, Orlando, FL, Nov.29-Dec.2, 1988.
- Zernov, N.N., and B. Lundborg,** An integral representation of the wave field in inhomogeneous media in terms of diffracting component waves, *Radio Sci.*, **31**, 67-80, 1996.