Faculty of Sciences
Department of Chemistry
Division of Physical and
Analytical Chemistry

Belgian Institute for
Space Aeronomy

Remote sounding of the atmosphere by the Occultation Radiometer: an original retrieval method and study of the relaxation of Mount Pinatubo stratospheric aerosols

Ph. D. Thesis
Philip Vanhellemont

Promotor
Prof. Dr. Jozef Peeters

Co-promotor
Dr. Didier Fussen

June 2001

Preface.

Although the study of the Earth's atmosphere has a long history, it was not until recent times that the interest in the physical and chemical properties of the atmosphere has grown so strong. Partially responsible for this rapidly growing scientific field are observed phenomena such as the Antarctic ozone hole and the greenhouse effect, which have received considerable attention in the scientific world and the media. On the other hand, technological progress has provided much better means to observe (satellite and balloon instruments, ground stations) and explain (computer models) the nature of the atmosphere.

More specifically, the birth of space technology has led to convenient ways for the exploration of the atmosphere in a fast and global manner. The main body of this thesis is formed by the processing and interpretation of data obtained from such an experiment. The *Occultation Radiometer* ORA was designed at the Belgian Institute for Space Aeronomy (BISA) to measure concentrations of ozone chemistry-related trace gases (O_3, NO_2, H_2O) and aerosols in the stratosphere, as well as O_3 concentrations in the mesosphere.

ORA is a fairly simple radiometer, equipped with 8 broadband detector channels in the UV/visible/near IR wavelength region. It was mounted on the European Retrievable Carrier EURECA, a space platform commissioned by the European Space Agency ESA to house a diversity of scientific experiments. In the period from August 1992 to May 1993, while in orbit, ORA measured the solar light intensity, attenuated and refracted by the Earth's atmosphere during subsequent orbital sunsets and sunrises. In the following period, much effort was put into the

construction of retrieval methods to derive the concentrations and other properties of the above-mentioned atmospheric species.

Entirely unexpected, in June 1991, during the final preparations for the mission, the violent eruption of Mount Pinatubo in the Philippines provided the largest stratospheric aerosol loading since the 1883 Krakatau eruption. While the ORA design was not prepared for such a disturbance, the eruption led to a major challenge, offering an exceptional opportunity to study the transport and dynamics of volcanic aerosol populations in the stratosphere, and a large part of our research was dedicated to this subject.

This thesis is structured in a more or less chronological way. We start from what is known: Chapter 1 gives a brief summary of the current knowledge on the physical and chemical properties of the atmosphere. More specifically, stratospheric chemistry, aerosols and their interrelation have been given enough attention to emphasize why experiments such as ORA are designed in the first place. This naturally leads to Chapter 2, which describes ORA, its carrier EURECA, and the general methodology of the occultation technique. Important details concerning the ORA experimental implementation are discussed, and the mission history and post-mission findings are presented.

Chapter 3 provides the basic physical theories needed to describe the interaction of light with atmospheric species. Solar light is attenuated in the atmosphere as a result of absorption and scattering by gas molecules and particles, and experiences refraction. These processes can be quantified by theoretical results obtained from quantum mechanics and electromagnetic wave theory.

Every remote sensing experiment has to face a so-called inverse problem: given the measurements, what was the state of the atmosphere when the incoming light interacted with the gas molecules and particles, before entering the detector channels of the experiment? In other words: how can we retrieve the properties (density, particle sizes) of the atmospheric species from the measurements? The answer is non-trivial, and usually one has to face additional problems that are related to the limited *information content* of the measurements.

Actually, the inverse problem can be subdivided in two partial problems. The first problem arises because measured light intensities are attenuated by species located at different altitudes in the atmosphere.

The associated *spatial inverse problem* is discussed in Chapter 4, together with the results and a comparison with results from another space experiment. Chapter 5 describes the *spectral inverse problem*: the light attenuation at a certain wavelength is composed of contributions from the different atmospheric species. A method to separate these contributions is discussed, together with the results.

Given the fact that ORA measured during an extraordinary time period (one year after the Mt. Pinatubo eruption), we have focussed our attention primarily on the stratospheric aerosols (of volcanic origin). Chapter 6 introduces a third inverse problem, solved by a method that is able to extract information on the density and size of these stratospheric particles. The results obtained have been interpreted in terms of atmospheric transport and particle dynamics, and important new insights concerning these issues have been gained.

Acknowledgements

In the final stage of my writings, I feel the need to thank many people that were, one way or another, involved in the experimental design of ORA, the subsequent data processing and the associated scientific thinking.

Let me start by thanking Prof. Dr. Jozef Peeters (promotor) of the Catholic University of Leuven for giving me the opportunity to prepare this text, and for the useful advice, suggestions, interesting discussions and reading.

Before I began working at the Belgian Institute for Space Aeronomy, my knowledge about space research was very limited. In the past few years, the pleasant and stimulating collaboration with Dr. Didier Fussen (co-promotor, data processing, scientific interpretation) has changed this situation thoroughly. The results presented in this thesis would not exist without his contribution, and I am truly grateful.

It goes without saying that the design, construction and testing of a space experiment is far from easy. Quite a few people have spend a considerable amount of their time on these issues. I wish to congratulate Dr. Etienne Arijs (instrument design, principle investigator) and Dr. ir. Dennis Nevejans (electronic design), together with Piet Frederick, Emiel Van Ransbeeck and all other co-workers for the magnificent job that they

have done. I specifically wish to thank Dr. Arijs for some very useful discussions that were necessary during the writing of this text.

My main task has been the processing and the interpretation of the ORA data. I therefore wish to thank Fabienne Leclère (data collection, telemetry/telecommand operations) and Johan Bulcke (computer maintenance) for providing me with the necessary digital food. I also thank Christine Bingen, for some very clarifying conversations about used methods and the comparison of the ORA and SAGE II data.

The mental conception of ORA started in the 80s through conversations between Dr. Baron Marcel Ackerman (BISA) and Professor Fred Taylor (University of Oxford, UK), and they should be credited as the initiators of the entire project. This thesis deals primarily with the results from the UV/visible part of ORA. However, the infrared part was fully integrated in the experiment, and existed thanks to the work of the Oxford team: Simon Calcutt, Chris Hepplewhite, Ian Burchell, Clive Rodgers, T.M. Pritchard and S. Werrett.

Essential work in the design and testing of the experiment was done by teams at the European Space Research and Technology Center (ESTEC, Noordwijk, the Netherlands), and many people from ESA were responsible for logistic support and launch activities.

Some external information was used in the data processing stage. I thank Colette Brogniez and Jacqueline Lenoble (University of Lille, France) for the SAGE II data, and the people at the United Kingdom Meteorological Office (UKMO) for some crucial numbers (pressure, temperature, wind velocity).

This work would not have been possible without the financial support and possibilities offered by the Belgian Federal Office for Scientific, Technical and Cultural affairs (OSTC). The same applies to the Belgian Institute for Space Aeronomy, with its current director, Prof. Dr. Paul Simon. My sincere gratitude.

Many thanks to Ruth Timmermans, Lieve Dhaene and especially Maria Desmond for their specific role in the necessary proof-reading of the final text.

To my family and friends (you all know who you are when reading this): thanks for bringing me back to Earth once in a while. Life can be lonely in space.

Brussels, June 2001

Contents

List of Figures

List of Tables

Chapter 1

The Earth's atmosphere

Before we will dive into the ORA experimental description and the complex problem of the data processing, it is necessary that we have a good and clear understanding of the properties and the chemical composition of the atmosphere that surrounds our planet. Therefore, in this first chapter, we will first examine the classification of the different atmospheric layers that is used in meteorology and aeronomy, and offer a brief description of the global atmospheric circulation and the altitude variation of pressure.

A full treatise on the chemical aspects of the atmosphere lies outside the scope of this thesis. We will focus our attention mainly on the chemical species that are relevant for us, that is, the species in the stratosphere that play an important role in the ozone-related chemistry. Stratospheric aerosols occupy a major part of this work, so they are introduced in a separate section, together with the history of the Mount Pinatubo eruption and its consequences for the stratosphere.

Knowledge about the atmosphere is far from complete, and new questions emerge very frequently. Therefore, the chapter concludes with a motivation for the development of new experiments, such as ORA.

1.1 Classification of atmospheric layers

The Earth's atmosphere exhibits properties that are very dependent on altitude. These properties are used to classify the atmosphere in

different layers, or *spheres*. The boundary that separates one layer from another is named *pause*. Three criteria that are commonly used to define the layers are composition, temperature and electrical properties. The classification that is finally used depends on the field of study.

1.1.1 ... according to composition

The largest part of the atmosphere consists of N_2, a gas that is in atmospheric conditions chemically inert, practically insoluble in water, and non-condensable. Since the existence of the Earth, it has accumulated in the atmosphere, and its source has to be sought in volcanic emissions. The second most frequently occurring molecule, oxygen (O_2), was produced through photosynthesis by biological species. O_2 is also relatively inert, and actually finds itself in a dynamic steady state, formed by the removal of O_2 through respiration and oxidation of organic carbon, and O_2 production by photosynthesis. On average, N_2 represents 78 % and O_2 21 % of the atmosphere, while the remaining part is occupied by noble gases such as Ar, and species that are highly variable in concentration, such as water vapor and CO_2. Water vapor is mostly confined to the lower atmospheric regions (the troposphere, see below), and its concentration is mainly determined by the physical equilibrium between evaporation from oceans and surface water, and removal through precipitation.

At extremely low concentrations, we find a large diversity in so-called trace gases. These minor constituents nevertheless have a profound effect on the solar radiation received on Earth, and the radiation coming from Earth itself. Trace gases therefore play a crucial part in atmospheric processes. Due to increasing industrialization and population, the concentration of certain trace gases has been altered significantly, and new environmental problems have arisen, of which the *ozone hole* and the *greenhouse effect* are two well-known examples.

The most abundant atmospheric species are thus N_2 and O_2, and their total concentrations are remarkably constant. Often, the mixture is referred to as the *neutral density*. Below roughly 100 km, the mixing fraction of 78 % to 21 % is constant, and as a consequence, the average molecular mass of air is also constant. The atmospheric layer, ranging from the ground up to 100 km is therefore named the *homosphere*. Above this layer, photodissociation of O_2 becomes important,

so the contribution of atomic oxygen increases. As a result, the average molecular mass of air decreases. This region is labeled the *heterosphere*.

1.1.2 ... according to temperature

The Earth's atmosphere is characterized by a variation in temperature as a function of altitude. The non-monotonic shape of the temperature profile results from the interaction of the molecules (and particles) with radiation. The most important contribution comes from the absorption of solar radiation by atmospheric molecules in the ultraviolet and visible ranges. Also, the heated Earth emits radiation in the infrared, which is absorbed by a variety of molecules that have strong absorption bands in this wavelength region, such as CO_2, H_2O, CFCs (chlorofluorocarbons) , and CH_4 - the *greenhouse gases*. Absorption can be followed by emission of secondary radiation, that can be absorbed again, etc. Through this process, a thermal equilibrium can be established, which presents a net warming of the Earth-atmosphere system. Without the presence of an atmosphere, the surface of the Earth would be roughly 30 degrees colder than it is now (see e.g. [Harries, 1994]).

The resulting temperature profile (Fig. 1.1) gives rise to a frequently used classification of layers, discussed hereafter.

The troposphere

The lowest atmospheric level starts at the ground level and ends at the so-called *tropopause*, which is situated at an altitude of 8 to 18 km, depending on the latitude and the time of year. The troposphere[1] is characterized by a temperature that decreases in an almost linear way with altitude. During the day, the heated Earth surface warms the air at ground level. This air becomes buoyant and starts to rise. This movement (convection) ensures a mixing of air throughout the entire troposphere. Since rising air, in a good approximation, adiabatically expands and cools (pressure goes down), the resulting temperature profile decreases in altitude. At the tropopause, the temperature falls to an average (in latitude and time) of 217 K (-56 degrees Celsius).

[1]The name troposphere was coined by a British meteorologist, Sir Napier Shaw, and is derived from the Greek word *tropos*, meaning *turning*, referring to the ceaseless movement of the air in this layer.

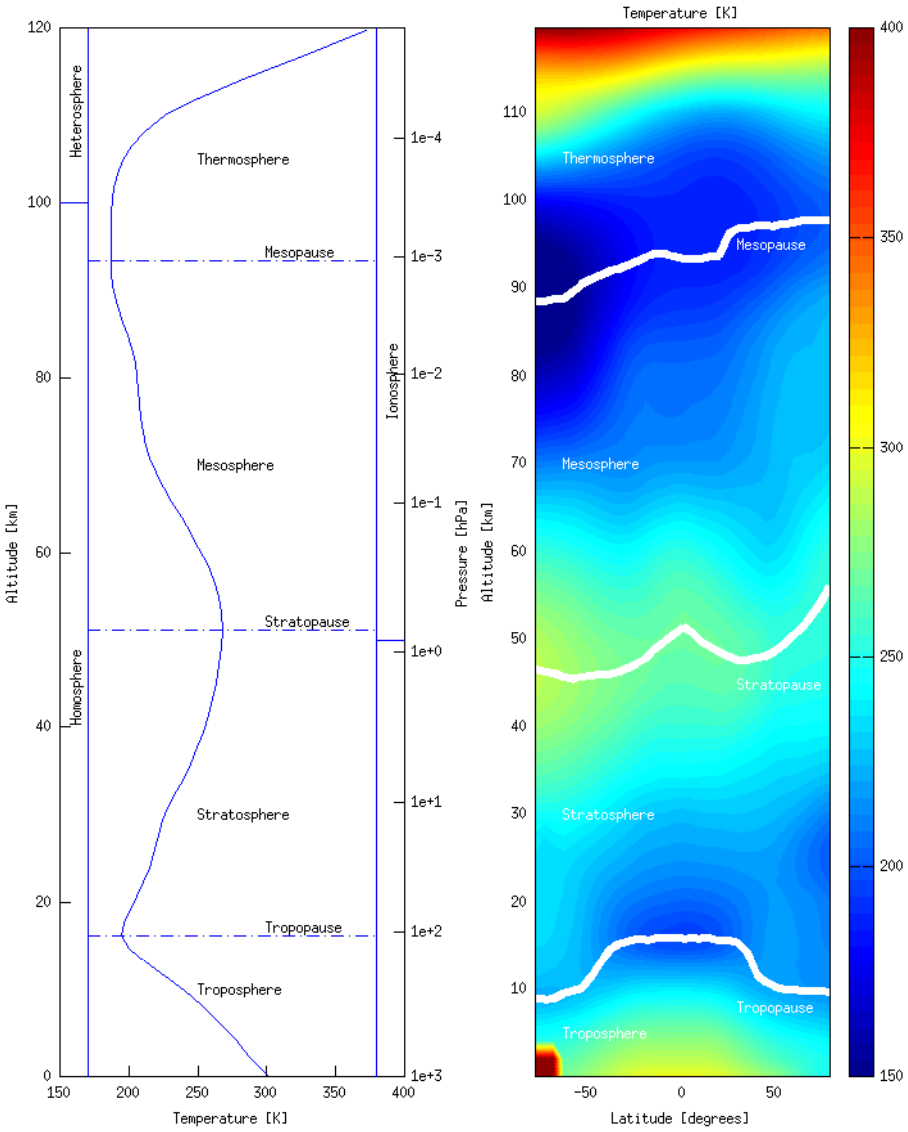

Figure 1.1: The layered structure of the Earth's atmosphere. Temperatures shown represent January values from the CIRA climatology. Left: temperature profile at the equator. Right: temperature vs. latitude and altitude.

Also, the troposphere contains about 80 % of the total atmospheric mass. Almost all of the atmosphere's water vapor is found in the troposphere, where it leads to the formation of clouds.

The actual definition of the tropopause level, according to the convention of the World Meteorological Organization (WMO) is: [Holton et al., 1995]

> The tropopause level is the lowest level at which the rate of decrease of temperature with height (temperature lapse rate) decreases to 2 K km^{-1} or less and the lapse rate averaged between this level and any level within the next two km does not exceed 2 K km^{-1}.

The tropopause height is at its maximum over the tropics, and slopes downward moving to the poles.

Since the tropopause marks the minimum of the temperature profile, convection ceases. There is no process left that can transport air (and chemicals) through the tropopause in an effective way. Therefore, the tropopause can be considered as a barrier for atmospheric constituents.

The stratosphere

The stratosphere[2] starts at the tropopause level and ends at the *stratopause*, which is located at an altitude of approximately 45 to 55 km. An isothermal region exists from 11 to 20 km at midlatitudes, but temperature increases from 20 to 50 km, with a final maximum temperature of 271 K on average. The increase in temperature with height is mainly caused by the presence of the stratospheric ozone layer, which absorbs UV radiation and releases heat in the following photodissociation to O_2 and atomic oxygen, and recombination of the O-atoms with O_2.

As already said, because of the absence of convection, transport of atmospheric species across the tropopause into the stratosphere is very slow[3]. For example, CH_4 (originating in anthropogenic activity) is ho-

[2]The stratosphere was discovered at the beginning of the 20th century by the French meteorologist Léon Philippe Teisserenc de Bort. Contrary to the popular belief of the day, balloon measurements showed that the temperature did not decrease steadily to absolute zero with increasing altitude, but stopped falling and remained constant at about 11 km. The region was named stratosphere, from the Latin word *stratum*, meaning *layer*.

[3]In the long term, diffusion into the stratosphere can take place, but *very* slowly.

mogeneously mixed in the troposphere but concentration levels decrease rapidly across the tropopause.

The mesosphere

Above the stratopause, heating by absorption of solar radiation becomes less important, and so the temperature decreases again with altitude. The mesosphere extends up to the *mesopause*, which is the coldest point in the atmosphere. It is located approximately at an altitude of roughly 90 km. Similar to the troposphere, vertical mixing is rapid because of convection caused by colder layers that are situated above warmer ones.

The thermosphere

Above the mesopause, the temperature once again goes up with altitude, due to the absorption of short wavelength radiation by N_2 and O_2. Depending on solar activity, temperatures can become as high as 2000 Kelvin.

The exosphere

At altitudes above 500 km, in the outermost region of the atmosphere, some gas molecules with sufficient energy can escape from the Earth's gravitational attraction. It is in this region that the Occultation Radiometer (ORA) orbits around the Earth at an altitude of 508 km.

1.1.3 ... according to electrical properties

In the atmosphere, some atoms and molecules can be ionised by solar radiation, cosmic radiation and high-energetic particles (protons, electrons). Above an altitude of 50 km, we define the *ionosphere*, where enough free electrons exist to influence the transmission and reflection of radio waves. In this layer, molecules and atoms such as NO, N_2 and O_2 deliver electrons through ionisation.

At high altitudes, the ionosphere gradually transforms into the *magnetosphere*. This is the region where charged particles are being controlled by the Earth's magnetic field. The magnetosphere is not a sphere in the strict sense, because it has a complex shape that is strongly determined by the interaction of the solar wind with the magnetic field

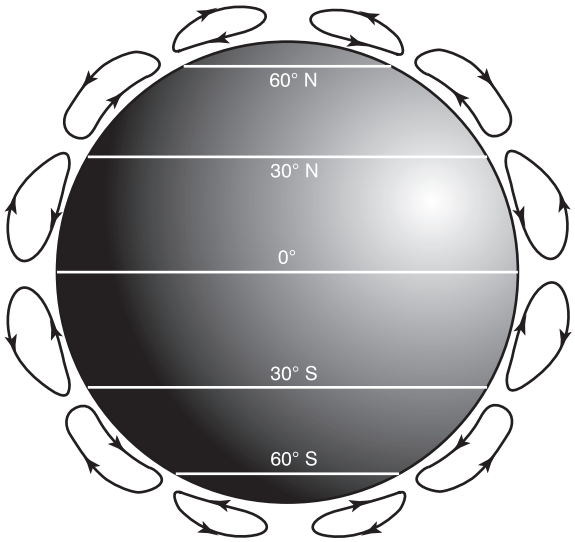

Figure 1.2: Schematic representation of the general circulation in the atmosphere.

of the Earth. Within the magnetosphere, other subregions are defined: the *plasmasphere*, and the famous *Van Allen* radiation belts. A detailed discussion of these regions, however, lies outside the scope of this text.

1.2 Global atmospheric transport

Globally, the Earth-atmosphere system is in radiative balance: the total amount of received solar radiation equals the amount of radiation that is lost to space. Locally, however, this is not true, and parameters such as the latitude, the nature of the surface (land, ocean) and the degree of cloudiness play a role. On average, in equatorial regions solar heating is more important, while in polar regions radiative cooling dominates. It is easy to see that this uneven distribution of energy leads to large-scale air motions: warm tropical air rises while cold polar air descends, with poleward and equatorward flows completing the circulation. In other words: a thermal convection cell.

The actual situation is more complex than this (see Fig. 1.2). The upper poleward flow loses heat through radiative cooling and starts to

descend already at a latitude of about 30°in both hemispheres, resulting in a (smaller) *Hadley cell*. The same situation occurs in the polar regions, in the [60°- 90°] latitude range. At the mid-latitudes, the influence of both polar and tropical cells is felt, resulting in a circulation in the opposite direction.

Furthermore, these air movements are subject to the *Coriolis force*: zonal winds (in the longitudinal direction), cyclones and anti-cyclones are formed.

The general circulation on Fig. 1.2 can be considered as the annual mean circulation of the atmosphere. This averaging period is taken because solar heating is the driving force for the global winds. On longer time scales, the *Quasi-Biennial Oscillation*[4] (QBO) and the *El Niño/Southern Oscillation* (ENSO) phenomena become important.

1.3 Variation of pressure with altitude

In a very good approximation, a volume of air finds itself in static equilibrium. This equilibrium is maintained by two forces that cancel one another: (1) Air located at an altitude z experiences a greater pressure than air at an altitude $z + \mathrm{d}z$, resulting in an upwards force. This force is exactly counterbalanced by (2) gravitational pull downwards. This mechanism is expressed by the *hydrostatic equation*:

$$\frac{\mathrm{d}p(z)}{\mathrm{d}z} = -\rho(z)\mathrm{g}$$

with $p(z)$ and $\rho(z)$ the pressure and mass density of air at altitude z, and g the gravitational acceleration. Also, the ideal gas law states:

$$\rho(z) = \frac{\mathrm{M_{air}}p(z)}{\mathrm{R}T(z)}$$

where $\mathrm{M_{air}}$ is the average molar mass of air (28.97 g mol^{-1}), R the molar gas constant (8.314 J mol^{-1} K^{-1}), and $T(z)$ the temperature (K) at altitude z. Combined, this results in:

$$\frac{\mathrm{d}p(z)}{\mathrm{d}z} = -\frac{\mathrm{M_{air}}\mathrm{g}p(z)}{\mathrm{R}T(z)}$$

[4]QBO is a reversal of wind direction in the stratosphere that occurs on roughly a 26 month cycle.

or, stated otherwise:

$$\frac{\mathrm{d}}{\mathrm{d}z}\ln p(z) = -\frac{1}{H(z)} \tag{1.1}$$

where we have defined the *atmospheric scale height*:

$$H(z) = \frac{\mathrm{R}T(z)}{\mathrm{M_{air}g}}$$

If we approximate T and g to be constant for all altitudes[5], then equation 1.1 can be integrated as:

$$p(z) = p_0 \mathrm{e}^{-\frac{z}{H}}$$

with p_0 the pressure at ground level (approximately 1000 hPa). Thus, the scale height is a quantity that reflects the (e-folding) rate of pressure drop as function of altitude. Typically, its value is in the range of 7 to 8 km.

We can apply the derived expressions to the chemical species separately, if we use partial pressure instead of total pressure. From the definition of $H(z)$, we see that heavy molecules have smaller scale heights than light molecules. This leads us to conclude that the heavy molecules are located more near the surface of the Earth, while the lighter ones have a larger extension up into space. The conclusion is incorrect: atmospheric dynamics ensure a ceaseless mixing in altitude (in the homosphere). Only at very high altitudes (> 120 km) is the conclusion more or less correct, with H and He extending to the top of the atmosphere, and heavy molecules (Ar, Xe, Kr) decaying more rapidly with altitude.

1.4 Stratospheric chemistry

1.4.1 Stratospheric ozone

Ozone (O_3) represents one of the most important trace gases in the atmosphere. About 90 % of all ozone is located in the stratosphere, and frequently it is referred to as the *ozone layer*. Because the gas absorbs strongly in the ultraviolet range, it plays a major role in the

[5] An assumption that is not valid if we want to make accurate calculations, for example in atmospheric spectroscopy!

stratospheric radiative balance. Maybe more important, the presence of stratospheric ozone is crucial for the existence of life on Earth. Without the ozone layer, ultraviolet radiation would reach the surface of the Earth unattenuated, making life impossible. A decrease in present-day ozone levels would have dramatic consequences (skin cancer, effects on phytoplankton, disruptions in the immune system, blindness, genetic mutation).

Ozone was discovered in the 19th century by C.F. Schönbein[6]. At the end of that century, measurements showed that the ozone concentrations in higher atmospheric layers was larger than on the surface. Early 20th century, Fabry and Buisson performed experiments with an optical method in the ultraviolet, and estimated the total amount of ozone in a vertical column to be 3 mm thick when compressed to standard pressure at standard temperature. Also worth mentioning is the British scientist Dobson[7], who developed a spectrophotometer to measure the total ozone column.

In 1930, the British scientist Sydney Chapman proposed a chemical cycle that was able to explain the existence of a stratospheric ozone layer [Chapman, 1930]. In this cycle, ozone was produced by photolysis of O_2 in the stratosphere, and removed only by reaction with atomic oxygen mainly coming from O_3 photodissociation. After a while, it became clear that the model predicted too much ozone, in comparison with observed stratospheric profiles. Extra chemical removal reactions would have to be taken into account. A first effort came from [Bates and Nicolet, 1950], who proposed catalytic cycles involving hydrogen radicals. In 1970, Paul Crutzen suggested that the NO_x (a collective noun for NO and NO_2) played an important part as catalyst for ozone destruction [Crutzen, 1970]. And research showed that NO_x emissions from planned stratospheric aircraft would be damaging for the ozone layer [Johnston, 1971].

Shortly thereafter (1974), Mario Molina and F. Sherwood Rowland[8] predicted that chlorine atoms, coming from chlorofluorocarbons (CFCs) play a very important role in ozone depletion. The final proof of extreme

[6]Schönbein came up with the name *ozone* after the Greek word $o\zeta\varepsilon\iota\nu$, which literally means *to smell*, because ozone has a very typical odour.

[7]The unit of total ozone column is the *Dobson Unit* (DU).

[8]For their research in the field of catalytic ozone depletion cycles, Crutzen, Molina and Rowland received the Nobel prize for chemistry in 1995.

ozone destruction in the stratosphere came in 1985, when a team led by the British scientist Farman reported the Antarctic ozone hole [Farman et al., 1985]. Since then, massive ozone depletion is being reported above the Antarctic continent every year, during the Antarctic spring (September to November). And in the nineties, it was found that areas outside the Antarctic region also are susceptible to stratospheric ozone depletion. Both ground stations and satellite experiments show a downward ozone trend in both hemispheres [Herman et al., 1991, Stolarski et al., 1992].

1.4.2 Homogeneous chemistry - the catalytic cycles

The Chapman mechanism

At altitudes above 30 km, photodissociation of O_2 takes place in the presence of ultraviolet ($\lambda < 242$ nm) solar radiation:

$$O_2 + h\nu \rightarrow O + O \tag{1.2}$$

In the presence of a third molecule (for example N_2 or O_2), ozone can be produced:

$$O + O_2 + M \rightarrow O_3 + M \tag{1.3}$$

This is the only reaction in the stratosphere that produces ozone. Also, ozone can become photodissociated ($240 < \lambda < 320$ nm):

$$O_3 + h\nu \rightarrow O_2 + O \tag{1.4}$$

A fourth reaction is given by:

$$O_3 + O \rightarrow O_2 + O_2 \tag{1.5}$$

which leads to net removal of ozone. Reaction 1.4 followed by 1.3 represents a very fast conversion of O_3 to O and back. O and O_3 are therefore often represented as one and the same constituent: O_x (odd oxygen). The ultimate concentration ratio of O_3 and O is determined by the concentration of the molecule M. The larger this concentration, the higher the frequency of reconversion of O to O_3 (reaction 1.3), and O_3 will be the preferred form, as is the case in the lower stratosphere. At high altitudes in the stratosphere, with low concentrations of M, O will have the preference.

As stated before, this mechanism predicts too much stratospheric ozone (up to 5 times in the upper stratosphere) , and so we have to take other O_3 removal processes into account. In order for an extra constituent to participate in ozone destruction, it will have to be either present in large mixing fractions, or (in the case of a trace gas) regenerated in a catalytic cycle.

The HO_x cycles (H, OH, HO_2)

As early as 1950, Bates and Nicolet [Bates and Nicolet, 1950] suggested that hydrogen radicals could give an explanation for the discrepancy between model and theory. Four catalytic cycles are important [Hunt, 1966]:

$$\begin{array}{rrcl} & H + O_3 & \rightarrow & OH + O_2 \\ & OH + O & \rightarrow & H + O_2 \\ \hline \text{Net :} & O_3 + O & \rightarrow & O_2 + O_2 \end{array} \tag{1.6}$$

$$\begin{array}{rrcl} & OH + O_3 & \rightarrow & HO_2 + O_2 \\ & HO_2 + O & \rightarrow & OH + O_2 \\ \hline \text{Net :} & O_3 + O & \rightarrow & O_2 + O_2 \end{array} \tag{1.7}$$

$$\begin{array}{rrcl} & OH + O & \rightarrow & H + O_2 \\ & H + O_2 + M & \rightarrow & HO_2 + M \\ & HO_2 + O & \rightarrow & OH + O_2 \\ \hline \text{Net :} & O + O + M & \rightarrow & O_2 + M \end{array} \tag{1.8}$$

$$\begin{array}{rrcl} & OH + O_3 & \rightarrow & HO_2 + O_2 \\ & HO_2 + O_3 & \rightarrow & OH + O_2 + O_2 \\ \hline \text{Net :} & O_3 + O_3 & \rightarrow & O_2 + O_2 + O_2 \end{array} \tag{1.9}$$

The relative importance of each cycle is of course connected with the relative concentrations of O, OH, and HO_2, and these are each dependent on the altitude that is considered.

The NO_x-cycles (NO, NO_2)

The NO_x in the stratosphere are mainly produced from N_2O (partly coming from anthropogenic sources, such as agriculture), through a re-

action with the singlet-D oxygen atom[9]:

$$O(^1D) + N_2O \rightarrow 2NO \tag{1.10}$$

The NO produced gives rise to three cycles:

$$\begin{array}{rrcl} & NO + O_3 & \rightarrow & NO_2 + O_2 \\ & NO_2 + O & \rightarrow & NO + O_2 \\ \hline \text{Net :} & O_3 + O & \rightarrow & O_2 + O_2 \end{array} \tag{1.11}$$

$$\begin{array}{rrcl} & NO + O_3 & \rightarrow & NO_2 + O_2 \\ & NO_2 + O_3 & \rightarrow & NO_3 + O_2 \\ & NO_3 + h\nu & \rightarrow & NO + O_2 \\ \hline \text{Net :} & 2O_3 & \rightarrow & 3O_2 \end{array} \tag{1.12}$$

$$\begin{array}{rrcl} & NO + O_3 & \rightarrow & NO_2 + O_2 \\ & NO_2 + h\nu & \rightarrow & NO + O \\ \hline \text{Net :} & O_3 + h\nu & \rightarrow & O_2 + O \end{array} \tag{1.13}$$

In this last cycle (the *null* cycle), overall nothing happens with respect to odd oxygen, but the NO_2 in these reactions is not available for the destruction of ozone in the cycles 1.11 and 1.12, so this cycle has to be accounted for when we analyze the NO_x influence on ozone depletion.

A very important reaction that is related with the NO_x cycles is the production of dinitrogen pentoxide:

$$NO_3 + NO_2 + M \rightarrow N_2O_5 + M \tag{1.14}$$

In reverse, N_2O_5 can redissociate to produce NO_3 and NO_2. So, reaction 1.14 does not represent a permanent loss of the NO_x, just a temporary removal. N_2O_5 can be considered as a NO_x *reservoir* molecule.

However, this reservoir can be disturbed by a heterogeneous reaction:

$$N_2O_5 + H_2O(s) \rightarrow 2HNO_3 \tag{1.15}$$

with $H_2O(s)$ a representation of a water molecule on the surface of a particle. In the gas phase, the reaction of N_2O_5 with water vapor is

[9]This is an electronically excited state of oxygen, that plays a very important part in the atmosphere. It is formed by photodissociation of O_3 as well as by photolysis of N_2O.

very slow, but on the surface of an aerosol particle, the reaction becomes very effective. More on heterogeneous chemistry will be discussed in subsection 1.4.3.

The ClO_x cycles

In the early seventies, the first measurements of man-made CFCs were performed, and the quantities of CFCs measured in the atmosphere appeared to be roughly equal to the total amount that were produced until then. In 1974, Molina and Rowland [Molina and Rowland, 1974, Rowland and Molina, 1975] realized that the CFCs, which are produced and used by man for a variety of technological applications such as refrigerants and aerosol spray propellants, have very long lifetimes (chemically inert, non-soluble) in the troposphere, and as a consequence can be transported to the stratosphere. Once there, they become photolyzed by high-energy ultraviolet radiation. For example, in the case of CFC-11 and CFC-12, we have:

$$CFCl_3 + h\nu \rightarrow CFCl_2 + Cl \tag{1.16}$$

$$CF_2Cl_2 + h\nu \rightarrow CF_2Cl + Cl \tag{1.17}$$

We see that these reactions produce free chlorine atoms. These atoms react very fast with O_3, and a rapid ozone depletion cycle is initiated, involving the chlorine monoxide (ClO) radical as catalytic agent [Stolarski and Cicerone, 1974, Wofsy and McElroy, 1974]:

$$\begin{array}{rrcl} & Cl + O_3 & \rightarrow & ClO + O_2 \\ & ClO + O & \rightarrow & Cl + O_2 \\ \hline \text{Net}: & O_3 + O & \rightarrow & O_2 + O_2 \end{array} \tag{1.18}$$

It is believed that CH_3Cl (methyl chloride) is the only natural source of chlorine atoms in the stratosphere. The tropospheric lifetime of the molecule is sufficiently long, so it is expected that some quantity will be transported to the stratosphere, along rising air streams in the tropics. Through oxidation or photolysis, again, chlorine is released, giving rise to ozone depletion by cycle 1.18.

Coupling of cycles and reservoirs

Until now, we have described catalytic cycles, and one could think that these cycles would go on forever in destroying ozone. However, when one of the catalysts considered (OH, NO, ClO, etc.) participates in yet another reaction, then the cycle can be interrupted (chain termination), because the catalyst is no longer available. If we know these additional reactions, then we can estimate the number of times an ozone destruction cycle occurs on average, before the chain will be terminated. A good example is cycle 1.18, that on average repeats 10^5 times, before Cl or ClO is removed through reaction with another constituent. So, on average, one chlorine atom will destroy 100,000 ozone molecules before it will be removed.

Two important reactions that remove catalysts are:

$$\mathrm{OH} + \mathrm{NO_2} + \mathrm{M} \rightarrow \mathrm{HNO_3} + \mathrm{M} \tag{1.19}$$

$$\mathrm{Cl} + \mathrm{CH_4} \rightarrow \mathrm{HCl} + \mathrm{CH_3} \tag{1.20}$$

Both HNO_3 and HCl are relatively stable species in the stratosphere, so some amount will be transported to the troposphere. The amount that stays in the stratosphere can be transformed back into the original reagents, for example:

$$\mathrm{HCl} + \mathrm{OH} \rightarrow \mathrm{H_2O} + \mathrm{Cl} \tag{1.21}$$

This means that HCl can be considered as a reservoir molecule for chlorine. Thus, Cl atoms can be made inactive temporarily, but can be released later, and their chain can be re-initiated. Another example of a reservoir constituent has already been discussed in reaction 1.14. Yet another important example is the reservoir $ClONO_2$:

$$\mathrm{ClO} + \mathrm{NO_2} + \mathrm{M} \rightarrow \mathrm{ClONO_2} + \mathrm{M} \tag{1.22}$$

The reagents can be released again in the reverse (two-way) photolysis process:

$$\begin{array}{rcl} \mathrm{ClONO_2} + \mathrm{h}\nu & \rightarrow & \mathrm{ClO} + \mathrm{NO_2} \\ & \rightarrow & \mathrm{Cl} + \mathrm{NO_3} \end{array} \tag{1.23}$$

Reactions 1.22 and 1.23 represent a coupling between the NO_x and ClO_x cycles. If the N_2O emission increases, then, through reaction 1.10,

the stratospheric NO concentration will increase also, and ozone destruction will increase by the NO_x cycles. If the emission of CFCs increases at the same time, ozone destruction by the ClO_x cycles will increase also. But an augmentation of NO_x will induce a rise in the $ClONO_2$ reservoir and a decrease in the ClO concentration, by the coupling 1.22. Netto, the ozone depletion caused by CFCs en N_2O emissions will be smaller compared to the effect of the two emissions separately.

The existence of reservoir constituents is crucial in the stratospheric chemistry, and more specifically for the depletion of ozone. HCl and $ClONO_2$ together form a reservoir that contains 99 % of the chlorine atoms that are responsible for ozone destruction. A small change in concentration of these reservoirs can have a huge impact on the effectiveness of the ClO_x cycles.

We have given an example of the coupling between NO_x and ClO_x cycles, but all other cycles are connected in a similar way. Another important example, involving the coupling between BrO_x and ClO_x cycles, produces both free Br and Cl atoms:

$$\begin{array}{rrcl} & BrO + ClO & \rightarrow & Br + Cl + O_2 \\ & & \rightarrow & Br + OClO \\ & & \rightarrow & BrCl + O_2 \\ & Br + O_3 & \rightarrow & BrO + O_2 \\ & BrCl + h\nu & \rightarrow & Br + Cl \\ & Cl + O_3 & \rightarrow & ClO + O_2 \\ \hline \text{Weighted net :} & O_3 + O_3 & \rightarrow & O_2 + O_2 + O_2 \end{array} \qquad (1.24)$$

While bromine species are present in small amounts in the stratosphere, compared to chlorine species, they are per molecule much more effective at ozone destruction, because the formation of bromine reservoirs such as HBr and $BrONO_2$ is much slower than their chlorine counterparts HCl and $ClONO_2$, and/or because the Br reservoirs are much less stable.

1.4.3 Heterogeneous chemistry - the ozone hole

In 1985, the scientific community was shocked by an alarming report. A scientific team, led by British scientist J. Farman, reported a massive depletion of stratospheric ozone concentration during the polar spring

(September to November) above the Antarctic continent, an observation that could not be explained by the known stratospheric chemistry. Moreover, the decrease appeared to become stronger every year. Further analysis of previous data showed that the downward trend was going on at least since the seventies.

In normal conditions, the atmosphere above the Antarctic continent has high ozone concentrations, compared with other latitudes. During the Antarctic spring, the air cools, descends, and forms a wind circulation, called the *polar vortex*. Inside this circulation, the air is extremely cold. During winter and early spring, the vortex is very stable, and this causes the center of the vortex to be effectively sealed off from the air outside the vortex. In this way, large concentrations of ozone can be trapped within the vortex for several months. In September, when the Sun returns, temperature increases, and the vortex weakens, up to the point when it completely disappears in November. The return of sunlight usually is accompanied by a small decrease in ozone levels.

The discovery of the Antarctic ozone hole, however, showed a spectacular decrease in ozone levels. Initially, it was suggested that the behavior could be explained on the basis of sun cycles and atmospheric dynamics. Neither explanation was consistent with observed ozone hole features.

Polar stratospheric clouds (PSCs)

In general, the stratosphere is very dry, and clouds cannot form. However, during the polar night, temperatures between 15 and 20 km altitude can drop to 183 Kelvin (-90 °C). This is cold enough for even the small amount of water present to condense to *polar stratospheric clouds* (PSCs).

Furthermore, the stratosphere contains a natural aerosol layer, composed of small sulfuric acid droplets with an average radius of 0.1 μm. In the midlatitude lower stratosphere, these droplets contain a solution with a weight mixing ratio of roughly 75 % H_2SO_4 to 25 % H_2O. In the polar vortex, at very low temperatures, the particles will absorb water to maintain equilibrium.

Based on observations and laboratory research, PSCs have been categorized into two types, Type I and Type II, of which Type I has been further subdivided into Type Ia and Type Ib. Type Ia PSCs have been

identified as crystals of $HNO_3.3H_2O$ (nitric acid trihydrate, or NAT). They form once temperatures fall below roughly 195 Kelvin. Type Ib PSCs contain super-cooled droplets consisting of a $HNO_3/H_2SO_4/H_2O$ solution. These PSCs form at about the same temperature as Type Ia PSCs, 195 Kelvin. And Type II PSCs contain non-spherical crystalline particles composed of frozen water ice, that form below the ice frost point, which is about 188 to 190 Kelvin in the lower stratosphere.

Despite this (broad) classification, knowledge about the composition, size and phase of PSCs is still very uncertain [Toon and Tolbert, 1995].

The role of PSCs in ozone depletion

We know now that heterogeneous chemistry plays the most important role in ozone depletion above the Antarctic continent [McElroy et al., 1986, Solomon et al., 1986, Molina, 1991]. The release of chlorine atoms from the HCl and $ClONO_2$ reservoir species occurs relatively slowly, but PSCs accelerate this process considerably. First, HCl is taken up by PSC particles, followed by a heterogeneous reaction of the $ClONO_2$ gas with the particle:

$$HCl(s) + ClONO_2 \rightarrow Cl_2 + HNO_3(s) \tag{1.25}$$

Here, (s) represents a constituent on the surface of the particle. In normal conditions, HCl has low solubility in stratospheric aerosols, but at extremely cold temperatures (lower than 200 Kelvin), these aerosols absorb water, and so increase the solubility of HCl. As a result, reaction 1.25 is possible. Two other heterogeneous reactions involving HCl are:

$$HOCl + HCl(s) \rightarrow Cl_2 + H_2O \tag{1.26}$$

$$N_2O_5 + HCl(s) \rightarrow ClNO_2 + HNO_3(s) \tag{1.27}$$

Also important are two heterogeneous reactions with H_2O:

$$ClONO_2 + H_2O(s) \rightarrow HOCl + HNO_3(s) \tag{1.28}$$

$$N_2O_5 + H_2O(s) \rightarrow 2HNO_3(s) \tag{1.29}$$

HNO_3, produced by reactions 1.25, 1.27, 1.28 and 1.29, remains in the ice particles. Eventually, the PSCs are removed from the stratosphere through sedimentation. This process, leading to an overall decrease of nitrogen oxides in the gas phase, is referred to as the *stratospheric denitrification*. At the end of the polar winter, the reservoir species N_2O_5 and $ClONO_2$ have disappeared almost entirely because of this denitrification.

The three reaction products HOCl, Cl_2 and $ClNO_2$ (nitryl chloride) are photochemically less stable than their reservoir associates, and can be photodissociated very quickly in the presence of sunlight[10]:

$$\begin{aligned} HOCl + h\nu &\rightarrow OH + Cl \\ Cl_2 + h\nu &\rightarrow 2Cl \\ ClNO_2 + h\nu &\rightarrow ClO + NO \end{aligned} \tag{1.30}$$

The produced ClO_x radicals once again lead to the destruction of O_3, this time, however, not through reaction 1.18. The reason for this is that the polar stratosphere, at the beginning of spring, does not contain sufficient amounts of atomic oxygen (O) for this cycle to proceed. The rapid decrease in ozone levels is explained with another catalytic cycle, involving the Cl_2O_2 dimer:

$$\begin{aligned} ClO + ClO + M &\rightarrow Cl_2O_2 \\ Cl_2O_2 + h\nu &\rightarrow Cl + ClO_2 \\ ClO_2 + M &\rightarrow Cl + O_2 + M \\ 2(Cl + O_3 &\rightarrow ClO + O_2) \\ \hline \text{Net :} \quad 2O_3 + h\nu &\rightarrow 3O_2 \end{aligned} \tag{1.31}$$

Notice that ClO is available because reaction 1.22 is slow as a result from the denitrification process.

From all these reactions we can conclude that the production of Cl increases considerably when heterogeneous reactions enter the arena. This Cl is finally responsible for the enormous ozone depletion that is being observed. Two conditions are necessary: extremely cold temperatures (for HCl to dissolve in PSC particles), and (some) sunlight (to photodissociate Cl_2, HOCl and $ClNO_2$). Both conditions are fulfilled in

[10]This photodissocation is rapid, even at the beginning of Antarctic spring, when the light intensity is low and has low UV levels.

the Antarctic spring, when the first sunlight returns after the cold polar winter.

In the Antarctic spring, as the polar air mass warms again, the remaining PSCs evaporate, and the small amount of HNO_3 still present is released again. Through photodissociation and reaction with OH, the NO_x are restored:

$$HNO_3 + h\nu \rightarrow OH + NO_2 \tag{1.32}$$

$$HNO_3 + OH \rightarrow H_2O + NO_3 \tag{1.33}$$

The released NO_2 starts to lock up the free ClO radicals again in the $ClONO_2$ reservoir, through reaction 1.22. Still, full recovery can only be established when the polar vortex breaks up entirely, causing the denitrified polar air to mix with NO_x-rich air from outside.

Antarctic vs. Arctic ozone loss

The ozone hole was first discovered in the Antarctic stratosphere, where it manifests itself spectacularly. What about the Arctic region? Several measurements leave no doubt that ozone depletion is also occuring in the Arctic stratosphere in early spring [Salawitch, 1993, Webster, 1993]. The observed depletion is, however, not so dramatic compared to its Antarctic counterpart. Since the Arctic polar vortex is less stable, it is less effective in isolating the inside air. Consequently, during winter the stratosphere is on average about 10 K warmer than in the Antarctic. PSCs are therefore less common, and they disappear sooner by evaporation. NAT PSCs therefore release their nitric acid content again in the gas phase, before they are able to disappear through sedimentation. This seems to be the major difference: while denitrification in the Antarctics is massive, it is rather moderate in the Arctics.

The lesser ability of the Arctic vortex to isolate the inside air also means that nitrogen-rich air and ozone from outside can penetrate the vortex, thereby further reducing the ozone hole phenomenon. This effect, however, also works in reverse: ozone-poor air can on occasion penetrate midlatitudes - with large human populations - and this is probably one of the reasons for the recently increased interest in the Arctic ozone hole.

Nonpolar (midlatitude) ozone loss

Analysis of ozone level data from the 80s and 90s shows a decrease in both hemispheres in all seasons and at at all latitudes [Herman et al., 1991, Stolarski et al., 1992]. The global average total ozone (between 60°S to 60°N) has decreased 3 to 4 % over the period from the late 70s to the late 90s [World Meteorological Organization, 1999][11]. This behavior once again was unexplicable, assuming only gas phase chemistry. In analogy with the polar cases, it was recognized that midlatitude ozone destruction should be associated with heterogeneous reactions that take place on the surface of sulfuric acid aerosols in the stratosphere [Solomon et al., 1996].

We have already mentioned the key reaction (1.15):

$$N_2O_5 + H_2O(s) \rightarrow 2HNO_3 \tag{1.34}$$

As shown by laboratory experiments, this reaction proceeds rapidly on the surface of sulfuric acid aerosols droplets [Mozurkewich and Calvert, 1988, Van Doren et al., 1991]. At midlatitudes, the HNO_3 produced is released immediately in the gas phase, due to undersaturation of the stratosphere with respect to HNO_3. Since the N_2O_5 reservoir is converted to the relatively stable HNO_3, NO_2 is removed from the catalytic NO_x system. Consequently, the ozone depletion decreases. However, this also implies a decrease in formation of the $ClONO_2$ reservoir, so more ClO is available for the ClO_x cycle. The net effect is an increase in ozone destruction. Inclusion of this mechanism has improved midlatitude model predictions of trace gases considerably.

Less dominant, but important, is another reaction that takes place in somewhat colder conditions [Prather, 1992]:

$$ClONO_2 + H_2O(s) \rightarrow HOCl + HNO_3(s) \tag{1.35}$$

which represents the conversion of the inactive $ClONO_2$ reservoir to the active HOCl molecule, that can be readily photodissociated to free chlorine (reaction 1.30). Once again: additional ozone depletion.

[11] Locally, the decrease can be stronger: between 25 °N and 60 °N, total ozone levels have decreased 4 to 6 % from 1990 to 1999 [World Meteorological Organization, 1999]. The global decrease is less strong because no significant ozone depletion is observed in the tropics.

It is clear now that there exists an important relation between stratospheric aerosol presence and ozone levels. Therefore, we need to take a closer look at these aerosols.

1.5 Aerosols - a constituent apart

The reason for a separate treatment of aerosols can be found in their phase. They are fluid or solid particles, of microscopic or sub-microscopic size, but larger than molecular dimensions. While a molecule can be entirely described by the atoms of which it is composed, the description of an aerosol particle demands specification of size, chemical composition, phase and morphology. Consequently, the behavior of such a particle differs considerably from that of one molecule of the sort which the particle is composed of. By behavior, we mean not only the chemical behavior, but also the dynamical behavior, and, last but not least, the optical behavior.

Typical values for atmospheric aerosol sizes are in the range from a few nanometers to tens of micrometers. Given a population of aerosols present in the air, they can change their size and composition by a diversity of processes: condensation of other gases on the particle, evaporation, coagulation with other particles, chemical reactions and activation in the presence of super-saturated water vapor to the formation of fog and cloud droplets. At the end, two mechanisms can remove the particles from the atmosphere: sedimentation on the surface (dry deposition) or incorporation into a cloud droplet during formation of precipitation (wet deposition).

Aerosols have an important impact on the climate. They act as nuclei for condensation of water vapor, and are in this way indispensable for the production of clouds. They are capable of reflecting solar radiation back into space, thereby causing a cooling of the surface of the Earth. At the same time, they can absorb infrared radiation coming from the Earth, resulting in a heating of the stratosphere. The role of aerosols in heterogeneous chemistry has already been discussed in the previous section.

1.5.1 Stratospheric aerosols

In the late 1950s and early 1960s, experiments confirmed that a stable aerosol layer is present in the stratosphere [Junge et al., 1961]. The altitude of this layer depends on latitude, but on average, the layer can be found at altitudes between 20 and 25 km. Often, the layer is referred to as the *Junge layer*.

A diversity of measurements showed that the aerosols in the stratosphere are composed of a water solution of 60 to 80 % of sulfuric acid (H_2SO_4). It is generally accepted that in normal conditions the source of this sulfuric acid can be found in the oxidation of OCS (carbonyl sulfide) (as was first suggested by [Crutzen, 1976]), a molecule that is produced in surface seawater by the photochemical degradation of dissolved organic matter [Flöck and Andreae, 1996]. OCS is very inert, practically insoluble, and has a long residence time in the troposphere. Therefore, it can be transported to the stratosphere. Once it arrives, it can be photodissociated by ultraviolet radiation, or an oxidation with oxygen can occur. Both ways deliver the ingredients necessary to produce sulfuric acid, the main material for stratospheric aerosols. Other molecules containing sulfur (SO_2, DMS, CS_2) cannot survive long enough in the troposphere to reach the stratosphere.

OCS is the main source for stratospheric aerosols in normal conditions. These normal conditions form the subject of a controversy in the scientific world, because many are convinced that the stratospheric aerosols always exceed this *background* level [Robock, 2000]. The reason for this can be found in the frequent occurrence of volcanic eruptions, that inject large amounts of SO_2 gas into the stratosphere. The sulphuric acid cloud that emerges after such an eruption can cover the entire Earth in a few months, and aerosol concentrations can be several orders of magnitudes larger than the normal background. Because stratospheric aerosols after a volcanic eruption have a relaxation period of several years, we can conclude that these aerosols are seldom in the state of a non-disturbed background (if it exists).

The most important chemical processes that lead to the production of sulfuric acid aerosols in the stratosphere can be described by the following reactions [Koziol and PudyKiewicz, 1998]:

$$\begin{array}{rcl} OCS + h\nu & \rightarrow & S + CO \\ OCS + O & \rightarrow & SO + CO \\ S + O_2 & \rightarrow & SO + O \\ SO + OH & \rightarrow & SO_2 + H \\ SO + O_2 & \rightarrow & SO_2 + O \\ SO + ClO & \rightarrow & SO_2 + Cl \\ SO_2 + OH & \rightarrow & HSO_3 \\ SO_2 + h\nu & \rightarrow & SO + O \\ HSO_3 + O_2 & \rightarrow & SO_3 + HO_2 \\ SO_3 + H_2O & \rightarrow & H_2SO_4 \end{array} \tag{1.36}$$

Since H_2SO_4 is a molecule with a large polarity (which means a large solubility in water), it will, after formation, immediately associate itself with water, and droplets will be produced on condensation nuclei (very small particles, ions).

1.5.2 The Mount Pinatubo eruption

In June 1991, when ORA had already been delivered to ESA for launch, an event occurred that was completely unexpected and that would put its stamp on the data processing of the experiment. After ten weeks of precursory activity, Mt. Pinatubo (Fig. 1.3) in the Philippines (15°08'N, 120°21'E) erupted. The plinian eruption was extremely violent, and was, after the 1912 Katmai-Novarupta eruption in Alaska, the second largest volcanic eruption in the 20th century in terms of magnitude, or volume of expulsed magma. More than 5 km^3 of magma was produced [Scott et al., 1996]. Eruption columns reached an altitude of more than 35 km and placed a giant umbrella-shaped cloud in the lower to middle stratosphere with a dimension of 1,100 km in diameter [Koyaguchi and Tokuno, 1993]. Within the cloud, about 17 Megaton of SO_2 gas[12] was injected into the stratosphere [Self et al., 1996], which almost immediately began to convert into H_2SO_4 aerosols. Since the Krakatau eruption (1883), no volcano has had such an enormous impact on the stratospheric aerosol layer.

[12]Besides SO_2, also CO_2, an estimated 250 to 500 Megaton of H_2O, and 3 Megaton of HCl was produced. The HCl, potentially dangerous for the stratospheric ozone, is however highly soluble in water, and was returned to the Earth surface very rapidly in the form of precipitation.

Figure 1.3: Eruption of Mt. Pinatubo, the Philippines (photo: US Geological Survey).

The full conversion of the SO_2 cloud produced about 25 Megaton of new sulfuric acid aerosols [McCormick and Veiga, 1992]. Measurements by the SAGE II and AVHRR instruments showed that these aerosols encircled the Earth (a latitude band between 30°N and 10°S) within a time span of 3 weeks [McCormick and Veiga, 1992, Stowe et al., 1992]. It took one year for the aerosol cloud to cover the entire Earth [Self et al., 1996].

The Pinatubo aerosol cloud induced several important effects that disturbed the normal state of the atmosphere and the surface.

- Optical effects. Large volcanic eruptions induce many optical effects, such as unusually colored sunrises and sunsets, and a hazy appearance of the Sun. The effects are caused by the ability of aerosols to scatter light (see Fig. 1.4). Optical thickness[13] measurements after the Pinatubo eruption showed peak values at mid-visible wavelengths of 0.4 [Stowe et al., 1992, Valero and Pilewskie, 1992]. After the widespread dispersal of the aerosols, global values were about 0.1 to 0.15, and values remained elevated for several

[13]The optical thickness will be defined in Chapter 3.

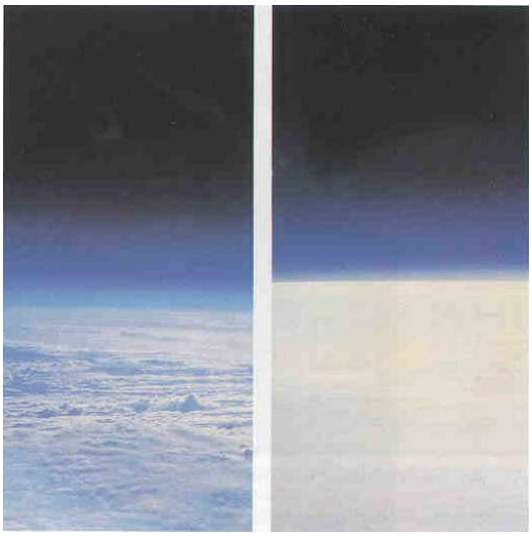

Figure 1.4: Photographs of the Earth's limb, taken above France in June 1980, before (left) and after (right) the Mount St. Helens eruption, Washington, USA, May 1980. The scattering effect of the volcanic aerosols is clearly visible [Ackerman et al., 1980].

years [Self et al., 1996]. The cloud caused a dramatic decrease in the amount of net radiation reaching the Earth's surface.

- Climatic effects. Immediately after the eruption, the lower stratosphere began to warm [Labitzke and McCormick, 1992], due to absorption of radiation by aerosols, resulting in a temperature peak in 1991. Since that time, the stratosphere started to cool to the lowest temperatures ever recorded [Christy and Drouilhet, 1994, Monastersky, 1994]. Important changes in atmospheric circulation resulted from this cooling, which may be related to aerosol induced ozone depletion in the lower stratosphere (mainly ozone is responsible for the normal stratospheric heating).

  A surface cooling of up to 0.5 to 0.7 °C was observed in the Northern Hemisphere [Dutton and Christy, 1992], equivalent to a hemispheric reduction in net radiation of 4 Wm^{-2}, and an overall cooling of as much as 0.4 °C occurred over large parts of the Earth in 1992 to 1993. Climate models have predicted this cooling with

a reasonable accuracy [Hansen et al., 1992]. In terms of climate forcing, the Pinatubo effect was stronger than the warming effects of the El Niño event or the anthropogenic greenhouse effect in the period 1991-1993 [Mo and Wang, 1994, Self et al., 1996].

- Stratospheric chemistry effects. Mid-latitude ozone concentrations reached their lowest level on record in the period of 1992 to 1993 [Gleason et al., 1993]. The total ozone amount was 2 to 3 % lower than in any earlier year. The Southern Hemisphere ozone hole increased in 1992 to an unprecedented size of 27×10^6 km^2, and depletion rates were faster than ever before recorded [Brasseur, 1992, Hofmann et al., 1992]. The levels of other chemical species were also affected (see e.g. [Kinnison et al., 1994]).

The scientific impact of the Pinatubo eruption can hardly be underestimated. The event has provided an exceptional natural laboratory for atmospheric scientists, and has stimulated a lively interest in the role of volcanic aerosols in climate change and atmospheric chemistry. Scientists were able to test and tune atmospheric models. And important new insights have emerged in scientific fields, such as aerosol dynamics, the radiative budget of the atmosphere, heterogeneous chemistry and global air circulation.

1.6 The need for new measurements

In this chapter, we have presented a general description of the Earth's atmosphere and, more specifically, an introduction to stratospheric chemistry and aerosols. However, our current picture is far from complete: although present-day models predict trace gas concentrations with reasonable accuracy, lots of issues remain unsolved. Furthermore, these concentrations are not constant: long-term trends for some constituents are observed.

Aerosols form a somewhat independent question. Although volcanic eruptions frequently provide new information, processes such as aerosol formation, dynamics and global dispersion are still poorly understood. In view of their importance (heterogeneous chemistry, climatic effects, impact on cloud and PSC formation), our understanding of aerosols has to improve.

The bottom line: we need more data. And indeed, much effort is made to design and construct new instruments, following a variety of experimental principles (absorption, emission and back-scattering of light, mass spectrometry, in situ sampling) and carriers (satellites, high-altitude balloons, rockets, air planes). This thesis presents the data processing and results from the Occultation Radiometer (ORA), an instrument designed to measure ozone chemistry-related trace gas concentrations (O_3, NO_2, H_2O) and stratospheric aerosols. The experimental set-up of ORA will now be described in the following chapter.

Chapter 2

The Occultation Radiometer onboard EURECA.

Now that we know what we want to measure in the atmosphere and why, it is time to put everything into practice: a space experiment. First, we will have a look at the EURECA platform. This device represents the *home* of our experiment: it offers electrical power, a comfortable temperature and communication facilities with the outside world. The Occultation Radiometer (ORA) itself will be introduced subsequently, first by outlining the actual method (the occultation technique), and then by describing the optics, detectors and electronics. The chapter ends with a report on the mission and the findings afterwards.

2.1 The EURECA carrier.

2.1.1 General description.

The *EUropean REtrievable CArrier* (Fig. 2.1) was proposed by the *European Space Agency* (ESA) in the early eighties and was somehow to be considered as an unmanned *Spacelab* follow-on: a multi-user platform that could carry a large number of experiments, with a strong emphasis on microgravity science. From the ESA viewpoint, the most important EURECA feature was that it could be used many times. It was designed to be retrieved from orbit at the end of the mission, and to be launched

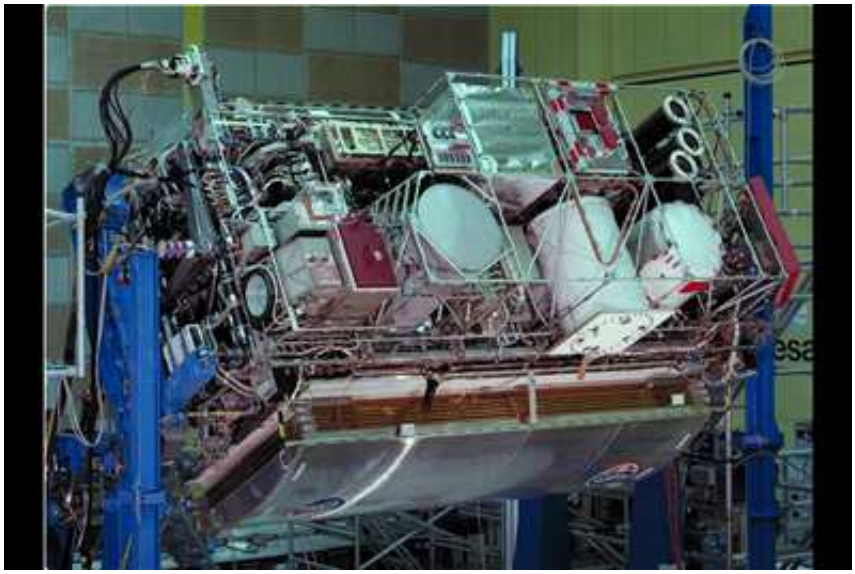

Figure 2.1: The European Retrievable Carrier, together with the experiments. ORA can be found on the bottom left of the front side [Daimler-Benz Aerospace, 1996].

again with a different payload. The ultimate purpose of such a platform, of course, is a reduction in design and manufacturing costs.

A retrievable platform is very interesting for microgravity scientists, because samples are returned after the space flight. For example, crystals can be grown in the (apparent) absence of gravity, and can be fully investigated in the lab afterwards. Other experiments (such as ORA) can be tested for changes in instrument characteristics, caused by the crude conditions in a space environment.

The actual project started in 1984. At the time of launch in 1992, EURECA was the largest spacecraft ever built and flown by ESA: it weighed 4500 kg and measured 3.5 $\times$ 20 m (fully deployed solar panels). The technologically highly advanced platform consisted of a diversity of subsystems:

A mechanical structure was needed, of course, to mount the different subsystems and experiments.

Electrical power was delivered by solar panels and stored in batteries. The power was delivered to the subsystems and experiments by distribution units.

Data traffic was taken care of by a Data Handling System (DHS), that controlled telecommand and telemetry data, received and sent by the ground stations, the onboard subsystems and payload. Interfacing between the payload and the DHS was provided by so-called *Remote Acquisition Units* (RAU) and *Process Interface Adapters* (PIA). Most of the time, when there was no contact between EURECA and the ground stations, data was temporarily stored in the 128 Mbit *Mass Memory Unit*.

Communication with ground stations was provided by a *Telemetry and Telecommand Subsystem* (TTS), containing the actual transmitting and receiving equipment, including antennas.

Thermal control is needed. An active freon cooling system took care of heat dissipation to space (heat generated by the payload and subsystems). Passive thermal control by multilayer isolation protected the equipment against solar heating.

Satellite movement could be controlled by thrusters. For large distances, hot gas thrusters (burners) were used, changing the orbit altitude at the start and end of the mission. Small translational and rotational corrections were made by cold gas thrusters. These thrusters were controlled by a system that monitored the attitude and orientation of the satellite.

Microgravity monitoring was performed by a system, accurately measuring the accelerations of the satellite.

2.1.2 Onboard experiments.

Let us now have a look at the inhabitants of the space platform. It is not easy to obtain experiment space on such a platform: power and spatial volume is limited, competing experiments are not scarce and

Payload mass	1000 kg
Payload volume	8.5 m^3
Payload power	
average	1000 W
peak (limited time)	1500 W
Sun inertial pointing	± 1 °
	(actual flight performance: ± 0.4 °)
Attitude measurement accuracy	± 0.25 °
Microgravity	nominal $< 10^{-5}$ g (f $<$ 1 Hz)
	linear for 1 Hz $<$ f $<$ 100 Hz
	$< 10^{-3}$ g for f $>$ 100 Hz
Orbital temperature	
Operational	0 to 40 °C
Non-operational	-10 to 40 °C
Data rate	
Average	1.5 kbps
Peak	43/21.5 kbps depending on interface

Table 2.1: EURECA payload facilities.

launch costs are high. Technical specifications for the EURECA payload accomodation are given in Table 2.1. For ORA, a very important number was the Sun pointing accuracy of $\pm$ 1 degree.

The largest part of the EURECA payload was occupied by microgravity experiments in the field of material sciences, fluid physics and biology. Radiation effects on biological organisms were also studied. Space and Earth Science experiments, together with three technological tests occupied the remaining payload. It lies outside the scope of this text to describe these experiments in full detail, but a summary is given in Table 2.2.

The EURECA experiments are the result of the work performed by scientists and engineers from 29 universities and institutions in European Community member states: Germany, France, Norway, Denmark, Belgium, The United Kingdom, the Netherlands and Italy. For more information, the reader can consult [De Beule, 1993, Innocenti, 1993, Minster et al., 1993, ESA, 1987]

Discipline	Experiment or device	Study / test
μ-gravity	Automatic Mirror Furnace	Crystal growth
	Solution Growth Facility	Crystal growth
	Protein Crystallization Facility	Protein crystal growth
	High Precision Thermostat	Critical point
	Surface Forces and Adhesion Instrument	Impact adhesion
	Multi-Furnace Assembly	Crystal growth
Space & Earth Science	Exobiology Radiation Assembly	Radiation exposure
	Solar Spectrum Instrument	Solar irradiance
	Solar Constant and Variation Instrument	Solar constant
	Occultation Radiometer	Trace gas and aerosol
	Wide Angle X-ray Telescope	X-, Gamma-ray bursts
	Timeband Capture Cell Experiment	Space microparticles
Technology	Radiofrequency Ion Thruster Assembly	Ion propulsion unit
	Inter-Orbit Communication Instrument	Data relay system
	Advanced Solar Gallium Arsenide Array	Solar Cell

Table 2.2: The instruments and devices on board EURECA.

2.2 The occultation technique

2.2.1 Description

Our goal is to measure concentrations of trace species as function of altitude. Furthermore, we want to take measurements relatively fast, for different parts of the Earth. This is the reason why we are going to space: a wide Earth coverage, and a large number of samples.

The Sun pointing capability of EURECA offers a very suitable way to perform so-called *Sun occultation* measurements. The approach of the method is to measure the attenuation of solar radiation, due to absorption and scattering by atmospheric constituents, during subsequent sunrises and sunsets, as seen by the experiment from its orbit around the Earth (Fig. 2.2). In the course of one occultation, the instrument line-of-sight scans successive layers of the atmosphere, because the grazing height of the solar rays changes as the satellite moves along its orbit. If we know the optical absorption and scattering properties of the atmospheric constituents, and if the measurements are performed at appropriate wavelengths, we can obtain information on the constituent concentrations as function of altitude, integrated along the atmospheric path of the solar light.

It is the task of specific numerical retrieval methods to convert this

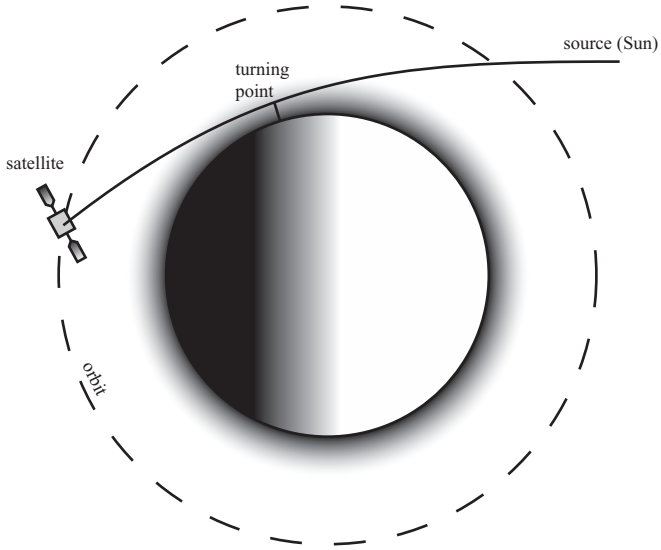

Figure 2.2: The solar occultation method.

integrated information into the required concentration data for a given altitude. Unfortunately, while the measurement method is fairly simple, these retrieval computations are somewhat cumbersome, and a large part of this thesis is dedicated to this subject. To make life easier, one usually assumes the Earth's atmosphere to consist of *homogeneous* spherical layers (for one occultation, of course). In such a case, the properties of one layer can be characterized at one reference point in the layer. The most convenient choice is the *turning point*, where the beam of solar radiation is closest to the Earth's surface.

2.2.2 Pros and cons of occultation methods

Many different types of remote sensing set-ups exist, ranging from *nadir*[1] viewing experiments to *limb emission*[2] techniques. These experiments are all optimized towards certain properties; however, this usually com-

[1] A nadir view is a view directed towards the center of the Earth.

[2] A limb view is directed towards the horizon of the Earth, but not necessarily towards the sunset or sunrise. For emission techniques the atmospheric layer itself is the radiation source.

promises the quality of other features of importance. Let us find out what is so special about occultation methods.

Advantages

Long optical paths. A view towards the Sun provides an optical path that is much longer than a nadir view, which results in an improved sensitivity of the measurements[3].

High S/N. The high-intensity solar light produces a strong signal, and as a consequence, the signal-to-noise ratio is very high. Therefore, wavelength selection filters can be used that are very narrow, and the resulting spectral separation of the different chemical species is more accurate.

Self-calibrating. Absolute values of the incoming irradiance do not have to be measured; the instrument is by definition self-calibrating, so the atmospheric transmittance is simply calculated from the *full-Sun* signal (no atmosphere measured), the *full-Earth signal* (the Sun is behind the Earth's horizon), and the signal values in between.

LTE-independent. While molecular emission of radiation is very dependent on the local temperature and the (perhaps faulty) assumption of local thermal equilibrium (LTE), molecular absorption is almost not. Since a transmission signal is formed by absorption (and also scattering), a solar occultation experiment is less sensitive to the atmospheric temperature and LTE, compared with an experiment that measures emission.

Good vertical resolution. One of the main advantages of solar occultation experiments is the good vertical resolution, because they measure the atmosphere from the side. Of course, this vertical resolution is dependent on the instrument field of view. This subject will be further discussed in the next subsection.

[3] At very low altitudes (troposphere), this actually becomes a disadvantage, because the sunlight that travels along the path through dense atmospheric layers is attenuated almost entirely. This altitude range is often labelled as the *saturation regime*.

Disadvantages

Low sampling. In the course of one orbital pass, only two geolocations are measured: those where the sunrise and sunset occurs[4]. This is a clear disadvantage with respect to a nadir experiment, which measures continuously.

Bad horizontal resolution. The measurement of atmospheric absorption and scattering along a long atmospheric path delivers data that have a relatively poor horizontal (latitude vs. longitude) resolution. However, if the horizontal gradients of the constituents are not too large (and this is mostly the case), this effect will not be very dominant. As a rough estimate, the 'effective' path length along which the largest part of the light attenuation occurs is about 500 km.

Tangential point displacement. During an occultation, the latitude and longitude of the tangential height of the solar beam changes, due to the movement of the satellite and the rotation of the Earth. This again causes a blurring of the horizontal measurement resolution.

Low time resolution. It is not a straightforward matter to interpret measured atmospheric constituents whose concentrations change rapidly at sunset and sunrise because of photochemical effects. An example: NO_2 concentrations change while a sunrise measurement is being taken, because photodissociation to NO and O is taking place at the same time.

Absence of occultation. When global coverage of the Earth is desired, the satellite orbit needs a high inclination with respect to the Earth's equator. In such cases, it is clear that there will be periods without any occultation at all. The Sun is then visible from all satellite positions along the orbit.

[4]When different stars are used as the source in an occultation experiment, the sampling is of course much better. However, because star light is weak, the signal-to-noise ratio becomes much worse. Example: GOMOS (Global Ozone Monitoring by Occultation of Stars).

2.2.3 The field of view

It is clear that the vertical resolution of an occultation experiment is strongly related to the field of view, i.e. the viewing angle from which the experiment collects light from the Sun. Advanced experiments, like ATMOS or SAGE II, have a viewing angle that is very small, in order to measure thin atmospheric layers separately. These experiments are equipped with an expensive Sun tracking device and a telescope, through which the experiment is able to measure different small parts of the Sun successively.

This setup was not possible for an experiment onboard EURECA, a platform that was primarily intended for microgravity research. Even a small acceleration of the satellite would cause unwanted vibrations. As a first consequence, the platform was equipped with solar panels that are immobile and always directed towards the Sun (with a pointing accuracy of $\pm$ 1 degree). Furthermore, ORA was not equipped with a Sun tracker, so other alternatives were necessary. Finally, a large field of view was chosen, large enough to view the full solar disc at all times, taking into account the pointing error of the satellite.

As a result, the *apparent* vertical resolution of ORA is very poor, because the full solar disc spans a vertical atmospheric distance of roughly 25 km at the turning point. However, as we will point out in Chapter 4, this low apparent vertical resolution is greatly improved by the very large signal to noise ratio of the instrument.

Another problematic fact emerged when the processing of the ORA data was already at an advanced stage. The instrument sees not only the image of the full solar disc, but also the surrounding atmosphere of the Earth. This means that additional straylight enters the optical modules. During the design period of ORA, it was believed that this straylight would be negligible compared to the strong solar light, an assumption that turned out to be questionable. More on this will be covered in Chapter 5.

2.3 ORA - experimental description.

ORA has a total weight of 17 kg and consists of four major parts: the mechanical structure, the electronics unit, the UV-visible module and the infrared unit (Fig. 2.3). The first three parts were developed at the

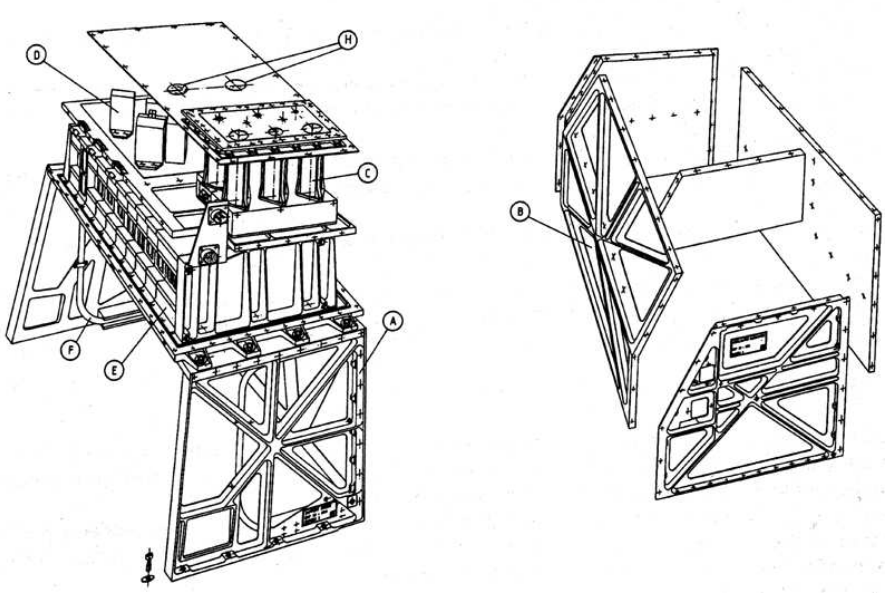

Figure 2.3: The different parts of ORA: a bridge (A), a casing (B), a UV-visible module (C), an infrared unit (D), and an electronics unit (E) [Arijs et al., 1994].

Belgian Institute for Space Aeronomy (BISA) by Etienne Arijs, Dennis Nevejans and co-workers [Arijs et al., 1994, 1995, Nevejans et al., 1990], while the infrared module was delivered by the Department for Atmospheric, Oceanic and Planetary Physics of the University of Oxford [Calcutt et al., 1993, Burchell, 1996]. BISA was also responsible for the qualification of the instrument flight model, communications with ESA and the project management.

2.3.1 Mechanical Structure.

In view of the limited space on the EURECA carrier, the ORA instrument had to be mounted above another instrument (the Danish WATCH experiment) by means of a support structure (bridge, Fig. 2.3, A). This bridge caused a number of mechanical problems with respect to vibration levels and required a careful mechanical design and simulation, the

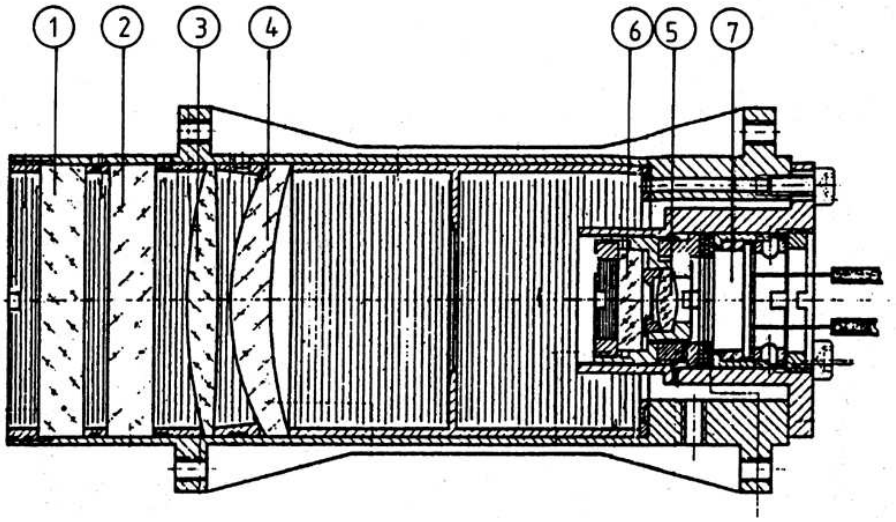

Figure 2.4: A typical ORA UV-visible module consists of a quartz window (1), an interference filter (2), lenses (3,4,5) and a photodiode (7). Some modules have an additional filter (6) [Arijs et al., 1994].

latter being performed by ESTEC[5] and the Belgian company LMS[6]. After three prototypes and several vibration tests, a final model emerged. To allow for the heat dissipation of the instrument, two heat pipes (Fig. 2.3, F), constructed by SABCA[7], were integrated in the bridge.

2.3.2 The UV-visible part.

This unit was designed to measure ozone, NO_2, water vapor and aerosols, through extinction measurements in the ultraviolet, visible and near-infrared part of the solar spectrum. It contains 8 similar modules, with the structure shown in Fig. 2.4.

Aperture.

In front of the different modules, appropriate diaphragms were mounted (Fig. 2.5) to limit the incoming light flux. In this way, the detectors can

[5]European Space Research and Technology Center, Noordwijk, The Netherlands.
[6]Leuven Measurement Systems.
[7]Société Anonyme Belge de Constructions Aéronautiques.

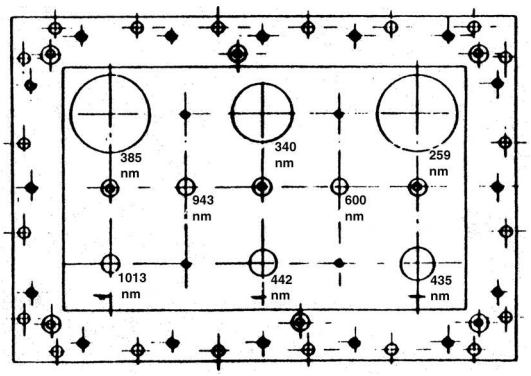

Figure 2.5: The apertures for the different UV-visible modules, located on the front panel of ORA. Notice that the large apertures are associated with short wavelengths [Sekkate, 1988].

operate in their linear domain. The actual diameter of the diaphragm is of course dependent on the wavelength. In general, at short wavelengths, atmospheric extinction is greater due to Rayleigh scattering of the air, so larger apertures are needed. However, the characteristics of other components (filter transmittance, detector sensitivity) also play an important role in the choice of aperture diameters.

Quartz windows.

To isolate the internal components of an optical module from space, each module has a quartz window, mounted at the entrance of the module. Quartz is the standard material for the UV-visible range, because of its good transmission properties.

Main filters.

To select the appropriate wavelength domain, Fabry-Perot interference filters were used (Fig. 2.6). The spectral features of most constituents are broad and therefore small filter bandwidths are not necessary. However, the two channels centered at 435 and 442 nm were chosen to perform a differential measurement on small-scale structures of the NO_2 absorption spectrum, so these channels were equipped with filters that have small bandwidths (a few nanometers).

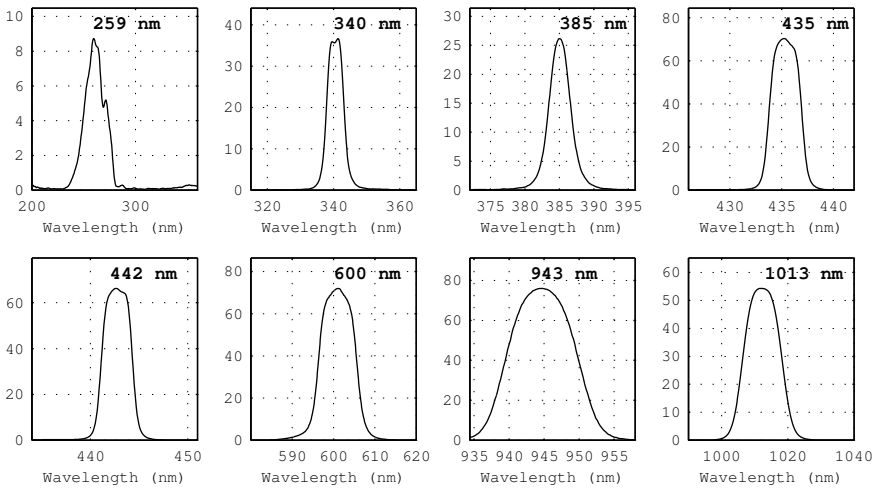

Figure 2.6: Main filter spectral transmittance (%) for the eight UV-visible modules.

Optics.

EURECA has a Sun pointing accuracy of $\pm 1^{\circ}$. Combined with the Sun's angular size (32 arcminutes), and a possible misalignment of ORA with the optical axis of the spacecraft, a view of $\pm 2^{\circ}$was chosen to guarantee a full view of the solar disc (Fig. 2.7, top). The reason for a limited field-of-view is clear: light coming from sources other than the Sun is not wanted (e.g. the Moon, atmospheric straylight, reflections of solar light on other EURECA experiments).

The actual field-of-view is determined by the interplay between the length of a module, the diameter of the entrance hole and the optics that are used. After careful simulations, an optimal design for the optics was chosen, consisting of three lenses and a diaphragm that cuts rays coming from larger viewing angles. Also taken into consideration was the need to spread the incoming light flux uniformly across the surface of the photodiode.

Measurements of the module angular dependence were taken after construction [Sekkate, 1988], and were in agreement with the requirements (Fig. 2.7, bottom).

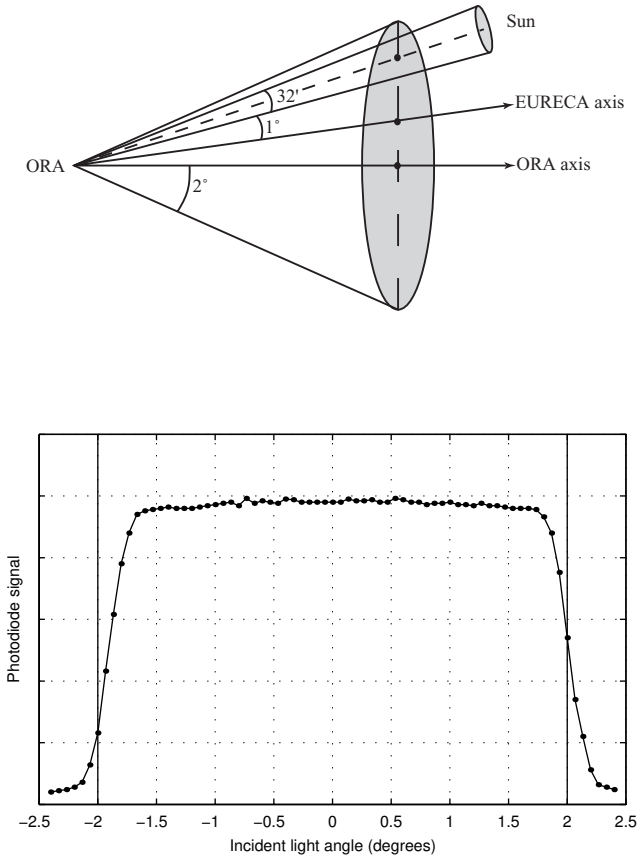

Figure 2.7: Top: With a viewing angle of $\pm 2°$, there is still room left for a possible misalignment between ORA and EURECA. Bottom: measured transmission of the 442 nm optical module as function of incident light angle.

Main interference filter			Photodiode type	Correcting filter
wavelength	band width (FWHM)	peak transmission		
259 nm	23 nm	8.4 %	Si	> 270 nm
340 nm	6 nm	36 %	GaP	none
385 nm	3.4 nm	26 %	GaP	none
435 nm	3.2 nm	70 %	GaP	none
442 nm	3.3 nm	66 %	GaP	none
600 nm	9.5 nm	72 %	Si	KG3
943 nm	11 nm	74.5 %	Si	KG3
1013 nm	12.2 nm	54 %	Si	none

Table 2.3: The filter characteristics and photodiode types of the ORA UV-visible modules.

Additional filters.

The measured transmission of the 259 nm module main filter showed a second unwanted peak at 271 nm (Fig. 2.6). To correct the filter characteristic, another interference filter was introduced just in front of the detector. Some other modules were equipped with a colored glass KG3 to remove infrared radiation entering the module through the (non-perfect) side wings of the main filter transmittance.

All the main and additional filters that were used were manufactured according to the BISA specifications by the French company MATRA.

Detectors.

To measure the radiation intensity, photodiodes were used. Depending on the wavelength (Fig. 2.8), either a UV-enhanced Silicon photodiode or a UV-sensitive GalliumPhosphide diode (covered with a fused silica window) was mounted in the back of the optical modules. Both are low-noise diodes, with a very large dynamic range.

The characteristics of the different UV-visible modules are summarized in Table 2.3.

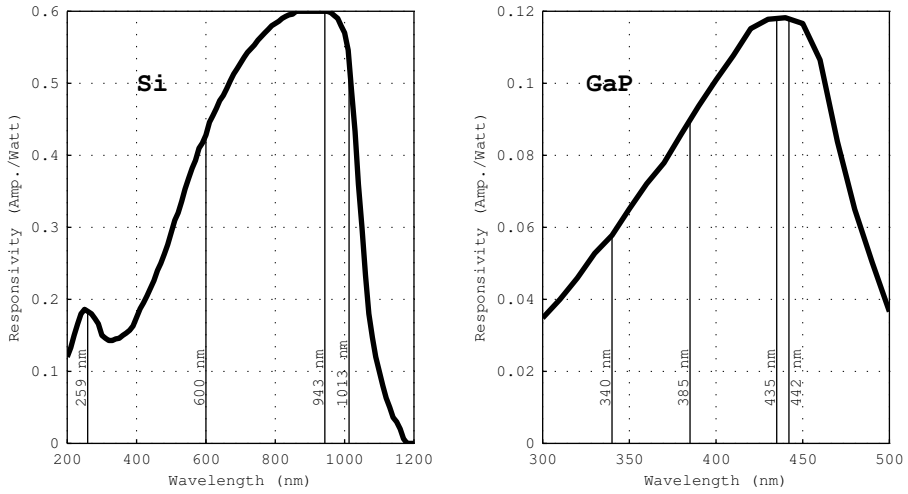

Figure 2.8: Spectral properties of the photodiodes (delivered current per unit of light intensity). ORA channels are also indicated.

The choice of wavelengths

Most filters were chosen to have their central wavelength in specific absorption bands of the species to be examined. At 259 nm, the ozone *Hartley bands* dominate absorption in the mesosphere, while the 600 nm *Chappuis band* maximum can deliver information about ozone concentrations in the stratosphere. The two channels at 435 and 442 nm deliver differential information on NO_2 (maximum and minimum of NO_2 absorption). At 943 nm, water vapor has absorption lines in the near-infrared.

Atmospheric scattering causes further extinction. At 340 and 385 nm, the stratospheric light attenuation is primarily caused by *Rayleigh* (air) scattering. And the 1013 nm module was chosen to retrieve stratospheric aerosols. We will take a closer look at absorption and scattering in Chapter 3. And, although every module is dedicated to the measurement of a specific constituent, overlapping of absorption and scattering by different species takes place in every channel. This problem is the subject of Chapter 5.

2.3.3 The infrared part.

This unit was developed by the Department for Atmospheric, Oceanic and Planetary Physics of the University of Oxford (UK). This text describes the UV-visible measurements and data processing, and therefore the IR unit will be discussed briefly, just for completeness. A full description can be found in [Burchell, 1996, Calcutt et al., 1993].

The infrared part of ORA (known in the UK as *EOR - EURECA Occultation Radiometer*) consists of two pairs of optical channels, one pair of which is used to measure water vapour, and the other carbon dioxide. As both pairs of channels are similar, we will describe only one here. The solar radiation, which enters the instrument through one of its entrance holes, is reflected by a polished aluminium mirror inclined at 45°. It then passes through an aperture stop and narrow interference filter which selects the relevant spectral region (infrared) and reaches a four-blade rotating chopper. The latter modulates the beam at 533.3 Hz and splits it in two portions. One beam passes through the chopper, bounces off another mirror onto a lead sulphide detector, while the other beam is reflected off the chopper blade, travels through a cell containing the gas to be measured (H_2O or CO_2), and falls onto another lead sulphide detector.

There are two ways of deriving the atmospheric constituent from the measurements. The simplest is based upon the signal analysis of the path with no gas cell, and is known as the *Wide Band method*. The second method is a differential technique, called the *Gas Correlation method*, and it uses the signal of light that is modified by passing through the gas cell. Light that is only partially absorbed by the relevant atmospheric constituent will be nearly totally absorbed by the gas cell. This means that no matter how much of the constituent is present in the atmosphere, the detector signal of the gas cell path will be sensitive only to other absorbers. The signal of the other detector (no gas cell) is however sensitive to all absorbers. Taking the difference between the two signals, one obtains the contribution of the considered constituent only.

2.3.4 Electronics unit.

The electronics unit was completely designed at BISA. Briefly, it consists of the following modules:

- Three integrated *ADC converter boards*, each consisting of 4 programmable gain stages (15 steps of 1.5 dB), 4 voltage to frequency converters and four counters (16 bit).

- A *microprocessor board*, containing a 80C86 microprocessor, 16 kbytes of EPROM, 32 kbytes of RAM, an address decoder, an interrupt controller and programmable timers. The processor runs a real-time multitasking operating system (VRTX from Ready Systems) that controls the experiment.

- A *command receiver board*, that accepts commands formatted in telecommand data packets, arriving from a Remote Acquisition Unit (RAU) connected to the EURECA Data Handling System (DHS).

- A *data transmitter board* that sends data formatted into telemetry or telecommand packets to a RAU connected to the DHS.

Total power consumption of the entire ORA experiment was 25 Watt (at 28 Volt) if all heaters were operating. For a detailed treatise, including the IR-unit electronics and software, see [Nevejans et al., 1990].

Let us have a look at one of the ADC board channels (Fig. 2.9). A UV-visible photodiode is connected to an *operational amplifier* with a feedback resistance of which the value can be logarithmically controlled to alter the gain. The output voltage of the amplifier is proportional to the photodiode current. This voltage is transformed into a series of pulses by a *voltage to frequency converter*. The frequency of the pulses is proportional to the output voltage of the amplifier. Finally, a 16-bit *counter* 'counts' the pulses and delivers a digitized output, ready for noiseless transmission to the Earth's surface.

A measured intensity sample is thus encoded in 16 bits, a state of the art format at the time of design. This means that the full occultation can be represented within a dynamic range from 0 to $2^{16} = 65536$ counts.

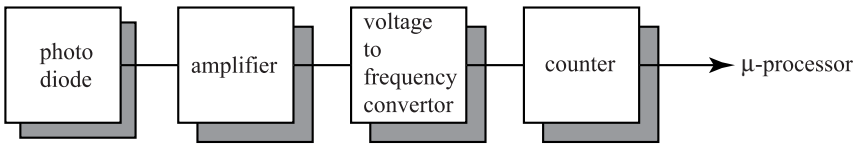

Figure 2.9: The amplification and digitization path of a UV-visible channel.

Figure 2.10: The ORA flight model [Daimler-Benz Aerospace, 1996].

2.4 The mission.

2.4.1 Final Tests

After testing of the inhouse made prototypes, the space compatible flight model was constructed and qualified (vibration and thermal testing). The ORA flight model (Fig. 2.10) was delivered to ESA in August 1989. After the assimilation of all experiments on the EURECA platform (Fig. 2.1), final testing was performed. ORA was tested on the highest possible level. Using a stimuli box, containing several light sources with the appropriate spectral output, sunsets and sunrises were simulated, and the instrument behavior and data handling were tested. Everything worked just fine.

2.4.2 A space odyssey.

Originally, EURECA was intended to be launched in early 1988. However, due to the CHALLENGER disaster in 1986, the entire Shuttle program had to be rescheduled, and the launch was postponed for 3 years. Finally, on 31 July 1992 at 13:56 GMT, EURECA was launched onboard the US Space Shuttle ATLANTIS (flight STS-46) from the launch platform at Kennedy Space Center, Florida [Young, 1992]. According to schedule, it should have been released about 18 hours after launch, but the deployment was delayed due to communication problems. Finally, EURECA was released from the Shuttle cargo bay on August 2, 1992 at 8:07 am BST, 426 km above the southern Pacific Ocean, northwest of New Zealand, by ESA astronaut Claude Nicollier (Fig. 2.11).

EURECA ran into problems when its thrusters fired that day to place it in its operational orbit, 508 km above Earth. After six minutes of thrust, EURECA began to pitch over unexpectedly, and the burn was aborted, leaving the carrier in an elliptical orbit with an average altitude of 442 km, some 65 km below the planned altitude. A few days later, problems were resolved, and the spacecraft was boosted to its final mission orbit on August 7, 1992, inclined at 28.5 degrees with respect to the Earth's equator.

Flight operations were conducted from the EURECA Operation Control Center (OCC) at the European Space Operations Center (ESOC) in Darmstadt, Germany. The actual telemetry data acquisition, telecommanding and tracking support was provided by two ground stations, one in Maspalomas, Gran Canaria, and the other in Korou, French Guiana, while a third station in Perth, Australia served as a backup station.

Scientific operations were started on August 10, 1992 and continued according to plan until the end of January 1993. On January 26, the freon cooling loop was deactivated, concluding the baseline scientific operations. However, some experiments, not critically depending on temperature (including the ORA UV-visible module) could remain active.

The return of the instrument was initiated on May 20, 1993. Until June 8, a series of orbital transfer maneuvers was performed to bring EURECA into a lower orbit. On June 21, the shuttle ENDEAVOUR (flight STS-57) was launched for the EURECA retrieval. Grappling of EURECA was performed on July 24, and on July 1, 1993, after almost 11

Figure 2.11: EURECA, as it is being released from the space shuttle ATLANTIS cargo bay [Daimler-Benz Aerospace, 1996].

months in space, EURECA landed safely on Earth at Kennedy Space Center. A few unforeseen events occurred during the mission, but in general the experimental operation was very successful.

2.4.3 ORA measurements - the data

While in orbit, several types of data (measurements, satellite position etc.) were suitably transformed into telemetry packets by the processor board, and transmitted to the ground station. Seven different types of telemetry packets were defined:

- Type 1 data packets, containing data concerning captured events (sunset and sunrise occultation data).

- Type 2 data packets containing a copy of the EURECA attitude data.

- Housekeeping data packets containing data about the health of the ORA hardware.
- Memory dump packets containing a copy of the contents of a block of ORA memory.
- Acknowledgement packets, containing information about the success of received telecommand packets.
- Exception packets to indicate abnormal behavior of ORA.
- Packets to report successful progress in the execution of ORA software.

2.4.4 The ORA performance.

ORA in action.

The ORA instrument was activated on August 11, 1992, and operated continuously for 10 months, taking measurements at almost every sunset and sunrise (a few times the measurements were interrupted due to operational requirements).

During the mission, a gradual signal decrease took place in three of the UV-visible channels at 259, 340 and 385 nm. This signal deterioration was not dramatic in view of the high resolution of the ADC boards. Furthermore, the self-calibrating property of the instrument ensures that the derived constituent profiles are not affected by this temporal signal variation, although the relative measurement error will increase.

At the end of January 1993, the freon cooling loop of EURECA was deactivated. This again resulted in a small signal decrease, probably caused by a temperature dependence of the electronics. Again, self-calibration ensures that this has no effect on the measurement results.

On May 14, 1993, EURECA was commanded to Dormant Mode, with a pointing accuracy of roughly $\pm 5^{\circ}$, so ORA operations were ended at this stage. In total, almost 7000 sunsets and sunrises were recorded.

Post-flight findings.

After the return to Earth, teams at ESTEC examined EURECA to investigate the effects of space radiation damage and dust/particle impact.

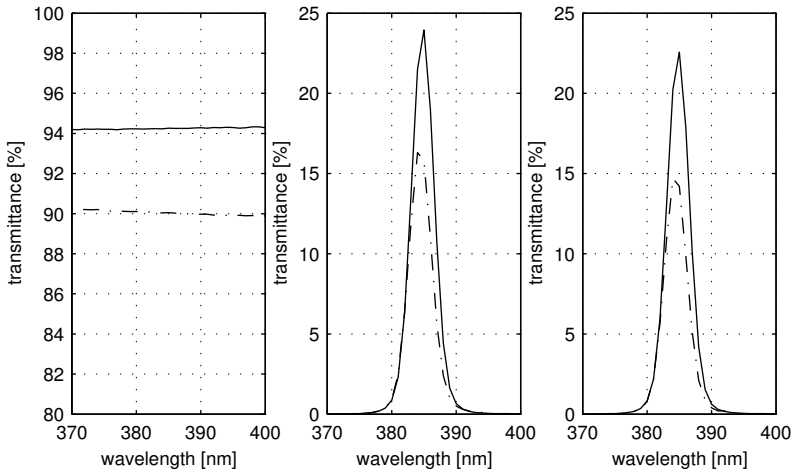

Figure 2.12: Transmittance degradation of the 385 nm components, as measured at ESTEC. Solid lines and dashed lines repectively represent pre-flight and post-flight values. Left: quartz window. Middle: main filter. Right: window plus filter.

For ORA, it was found that the signal decrease in the ultraviolet channels was mainly caused by a decrease in transmittance of the quartz windows and the main interference filters (Fig. 2.12). A *shift* in the central wavelength of some filters was also observed. Luckily, the deviations (usually less than 1 nm) were small compared with the filter bandwidths. Most likely, the degradation has been caused by radiative effects on the glue that was used in the construction of the windows and filters.

2.5 A final remark.

ORA is an experiment in the true sense of the word. It measures the properties of the Earth's atmosphere. But we have to highlight another aspect. It is also an experiment because it is an *experimental device*. A large part of the labor and headaches of the scientists and engineers involved in the ORA project concerned the question: *will it work?* Can it deliver quality data, considering the limitations? We have to keep in mind that ORA is relatively simple, and therefore low-cost: compared

with other occulation experiments (with mind-dazzling price tags), ORA is reduced to the basics: no Sun tracking device, no high-resolution CCD arrays, and a large field of view.

So, will it deliver good data? The answer lies ahead, but we are not there yet. First we need the basic theoretical tools to start the data processing, and this forms the content of the next chapter.

Chapter 3

General scientific concepts.

The ORA experiment is a solar occultation experiment. This means that it measures solar radiation, after it has passed through the Earth's atmosphere. In order to derive the atmospheric quantities, we have to understand the properties of this radiation and the atmospheric processes that alter these properties.

When solar radiation penetrates the outer layers of the Earth's atmosphere it becomes attenuated, through atmospheric absorption, scattering and refractive dilution, while the direction of the beam is altered by atmospheric refraction. Some part of the solar radiation remains and leaves the atmosphere. It is this transmitted part that finally reaches the ORA detector where it is measured.

But before we start describing the radiative processes, it is perhaps useful to repeat some concepts that are being used in modern electromagnetic wave theory to formulate radiation characteristics in a mathematical way.

3.1 Electromagnetic waves

The well-known theory of James Clerk Maxwell was the first unification theory, identifying electric and magnetic forces as two different appearances of the same electromagnetic force. Furthermore, the theory predicts the existence of electromagnetic waves: perpendicular electric and magnetic fields (**E**, **H**), varying in time and space, and characterized by

a *frequency* ν (or *angular frequency* $\omega = 2\pi\nu$), the *wavenumber vector* $\mathbf{k}$ and field *amplitudes* $\mathbf{E_0}$ and $\mathbf{H_0}$.

In order to describe the interaction of matter and radiation, the Maxwell theory[1] is not sufficient. We need supplementary *constitutive relations*, describing the response of matter to electromagnetic fields. These responses are quantified by the *conductivity* σ, the *magnetic permeability* μ and the *electric permittivity* ϵ: the so-called *phenomenological coefficients*[2].

When the electric and magnetic fields depend on time, these phenomenological coefficients are assumed to be depending on frequency. This is a general physical phenomenon: the response of a mechanical system to a periodic driving force is frequency dependent.

Matter needs time to respond to a time-varying applied electromagnetic field. Therefore, the response is not direct: the matter experiences a time phase lag. The easiest way to describe this mathematically is by introducing *complex* phenomenological coefficients. By doing so, we also introduce *absorption* of electromagnetic energy in matter: at a resonant frequency, the response of the system will not become infinite.

We have not yet spoken about the velocity of electromagnetic radiation. In vacuo, the velocity equals $c_0 = 1/\sqrt{\epsilon_0\mu_0}$. In a medium, we can relate the velocity of radiation to c_0 by introducing the *refractive index*:

$$N = \frac{c_0}{c} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \tag{3.1}$$

and since ϵ and μ are complex, N is also complex[3]:

$$N = n + ik \tag{3.2}$$

with $n, k > 0$. A spatial gradient in the real part n leads to the actual refraction (the bending of light caused by a change in velocity). The imaginary part k on the other hand represents absorption in the medium (see the above discussion).

[1] The Maxwell equations provide 4 field vector equations, relating the electric field $\mathbf{E}$, the electric displacement $\mathbf{D}$, the magnetic induction $\mathbf{B}$, the magnetic field $\mathbf{H}$, the charge density ρ_F and the current density $\mathrm{J_F}$.

[2] Alternatively, μ and ϵ can be replaced by the magnetic and electric susceptibilities χ_M and χ_E

[3] A notation issue: while wavenumbers will be represented by an upright k, the imaginary part of N is written in italics: k.

Consequently, the wavenumber will also be complex:

$$\mathrm{k} = \mathrm{k}' + i\mathrm{k}'' = \frac{\omega}{c} = N\frac{\omega}{c_0} = N\mathrm{k}_0 \tag{3.3}$$

Let us now discuss the energy of an electromagnetic wave. It is customary to introduce the *Poynting vector*:

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}$$

The Poynting vector comes in handy when we want to describe the propagation of electromagnetic energy. $\mathbf{S}$ points in the direction of propagation, and the magnitude of $\mathbf{S}$ is defined as the *irradiance*[4] I, with a dimension of energy per time and area [Wm^{-2}]:

$$I = |\mathbf{S}| \tag{3.4}$$

Before we continue, one remark: from now on, we will implicitly assume that the Poynting vector and the irradiance are time-averaged quantities. Indeed, $\mathbf{S}$ is a vector that fluctuates very rapidly in time, and it is impossible for common detectors to measure this fluctuation. A photodiode always measures time-averaged irradiances.

The *energy*, carried per time unit through an imaginary surface A is given by the *flux* of the Poynting vector:

$$W = \int_A \mathbf{S} \cdot \mathrm{d}\mathbf{A}$$

3.2 Solar irradiance

3.2.1 The Sun approximates a black body

The Sun is a gaseous sphere, with a radius of $6.96\,10^5$ km. It consists of three parts H and one part He. Nuclear fusion is the heating source, and at present, the main mechanism for energy transfer from the core to the surface is believed to be electromagnetic radiation. Most of the radiation, received on Earth, is emitted by the outer layer, the *photosphere*, which has a thickness of approximately 500 km.

The solar radiative output is well approximated by *blackbody* radiation. This theoretical concept expresses the thermodynamic equilibrium

[4] Also frequently labelled as the *intensity* of a wave.

of radiation and matter in coexistence. In this equilibrium, all emitted radiation is assumed to be absorbed again at an equal rate. The equilibrium (monochromatic) irradiance is expressed by the *Planck function* [$\mathrm{Wm^{-2}m^{-1}}$], that is independent of the constituency of the body and depends only on the wavelength and the temperature:

$$I_{[\lambda,\lambda+\mathrm{d}\lambda]} = \frac{2\pi c^2 h \lambda^{-5}}{e^{hc/\lambda k_b T} - 1}$$

where c is the speed of light, h the Planck constant, λ the wavelength, k_b the Boltzmann constant and T the temperature.

Actually, while blackbody radiation describes a static equilibrium, the Sun is in a dynamic equilibrium: it has a source (the core) and a sink (the photosphere). Nevertheless, in the largest part of the solar body (the layers between the core and the photosphere), an emitted photon will be absorbed again. It is this 'soup' of radiation-matter that finally determines the spectral properties of the solar irradiance.

While the blackbody radiation is able to model the broad-scale spectral output of the Sun, it fails to do so on a small scale. Here we find the different absorption and emission lines of the species that are present in the Sun.

The solar irradiance and the 5777 K blackbody irradiance is shown in Fig. 3.1, together with the 300 K blackbody irradiance, which closely resembles the Earth's (surface + atmosphere) radiative emission. As we can see, most of the solar radiation is concentrated in the visible part of the spectrum (400 to 700 nm), with a maximum at approximately 500 nm. On the other hand, the blackbody emission of the Earth is located in the infrared, with a peak at approximately 10 μm. The important conclusion is that an instrument which measures in the UV-visible (such as ORA), only receives solar radiation (but also an amount of atmospheric straylight).

The total emission [$\mathrm{Wm^{-2}}$] is the monochromatic irradiance, integrated for all wavelengths. The fraction of this irradiance, received on a surface perpendicular to the solar beam at the Earth orbit is called the *solar constant* and it has a value of about 1370 $\mathrm{Wm^{-2}}$ [Seinfeld and Pandis, 1998].

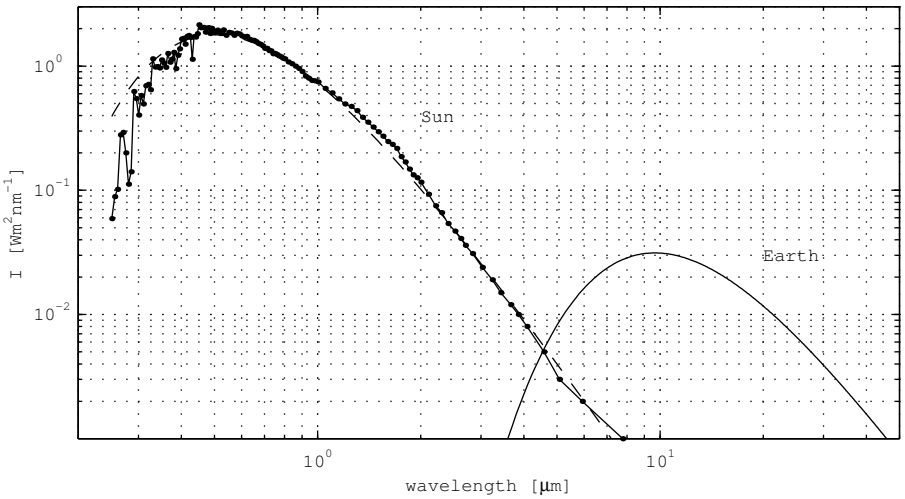

Figure 3.1: Solar irradiance, and 5777 K blackbody irradiance, received at the top of the Earth atmosphere. For comparison, the Earth's 300 K blackbody irradiance is also shown.

3.2.2 Solar variability

Is the solar constant really a constant? Not quite. First, long term variations (several thousands of years) exist because of the changing Earth orbit (excentricity, precession), and these variations induce long-term climatological changes called *Milankovitch cycles*. Also, on a much shorter timescale, we have the 11-year solar cycle, associated with solar activity and the presence of Sun spots. Superimposed on this cycle is the 27 day rotation period of the Sun. During one period of rotation, active regions are successively hidden or visible from the Earth. The short-wavelength irradiance (x-rays, UV $<$ 190 nm) can vary extremely because of these cycles. At longer wavelengths , however, variations become very small. The spectral range between 300 nm and 10 μm occupies 99 % of the total solar constant. Long term measurements show a very small variability (0.1 % for the solar constant in the period of 1979 to 1990). For our purposes in the UV-visible, it is therefore safe to assume that the solar output is constant during one occultation.

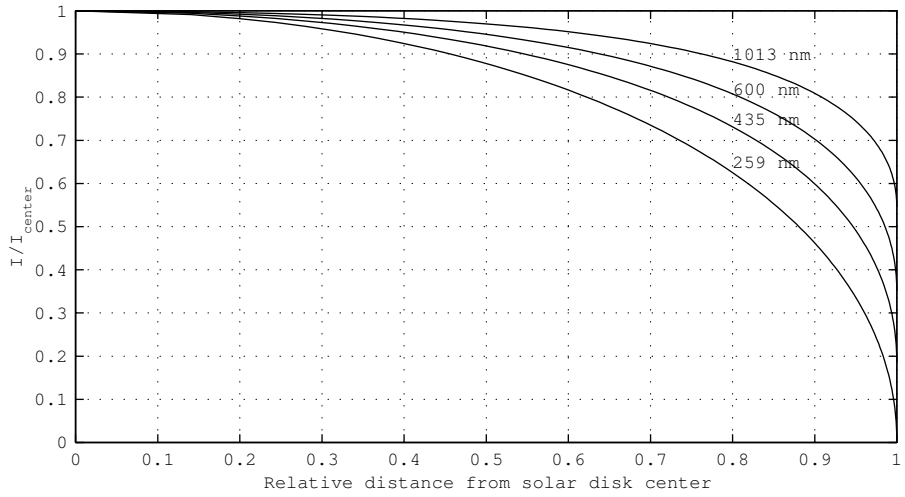

Figure 3.2: Relative solar irradiance versus normalized radial distance from solar disk center, for some ORA wavelengths.

3.2.3 Solar limb darkening

When we look at the Sun, the solar irradiance that we receive is not uniformly distributed across the solar disk. The effect is caused by the temperature gradient in the outer layers of the Sun, combined with geometrics. When we look at the center of the disk, we receive radiation coming from deeper (hotter) layers, compared with a view towards the edge of the solar disk. According to the Planck function, hot layers emit more radiation, and this radiation is shifted towards smaller wavelengths. The result: a gradual, wavelength-dependent decrease of solar output when we shift our focus from the center to the edge of the solar disk (Fig. 3.2, [Allen, 1985]).

3.3 Atmospheric absorption and scattering

3.3.1 Extinction = scattering + absorption

When solar radiation penetrates the Earth's atmosphere, an interaction occurs. First, molecules and particles *absorb* some part of the radiation, and convert it into *heat* (vibrational, rotational and translational

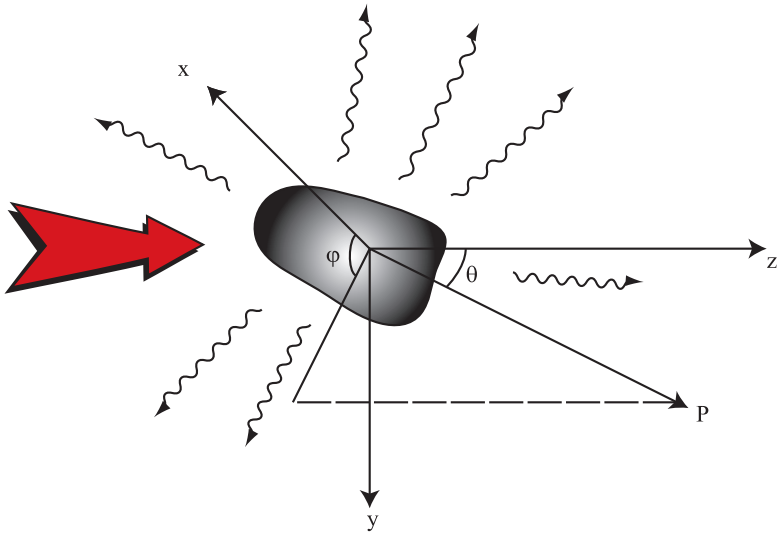

Figure 3.3: Extinction of incident light by a particle/molecule is formed by a combination of absorption inside the particle/molecule and scattering in all directions. The scattering angle θ and azimuth angle ϕ are indicated.

motion). Equally important is the *scattering* process: the incoming radiation forces molecules and particles to emit secondary, scattered waves (see Fig. 3.3). The combined process is called *extinction*. In summary, the presence of a particle or molecule removes energy from incident radiation by absorption and scattering, and this energy rate is equal to:

$$W_{\text{ext}} = W_{\text{abs}} + W_{\text{sca}} \tag{3.5}$$

We can now quantify these effects by introducing the extinction, absorption and scattering cross-section for the particle or molecule:

$$C_{\text{ext}} \equiv \frac{W_{\text{ext}}}{I_{\text{i}}} \quad C_{\text{abs}} \equiv \frac{W_{\text{abs}}}{I_{\text{i}}} \quad C_{\text{sca}} \equiv \frac{W_{\text{sca}}}{I_{\text{i}}} \tag{3.6}$$

with I_{i} the irradiance of the incoming radiation. Cross sections have the dimension of an area. From Eqns. 3.5 and 3.6, it is clear that

$$C_{\text{ext}} = C_{\text{abs}} + C_{\text{sca}} \tag{3.7}$$

The definition of a cross-section forms an analogy with the removal of radiation from a beam, incident on an object. The size of the *shadow*

that is cast by the object is proportional to the geometrical cross-section area of the object, i.e. the area of the projection of the object on a surface, perpendicular to the beam. Note, however, that this is only an analogy: we will discuss later how the extinction cross section of a particle can be much smaller or larger than its geometrical cross section. The analogy between the macroscopic and the (sub)microscopic should not be taken too strictly. Although similar, scattering cannot be described by simply assuming reflection and transmission processes.

Nevertheless, it is customary (for particles) to use cross section efficiencies, that express the ratio of the optical cross sections to the geometric cross sections. In the case of a spherical particle with radius a, we have:

$$Q_{\text{ext}} \equiv \frac{C_{\text{ext}}}{\pi a^2} \quad Q_{\text{abs}} \equiv \frac{C_{\text{abs}}}{\pi a^2} \quad Q_{\text{sca}} \equiv \frac{C_{\text{sca}}}{\pi a^2} \tag{3.8}$$

and, by the use of Eqn. 3.7, this leads to

$$Q_{\text{ext}} = Q_{\text{abs}} + Q_{\text{sca}} \tag{3.9}$$

Particles and molecules in general do not scatter light isotropically, and the scattered intensity will thus depend on the *scattering angle* θ and the *azimuth angle* ϕ (Fig. 3.3). The angular dependence of scattering is expressed by the *phase function*:

$$p(\theta, \phi) \equiv \frac{I_{\text{sca}}(r, \theta, \phi)}{\int_\Omega I_{\text{sca}}(r, \theta, \phi)\, \mathrm{d}\Omega} \tag{3.10}$$

with $I_{\text{sca}}(r, \theta, \phi)$ the irradiance scattered in a direction (θ, ϕ) at a distance r from the scatterer, and $\mathrm{d}\Omega = \sin(\theta)\, \mathrm{d}\theta\, \mathrm{d}\phi$ an infinitesimal solid angle around the direction (θ, ϕ). Integrating the phase function on a sphere, we see that it is properly normalized:

$$\int_\Omega p(\theta, \phi)\, \mathrm{d}\Omega = 1 \tag{3.11}$$

3.3.2 The optical extinction theorem

All kinds of seemingly different scattering types, whether it concerns acoustic waves, electromagnetic waves or elementary particles, show a very curious phenomenon (mathematically expressed by the optical theorem): extinction depends only on the scattering amplitude *in the forward direction* ($\theta = 0°$). This appears to be in conflict with what we have

just described, because extinction is the combined effect of absorption in the particle or molecule, and scattering in *all* directions.

But extinction really can be seen as formed by the destructive interference between the incident and the forward scattered light. Conservation of energy then requires that the light removed from the incident beam by this interference is accounted for by scattering in all directions and absorption in the particle.

An important consequence of this phenomenon is, that extinction is a well-defined observable quantity: we only need to perform a measurement in the forward direction to fully derive the extinction. However, in general an extinction measurement alone cannot deliver information on the relative importance of absorption and scattering. Other measurements (at different angles) or prior information are needed in order to infer the separate contributions of the two processes.

3.3.3 Extinction by multiple particles or molecules

Given the extinction behaviour of one particle or molecule (characterized by C_{ext}), what is the behaviour of the radiation in the forward direction when it travels through a collection of particles or molecules? We have to make a subdivision at this point.

Identical particles/molecules

Consider a slab of identical particles/molecules with thickness $\mathrm{d}x$, that is illuminated by a beam of radiation (Fig. 3.4). The energy that is transmitted per second through a surface area A at $x + \mathrm{d}x$ equals:

$$W(x + \mathrm{d}x) = W(x) - W_{\text{ext}}^{\text{tot}} \tag{3.12}$$

or, using Eqn. 3.6,

$$W(x + \mathrm{d}x) = W(x) - NC_{\text{ext}}I(x) \tag{3.13}$$

where N equals the number of particles/molecules within the cylinder. Using the expression for energy rate, $W = IA$, we obtain

$$(I(x + dx) - I(x))A = -NC_{\text{ext}}I(x) \tag{3.14}$$

or

$$\frac{I(x + dx) - I(x)}{\mathrm{d}x} = -\mathcal{N}C_{\text{ext}}I(x) \tag{3.15}$$

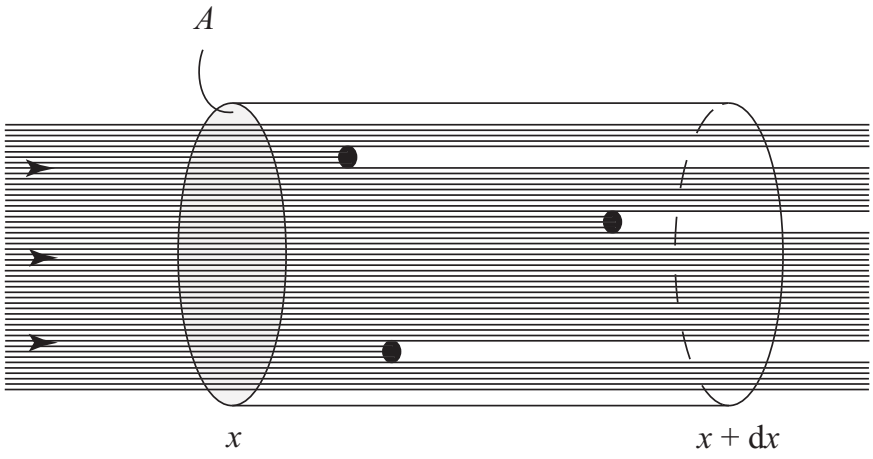

Figure 3.4: Extinction by a number of particles and/or molecules

where we have used the number density $\mathcal{N} = N/V$ [molecules/particles m^{-3}], and the cylinder volume $V = A\mathrm{d}x$. Using derivatives, this leads to the differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln I(x)) = -\mathcal{N}C_{\mathrm{ext}} \tag{3.16}$$

We now can define the *extinction*, *absorption*, and *scattering coefficients* $[\mathrm{m}^{-1}]$:

$$\beta_{\mathrm{ext}} \equiv \mathcal{N}C_{\mathrm{ext}} \quad \beta_{\mathrm{abs}} \equiv \mathcal{N}C_{\mathrm{abs}} \quad \beta_{\mathrm{sca}} \equiv \mathcal{N}C_{\mathrm{sca}} \tag{3.17}$$

that are related by

$$\beta_{\mathrm{ext}} = \beta_{\mathrm{abs}} + \beta_{\mathrm{sca}} \tag{3.18}$$

Using the definition for the extinction coefficient, and integrating Eqn. 3.16 between x_1 and x_2, we have

$$I(x_2) = I(x_1) \exp\left[-\int_{x_1}^{x_2} \beta_{\mathrm{ext}}(x)\mathrm{d}x\right] \tag{3.19}$$

The *optical thickness* (no dimension) is defined as the integrated extinction along the optical path:

$$\delta \equiv \int_{x_1}^{x_2} \beta_{\mathrm{ext}}(x)\mathrm{d}x \tag{3.20}$$

If the medium that contains the particles/molecules is homogeneous, Eqn. 3.19 simplifies to

$$I(x_2) = I(x_1) \exp\left[-\beta_{\text{ext}}(x_2 - x_1)\right] \tag{3.21}$$

which is the well-known attenuation law of *Beert-Lambert*.

We can now express these concepts in a directly observable quantity. The *transmittance* (no dimension) is defined as the fraction of the radiation that is not removed from the incident beam:

$$\tau \equiv \frac{I(x_2)}{I(x_1)} = \exp(-\delta) \tag{3.22}$$

By definition, the transmittance has a value between 0 and 1.

A diversity of molecular species

If l different types of molecules are present, then the second term on the right hand side of Eqn. 3.13 is not valid anymore. The total energy that is removed per second is now:

$$\begin{aligned} W_{\text{ext}}^{\text{tot}} &= N_1 W_{\text{ext}}^1 + \cdots + N_l W_{\text{ext}}^l \\ &= (N_1 C_{\text{ext}}^1 + \cdots + N_l C_{\text{ext}}^l) I_{\text{i}} \end{aligned} \tag{3.23}$$

Consequently:

$$\begin{aligned} \beta_{\text{ext}}^{\text{tot}} &= \mathcal{N}_1 C_{\text{ext}}^1 + \cdots + \mathcal{N}_l C_{\text{ext}}^l \\ &= \beta_{\text{ext}}^1 + \cdots + \beta_{\text{ext}}^l \end{aligned} \tag{3.24}$$

Therefore, extinction coefficients are *additive*.

A diversity of particles

The previous formalism is discreet: every term in Eqn. 3.24 respresents the extinction coefficient for one molecular constituent. For particles, we have to think in terms of continuous populations: the diversity in particle sizes[5] is in general huge. Therefore, we move to the continuous case:

$$\beta_{\text{ext}}^{\text{tot}} = \int_0^\infty \frac{\mathrm{d}\mathcal{N}(r)}{\mathrm{d}r} C_{\text{ext}}(r)\, \mathrm{d}r \tag{3.25}$$

[5]Not only sizes differ within particle populations. Chemical composition and shape can also vary greatly from one particle to another. But we will not treat these cases for reasons that will become clear later.

where $\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}r}$ (formally) represents the number density of particles with size in the interval from r to $r + \mathrm{d}r$. This function is known as the *particle size distribution*.

Validity of the Beert-Lambert law

In the previous derivations, we have implicitly used an important assumption: the results are valid only in the *single-scattering* regime. In general, a particle or molecule feels the incident field *and* the secondary field, scattered by all other particles and molecules. Single scattering means that this secondary field is negligible compared to the incident field.

In practice, a clear criterion to check if this assumption is valid does not exist. Sometimes there is no doubt: light extinction in clouds, for example, cannot be adequately described by the Beert-Lambert law, because *multiple scattering* plays an important role. As a rule, multiple scattering can be neglected when the number of particles and the scattering cross-sections are locally relatively small (Bohren and Huffman [1993]), or otherwise stated, when $\delta \ll 1$ across small distances, a sufficient condition that is fulfilled for most atmospheric species.

3.3.4 Conclusion

The concepts that we have described in this subsection build the frame in which we can model the interaction of radiation with atmospheric particles and molecules. Only three quantities (of which one is redundant) remain to be specified: the extinction, scattering and absorption cross sections. In the case of molecules, they are exclusively determined by the building stones (atoms) and the geometry of the molecular structure. For particles, the situation is more complex, with particle size and shape as additional factors.

3.4 Molecular absorption

Absorption (and emission) of radiation by molecules is —on a basic level— identical to absorption by single atoms. A photon hits a molecule, which absorbs the photon. After the process, the molecule finds itself in an excited state with higher energy than the initial state. Since

the energy levels of bound states are discrete, the absorption spectrum will also consist of discrete lines. Excitation towards an unbound state (dissociation of the molecule or ionisation), on the other hand, results in a continuous spectrum.

The analogy between absorption by atoms and molecules ceases to exist at this point. Atomic spectra are completely described by electronic and nuclear transitions. The situation is more complex with the case of molecules, because they possess a spatial geometrical structure. This structure leads to additional energy states.

3.4.1 Molecular transitions

Rotational transitions

First of all, in the gas phase, molecules can rotate. In quantum physics, angular momentum is quantized, resulting in discrete rotational energy levels. These levels can be degenerate, and splitting occurs in the presence of an external field (e.g. the Stark effect), and the distortion of the molecule by centrifugal forces. The necessary pre-condition for a transition between two rotation levels to occur is the presence of a permanent molecular dipole moment. Pure rotational absorption bands can be found in the microwave region of the electromagnetic spectrum.

Vibrational transitions

Second, the nuclei that form molecules can vibrate. The associated energy levels are also discrete (compare with the energy levels of a harmonic quantum oscillator). Transitions between vibration energy levels occur at the condition that the molecule can have a changing dipole moment (a permanent dipole is not necessary). Worth mentioning is that a vibration transition usually induces a rotation transition. In a classical way, this can be understood by considering that a vibration alters the moment of inertia of the molecule and hence its rotation speed. Usually then, vibration spectra are accompanied by rotation branches. Vibration bands are located in the infrared and near infrared spectrum range.

Electronic transitions

And third, the redistribution of electrons that surround the nuclei of the molecule results in electronic transitions. Because the mass of an electron is much smaller than the mass of a nucleus, the latter can be considered stationary during an electronic transition (the Franck-Condon principle). After the transition, however, the Coulomb force distribution inside the molecule has changed, inducing vibrations of the nuclei around their stationary location. So, an electronic transition can lead to a vibrational transition, which in turn can cause a rotational transition. Electronic transitions produce absorption bands in the UV-visible.

Line broadening

Until now, we have discussed discrete absorption lines. However, these lines are not infinitely thin, but are broadened by three mechanisms:

- Doppler broadening. A molecule moves with respect to the light source, and by Doppler shift, radiation with a slightly different wavelength than the transition wavelength can be absorbed by the molecule. It is clear that the Doppler broadening of a spectral line is temperature dependent.

- Pressure (collisional) broadening. The lifetime of an excited state is reduced through collisions with other molecules, and this increases the uncertainty on the energy of the level. Result: a broadened line. Since the broadening is induced by collisions, it is pressure dependent.

- Natural broadening. The Heisenberg principle gives energy levels a natural quantum uncertainty: $\Delta E \approx \hbar/\Delta t$.

Although natural broadening is present, the dominant mechanisms are pressure and Doppler broadening.

Line intensities

One last remark concerning absorption line strengths. The actual intensity of a line in a sample is determined by how well the initial energy

state of the transition is populated. Statistically, the energy levels of the molecules in the sample are populated according to the Boltzmann distribution. Therefore, line intensities are in general temperature dependent.

Unbound states

While transitions between bound energy states produce lines and are governed by selection rules, transitions to unbound states are not. If the frequency of the incoming photons is high enough, the molecule can be dissociated or ionized, and the resulting spectra are continuous (photon energy is converted to non-quantized translational energy of the molecule fragments or electrons).

Curiously, dissociation can take place at lower frequencies. Some excited electronic states possess the same amount of potential energy as an unexcited dissociative state. Excitation to this level results in *predissociation*. Thus, at energies that are lower than the dissociation energy, the discreet spectral lines are heavily blurred by predissociation.

3.4.2 UV-visible spectra

Let us now focus on the spectral range of our interest. We have seen that electronic transitions are accompanied by vibrational and rotational transitions. The result is an incredible maze, built up from a massive amount of possible transitions. The associated spectra therefore consist of a huge number of discrete lines, broadened by the already mentioned mechanisms, and densely packed together into continuous bands. Spectral lines in the infrared (vibration) and microwave (rotation) regions are more or less distinguishable, and are in practice modeled using theoretical line shapes. Theories for UV-visible (electronic) transitions exist but are very complex. Rather, absorption cross-sections (as defined in the previous section) are experimentally determined, usually at different temperatures. Several databases with absorption cross sections are available, of which HITRAN[6] is probably the best known example.

[6]HIgh resolution TRANsmission molecular absorption database.

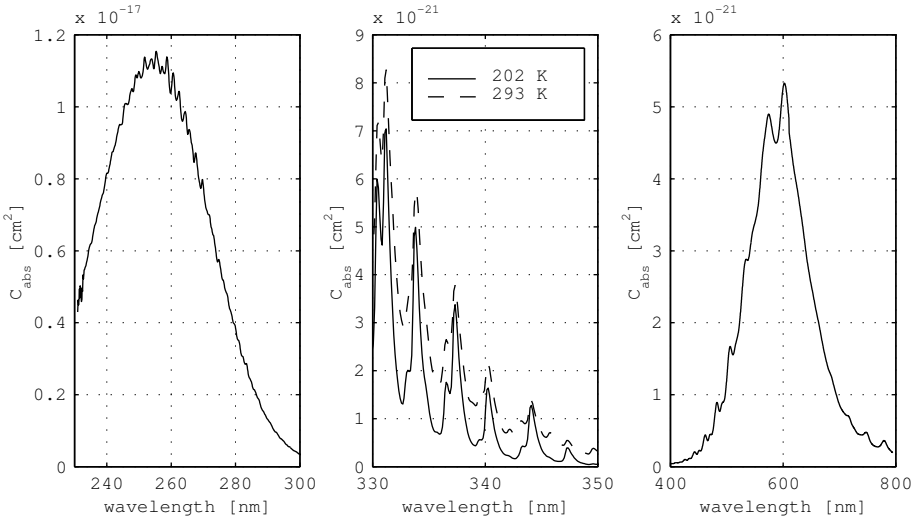

Figure 3.5: Ozone cross sections in the ultraviolet and visible wavelength regions. Left: the Hartley bands. Center: the Huggins bands. Right: The Chappuis bands.

Ozone

In the ultraviolet range, between 200 and 300 nm, we find the *Hartley band*, centered at 255.3 nm. Practically all of the UV radiation is absorbed by this band in the stratosphere. It is weakly dependent on temperature. The smooth structure originates from the fact that absorption by ozone below 300 nm is associated with dissociation and predissociation. From 300 to 340 nm, the strongly temperature dependent *Huggins band* can be observed. The *Chappuis band* is located in the visible region, centered at roughly 600 nm. Although relatively weak, the Chappuis absorption band is more important for the heating of the middle and lower atmosphere, because the solar irradiance peaks at this wavelength region. Ozone cross-sections are shown in Fig. 3.5.

NO_2

Nitrogen dioxide is an extremely important trace gas, because it absorbs in the entire UV-visible region in the lower atmosphere. Between 300 and 370 nm, more than 90 % of the molecules will dissociate into NO and O.

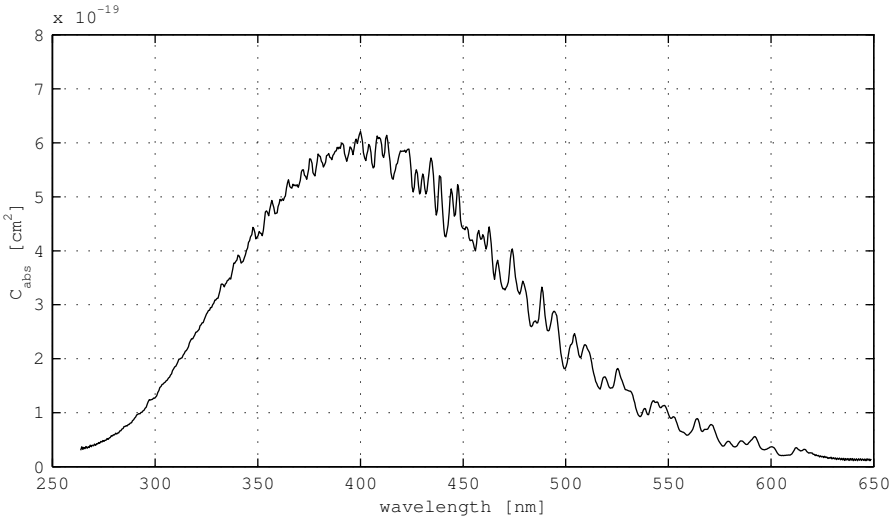

Figure 3.6: NO_2 cross sections in the ultraviolet and visible wavelength regions.

This amount decreases slowly with wavelength, and above approximately 420 nm, dissociation does not occur. NO_2 cross-sections are shown on Fig. 3.6.

H_2O

In the spectral range from 932 nm to 961 nm, a vibration-rotation spectrum of water vapor is found (the so-called $\rho\sigma\tau$ bands), consisting of thousands of discrete absorption lines, with differing absorption strengths and half-widths. It is one of the relatively strong H_2O absorption bands in the near-infrared region.

3.5 Molecular (Rayleigh) scattering

We have seen in Chapter 1 that air consists mainly of N_2 and O_2 (let us forget about trace gases for the time being). These are linear homoatomic molecules, which implies that they do not possess a permanent dipole moment. However, a static uniform electric field can *induce* a dipole moment by separating the charges inside the molecules. The

induced dipole moment $\mathbf{p}$ can be related to the electric field by the introduction of the *polarizability* α (dimension: volume):

$$\mathbf{p} = \alpha \mathbf{E}_{\mathbf{eff}} \tag{3.26}$$

The effective field $\mathbf{E}_{\mathbf{eff}}$ represents the combination of the applied field and the field, caused by the polarization of the molecules that surround the considered molecule. Furthermore, by introducing the number density of the molecules, $\mathcal{N}$, we obtain the total electric dipole moment $\mathbf{P}$ per unit volume:

$$\mathbf{P} = \mathcal{N}\mathbf{p} = \mathcal{N}\alpha \mathbf{E}_{\mathbf{eff}} \tag{3.27}$$

It is very straightforward [Born and Wolf, 1980] to derive the *Lorentz-Lorenz equation*[7] (for non-magnetic substances, such as air):

$$\alpha = \frac{3}{4\pi\mathcal{N}} \frac{n^2 - 1}{n^2 + 2} \tag{3.28}$$

with n the real part of the gas refractive index. This equation connects Maxwell's phenomenological theory with the atomistic theory of matter.

What happens when the applied electric field varies in time (radiation)? Let us assume that the wavelength of the incident field is much larger than the dimension of the dipole. The field as such can still be considered uniform, and the results above still apply. However, the direction of the applied field changes with time, and so does the dipole moment direction. The induced field, caused by the dipole, therefore also changes in time. Near the molecule, this time-varying field can be considered uniform in space, with the typical $1/r^3$ dipole field character. At long distances, retardation effects come into play, and the field oscillates in space and time — a secondary, scattered wave. The scattered wave has the same frequency as the applied wave (*elastic* scattering), and an expected $1/r$ dependence.

We can now derive this electric field at a far distance r from the dipole at a direction (θ, ϕ). The Poynting vector and the irradiance is then calculated from the electric (and associated magnetic) field. For unpolarized incident light, it turns out that the irradiance is independent

[7] H. A. Lorentz and L. Lorenz discovered the relation independently and practically at the same time (1880, 1881). Quite a coincidence.

of the azimuth angle ϕ. We get:

$$I_{\text{sca}} = \frac{9}{2}\frac{\pi^2}{\lambda^4 \mathcal{N}^2 r^2}\left(\frac{n^2-1}{n^2+2}\right)^2 (1+\cos^2\theta) I_{\text{i}} \tag{3.29}$$

Integrating this equation over a sphere with radius r, we can calculate the total scattered energy, and dividing by the incoming intensity I_{i}, we finally come to the scattering cross-section

$$C_{\text{sca}} = \frac{24\pi^3}{\lambda^4 \mathcal{N}^2}\left(\frac{n^2-1}{n^2+2}\right)^2 \tag{3.30}$$

which is the well-known *Rayleigh law*, with its λ^{-4} wavelength dependence. The physical consequences: in the morning or evening, when we look directly at the Sun (forward direction), we observe a red color, because the smaller wavelengths (blue, violet) are partially scattered away in other directions. Looking at the air above, however, we see that the sky is blue because we observe only scattered light.

The associated phase function, calculated according to Eqn. 3.10 is given by:

$$p(\theta,\phi) = \frac{3}{16\pi}(1+\cos^2\theta) \tag{3.31}$$

and is depicted in Fig. 3.7. The pattern is very familiar, because it resembles the radiation pattern of a linear radio antenna.

To make things just a little bit more complicated, remark that Eqn. 3.26 implicitly assumes the polarizability of a molecule to be *isotropic*, that is to say independent of the orientation of the molecule with respect to the incoming beam. This is in general not true. The anisotropy is taken into account by introducing the *anisotropy factor* ϱ:

$$C_{\text{sca}} = \frac{24\pi^3}{\lambda^4 \mathcal{N}^2}\left(\frac{n^2-1}{n^2+2}\right)^2 \varrho \tag{3.32}$$

and for air, it has a value of 1.06 [Lenoble, 1993].

Why did we neglect to speak about trace gases? In general, of course, trace gases will scatter. However, C_{sca} is proportional to α^2, and the polarizability is in a very good approximation constant. Therefore, the scattering coefficient β_{sca} is proportional to the number density of the gas. Because we are talking about *trace* gases, their scattering coefficients will therefore be completely negligible compared to the air scattering coefficient.

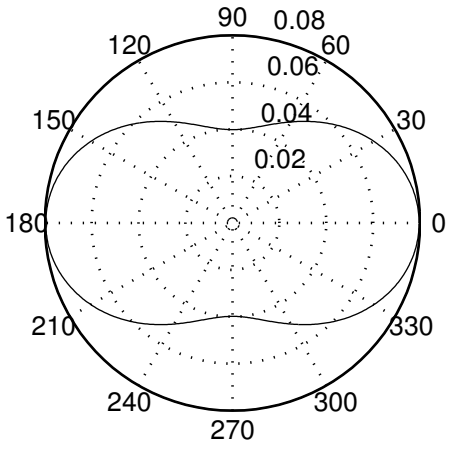

Figure 3.7: The phase function for Rayleigh scattering. Incident (unpolarized) light enters from the left.

3.6 Particle absorption and scattering

3.6.1 The Mie theory

Let us now consider the general case of a particle with a shape, size and composition, embedded in a medium. No approximations concerning the ratio of the size to the wavelength of the incident radiation will be made (as was the case with Rayleigh scattering). Many things can happen: absorption and internal reflection inside the particle, refraction and reflection at the surface of the particle, diffraction at the edge of the particle, etc.[8]. One can intuitively sense that the combination of all these processes is incredibly complex.

Nevertheless, the problem has been solved exactly for *spherical* particles, and the accompanying theory is usually referred to as the *Mie theory*[9].

[8]The concepts 'reflection', 'diffraction' etc. are strictly speaking only valid if we specify the wavelength with respect to the particle size (what does 'reflection' mean if we have a particle that is much smaller than the wavelength?).

[9]After Gustav Mie, who developed it in 1908. Actually, Peter Debye derived similar results at the same time. Neither of them was the first to solve the problem, however, and it is perhaps Lorenz who earns the credits.

We will not give the derivation, which is too complicated. Excellent treatises can be found in [Bohren and Huffman, 1993, van de Hulst, 1981]. Briefly, the calculation is performed on the electromagnetic field ($\mathbf{E_2}$,$\mathbf{H_2}$) outside the particle, and the internal field ($\mathbf{E_1}$, $\mathbf{H_1}$). The external field is the superposition of the incident and the scattered field: ($\mathbf{E_2}$,$\mathbf{H_2}$) = ($\mathbf{E_i}$,$\mathbf{H_i}$) + ($\mathbf{E_{sca}}$,$\mathbf{H_{sca}}$). The Maxwell equations are then solved inside and outside the particle, and conservation of energy is introduced by applying boundary conditions at the surface of the sphere (the tangential components of the fields have to be continuous at the surface).

The Mie theory is exact. No approximations are made regarding the ratio of particle size a to wavelength λ, a quantity that is expressed by the *size parameter*:

$$x = \mathrm{k}a = \frac{2\pi N a}{\lambda_0} = N\mathrm{k}_0 a \tag{3.33}$$

We also need the *relative refraction index*:

$$m = \frac{N_1}{N} \tag{3.34}$$

where N_1 and N are the refractive indices of particle and medium, respectively. The results for the cross sections are:

$$C_{\mathrm{sca}} = \frac{2\pi}{\mathrm{k}^2} \sum_{n=1}^{\infty} (2n+1)(|a_n|^2 + |b_n|^2) \tag{3.35}$$

$$C_{\mathrm{ext}} = \frac{2\pi}{\mathrm{k}^2} \sum_{n=1}^{\infty} (2n+1)\Re(a_n + b_n) \tag{3.36}$$

$$C_{\mathrm{abs}} = C_{\mathrm{ext}} - C_{\mathrm{sca}} \tag{3.37}$$

with the coefficients

$$a_n = \frac{m\psi_n(mx)\psi_n'(x) - \psi_n(x)\psi_n'(mx)}{m\psi_n(mx)\epsilon_n'(x) - \epsilon_n(x)\psi_n'(mx)} \tag{3.38}$$

$$b_n = \frac{\psi_n(mx)\psi_n'(x) - m\psi_n(x)\psi_n'(mx)}{\psi_n(mx)\epsilon_n'(x) - m\epsilon_n(x)\psi_n'(mx)} \tag{3.39}$$

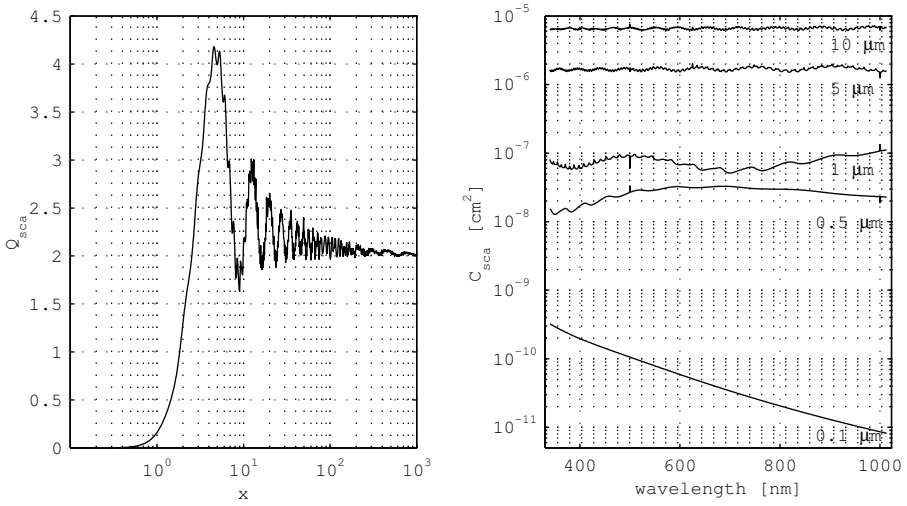

Figure 3.8: Left: scattering cross section efficiencies versus size parameter x for particles embedded in air with relative refractive index $m = 1.43 + i\,0.0$. Right: same data, now expressed as cross-section versus wavelength for different particle sizes a.

The primes ($'$) indicate differentiation with respect to the argument within parentheses. The coefficients are expresses in terms of the *Riccati-Bessel* functions:

$$\psi_n(\rho) = \rho j_n(\rho) \qquad \epsilon(\rho) = \rho h_n^{(1)}(\rho) \tag{3.40}$$

with $j_n(\rho)$ the *Gegenbauer generalization of Poisson's integral* and $h_n^{(1)}$ a *spherical Bessel function of the third kind* (also known as a spherical Hankel function).

A first glance at Eqns. 3.35 and 3.36 does not give much physical insight. This is the drawback of the Mie theory. However, with modern computers, it is very easy to calculate cross-sections relatively fast.

3.6.2 General scattering features

A typical example of a scattering cross-section efficiency (Eqn. 3.8) versus size parameter x can be observed on Fig. 3.8. For a small x, the function is very smooth. As x increases, Q_{sca} exhibits subsequent peaks and valleys that become gradually weaker and finally disappear.

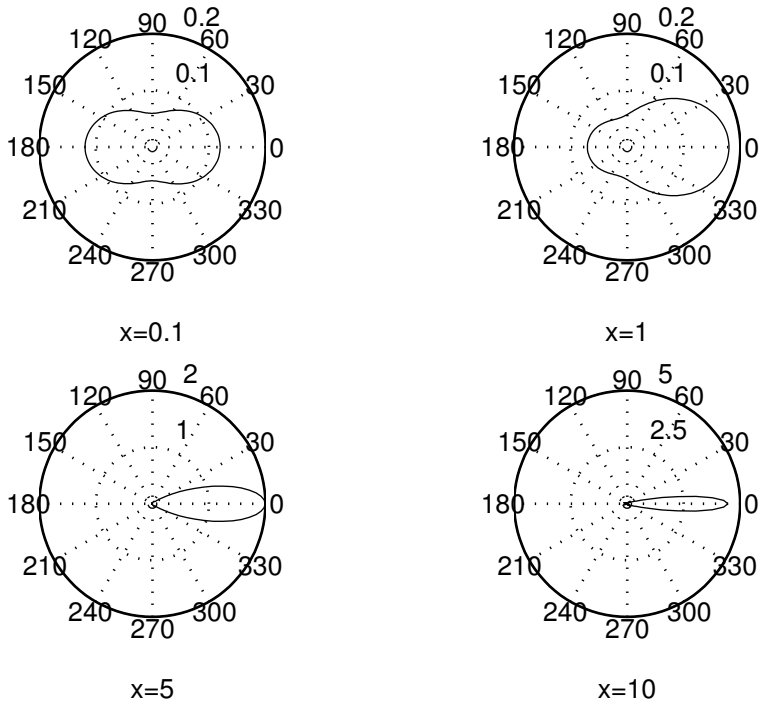

Figure 3.9: Phase functions for different size parameters. Incident (unpolarized) light enters from the left (data as in the previous figure).

We can make things more clear by analyzing the scattering cross-section as a function of wavelength for different particle sizes a (also Fig. 3.8). Small particles have a smooth cross-section that is strongly dependent on wavelength in a monotonic way. Large particles on the other hand exhibit a scattering behaviour that has a lot of fine structure, but a weak global wavelength dependence. Larger particles are thus associated with *flatter* curves.

Fig. 3.9 shows various phase functions for different size parameters. As can be seen, scattering by small particles is very similar to molecular scattering. In general, when size increases, the forward peak becomes more pronounced and can become very large. This is what is observed when looking at the Sun through a dirty car window: forward scattered light of blinding intensity.

3.6.3 Two interesting limits

While the physical content of the Mie formalism is somewhat hidden, it is possible to derive approximations for limits of the size parameter, thereby revealing the physics of the extinction process.

The small particle limit: Rayleigh scattering

Let us consider the case where $|m|x \ll 1$ (a *soft* sphere that is much smaller than the wavelength). The scattering cross section becomes [Bohren and Huffman, 1993]:

$$C_{\text{sca}} = \frac{8}{3}\pi a^2 x^4 \left|\frac{m^2-1}{m^2+2}\right|^2 \tag{3.41}$$

and we immediately recognize the λ^{-4} Rayleigh law. This confirms what we have already discussed: small particles scatter as if they were molecules.

The large particle limit: a paradox

For $x \gg 1$ (particles that are large compared with the wavelength), we have (as can be observed on Fig. 3.8) the curious result:

$$\lim_{x\to\infty} C_{\text{ext}} = 2\pi a^2 \qquad \lim_{x\to\infty} Q_{\text{ext}} = 2 \tag{3.42}$$

Thus: particles much larger than the incident radiation wavelength remove *twice* as much radiation than is (geometrically) incident upon them. This result contradicts common sense, and it is frequently referred to as the *extinction paradox*. The phenomenon, however, is real. It results from the fact that, no matter how large the object (particle), it always has an edge that diffracts radiation beyond the physical boundaries of the object. This additional extinction is exactly equal to the *blocking* by the geometrical cross section. Hence the factor 2. Why then is it a paradox? Because in everyday life, we never observe the full extinction of an object. We have to remember that the total extinction is completely determined by the forward direction ($\theta = 0^\circ$), and *only* this direction. For a detector, this can only be accomplished by measuring from a far distance (as is the case with satellite experiments).

Eqn. 3.42 confirms the fact that extinction by large particles is in effect wavelength-independent, as was already pointed out.

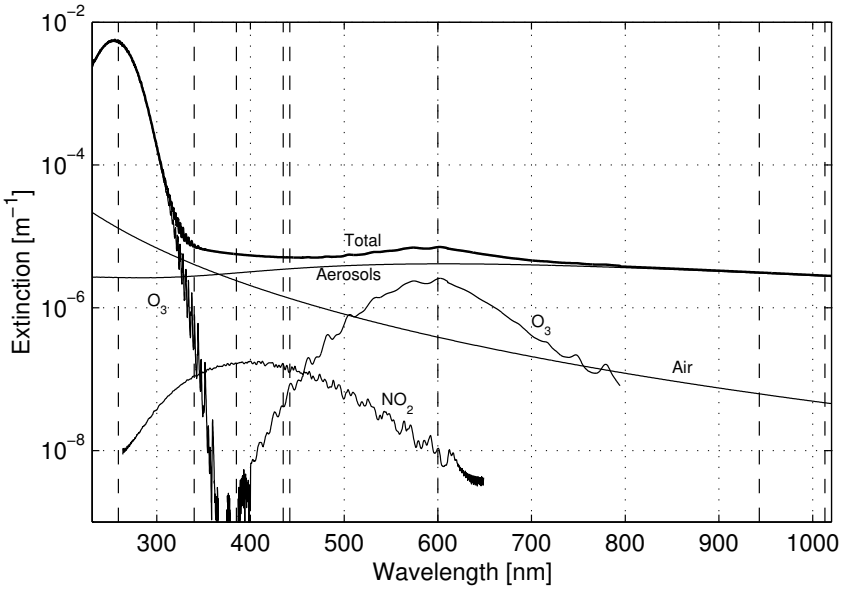

Figure 3.10: Extinction versus wavelength, calculated from gas densities and aerosol properties that are representative for an altitude of 22 km, in high-volcanic conditions. Total extinction is also shown (bold solid line) while ORA channels are indicated by dashed lines.

3.7 The choice of wavelengths - revisited

On Fig. 3.10, the spectral features of extinction by the different species is shown. The trace gas extinctions are calculated from their cross sections and typical density values for an altitude of 22 km. The aerosol values are calculated from the Mie theory, combined with a particle size distribution (see Chapter 6) that is typical for post-volcanic conditions with high aerosol loading in the stratosphere.

Clearly, channel λ_{259} is aimed at measuring O_3 absorption in the Hartley bands. Notice however that this absorption is very strong, and in fact, the channel is only usable when ozone levels are relatively low, i.e. in the mesosphere.

Channels λ_{340} and λ_{385} are primarily sensitive to air and aerosol scattering. The same can be said for λ_{435} and λ_{442}; however, they are located at a maximum and minimum for NO_2-absorption, respectively,

and as a consequence their main purpose is to increase the sensitivity to NO_2 (from 435 to 442 nm, the contributions from other species remain relatively constant).

The purpose of channel λ_{600} is clear: ozone absorption peaks at this wavelength in the Chappuis bands, making this channel suitable to measure ozone levels in the stratosphere.

In the near-infrared region, channel λ_{943} is constructed to measure H_2O absorption, while channel λ_{1013} was mainly intended to measure aerosols.

Two final remarks here. First, the situation, depicted in Fig. 3.10 is valid for an altitude of about 22 km, where the aerosol contribution dominates the total extinction. At larger altitudes, aerosol and air densities are much lower, with the result that the NO_2 and O_3 contributions are more pronounced.

Second, a volcanic eruption is impossible to anticipate when designing a space experiment. It is clear that the (unexpected) large aerosol contribution in Fig. 3.10 highly reduces the experiment sensitivity to the other species.

3.8 Atmospheric refraction

3.8.1 Refraction versus wavelength and altitude

In general, when a ray of light crosses the boundary between two different homogeneous layers, the ray will be bent according to the well-known law of Snell. The amount of bending is dependent on the ratio of the refraction indices associated with the two layers.

Clearly, the refractive index is connected with the ability of molecules to respond to radiation. This was already pointed out when we discussed the Lorentz-Lorenz relation (Eqn. 3.28). Since the polarization α is a constant for a certain type of molecule at a certain wavelength, we can now state:

$$\frac{n^2-1}{n^2+2} = \frac{\mathcal{N}}{\mathcal{N}_0}\frac{n_0^2-1}{n_0^2+2}$$

with $\mathcal{N}_0$ and n_0 the number density and the (real) refractive index at standard temperature and pressure (STP): $T = 288.16$ K, $p = 1013.25$ hPa.

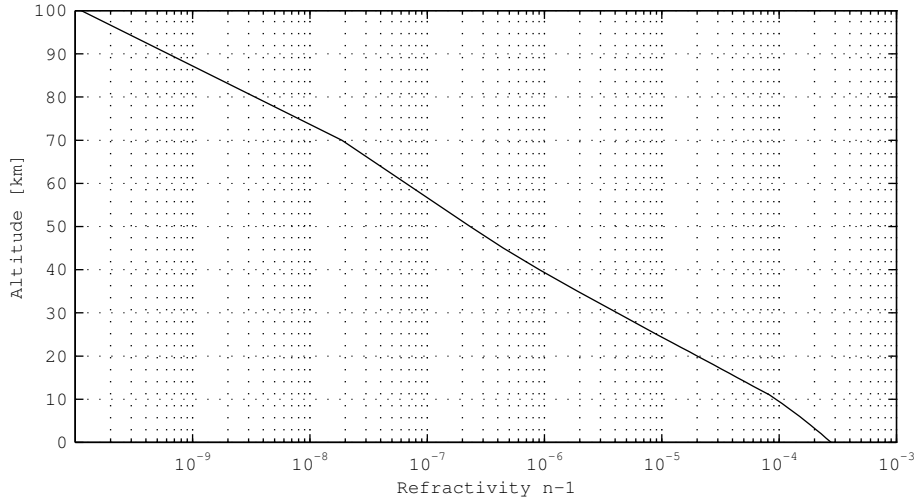

Figure 3.11: Refractivity of air at $\lambda = 600$ nm as function of altitude, for a standard atmosphere (US76).

The refractive index of air is approximately unity. Combined with the ideal gas law, we get:

$$n - 1 \simeq \frac{T_0}{p_0}\frac{p}{T}(n_0 - 1) \tag{3.43}$$

Using this equation, if we know temperature and pressure at a certain point in the atmosphere, we can calculate the air refraction index (or *refractivity* $n - 1$) at this point from the refractive index at STP (Fig. 3.11). The latter is given by the *Edlen formula* (for dry air) [Edlen, 1953], that also includes the wavelength dependence:

$$(n_0 - 1).10^8 = 6432.8 + \frac{2949810}{146 - 1/\lambda^2} + \frac{25540}{41 - 1/\lambda^2}$$

with λ expressed in micrometer. A typical example is shown in Fig. 3.11.

3.8.2 Refraction in an atmosphere with radial dependence

Since the refraction index is dependent on T and p, and since we will assume a radial dependence for all atmospheric quantities at a certain

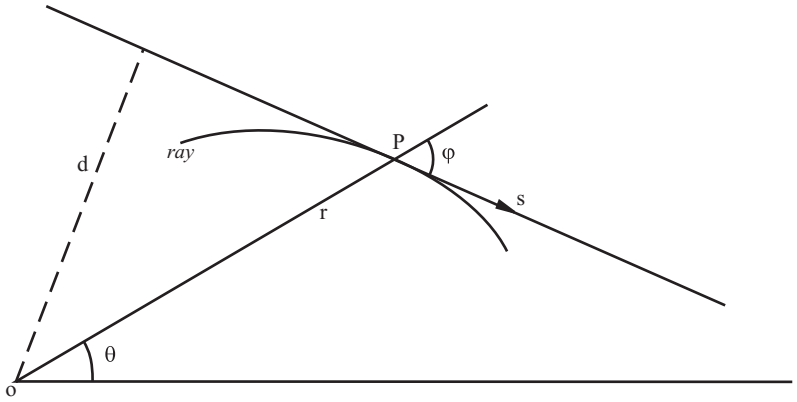

Figure 3.12: Coordinates of a refracted light ray in a radial dependent atmosphere

geolocation, we can derive an expression for the coordinate of every point on the refracted light ray. For a medium with spherical symmetry (as the Earth's atmosphere) it is fairly straightforward to derive [Born and Wolf, 1980]:

$$\mathbf{r} \times \mathbf{s} = \mathrm{c} \qquad \text{(constant)}$$

with $\mathbf{s}$ a pathlength vector lying in the line that is tangent to the ray at the point P with coordinate r (Fig. 3.12). The expression implies that a ray will trace a curve, situated in a plane through the origin O (thus, polar coordinates are suitable). It also leads to the *formula of Bouguer*, $nd =$ constant, an optical analogue to the conservation of angular momentum for a mass moving under the action of a central force.

Using geometrical considerations, we finally come to the differential equation of the light ray:

$$\frac{\mathrm{d}r}{\mathrm{d}\theta} = \pm\frac{r}{\mathrm{c}}\sqrt{n(r)^2 r^2 - \mathrm{c}^2} \tag{3.44}$$

with c the constant already mentioned.

The point where $\frac{\mathrm{d}r}{\mathrm{d}\theta} = 0$ is defined as the *turning point* $(r_\mathrm{t}, \theta_\mathrm{t})$: it is the point on the ray that is closest to the Earth surface. An incident ray will initially be refracted towards the Earth; after the turning point has been passed, the altitude of the ray will increase again. From Eqn.

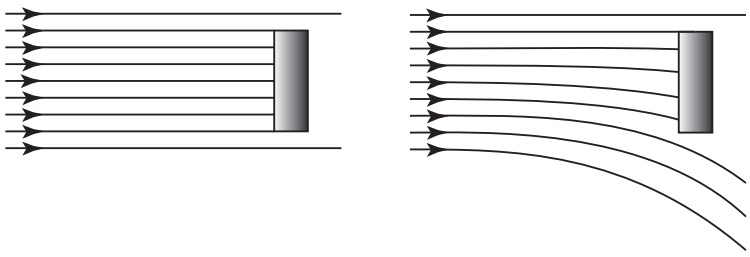

Figure 3.13: The refractive dilution effect. When refraction is present (right), the received flux on the detector is smaller due to the spreading of the light rays.

3.44, it is clear that:

$$\mathrm{c} = n(r_\mathrm{t})r_\mathrm{t}$$

The ray also has a symmetrical shape with respect to the direction of the turning point:

$$r(\theta_\mathrm{t} + \theta) = r(\theta_\mathrm{t} - \theta) \tag{3.45}$$

Eqn. 3.44 does not have an analytical solution, and has to be integrated numerically in order to obtain the path of the light ray.

3.8.3 Atmospheric refraction effects

Light rays that enter the atmosphere will bend *downwards* (towards regions with larger refractive indices). This of course implies that the altitude of the turning point is always larger than the geometrical turning point altitude, associated with the straight line connecting the Sun and the satellite. Thus, the turning point experiences an *upward shift*.

This shift causes the occultation time period to be longer than in the absence of refraction. Also, because of the curvature of the rays, the optical path length in the absorbing atmosphere is increased.

A ray that originates at the 'bottom' of the Sun experiences a larger bending than the 'top' ray, because it has to travel through denser atmospheric layers. This results in *Sun flattening* [Garriott, 1979]. Typical values are 15 %, 38 % and 69 % at turning point altitudes of 30 km, 20 km and 10 km respectively.

Sun flattening is associated with a less intuitive phenomenon: *refractive dilution*. Fig. 3.13 shows (schematically) that the non-uniform

refractive bending causes the density of light rays to decrease. In other words: the flux of light rays (and hence the irradiance) decreases. It is possible to show that the received irradiance is proportional to the *apparent* size of the Sun ([Fussen et al., 1997, Burchell, 1996]). Thus for a solar occultation experiment, the reference irradiance to compute the transmittance has to be corrected in order to avoid that refractive dilution is interpreted as atmospheric extinction.

Chapter 4

The spatial inverse problem.

During an occultation, ORA receives the fraction of solar radiation that has been transmitted through the atmosphere. Travelling along the atmospheric path, the sounded altitude (measured from the Earth's surface) changes. Therefore, a measurement is made up by contributions from absorbing and scattering species situated at different altitudes. It is the task of the spatial inversion algorithm to answer the question: given a measurement (transmittance profile), what is the total extinction at every altitude?

The situation becomes more complex because we are not dealing with one ray of light, but with a bundle of light coming from the full solar disk. Furthermore, we have to deal with additional problems of a more fundamental nature that are associated with the *information content* of the measurement.

Discussion of these problems forms the main body of this chapter. We will describe the inversion algorithm (NOPE), and present the results obtained. The results will then be validated by a comparison with another experiment (SAGE II), and the chapter ends with the description of an extra stabilizing algorithm that improves the total extinction profiles using statistical considerations.

Let us now commence with a (brief!) discussion of geometrical issues: the satellite orbit, the shape of the Earth and the geolocation (latitude, longitude) above which the occultation takes place.

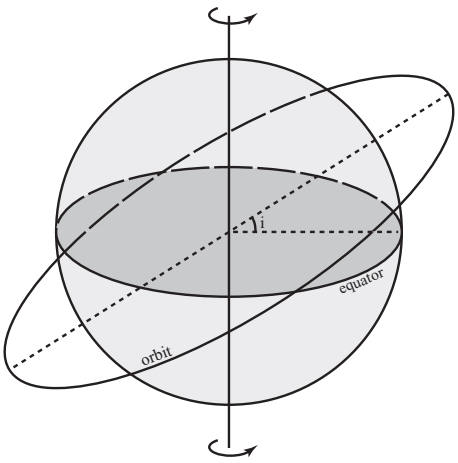

Figure 4.1: A satellite orbit with inclination angle i.

4.1 Experimental geometry.

4.1.1 The orbit

The orbit of a satellite is specified by its altitude, its angle of inclination (angle i between orbital plane and Earth equatorial plane, Fig. 4.1) and its degree of asymmetry (excentricity). We clearly see that the number of choices for an orbit is infinite.

In practice, the choice of orbit is nearly always based upon considerations about the Earth *coverage*, resolution issues and launch cost. For EURECA, the choice was quite simple: since microgravity studies formed the primary target of the mission, coverage and resolution were not an issue at all, and an orbit with a low inclination (28°) was chosen, because these orbits are easily accessible from the tropical launch sites. The nearly-circular orbit had an altitude of 508 km, which is enough to ensure that the atmospheric drag on the satellite does not disturb the microgravity environment.

Since gravitational and centrifugal forces cancel one another for a steady orbit, the velocity of the satellite is given by:

$$v = \sqrt{\frac{GM}{r}} = 7.6 \text{ km s}^{-1}$$

with G the gravitational constant, M the mass of the Earth and r the distance between the Earth's center and the satellite. At this velocity, the satellite needs about 95 minutes to travel a full orbit around the Earth. Within a time span of 24 hours, 15 orbits have been travelled, leading to 30 occultation events per day.

4.1.2 Occultation geometry

While the occultation is taking place, ORA records the irradiance as function of time: $I(t)$. At time t, the spatial coordinates of the satellite and a point on the solar disk *uniquely* determine a light ray emerging from the solar point and arriving at the ORA detectors. At high altitudes, a ray follows a simple straight line. At lower altitudes, refraction comes into play, and the ray has a curvature. Either way, there is one point on the ray that is closest to the Earth's surface. We have already encountered it in the previous chapter: the turning point. The altitude of the turning point is named the *tangential height* h.

The one-to-one relationship between occultation time and tangential height gives us a convenient way to switch variables. Measured signals can now be expressed as function of h_c, the tangential height of the solar ray that originates in the *center* of the refracted solar disk (Fig. 4.2).

4.1.3 Occultation geolocations

Given the satellite state vector (provided by ESOC) and the Sun coordinate (from astronomical ephemerides) in a geocentric reference frame, the central ray provides the tangential height h_c. At the end (sunset) or beginning (sunrise) of an occultation, we have $h_c = 0$. The geolocation of this point has special meaning. We have defined *the* latitude and *the* longitude of the occultation event as the coordinates of this point[1].

4.1.4 The Earth is not spherical

Due to rotation, the shape of the Earth has evolved to an oblate spheroid, with an equatorial radius of 6378 km and a polar radius of 6357 km.

[1]The geolocation of the turning point moves by a few degrees during the occultation, due to the orbital plane inclination with respect to the ecliptic, and the rotation of the Earth.

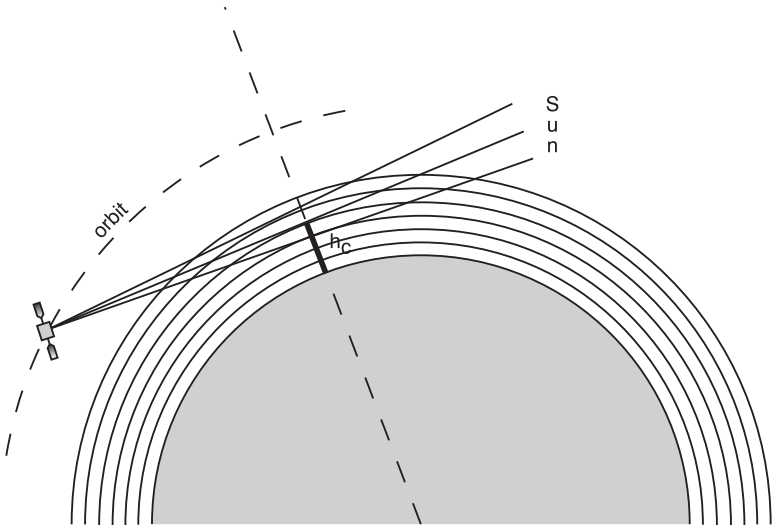

Figure 4.2: The ORA occultation geometry (schematically, refraction is not shown). The light ray originating in the center of the solar disk defines the tangential height h_c.

Assuming an atmosphere consisting of spherical layers is however a good *local* approximation. This was accomplished by representing the Earth's surface at the occultation latitude and longitude by a sphere, tangent to the (true) spheroid, with the same curvature.

4.1.5 Movement of occultation locations

From one sunrise (or sunset) to the next, the Earth rotates. It is easy to see that this movement primarily causes a shift of about 23.6°in the longitude of the occultation point: we have longitudinal scanning of the Earth's atmosphere.

Two mechanisms provide a modulation in the latitude of subsequent occultation points. First, the Earth rotates around the Sun in one year, during which the polar axis of the Earth remains fixed towards a certain point. More important however is the effect of satellite orbital plane *precession*, caused by the non-spherical shape of the Earth. During one precession, the inclination of the orbit is constant, but the orientation of the axis perpendicular to the orbital plane rotates around the Earth's

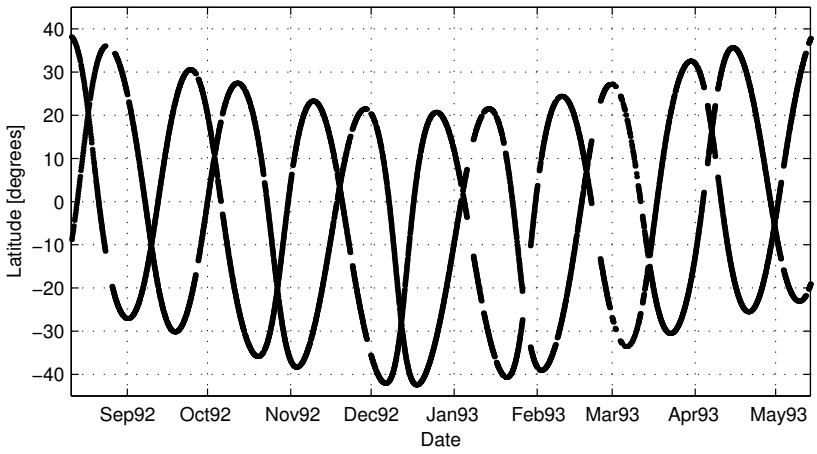

Figure 4.3: Latitudes for all subsequent ORA occultations versus time of year. Gaps in the data result from brief operational interruptions. The two oscillating curves represent respectively sunsets and sunrises.

rotation axis by about 6°per day.

The calculated latitudes for all occultation events are depicted in Fig. 4.3, on which the two modulations are clearly visible. All occultations are situated in a latitude band between 40°N and 40°S.

4.2 Exploitable channels

Although ORA is equipped with 8 spectral channels, two of them are excluded from this discussion. First of all, it was found that channel λ_{259} had a serious defect: an imperfect design of the main filter caused extra light to enter the channel in a spectral range centered around 340 nm (see Chapter 2, Fig. 2.6). In this range, the atmospheric transmittance of solar light is much greater than the 259 nm transmittance, and it is strongly dominated by Rayleigh scattering. Furthermore, the filter/window transmittance of this channel decreased strongly during the mission. It was therefore decided to process the data from channel λ_{259} (aimed at the measurement of mesospheric ozone) with a specific method. More details and results can be found in Appendix A.

The main reason for excluding channel λ_{943} lies in the nature of its

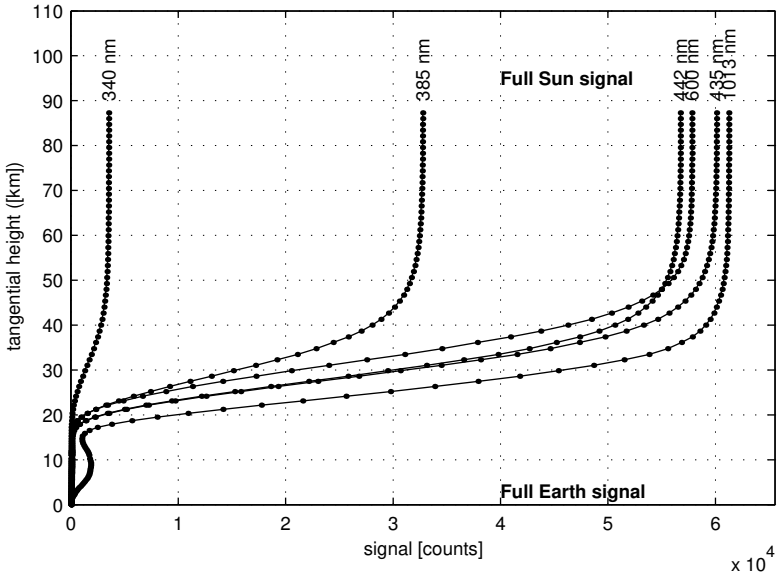

Figure 4.4: Measured occultation signals (Oct. 10, 1992, longitude = 13.58°, latitude = 27.06°). Notice the low sensitivity of the UV-channels, caused by the filter and window degradation.

target constituent: in contrast with the other trace gases (with continuous cross sections), the thousands of discrete water vapor spectral lines demand a special treatment [Chu et al., 1993, Kyle, 1993, Pruvost et al., 1993]. Absorption characteristics (line shape, intensity and half-width) are *strongly* dependent on the atmospheric pressure and temperature in a non-linear way. Thus, absorption calculations along the optical paths have to be performed on a line-by-line basis, a task with formidable computation costs.

In conclusion, from now on this thesis will discuss the processing of data obtained from the reduced set of 6 channels: λ_{340}, λ_{385}, λ_{435}, λ_{442}, λ_{600} and λ_{1013}.

4.3 Measured transmittance calculation

Let us now find out how to calculate the transmittance from the digitized signal measurements. We know that the photodiode and amplifier

deliver an electronic signal S that is proportional to the incoming irradiance I in a linear way:

$$S = aI + b \tag{4.1}$$

Typical examples for measured signals are shown in Fig. 4.4.

Realizing that the maximum of the signal $S_{\max}$ corresponds to the full Sun irradiance I_{full} (no atmospheric extinction), and the minimum of the signal $S_{\min}$ with the full Earth irradiance $I_0 = 0$ (no light received), Eqn. 4.1 leads to the measured transmittance:

$$\tau = \frac{I}{I_{\text{full}}} = \frac{S - S_{\min}}{S_{\max} - S_{\min}} \tag{4.2}$$

What is the measurement error on the signal and the transmittance? To calculate this, one has to take into account the induced thermal noise in the electronics and amplification factors. For the case of ORA, this was not necessary. After investigation of the measured signals, it was clear that no random components could be found, except an occasional 'jump' of one count. There are two reasons for this. First, ORA measures exceptionally large irradiances (the entire solar disk at once), and therefore the signal to noise ratio is very high. Second, the small amount of random noise present is completely overshadowed by the limited digitization resolution. To put it in another way: the 16 bit AD-convertors are not able to distinguish the peaks and valleys of the random noise.

Therefore, digitization errors are the only source of measurement uncertainty. If a measurement S has an uncertainty of 1 count, then it is very easy to derive the transmittance error:

$$\sigma_\tau = \frac{\sqrt{2}}{S_{\max} - S_{\min}} \sqrt{\tau^2 + \tau + 1} \tag{4.3}$$

These are very low error levels, as can be seen on Fig. 4.5.

Inspecting the transmittance, we can observe a peculiar feature. At lower tangential heights (roughly 15 km), we see *bumps* in the transmittance, and they grow stronger with increasing wavelength. At first, they suggested the presence of relatively high-altitude clouds, effectively blocking the incoming light (see e.g. [Kent et al., 1995]). However, as more transmittance profiles were measured, we concluded that the bumps are a visual signature of the stratospheric aerosol layer, because

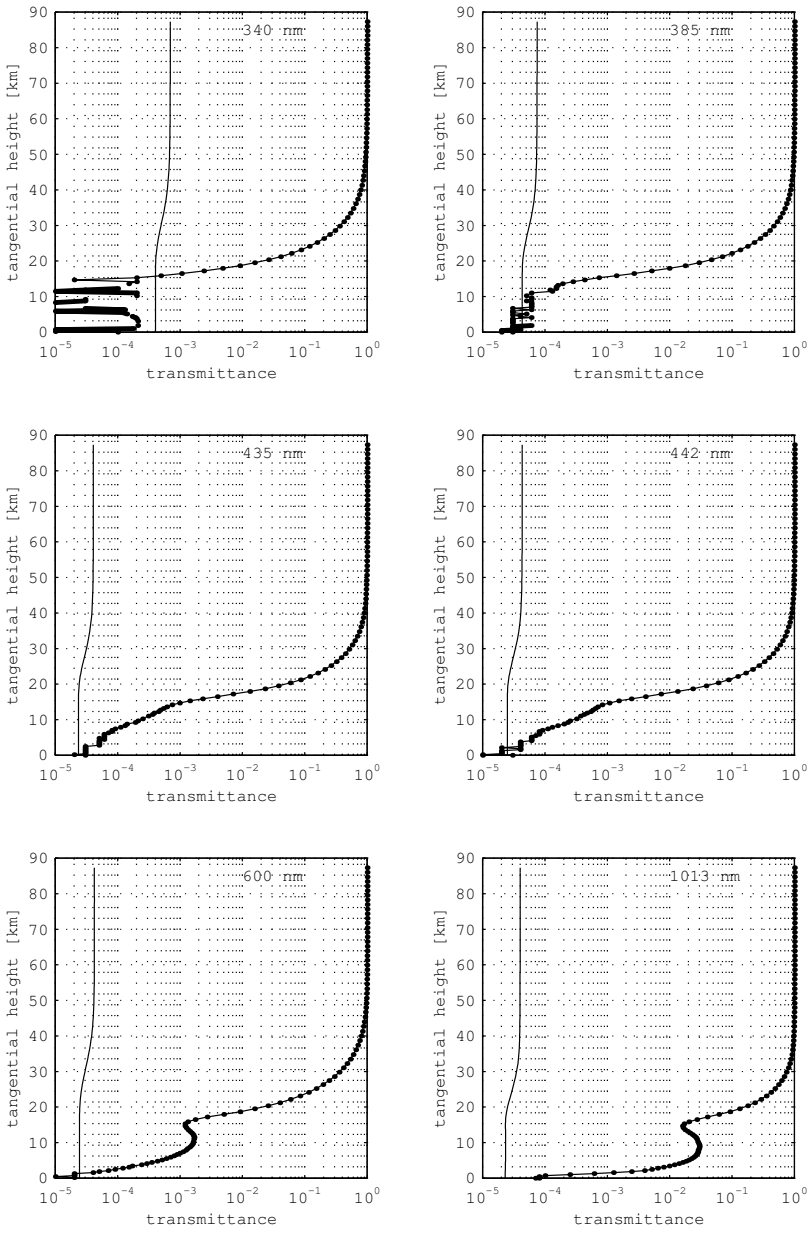

Figure 4.5: Transmittances and digitization error levels, calculated from the data in Fig. 4.4.

(1) they are not visible in transmittance profiles measured in normal non-volcanic background conditions, (2) because the bump magnitudes become smaller in time (relaxation of the aerosols), and (3) because the bump tangential heights correspond with the altitude of the Junge layer.

4.4 Signal description

A transmittance measurement in itself does not produce the desired information on constituent densities. We need a model that connects the measurements with the unknown data. The building stones for this model have already been described in Chapter 3.

At a certain tangential height h_c, the measured transmittance is the result of extinction at all altitudes along the optical paths of the received light rays. This chapter deals with the retrieval of *total extinction* as a function of altitude (separation of the contributions by the different species is handled in Chapter 5). Therefore, our model will have the following form:

$$\tau(h_c) = F\left(\beta_{\text{ext}}^{\text{tot}}(z)\right) \tag{4.4}$$

In the construction of this model, we are going to take a synthetic approach. Let us start from the most basic process: a ray of light with wavelength λ departs from a surface element Ω on the Sun, travels through the Earth's atmosphere and reaches the detector. The path depends on Ω because of geometry, and on λ because refraction is wavelength dependent. Along the optical path $s_{\lambda,\Omega}$, extinction occurs, quantified by the total extinction coefficient $\beta_{\text{ext}}^{\text{tot}}$. This coefficient is of course wavelength dependent. But it also changes along the path, because different points on the path are located at different altitudes in the atmosphere, with different atmospheric states (molecule and particle densities, temperature, pressure).

Inside the detector, the irradiance is affected by the transmission characteristics of the module window and filters, and the photodiode responsivity. The product of window, filters and diode characteristics is the *instrument response function* $F_{\text{N}}(\lambda)$, conveniently normalized:

$$\int_0^\infty F_{\text{N}}(\lambda)\,\mathrm{d}\lambda = 1 \tag{4.5}$$

The law of Beert-Lambert now gives us the received monochromatic irradiance:

$$I(\lambda,\Omega) = I_0(\lambda,\Omega)F_{\mathrm{N}}(\lambda)\exp\left(-\int_{s_{\lambda\Omega}}\beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda,s_{\lambda\Omega})\mathrm{d}s_{\lambda\Omega}\right) \tag{4.6}$$

This equation has to be integrated over the wavelength interval, spanned by the instrument response function:

$$I(\Omega) = \int_{\lambda} I_0(\lambda,\Omega)F_{\mathrm{N}}(\lambda)\exp\left(-\int_{s_{\lambda\Omega}}\beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda,s_{\lambda\Omega})\mathrm{d}s_{\lambda\Omega}\right)\mathrm{d}\lambda \tag{4.7}$$

The total irradiance, received by the instrument is obtained by integrating Eqn. 4.7 over the entire solar disk:

$$I = \frac{\int_{\Omega}\int_{\lambda} I_0(\lambda,\Omega)F_{\mathrm{N}}(\lambda)\exp\left(-\int_{s_{\lambda\Omega}}\beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda,s_{\lambda\Omega})\mathrm{d}s_{\lambda\Omega}\right)\mathrm{d}\lambda\,\mathrm{d}\Omega}{\int_{\Omega}\mathrm{d}\Omega} \tag{4.8}$$

The full Sun irradiance (no extinction) is obtained from this last equation:

$$I_{\mathrm{full}} = \frac{\int_{\Omega}\int_{\lambda} I_0(\lambda,\Omega)F_{\mathrm{N}}(\lambda)\mathrm{d}\lambda\,\mathrm{d}\Omega}{\int_{\Omega}\mathrm{d}\Omega} \tag{4.9}$$

and can be rewritten as

$$I_{\mathrm{full}} = \int_{\lambda} F_{\mathrm{N}}(\lambda)I_{\mathrm{disk}}(\lambda)\,\mathrm{d}\lambda = \tilde{I}_{\mathrm{disk}} \tag{4.10}$$

where $\tilde{I}_{\mathrm{disk}}$ represents the irradiance of the entire solar disk, convolved with the instrument response function.

Finally the ORA transmittance, according to Eqn. 4.2, equals:

$$\tau = \frac{\int_{\Omega}\int_{\lambda} I_0(\lambda,\Omega)F_{\mathrm{N}}(\lambda)\exp\left(-\int_{s_{\lambda\Omega}}\beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda,s_{\lambda\Omega})\mathrm{d}s_{\lambda\Omega}\right)\mathrm{d}\lambda\,\mathrm{d}\Omega}{\tilde{I}_{\mathrm{disk}}\int_{\Omega}\mathrm{d}\Omega} \tag{4.11}$$

This is the exact formulation. We know τ, the measured transmittance. Unknowns are the atmospheric state quantities, such as gas and particle densities, temperature and pressure, and the exact refracted optical paths, that depend on the local atmospheric states. Clearly, from Eqn. 4.11 it is impossible to solve the problem algebraically, i.e. to

write these unknowns as function of the measured τ. We intuitively feel that the unknowns can only be found in an iterative way, through some kind of optimization method. This is the approach: assume initial values for the unknowns, and calculate τ with Eqn. 4.11. Compare the measured and calculated transmittances. If they do not match, adjust the unknowns, and calculate a new transmittance. Repeat this sequence until the difference between measured and calculated transmittance is small enough, according to some criterion. The unknowns used in the final iteration then form a possible solution of the problem.

If we would follow this road, the inversion of all ORA measurements would literally take *years*. One iteration only involves the calculation of the refraction path for every solar surface element for every wavelength within the instrument response range, and this has to be calculated for every measurement taken within the occultation. Several iterations would take an incredible amount of time. Furthermore the calculation has to be performed for six spectral channels. And, of course, this has to be repeated for several thousands of occultation events.

Clearly, we have to make some approximations and reduce the computational cost considerably.

4.4.1 Approximations

Refraction wavelength dependence

Within the instrument response wavelength range, the refractivity of air does not change much (less than 0.2 %). Let us consider it to be constant. The optical path is then independent of wavelength, and the transmittance is given by:

$$\tau = \frac{\int_\Omega \int_\lambda I_0(\lambda,\Omega) F_\mathrm{N}(\lambda) \exp\left(-\int_{s_\Omega} \beta_\mathrm{ext}^\mathrm{tot}(\lambda, s_\Omega) \mathrm{d}s_\Omega\right) \mathrm{d}\lambda \, \mathrm{d}\Omega}{\tilde{I}_\mathrm{disk} \int_\Omega \mathrm{d}\Omega} \tag{4.12}$$

Solar limb darkening wavelength dependence

The limb darkening function differs considerably from one ORA channel to another. However, within the instrument response range of one channel, the differences are tenuous. It is thus possible to write the wavelength and geometrical dependency of the limb darkening in two separate functions:

$$I_0(\lambda, \Omega) = G(\lambda) H(\Omega) \tag{4.13}$$

Integration over the entire solar disk yields:

$$I_{\text{disk}}(\lambda) = \frac{\int_\Omega I_0(\lambda, \Omega)\, \mathrm{d}\Omega}{\int_\Omega \mathrm{d}\Omega} \tag{4.14}$$

$$= G(\lambda) \frac{\int_\Omega H(\Omega)\, \mathrm{d}\Omega}{\int_\Omega \mathrm{d}\Omega} \tag{4.15}$$

and so we have:

$$G(\lambda) = I_{\text{disk}}(\lambda) \frac{\int_\Omega \mathrm{d}\Omega}{\int_\Omega H(\Omega)\, \mathrm{d}\Omega} \tag{4.16}$$

Eqn. 4.13 now becomes:

$$I_0(\lambda, \Omega) = I_{\text{disk}}(\lambda) H_{\text{N}}(\Omega) \int_\Omega \mathrm{d}\Omega \tag{4.17}$$

where we have used the normalized geometrical dependence:

$$H_{\text{N}}(\Omega) = \frac{H(\Omega)}{\int_\Omega H(\Omega)\, \mathrm{d}\Omega} \tag{4.18}$$

Let us also define the normalized convolved solar irradiance:

$$I_{\text{N,disk}}(\lambda) = \frac{I_{\text{disk}}(\lambda) F_{\text{N}}(\lambda)}{\tilde{I}_{\text{disk}}} \tag{4.19}$$

Eqn. 4.12 then reduces to:

$$\tau = \int_\Omega \int_\lambda I_{\text{N,disk}}(\lambda) H_{\text{N}}(\Omega) \exp\left(- \int_{s_\Omega} \beta_{\text{ext}}^{\text{tot}}(\lambda, s_\Omega) \mathrm{d}s_\Omega \right) \mathrm{d}\lambda\, \mathrm{d}\Omega \tag{4.20}$$

The Sun is a 2D disk

For the calculations, it is convenient to assume that the Sun is a flat disk: we replace the spherical source by its 2-dimensional image, as seen by the observer (this is not really an approximation, just a simplification). We can now use plane coordinates:

$$H_{\text{N}}(\Omega) \to H_{\text{N}}(x, y) \qquad \text{and} \qquad \mathrm{d}\Omega \to \mathrm{d}x\, \mathrm{d}y \tag{4.21}$$

The y-axis lies in the plane containing the Sun, the center of the Earth and the satellite, and is normal to the observed ray coming from the center of the solar disk.

The transmittance equation now transforms to:

$$\tau = \int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \int_{y}\int_{x} H_{\mathrm{N}}(x,y) \exp\left(-\int_{s_{xy}} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_{xy})\mathrm{d}s_{xy}\right) \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\lambda \tag{4.22}$$

The solar disk consists of layers

In a very good approximation, rays that originate in points with identical y coordinates travel through the Earth's atmosphere at the same altitudes. This results from the fact that the Sun angular size is very small, as compared to the horizontal gradients of atmospheric layers. Therefore, we can state that the optical paths are not depending on x:

$$\tau = \int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \int_{y} \left[\int_{x} H_{\mathrm{N}}(x,y)\,\mathrm{d}x\right] \exp\left(-\int_{s_y} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_y)\mathrm{d}s_y\right) \mathrm{d}y\,\mathrm{d}\lambda \tag{4.23}$$

We can now split the integral along y into N piecewise integrals:

$$\tau = \sum_{i=0}^{N-1} \int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \int_{y_i}^{y_{i+1}} \left[\int_{x} H_{\mathrm{N}}(x,y)\,\mathrm{d}x\right] \exp\left(-\int_{s_y} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_y)\mathrm{d}s_y\right) \mathrm{d}y\,\mathrm{d}\lambda \tag{4.24}$$

with $y_i = -1 + i\Delta y$ and $\Delta y = \frac{2}{N-1}$.

If we take the number of terms N large enough (vertically thin layers), then we can assume that the optical paths originating in one y-layer are essentially the same:

$$\tau = \sum_{i=0}^{N-1} \int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \left(\int_{y_i}^{y_{i+1}} \int_{x} H_{\mathrm{N}}(x,y)\,\mathrm{d}x\,\mathrm{d}y\right) \exp\left(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_i)\mathrm{d}s_i\right) \mathrm{d}\lambda \tag{4.25}$$

where s_i is the optical path coming from layer i.

Define the *layer contribution coefficients*:

$$r_i = \int_{y=-1+i\Delta y}^{-1+(i+1)\Delta y} \int_{x=-\sqrt{1-y^2}}^{\sqrt{1-y^2}} H_{\mathrm{N}}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{4.26}$$

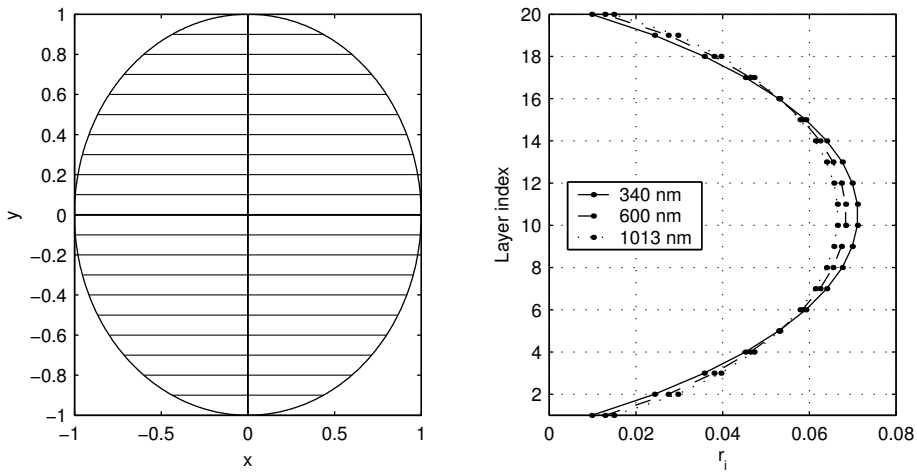

Figure 4.6: Left: geometry for solar disk discretization. Right: calculated layer contribution coefficients for some ORA channels.

that have the property

$$\sum_{i=0}^{N-1} r_i = 1 \tag{4.27}$$

These coefficients represent the contribution of a solar layer, and this contribution is determined by the relative surface of a layer, and the solar limb darkening. Fig. 4.6 shows some examples.

The transmittance can now be written as:

$$\tau = \sum_{i=0}^{N-1} r_i \int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \exp\left(- \int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_i) \mathrm{d}s_i \right) \mathrm{d}\lambda \tag{4.28}$$

Instrument response deconvolution

Eqn. 4.28 still contains integrals over the wavelength range of the instrument response. The situation can be simplified if we write every term with respect to the center wavelength λ_c of the channel. This can be achieved by the introduction of a correction factor g. The i-th term

becomes:

$$\int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \exp\left(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_i)\,\mathrm{d}s_i\right)\,\mathrm{d}\lambda = g_i \exp\left(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda_c, s_i)\,\mathrm{d}s_i\right) \tag{4.29}$$

The correction factors are thus defined as:

$$g_i = \frac{\int_{\lambda} I_{\mathrm{N,disk}}(\lambda) \exp\left(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda, s_i)\,\mathrm{d}s_i\right)\,\mathrm{d}\lambda}{\exp\left(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda_c, s_i)\,\mathrm{d}s_i\right)} \tag{4.30}$$

Using the US76 standard atmosphere, we have numerically calculated the g_i factors. They all have values close to unity, which of course means that the ORA channels are indeed nearly monochromatic. Eqn. 4.28 transforms to:

$$\tau = \sum_{i=0}^{N-1} r_i g_i \exp(-\int_{s_i} \beta_{\mathrm{ext}}^{\mathrm{tot}}(\lambda_c, s_i)\,\mathrm{d}s_i) \tag{4.31}$$

Refraction tables

Through Eqn. 3.43, we know that the actual refraction of a solar ray depends on the local atmospheric state, given by the pressure p and the temperature T, or alternatively, on the air density. But this is one of the constituents that needs to be derived. Consequently, we chose to calculate the refracted optical paths from the US76 standard atmosphere values [National Oceanic and Atmospheric Organization, 1976]. To be really correct, it would be necessary to calculate new optical paths from the retrieved air density, and start a new iteration. However, this is simply too time-consuming, and nothing guarantees that the iteration process would converge. Fortunately, we know that the atmospheric air density profiles do not vary a lot, and from this we can conclude that refraction paths are essentially identical for two different occultations.

The optical path calculation was performed for 100 different orbit positions occupied during each occultation. The 2 Mbyte data file was stored, ready for use in the inversion algorithm.

Summarized

We have come a long way to reduce the very complex (exact) transmittance model to the relatively simple approximation, represented by

Eqn. 4.31, a transmittance that is simply a weighted sum of partial transmittances. Now is a good time to start a discussion on inverse problems.

4.5 Inverse problems: general aspects

Let us consider a hypothetical experiment, designed to obtain data on some quantity x. Quite often, the situation forbids a direct measurement of x. Instead, a quantity is measured that is a function of the desired quantity: $y = f(x)$. When the experiment is performed, we are left with the task of retrieving x from the y-data: the *inverse problem*. Examples of such measurements are widespread. In the medical world, the transmittance of electromagnetic or acoustic waves delivers the density of human tissue as function of space (a *tomography scan*). Geology: given the time an acoustic pulse needs to travel through layers of rock, what is the composition of the layers? Atmospheric studies (e.g. ORA): given the transmittance of radiation through the atmosphere, what is the density of the atmospheric species as function of spatial coordinates?

Quite often, the inverse problem is trivial, when an algebraic inversion of the function f is possible. A measurement of an acoustic pulse time delay t through a medium of known thickness d directly produces the sound velocity v in the medium:

$$\begin{array}{lrcl} \text{Quantity} & x & \Longleftrightarrow & v \\ \text{Measurement} & y & \Longleftrightarrow & t \\ \text{Model} & y = f(x) & \Longleftrightarrow & t = \frac{d}{v} \\ \text{Inversion} & x = f^{-1}(y) & \Longleftrightarrow & v = \frac{d}{t} \end{array} \tag{4.32}$$

However, very often the inversion is far from trivial. Eqns. 4.11 or 4.31 are good examples. Specific numerical methods are then required to obtain a solution for x. Even so, one still has to be very careful for other reasons that will be discussed now.

4.5.1 Definitions

Suppose our experiment takes m measurements of y. The full measurement can then be expressed as a vector with m components:

$$\mathbf{y} = [y_1 \cdots y_m]^T \tag{4.33}$$

and $\mathbf{y}$ is called the *measurement vector*. Example: the ORA transmittance consisting of m measurements taken at subsequent orbit positions during an occultation. The mathematical space, spanned by all possible vectors $\mathbf{y}$ is the m-dimensional *measurement space*.

In the same way, the quantity that we want to retrieve can be expressed as a vector with n components, the *state vector*:

$$\mathbf{x} = [x_1 \cdots x_n]^T \tag{4.34}$$

Example: a vector containing the total extinction coefficients at n different altitudes in the atmosphere. The mathematical space, spanned by all possible vectors $\mathbf{x}$ is the n-dimensional *state space*.

Nature ensures the existence of an intrinsic relation that links the measurement with the state. This relation is expressed as the *forward function* f:

$$\mathbf{y} = f(\mathbf{x}) \tag{4.35}$$

We can now introduce the *forward model* F, which is the best model for f that we can construct, taking into account practical considerations (e.g. computational cost):

$$\mathbf{y} = F(\mathbf{x}) \tag{4.36}$$

Example: the right-hand side of Eqn. 4.31, that approximately relates transmittance to the total extinction.

F has been introduced in a general way. For the sake of clarity, we now assume it to be a linear model. The description that follows can always be generalized to the case of non-linear functions (by linearization in the vicinity of a realistic domain). Thus, F is represented by a $(m \times n)$ linear matrix $\mathbf{K}$:

$$\underset{(m \times 1)}{\mathbf{y}} = \underset{(m \times n)}{\mathbf{K}} \cdot \underset{(n \times 1)}{\mathbf{x}} \tag{4.37}$$

The matrix $\mathbf{K}$ can always be looked upon as a set of m row vectors or n column vectors:

$$\mathbf{K} = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_m \end{bmatrix} \qquad \mathbf{K} = [\mathbf{c}_1 \cdots \mathbf{c}_n] \tag{4.38}$$

The rows $\mathbf{r}_i$ of $\mathbf{K}$ are vectors with n components, and can therefore be considered as lying in the state space. Now it becomes clear what a measurement really means: the i-th measurement y_i is the orthogonal projection of the state $\mathbf{x}$ onto the i-th row vector $\mathbf{r}_i$:

$$y_i = \mathbf{r}_i \cdot \mathbf{x} \tag{4.39}$$

We have m row vectors $\mathbf{r}_i$ that lie in the state space. In general, some of them will be linearly dependent. Assume that p row vectors are linearly independent. These vectors span a p-dimensional subspace of the state space, called the *row space*.

Similarly, from the n column vectors of $\mathbf{K}$, p vectors are linearly independent, and they span a p-dimensional subspace of the measurement space, called the *range*. The number p is termed the *rank* of $\mathbf{K}$ and it plays a crucial role in inverse problem theory[2]. We always have:

$$p \leq \min(m, n) \tag{4.40}$$

The part of the state space that is not occupied by the row space is the $(n - p)$-dimensional *null space*. All vectors in the null space are orthogonal to the vectors in the row space. Therefore, components of $\mathbf{x}$ that are situated in the null space do not contribute to the measured $\mathbf{y}$ vector, because the orthogonal projection in Eqn. 4.39 is zero.

If we express an arbitrary state vector $\mathbf{x}$ as the sum of a row space vector and a null space vector:

$$\mathbf{x} = \mathbf{x}_{\text{row}} + \mathbf{x}_{\text{null}} \tag{4.41}$$

then we have:

$$\mathbf{y} = \mathbf{K}\mathbf{x} = \mathbf{K}\mathbf{x}_{\text{row}} + \mathbf{K}\mathbf{x}_{\text{null}} = \mathbf{K}\mathbf{x}_{\text{row}} \tag{4.42}$$

A simple example can make the theory above very clear. Suppose we have measured a quantity $y = 2$. The measurement is connected with the state vector $\mathbf{x}$ by the (underdetermined) model:

$$y = \mathbf{K} \cdot \mathbf{x} = [1\ 2] \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.43}$$

[2]From linear algebra theory, we know that the row rank equals the column rank. In other words: the number of linearly independent row vectors equals the number of independent column vectors.

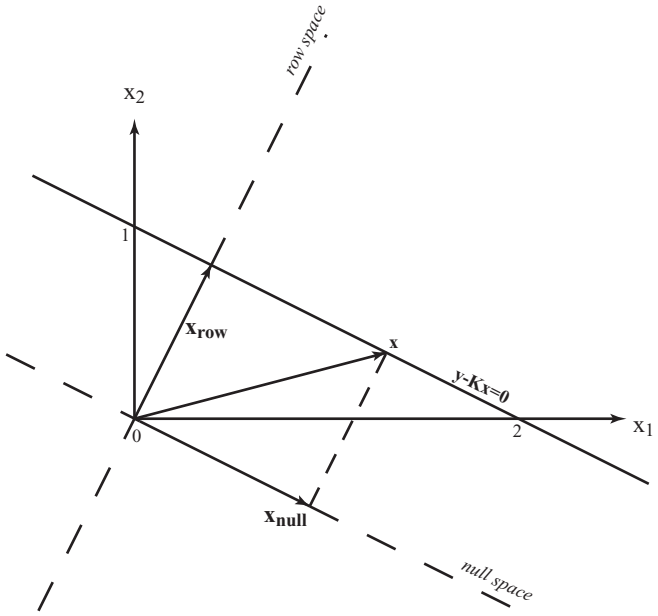

Figure 4.7: The two-dimensional state space for the inverse problem $y = x_1 + 2x_2$, with measurement $y = 2$.

The inversion problem consists of finding the vector $\mathbf{x}$ from the measurement y. We have: $m = 1$, $n = 2$ and $p = 1$, thus $n - p = 1$: a 1-dimensional null space is present.

Fig. 4.7 graphically represents the state space. The solid line indicated by $y - \mathbf{Kx} = 0$ represents the set of all state vectors $\mathbf{x}$ that form a solution to the inverse problem. Every $\mathbf{x}$ is composed of a *unique* row space component $\mathbf{x}_{\text{row}}$ and an arbitrary null space component $\mathbf{x}_{\text{null}}$. There is an infinite number of valid choices for $\mathbf{x}_{\text{null}}$, but they can *never* be derived from the measurement y.

4.5.2 Null space problems

We have come to a very important point here: no matter how well-designed an experiment can be, if the model has a null space, the retrieved result will be incomplete, because only row space components can contribute to the measurement. We can also state this in another way: due to some cause, the retrieved result can have an arbitrary null space

component, *that is not derived from the measurement*, and therefore is *meaningless*. The concept of the null space therefore represents the fundamental inability of a given experiment to measure the full content of a state.

We can now investigate the cases when a null space will be present.

- Overdetermined systems: more measurements than unknowns. We have the case: $m > n$. There is no null space if all row vectors are linearly independent ($p = n$). In the other case ($p < n$) there will be a null space.

- Underdetermined systems: more unknowns than measurements. Now we have $m < n$ and $p < n$. Underdetermined systems always have a null space.

- Well determined systems: equal numbers of measurements and unknowns. We have $m = n$. If all row vectors are linearly independent ($p = m = n$), the system is well-posed, and has a unique solution. If not ($p < m = n$), there is a null space.

Summarized, inverse problems will suffer mainly from underdeterminacy, and from the interdependence of several row vectors (measurements).

4.5.3 The role of error

So far, we have discussed *mathematical* inverse problems. A real-life experiment always has to face errors of different kinds: measurement uncertainty, model approximations, model parameter uncertainties, etc. So we need to write:

$$\mathbf{y} + \epsilon = \mathbf{K} \cdot \mathbf{x} \tag{4.44}$$

The solution to this inverse problem is *formally* written as:

$$\mathbf{x}_{\text{sol}} = \mathbf{K}^{-1} \cdot \mathbf{y} + \mathbf{K}^{-1} \cdot \epsilon \tag{4.45}$$

The notation is formal, because $\mathbf{K}$ is in general not square, and does not have an inverse. Recognizing the first right-hand term as the true solution, we have:

$$\mathbf{x}_{\text{sol}} = \mathbf{x}_{\text{true}} + \mathbf{K}^{-1} \cdot \epsilon \tag{4.46}$$

Let us now introduce the *Singular Value Decomposition* (SVD)[3] of **K** [Golub and Van Loan, 1996]:

$$\begin{array}{ccccc} \mathbf{K} & = & \mathbf{U} & \mathbf{S} & \mathbf{V}^T \\ (m \times n) & & (m \times p) & (p \times p) & (p \times n) \end{array} \tag{4.47}$$

U and **V** are unitary[4] matrices of which the columns form an orthonormal basis for the range and row space, respectively. **S** is a diagonal matrix, containing p non-zero *singular values*, of which some can be very small, nevertheless. Eqn. 4.46 now becomes:

$$\mathbf{x}_{\text{sol}} = \mathbf{x}_{\text{true}} + \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^T\epsilon \tag{4.48}$$

The error term now contains the inverse of **S**. If **S** contains very small values, then $\mathbf{S}^{-1}$ will contain very *large* values. The effect can be terrible: for small singular values, the error term will blow up to huge proportions, making the solution of the inverse problem worthless.

What causes small singular values to be present? At this point, we have to make a distinction between the *mathematical* rank p and the *effective* rank. As we saw, p represents the number of independent rows (or measurements). When p is smaller than n, there is a null space, spanned by base vectors with associated singular values that are exactly zero. When errors are added, some rows (or measurements) can become almost linearly dependent within the error range. Therefore, we have an effective rank $\tilde{p}$ that is smaller than p. Result: an effective null space of dimension $n - \tilde{p}$ that is larger than the mathematical null space, with associated singular values that are very small or zero.

Let us finish with an example. Consider the (mathematically) well-posed inverse problem ($m = n = p = 2$):

$$\begin{bmatrix} 3 \\ 3.01 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2.01 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.49}$$

The solution is clearly $[1\ 1]^T$. Let us now add a small error component:

$$\begin{bmatrix} 3.1 \\ 3.02 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 2.01 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.50}$$

[3]The SVD is a generalization of the eigenvalue decomposition for non-square matrices.

[4]A matrix **U** is unitary when $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}$.

Now the solution becomes $[19.1 \quad -8]^T$, clearly showing the huge error amplification of the system.

4.5.4 Regularization methods

When null space problems arise, we have to use specific methods to (partially) cure the disease. On a fundamental level, there are two approaches:

A priori information. If we have some advance knowledge of the state, we can include it in our system. In other words: from the infinite number of possible **x**-components in the null space, pick one that gives us a solution **x** within reasonable agreement with our prior information. A priori information can come in various guises. For example: a mean state, as measured by other experiments (climatologies). Or: a *smoothness* constraint.

Reducing the number of unknowns. If we have trouble with a null space of dimension $n - p$, we can make it disappear by reducing n until $n - p = 0$. This is the classical way of constraining a system by decreasing the number of degrees of freedom.

At present, a vast amount of literature concerning regularization is available, originating from a wide variety of scientific fields in which inverse problems play a significant role (see for example [Rodgers, 2000, 1990, Roble and Hays, 1972, Tarantola, 1987, Golub and Van Loan, 1996, Twomey, 1985, Lumpe et al., 1991]).

4.5.5 The concept of 'merit function'

Let us now return to the case of general (linear and non-linear)inverse problems. Since we want to compare the measurement with the forward model, we need some way to quantify the distance between the two. Usually, inverse problems deal with vectors for measurement and state, and therefore a vector norm has to be defined. In practically all cases, the *2-norm* is the preferred form:

$$\|\mathbf{x}\| = [\sum_i (x_i)^2]^{\frac{1}{2}} = (\mathbf{x^T x})^{\frac{1}{2}} \tag{4.51}$$

We will sometimes use a notation with brackets (Dirac notation) because of its clarity:

$$\|\mathbf{x}\| = \langle x|x\rangle^{\frac{1}{2}} \tag{4.52}$$

The *merit function* is defined as a scalar-valued function that contains the distance between measurement and forward model. For ordinary least squares, it is:

$$M = \Big\langle \mathbf{y} - F(\mathbf{x}) | {\mathbf{C}_\mathbf{y}}^{-1} | \mathbf{y} - F(\mathbf{x}) \Big\rangle \tag{4.53}$$

where ${\mathbf{C}_\mathbf{y}}^{-1}$ represents the measurement *covariance matrix*. A more detailed discussion on covariance matrices will be postponed until Chapter 5.

4.5.6 Merit function topology: multiple minima

Sometimes, the merit function of an inverse problem can be quite complex. In the case of non-linear inverse problems, merit functions with multiple minima cause much trouble. As an example, consider the (well-posed!) inverse problem ($m = n = p = 1$):

$$y = f(x) = \frac{1}{2}x + \cos(x) \tag{4.54}$$

It is clear from Fig. 4.8 that apart from the global minimum (labelled B), the merit function also exhibits other local minima. If we search for a solution in an iterative way, the obtained solution x_{sol} will be strongly dependent on the initial choice of x: on its way to the true minimum, the algorithm will become trapped in one of the secondary minima if the initial choice is too far from the true solution. It is therefore very important to choose an initial solution that is *not too far from what is expected*: once again, a priori information comes into play.

Suffice to say that these kinds of problems have nothing to do with ill-posedness. They do occur quite frequently, however. The globality of the solution should always be checked one way or another, e.g. by testing the sensitivity of the final solution to the initial choice.

4.6 The NOPE inversion method

It is now time to construct a spatial inversion algorithm, capable of producing the desired total extinction altitude profiles.

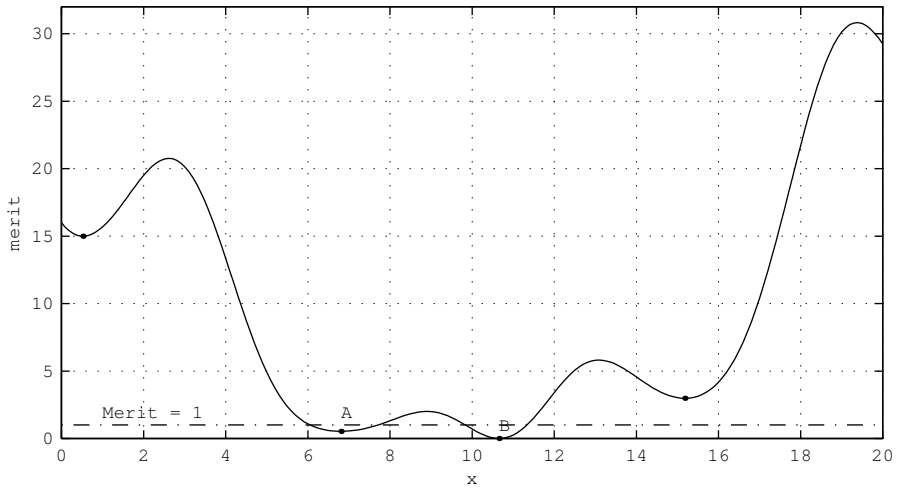

Figure 4.8: Merit function for the inverse problem $y = f(x) = \frac{1}{2}x + \cos x$, with measurement $y = 5 \pm 1$. Multiple minima are present. Notice: for minima A and B, measurement agrees with model within the measurement error.

4.6.1 The fundamental problems

At this point already, we can state that an occultation experiment implies an associated inverse problem that is fundamentally ill-posed. Indeed, we measure transmittances at a limited number of discrete points along the orbit, while these transmittances are the result of absorption and scattering by species with densities that are continuous functions of altitude. The inverse problem is underdetermined!

However, the situation is even worse than this. Because ORA measures the full solar disk, its vertical view is quite large. Two adjacent measurement points therefore deliver virtually the same information. Consequently, the 'rows' in our model 'matrix' are highly overlapping, and it is very likely that a large part of the measurements are interdependent. We are dealing with a serious case of 'null space'. This was confirmed when first inversion try-outs turned out to produce highly unstable solutions, while the fit between measurement and model was acceptable.

Another problem concerns the way we want to handle the inversion: locally, which means altitude per altitude, or globally, using all

altitudes at the same time? Classically, people working on occultation measurements use local methods such as *onion peeling* [Rodgers, 2000] or the *Chahine* iterative method [Twomey, 1975, Chahine, 1968, 1970]. Again, they produced very unstable solutions, or suffered from upward or downward error propagation along the total extinction profile.

4.6.2 NOPE: Natural Orthogonal Polynomial Expansion

After lots of trial and error, it was finally decided to follow a strategy, based on three ideas [Fussen, 1995, Fussen et al., 1997].

A global method

Every time an iteration is performed, the entire extinction profile is adjusted at once. In this way, we avoid the above-mentioned upward or downward error propagation.

Profile parametrization

As already stated in section 4.5, one way to stabilize an ill-posed system is to reduce the number of unknowns. And this is indeed possible. If we have an altitude profile evaluated at, say, 50 atmospheric layers, this does not necessarily mean that we have to determine 50 parameters (the function values itself). An appropriate description using a smaller parameter set can meet the objectives for required accuracy. One has to ask the question: which description combines a small set of parameters with a maximum information content? The answer: orthogonal polynomials. Also, we have to realize that total extinction profiles change by several orders of magnitude within the considered altitude range. It is therefore better to 'scale' the problem by introducing an extra factor, obtained from some a priori information. The following expansion was chosen:

$$\beta_{\text{ext}}^{\text{tot}}(z) = \tilde{\beta}_{\text{ext}}^{\text{tot}}(z) \sum_{i=0}^{n} \alpha_i P_i(z) \tag{4.55}$$

with $P_i(z)$ an orthogonal polynomial of the order i, α_i its associated coefficient, and $\tilde{\beta}_{\text{ext}}^{\text{tot}}(z)$ an a priori extinction profile, constructed for example from a climatology and the knowledge of extinction cross sections.

A large enough number of terms n is then chosen to ensure that the difference between model and measurement is acceptable, but small enough to make sure that the ill-posedness is resolved. An additional advantage of a small parameter set is of course the reduction of computation time.

Choice of polynomials

We still need to specify the type of polynomial that we will use. They need to be orthonormal:

$$\int_0^\infty P_i(z)P_j(z)w(z)\,\mathrm{d}z = \delta_{ij} \tag{4.56}$$

with $w(z)$ the *weight function* [Abramowitz and Stegun, 1972], defining the type of polynomials. Since $w(z)$ can be chosen arbitrarily, we have access to an infinite class of orthogonal polynomial sets. The weight function that we chose has special meaning:

$$w(z) = \frac{\tilde{\beta}_{\mathrm{ext}}^{\mathrm{tot}}(z)}{\int_0^\infty \tilde{\beta}_{\mathrm{ext}}^{\mathrm{tot}}(z)\,\mathrm{d}z} \tag{4.57}$$

The set of polynomials that are defined by this function is very convenient: it is very sensitive to altitude regions where structures are expected to occur. In this way, we use a priori information, however in a rigorous mathematical context: the special choice of $w(z)$ is nothing more than a convergence accelerator.

Important remarks

If $\beta_{\mathrm{ext}}^{\mathrm{tot}}(z)$ equals $\tilde{\beta}_{\mathrm{ext}}^{\mathrm{tot}}(z)$, the solution vector is $\alpha = [1\ 0\ \cdots\ 0]$. Thus, it is expected that the actual solution has a decreasing series of α_i if the actual atmosphere can be considered as a perturbation of the a priori standard atmosphere. If it is very different from $\tilde{\beta}_{\mathrm{ext}}^{\mathrm{tot}}(z)$, it can be systematically detected by inspecting the convergence rate. In that case, we have to correct $\tilde{\beta}_{\mathrm{ext}}^{\mathrm{tot}}(z)$ in the appropriate way by using extra information, or by increasing n.

Because the polynomials are orthonormal, the information content carried by α_i is independent (and usually smaller than) that carried by α_{i-1}. If we increase the number of basis functions n, the last terms of

the expansion are just small corrections relative to the first terms which contain the largest structures.

Concerning the typical problems with multiple minima in the merit function: there is really no way to guarantee that the global minimum of the merit function has been reached after the final iteration. This would require a systematic exploration of the merit function, which is a scalar field depending on n variables. However, the solution obtained by NOPE should be the closest one to the reasonable mean state of the atmosphere defined by $\tilde{\beta}_{\text{ext}}^{\text{tot}}(z)$.

4.6.3 The a priori extinction profile

The total extinction profile results from the contribution of the different absorbing and scattering species:

$$\tilde{\beta}_{\text{ext}}^{\text{tot}}(z) = \tilde{\beta}_{\text{ext}}^{\text{air}}(z) + \tilde{\beta}_{\text{ext}}^{\text{O}_3}(z) + \tilde{\beta}_{\text{ext}}^{\text{NO}_2}(z) + \tilde{\beta}_{\text{ext}}^{\text{aero}}(z) \tag{4.58}$$

where the superscripts of the different contributions refer to air (Rayleigh scattering), ozone, NO_2 and aerosols. For ozone, air and NO_2, we can simply use climatological values for the gas densities and multiply them with the extinction cross section at the considered wavelength. But for aerosols, we face a big problem because ORA measured in the aftermath of the Mount Pinatubo eruption. Other measurements confirm that the atmosphere was not in a typical non-volcanic state. Therefore, we can not use climatological values for aerosol densities. Also, we know nothing about particle size distributions, so the wavelength dependence of aerosol is a priori unknown. Furthermore, aerosol altitude profiles may exhibit large variabilities over short time periods or geographical ranges. How then can we determine an a priori aerosol extinction profile?

As already said, a striking feature for many measured transmittances was the presence of a 'bump' at lower altitudes. The bumps in the measured transmittance profiles are a clear signature of aerosols located in the lower atmosphere. This knowledge naturally led us to the construction of a method that provided us with the required aerosol extinction profiles.

First, we have defined a more elaborate generic aerosol profile:

$$\tilde{\beta}_{\text{ext}}^{\text{aero}}(z) = c_0 \exp\left(-\frac{c_1\delta^2 + \delta^3}{c_2 + c_3\delta^2}\right) \tag{4.59}$$

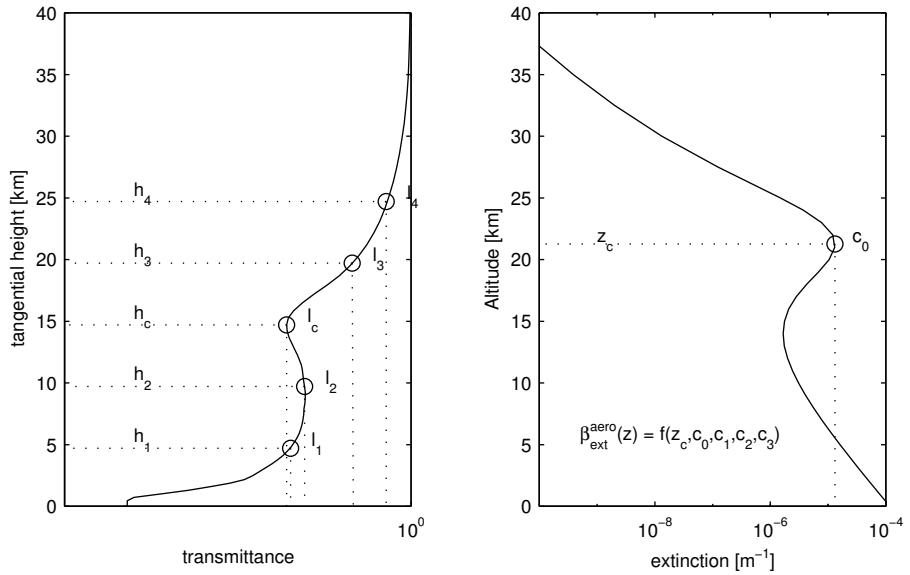

Figure 4.9: The mapping method for the construction of an 'a priori' aerosol extinction profile.

with $\delta = z - z_c$. This analytic form was chosen because it is able to show a maximum at $z = z_c$, while at the same time it is always asymptotic to a pure exponential behavior at high altitudes. The profile is thus parametrized by a vector containing five parameters:

$$\mathbf{v} = [z_c \; c_0 \; c_1 \; c_2 \; c_3]^T$$

We can now vary these parameters over a sufficiently large range, and calculate the aerosol profile for every set of parameters $\mathbf{v}$. Using Eqn. 4.59, we then construct the total extinction profile and simulate a synthetic transmittance with Eqn. 4.31. About 2700 synthetic occultations were calculated. On each of the synthetic transmittances, we identified the extremum at $h = h_c$, and four symmetrical nodes h_1, h_2, h_3 and h_4 (separated by a tangential distance of 5 km), together with the logarithm of the transmittance at these nodes, l_1, l_2, l_3 and l_4 (see Fig. 4.9). A new vector of parameters can then be defined:

$$\mathbf{u} = [l_c \; l_1 \; l_2 \; l_3 \; l_4]^T$$

There should be an intricate, unknown relation between the vector **v** and **u**. We expressed this relation as a full cubic form of five variables:

$$v^i = \mathbf{\Theta}^i \mathbf{u} \qquad i = 1, \cdots, 5 \tag{4.60}$$

The matrix $\mathbf{\Theta}$ is then determined by a linear least-squares fit using the entire synthetic signal data set.

For every occultation measurement, we can now directly retrieve a crude aerosol extinction profile by identifying the nodes l_i on the transmittance profile, and the use of Eqns. 4.60 and 4.59. However, to further reduce the computation time, we decided to do this only for channel λ_{1013}. We can expect this channel to be primarily affected by aerosol extinction (ozone, water vapor and NO2 cross sections are simply too small at this wavelength, and Rayleigh scattering becomes small at high wavelengths). Furthermore, the bumps are more difficult to identify at shorter wavelengths. The aerosol contributions at the other wavelengths were therefore extrapolated, using the simple exponential law obtained from the Mie scattering computations of Yue [Yue, 1986]:

$$\beta_\lambda = \beta_{1013} \exp[c(\lambda - 1013)] \tag{4.61}$$

with $c = -2.4\ \mu\text{m}^{-1}$.

One last remark: it would be wrong to conclude that this approximate aerosol extinction is *the* solution for aerosols. It is true that the profile is derived from the measured transmittances. But it should be interpreted as an extra term in the a priori total extinction profile, that permits the algorithm to capture the aerosol 'bump' structure. One wavelength channel is simply not sufficient to derive the actual aerosol extinction. For this, we have to use more channels, as will be described in the next chapter.

A typical example for the resulting a priori total extinction profiles has been given in Fig. 4.10.

4.6.4 The actual inversion calculation

To summarize, the final flowchart for the vertical inversion algorithm has been given in Fig. 4.11. From the satellite data and astronomical ephemerides, optical paths can be identified that have been calculated and stored in advance. These optical paths deliver the turning point with

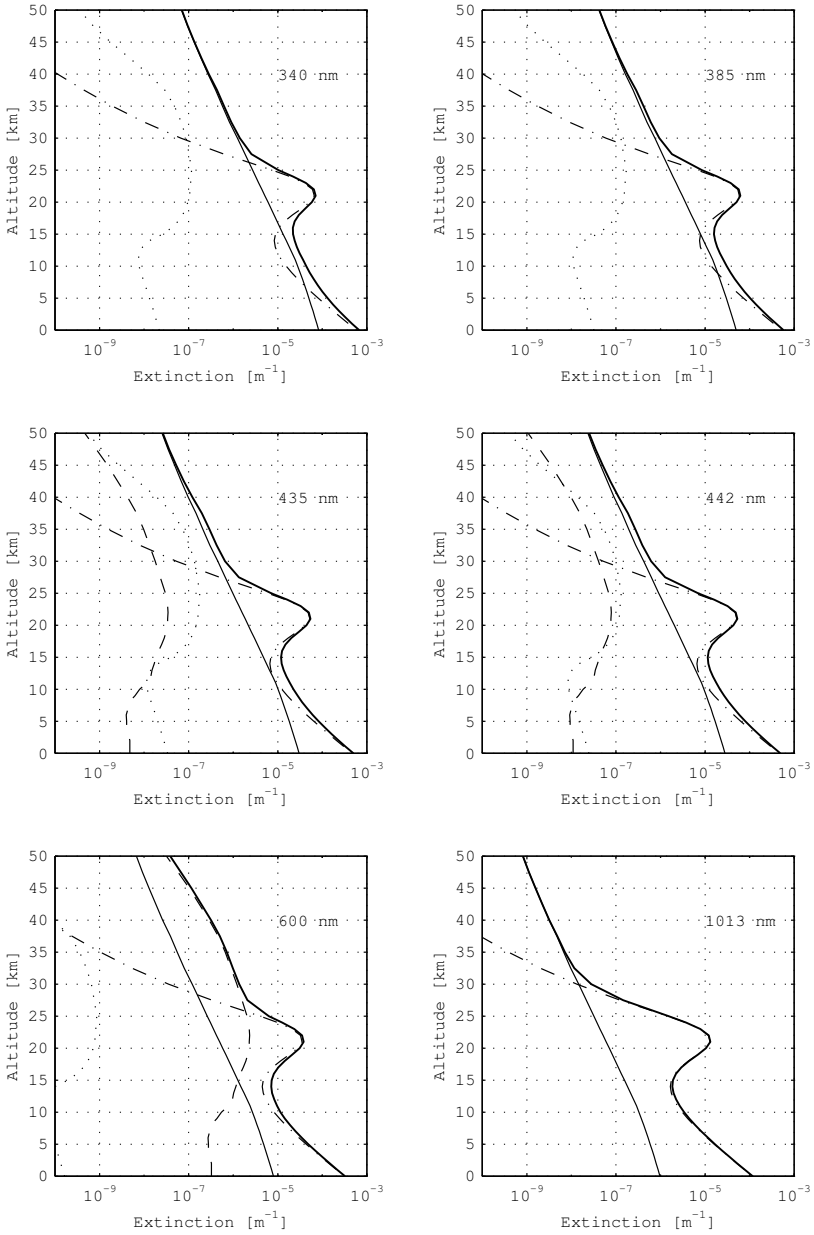

Figure 4.10: The a priori total extinction profile (bold solid) is the sum of contributions by air scattering (–), ozone absorption (- -), NO_2 absorption (·), and an 'aerosol' profile (-·) obtained by the mapping method.

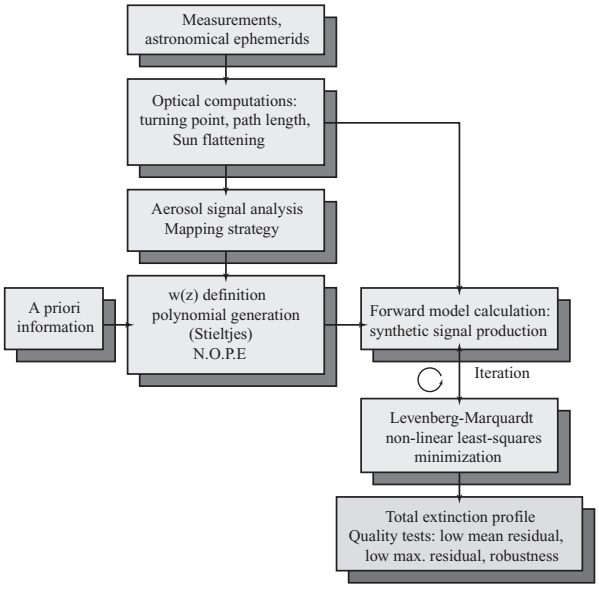

Figure 4.11: Flowchart for the spatial inversion of the ORA transmittances.

associated tangential height, and the atmospheric path length. The Sun flattening effect (that is also calculated in advance) can be taken into account. On the $\tau(h)$ transmittance profile, the previously mentioned nodes are identified, the vector $\mathbf{u}$ is calculated and an 'a priori' aerosol extinction profile is calculated. It is added to the a priori extinction profiles of the other constituents. We now have the total a priori extinction $\tilde{\beta}_{\text{ext}}^{\text{tot}}(z)$ which is used to define the weight function $w(z)$. Using the recursive *Stieltjes procedure* [Press et al., 1992], 10 orthonormal polynomials are then constructed. The first guess polynomial coefficients were taken to be $\alpha = [1\ 0\ \cdots\ 0]$, which means that the first guess total extinction equals the a priori extinction. This first guess, combined with the optical paths, are used to calculate a synthetic transmittance, using the forward model Eqn. 4.31. Model τ_M and measurement τ are compared by calculating the merit function (of which the value is simply the *chi-squared* of the inverse problem):

$$M = \left\langle \tau - \tau_{\text{M}} | \mathbf{S}_{\tau}^{-1} | \tau - \tau_{\text{M}} \right\rangle$$

with $\mathbf{S}_{\tau}$ the covariance matrix of the measured transmittance.

Using a standard *Levenberg-Marquardt* minimization procedure [Press et al., 1992], the solution vector α is iteratively adapted, until convergence. From the final α, the total extinction profile is calculated. We have a solution!

This procedure had to be repeated six times (for every wavelength channel). The processing of one occultation required about 10^9 flops, and the entire ORA data set inversion took about three months to complete. These are very large figures, and give an indication of the inversion complexity that we had to face.

4.6.5 Results

A typical inversion result is presented in Figs. 4.12 and 4.13. The mismatch between measurement and model can be expressed by the use of the sum of the squares of the inversion residuals, dividing it by the number of measurement points m and taking the square root. In this way, we derive the mean residual:

$$r = \sqrt{\frac{\sum_{i=1}^{m}(\tau_i - \tau_{\mathrm{M},i})^2}{m}}$$

The residual was found to increase slightly with wavelength (for example, $r \approx 0.0010$ for λ_{340} and $r \approx 0.0026$ for λ_{1013}). The value of r is greater than the experimental (digitization) noise (typically $\approx 10^{-4}$). This means that the algorithm is unable to fit the measurement within the very small error level, and this of course indicates that we are dealing with *modelization errors*, caused by the approximations in the transmittance model, and perhaps convergence of the iteration towards a local minimum that differs from the global one. Modelization errors come in different shapes: approximations in the model, uncertainty in model parameters, and of course the neglect of additional processes, such as air glow, reflections on the Earth surface, straylight, etc.

At high altitudes, one should be aware that the retrieved extinction profiles sometimes become erratic, because transmittance is nearly equal to one, which means very small extinction, so the measured transmittance contains practically no information at all. The same thing happens at low altitudes (the saturation regime), where almost all light is attenuated in the atmosphere and transmittance is almost zero.

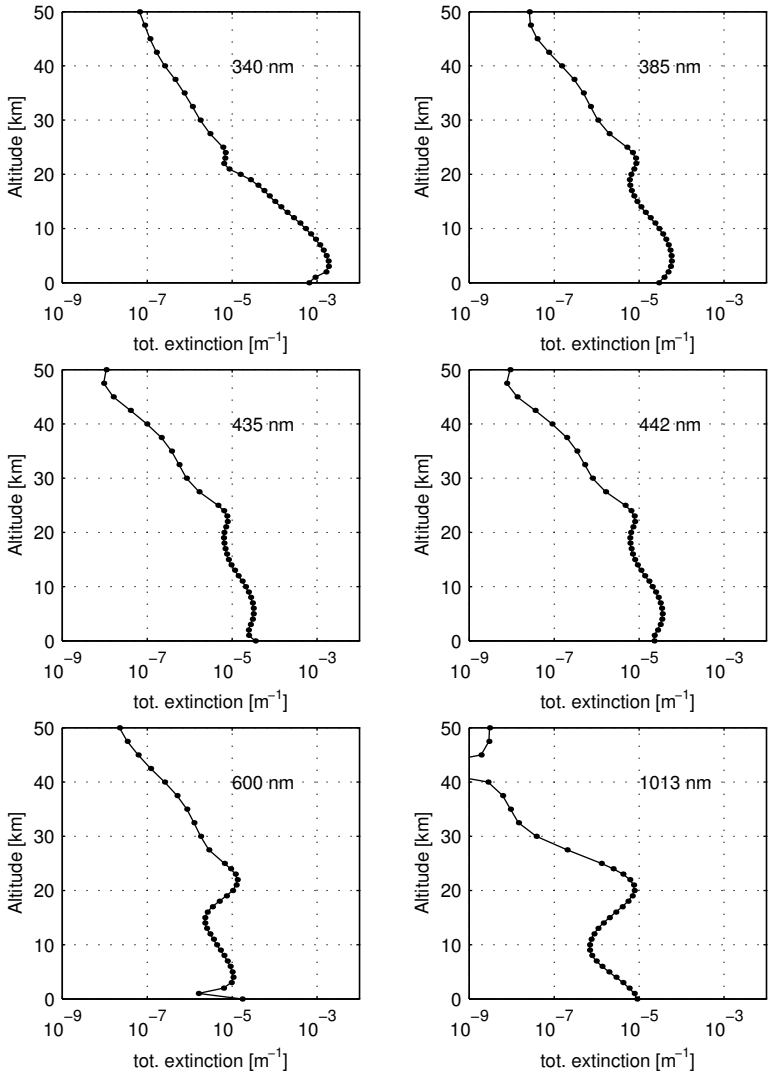

Figure 4.12: Retrieved total extinction profiles. Event: Oct. 10, 1992, longitude = 13.58°, latitude = 27.06°.

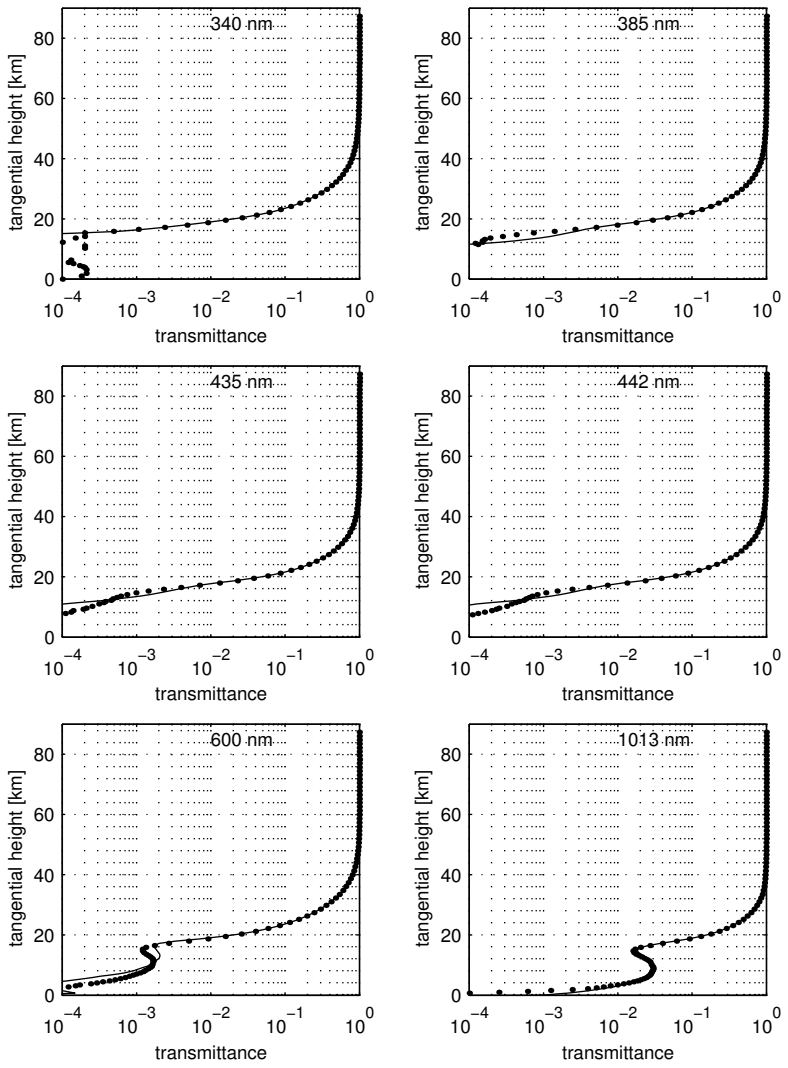

Figure 4.13: Comparison of measured transmittance (dots) and modelized transmittance (lines), for the same occultation event as Fig. 4.12

A quick inspection of Fig. 4.12 reveals that the aerosol extinction signature decreases with decreasing wavelength. This is of course caused by the increasing contribution of Rayleigh scattering at small wavelengths. At λ_{600}, the total extinction profile exhibits a shoulder, caused by the ozone contribution in the Chappuis band. It is the task of the spectral inversion to separate the different contributions of the constituents, as will be discussed in the next chapter.

4.7 Validation of the total extinction

4.7.1 The SAGE II experiment

While the Stratospheric Aerosol and Gas Experiment II (SAGE II) is designed to retrieve the same atmospheric species as ORA does, in the same wavelength region, the practical implementation differs [Mauldin et al., 1985]. Although also an occultation experiment that observes the Sun, it does so by using a Sun tracker, consisting of a telescope and a scanning system. Different regions of the solar disk are measured subsequently, which is possible because the field of view is very small (0.5 arc minutes, about a fraction of 1/60 times the angular size of the Sun). This excellent vertical resolution is counterbalanced by a signal to noise ratio that is much worse than ORA's. But this disadvantage is *hugely* compensated for by the simplicity of the inversion algorithm. Indeed, we can now assume that the instrument receives just one ray[5]:

$$\tau = \exp\left[-\int \beta_{\text{ext}}^{\text{tot}}(s)\,\mathrm{d}s\right] = \mathrm{e}^{-\delta} \tag{4.62}$$

By taking the logarithm of the measured transmittance, one immediately obtains the tangential path total optical thickness. Spectral inversion is then performed[6], followed by a spatial inversion on the partial (species) optical thickness. More details can be found in [Chu et al., 1989].

The reason for the choice of SAGE II as a comparison reference is simple: ORA and SAGE II have an overlapping Earth coverage, their measurement principle is similar, and three of the SAGE chan-

[5]Here, we have neglected the instrument response function for simplicity.

[6]Usually it is preferred to perform the spectral inversion first because the associated errors are smaller than those induced by the spatial inversion.

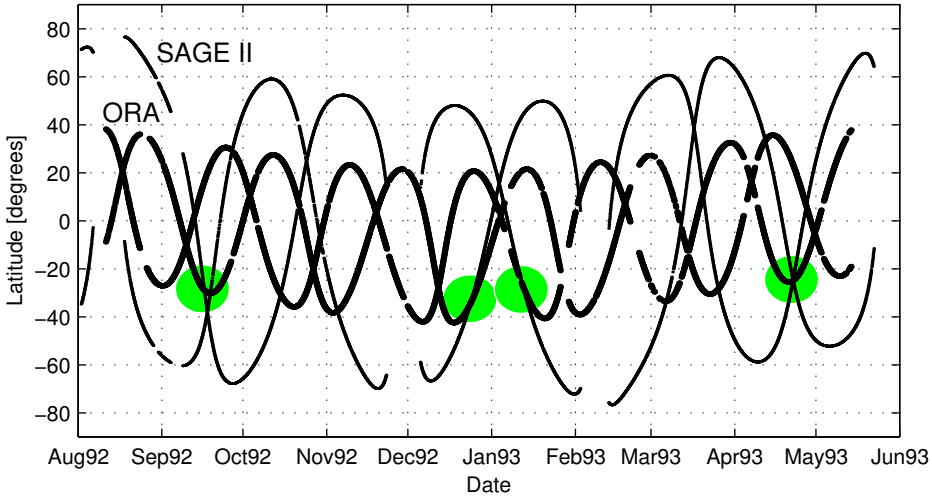

Figure 4.14: Time evolution of the ORA and SAGE II occultation latitudes. The low orbit inclination of ORA is responsible for the smaller amplitude. Not all crossings between curves represent potential comparison occultations, because of a possible difference in longitude. The selected events are indicated with gray circles.

nels (λ_{385},λ_{600},λ_{1020}) have nearly equal wavelengths as ORA (λ_{385}, λ_{600}, λ_{1013}).

SAGE II has been extensively validated since its launch in 1984 [Lenoble, 1989]. However, to our knowledge SAGE II data had never been compared with data from experiments with the same experimental set-up for the same nominal wavelengths (except for the comparison with the SAM II experiment[7] at λ_{1020}) [Yue et al., 1989]. Of course, this situation provided an exceptional opportunity for us.

4.7.2 Overlapping occultations for ORA and SAGE

While the latitudinal coverage for the two experiments is quite different, several synchronous occultations were very closely located within the tropical range of ORA (Fig. 4.14).

The event selection was a trade-off between maximizing the number of comparisons for statistical significance and narrowing the spatiotem-

[7]Stratospheric Aerosol Measurement II.

poral window where SAGE and ORA measurements are compared. The time difference Δt between the two sets of measurements at the 'same' location was always small, which is not surprising because both experiments made observations above the same region at local twilight or dawn.

The latitude-longitude window size was more crucial. We observed a significant decrease of comparison statistics as $\Delta\phi$ and $\Delta\theta$ were increased to 10°(1000 km at surface) and 20°(2000 km at surface) respectively, with 270 of (6821) events that took place within a one-hour window. By considering that the mean effective optical path (around the turning point) has a length of about a few hundred km and that possible cloud contamination could occur over larger distances, we decided to restrict the selection criteria to

$$\Delta t = 6 \text{ minutes} \quad \Delta\phi = 2^\circ \quad \Delta\theta = 2^\circ \tag{4.63}$$

We found 26 events that satisfy criteria 4.63, but we neglected one of them because NOPE had clearly converged toward an unrealistic solution (with a large residual). These events are naturally grouped into four time periods (see Fig. 4.14): September 1992, December 1992, January 1993 and April 1993, which can be considered representative of the entire ORA mission.

The natural variability of ozone and aerosol profiles over the domain defined by criterion 4.63 was studied by [Chu and McCormick, 1979] who evaluated the effects of possible horizontal inhomogeneities. They concluded that the perturbation on inversion accuracies (below 10 % for all constituents for tangent altitudes of 12-42 km) that is due to the horizontal gradient is small even for midlatitudes where the most important inhomogeneities are expected to occur. On the other hand it is well known that significant variations of ozone-aerosol profiles over short scales have been observed [Murphy et al., 1993] in the near-tropopause region, and one can estimate this variability at low altitudes by inspecting the scatter plots in Fig. 4.16 (see further). Therefore, both the SAGE II and ORA experiments retrieved extinction coefficients at a given altitude that must be interpreted as averaged values over a 200- to 300 km optical path length from an equivalent homogeneous atmosphere.

Coordinate	ORA 1	SAGE 1	ORA 2	SAGE 2	ORA 3	SAGE 3
Date	Sep16,92	Sep16,92	Dec24,92	Dec24,92	Apr22,93	Apr22,93
Time	10:46:33	10:47:43	14:53:54	14:55:37	11:48:54	11:47:18
Latitude	-28.8°	-28.0°	-33.0°	-32.2°	-24.9°	-24.0°
Longitude	105.7°	105.6°	62.9°	62.3°	-81.8°	-81.9°
Location	Indian Ocean near Australia		Indian Ocean near Madagascar		Pacific Ocean near Chile	

Table 4.1: Spatiotemporal coordinates of the three compared events.

4.7.3 SAGE II data

First, the SAGE II total extinction profiles had to be constructed from the separate species extinction profiles. The aerosol extinction profiles produced by the SAGE operational algorithm at 1020 and 385 nm together with the air density profiles were obtained from the NASA Langley Research Center. For the 385 nm channel we have added a mean NO_2 profile because it is only a small correction to the total extinction coefficient, the reconstruction of which is trivial.

The aerosol extinction at 600 nm is not given in the SAGE II data set. An independent inversion algorithm, developed by C. Brogniez and J. Lenoble[8] [Chu et al., 1989], which was found to be in fair agreement with the actual SAGE II algorithm, provided these data, and a total extinction profile could be calculated.

4.7.4 Results

Fig. 4.15 shows comparisons of three individual extinction profiles at 1013 (ORA) and 1020 (SAGE II) nm, which is usually the worst channel for ORA in terms of residuals because the dominant aerosol contribution produces a a 'bump' in the transmittance profiles that is difficult to fit. In Table 4.1 we summarize the spatio-temporal coordinates of the events compared.

Inspection of Fig. 4.15 should make clear that the NOPE is a global inversion method (not working layer per layer) and always produces an extinction profile down to the ground. The SAGE II profiles are given down to a cutoff altitude[9], at which the transmittance becomes too

[8]Laboratoire d' Optique Atmosphérique, University of Lille, France

[9]This altitude clearly decreases as the stratospheric Pinatubo aerosol presence

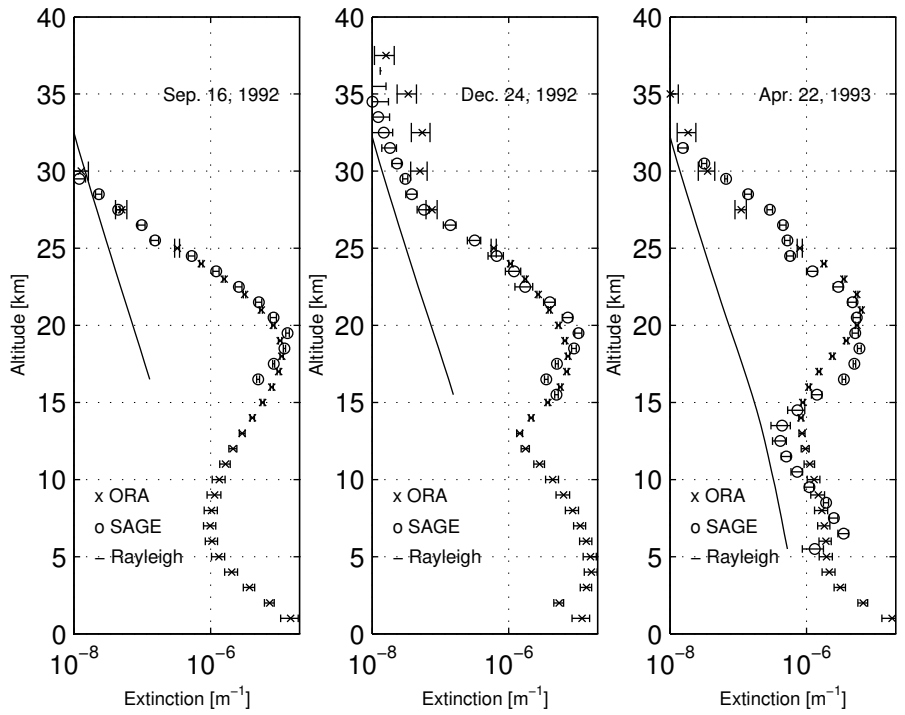

Figure 4.15: Comparison of three total extinction profiles at λ_{1020}: ORA, SAGE and Rayleigh (air density). Error bars are given for both ORA and SAGE values.

low to deliver information. In view of the remarks on saturation made previously, it is clear that the accuracy of the lower part (below 10-15 km) of the ORA profiles is questionable for high optical thicknesses.

The figure also shows that that profiles measured by the ORA and SAGE algorithms are in fair agreement. ORA profiles tend to be more smooth, as expected, because they are constructed from a set of smooth orthogonal functions. Profile 2 (labeled Dec. 24, 1992) also exhibits some extrema above 30 km that are probably spurious and due to critical cancellation effects between the α_i coefficients for low extinction values.

To get a more significant comparison between the two experiments we computed the mean profiles from the 25 coincident events. The av-

decreases.

	Altitude [km]						
channel	15	20	25	30	35	40	45
385 nm	-98	25	-15	1	8	-1	21
600 nm	-33	28	7	4	14	7	-16
1013 nm	16	-25	-12	-46	-8	6	90

Table 4.2: Mean relative differences (%) between total extinctions measured by the ORA and SAGE II experiments for the 25 coincident events.

eraged profiles for the three channels are shown in Fig. 4.16 together with the associated scatter plots that show the data dispersion. General agreement remains fair, but it is clear that the ORA profile structures are somewhat smoothed out with respect to those of the SAGE experiment (in particular at 1013 nm). We interpret this result as an incomplete recovery of the spatial inversion even though convergence was obtained. The aforementioned model errors such as solar disk discretisation, refraction calculations etc. might be implicated. Table 4.2 lists the mean relative differences between the ORA and the SAGE II experimental results.

It should be kept in mind that the number of compared events at 15 and 20 km may be fewer than 25 because of the unavailability of SAGE values in the case of cutoff. Furthermore, cloud contamination (i.e. clouds in the lower part of the solar disk interfering with pure aerosol in the upper part) is expected to occur sometimes in this altitude range because of the ORA large field of view. Such a contamination could lead the algorithm to converge toward a local solution that tends to overestimate the total extinction at low altitudes.

4.7.5 Conclusions

Validation of the SAGE and ORA algorithms is a necessary condition before spectral inversion processing is undertaken. In particular, the ORA instrument has been proved to resolve much sharper profile structures than one would expect from looking at the solar disc. This is the case for the aerosol signature which is satisfactorily retrieved in amplitude and width even if the ORA profiles tend to be slightly smoother. Over the entire spectral and vertical ranges (15-45 km), a typical difference of 20 % can be found between the total extinction profiles retrieved by

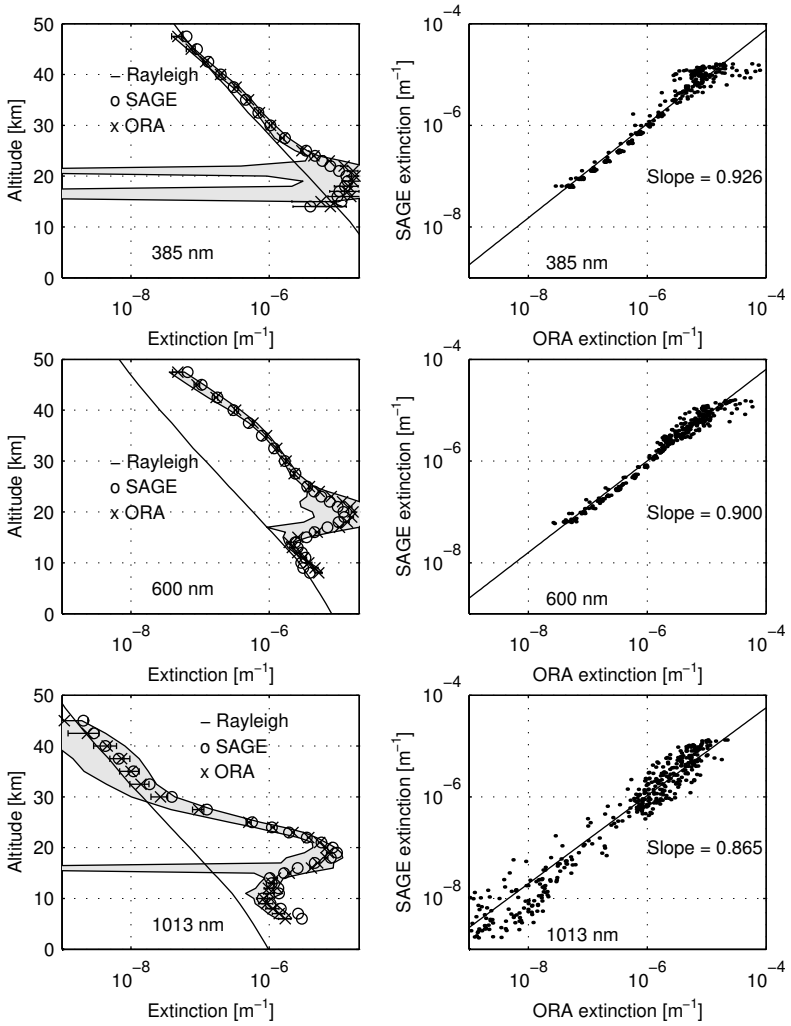

Figure 4.16: Comparison of extinction profiles for the set of closest 25 events (averaged) at three wavelengths. Left: averaged total extinction for ORA and SAGE together with the climatological Rayleigh component. The shaded area corresponds to the standard deviation around the ORA mean values (for clarity, only ORA error bars are given). Right: scatter plot associated with the averaged distributions. The scattering of the points reflects both the differences in measured values and the natural variability. The slope is indicative of the mean correlation.

the SAGE II and the ORA instruments. This value may be considered acceptable because of the imperfect correlation in time, geolocation and viewing geometry that has been remotely sensed.

4.8 The Direct Method: a correction

4.8.1 The basic idea

While the NOPE extinction profiles are reasonable, there is still some chaotic behavior present above 35 km and below 15 km, although not systematically. The reason is of course that transmittance becomes almost one or zero at high and low tangential heights, which means that the information content of the signal becomes very low.

In the course of our research, we decided to construct a method to improve the total extinction profiles, certainly at altitudes where they turned out to be highly unstable. But how are we supposed to do this? Haven't we used all the information that is present in the measurements already? Not quite. We now have thousands of NOPE total extinction profiles, together with their associated transmittance profiles. One way or another, they should be connected directly with each other through some 'atmospheric' function. We already discussed this function $\tau(h) = F_{h,z}(\beta_{\text{ext}}^{\text{tot}}(z))$. It is the forward model that was used in the NOPE method. But it is the *other* direction that we are interested in now:

$$\beta_{\text{ext}}^{\text{tot}}(z) = G_{z,h}(\tau(h)) \tag{4.64}$$

with $G_{z,h} = (F_{h,z})^{-1}$. We also know that it cannot be calculated analytically (otherwise the NOPE method would be unnecessary). The following discussion will explain how we found this function, in a *statistical* way.

4.8.2 The Direct Method

Fig. 4.17 represents a schematic view of the occultation geometry. The transmitted signal at satellite position 2 is mainly influenced by the value of the total extinction at $z = h$ and also by the rays issued from the altitudes ranging approximately from $h_{\min}$ to $h_{\max}$. Inversely, it is natural to consider that the value of the total extinction at $z = h$ can

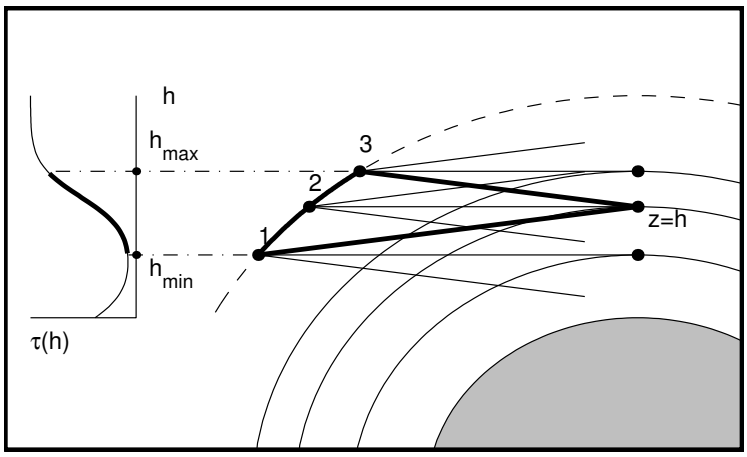

Figure 4.17: Occultation geometry (schematically, refraction is not shown). The measured transmittance (inset) is a smooth function of the tangential altitude. At satellite position 2, the signal is mainly influenced by the atmospheric layers defining the Sun's apparent size (between $z = h_{\min}$ and $z = h_{\max}$). Reciprocally, the value of the extinction coefficient at $z = h$ is strongly determined by the value of the transmittance recorded between satellite position 1 and 3.

be formally determined from the evolution of the total transmittance between points 1 and 3.

Using a scale factor (S=500 km) for the optical path, we now introduce a new effective transmittance function:

$$\tilde{\tau}(z) = \exp[-(\beta_{\text{ext}}^{\text{tot}}(z) - \beta_{\min}^{\text{air}}(z))S] \tag{4.65}$$

The $\beta_{\min}^{\text{air}}$ term makes sure that the total extinction profiles that will be calculated later are constrained to be larger than some minimum threshold:

$$\begin{array}{rcll} \beta_{\min}^{\text{air}}(z) & = & 0.9\tilde{\beta}_{\text{ext}}^{\text{air}}(z) & z \leq 30\text{km} \\ & = & (0.9 - 0.01(z-30))\tilde{\beta}_{\text{ext}}^{\text{air}}(z) & z \leq 50\text{km} \end{array} \tag{4.66}$$

where $\tilde{\beta}_{\text{ext}}^{\text{air}}(z)$ is the Rayleigh extinction profile for US76 Standard Atmosphere [National Oceanic and Atmospheric Organization, 1976]. This minimum threshold was constructed from considerations about the natural variability of air density. Total extinction always has to be larger than the lowest possible amount of Rayleigh scattering.

Eqn. 4.65 can be rewritten as:

$$\beta_{\text{ext}}^{\text{tot}}(z) = -\frac{\ln(\tilde{\tau}(z))}{S} + \beta_{\text{min}}^{\text{air}}(z) \tag{4.67}$$

To make sure that $\tilde{\tau}(z)$ is really a function with values in the range from zero to one, we introduce a tangent hyperbolic form:

$$\tilde{\tau}(z) = \frac{1}{2}[\tanh(\zeta(z) - \frac{1}{2}) + 1] \tag{4.68}$$

Now it is time to express $\zeta(z)$ as a linear combination of the measured transmittances that contribute to the extinction $\beta_{\text{ext}}^{\text{tot}}(z)$. All contribution terms are weighted by some unknown function f:

$$\zeta(z) = \int_{h_{\text{min}}(z)}^{h_{\text{max}}(z)} f(h; z)\tau(h)\,\mathrm{d}h \tag{4.69}$$

By applying a change in variable

$$x(h) = \frac{h - h_{\text{min}}(z)}{\Delta h} - 1 \qquad \text{with} \qquad \Delta h \equiv \frac{h_{\text{max}}(z) - h_{\text{min}}(z)}{2} \tag{4.70}$$

we can write:

$$\zeta(z) = \int_{-1}^{1} f(h(x); z)\tau(h(x))\,\mathrm{d}x \tag{4.71}$$

Eqns. 4.67, 4.68 and 4.71 represent the link between our measured transmittance and calculated extinction profiles. The function f is still unknown, however. Let us develop it (for every altitude z) on a basis of the first eleven Chebyshev polynomials (on the interval $[-1\ 1]$):

$$f(x, z) = \sum_{j=0}^{10} a_j(z)\cos(j\arccos(x)) \tag{4.72}$$

This concludes the mathematical framework. We have selected 1000 occultation events, and used the measured $\tau(z)$ and associated NOPE extinction profiles to determine the $a_j(z)$ coefficients with the aid of a linear least-squares procedure. Values for $h_{\text{min}}(z)$ and $h_{\text{max}}(z)$ were determined from the optical path refraction tables that were also used by the NOPE method.

The function f is thus known. The task of correcting the NOPE profiles is now simple: the new (corrected) extinction profiles are calculated by inserting the measured transmittances in the Eqns. above. The errors on the new profiles were calculated from the fitting errors on the $a_j(z)$ coefficients, the digitization errors σ_τ and the model errors from the NOPE method.

4.8.3 Discussion

The Direct Method is in fact a non-linear mapping between transmittance profiles and extinction profiles. It presupposes that the effective inversion scheme is only determined by the occultation geometry which does not vary from one event to the next.

In Fig. 4.18, we present a typical result of both NOPE and DM methods. In the upper part of the Figure, the NOPE total extinction profile is reported by circles.

At λ_{1013}, some erratic behavior still occurs above 40 km. Notice however that we are dealing with a huge dynamic scale. In the lower part of the figure, the NOPE associated best-fit transmittance is plotted against the measured transmittance. A close examination reveals that the NOPE modelized transmittance does not perfectly fit the dip at 15 km, which is associated with the Junge layer maximum at 20 km. Below 5 km, the transmittance seems to be slightly overestimated by NOPE. Observing the extinction profiles, all these unwanted effects seem to have been improved by the application of the DM (an overestimated NOPE transmittance should be corrected by an increased extinction).

At λ_{600} we observe a very good fit between measurement and NOPE model in the altitude range from 20 to 40 km. The DM does *not* introduce a big correction at these altitudes, as we would expect for coherency. Below 20 km, where the transmittance is almost zero, the total extinction receives extra stabilization.

4.8.4 Conclusion

The Direct Method has increased the accuracy of the total extinction profiles considerably, primarily in altitude regions where a low amount of information is present in the measurements, or where NOPE was unable to show a good fit with the measurement.

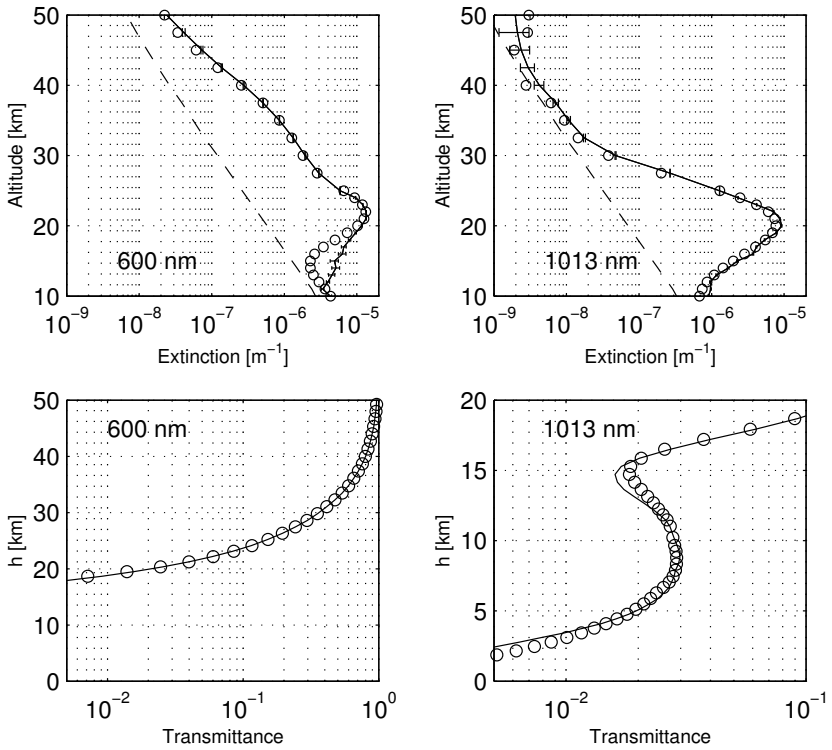

Figure 4.18: Top: NOPE total extinction (circles) and corrected Direct Method profiles (solid lines). The dashed lines represent Rayleigh extinction. Bottom: Measured transmittance (solid lines) and modelized NOPE transmittance (circles).

4.9 Resolution

The results from the Direct Method can be conveniently used to give us an estimate of the ORA vertical resolution. Usually, in *linear* inverse methods, the resolution can be obtained by investigating the matrix of *averaging kernels* $\mathbf{R}$, that relates the a priori profile , the retrieved profile and the true profile:

$$\beta_{\text{ext}}^{\text{tot}} - \tilde{\beta}_{\text{ext}}^{\text{tot}} = \mathbf{R}(\beta_{\text{true}} - \tilde{\beta}_{\text{ext}}^{\text{tot}}) \tag{4.73}$$

At a given altitude, the averaging kernel is then represented by the corresponding row of $\mathbf{R}$. Usually, the kernel forms an altitude profile with a

central lobe, of which the FWHM[10] represents the vertical resolution at the considered altitude. In most of the cases, the averaging kernel also contains negative lobes, making the interpretation of vertical resolution difficult. Therefore, we followed a slightly different approach. Eqn. 4.73 written in continuous form gives:

$$\beta_{\text{ext}}^{\text{tot}}(z) - \tilde{\beta}_{\text{ext}}^{\text{tot}}(z) = \int_0^\infty R(z, z')[\beta_{\text{true}}(z') - \tilde{\beta}_{\text{ext}}^{\text{tot}}(z')]\,\mathrm{d}z' \tag{4.74}$$

The kernel function R can now be constrained to be positive, by choosing

$$R(z, z') = \frac{\gamma(z)}{\delta(z)\sqrt{\pi}} \exp\left[-(\frac{z' - z_\star(z)}{\delta(z)})^2\right] \tag{4.75}$$

with γ the amplitude, δ the width and $z_\star$ the mean of the (Gaussian) kernel.

At each altitude z, Eqn. 4.74 has been numerically solved with a least-squares method for a large number of Direct Method and NOPE extinction profiles (using a mean Direct Method profile as 'a priori' extinction) at λ_{600}, producing values for γ, δ and $z_\star$.

As can be observed in Fig. 4.19, the ORA vertical resolution (δ) is about 2 km, except below 20 km where it increases up to 5 km (most probably due to the low signal). Also, from the values of $z_\star$ we can conclude that the resolution above 40 km is mainly determined by the lower edge of the solar disk. The information content at these altitudes is low (no extinction); most of it comes from the lowest rays in the solar beam.

4.10 The spatial inversion: conclusion

While experiments with a small field of view allow a spatial inversion that is relatively straightforward, things become quite complicated in our case: the mathematical description of the measured signal is far from trivial, and the associated inverse problem is very ill-posed. The NOPE method nevertheless was able to produce total extinction altitude profiles that were acceptable in comparison with those from the SAGE II

[10]Full Width at Half the Maximum.

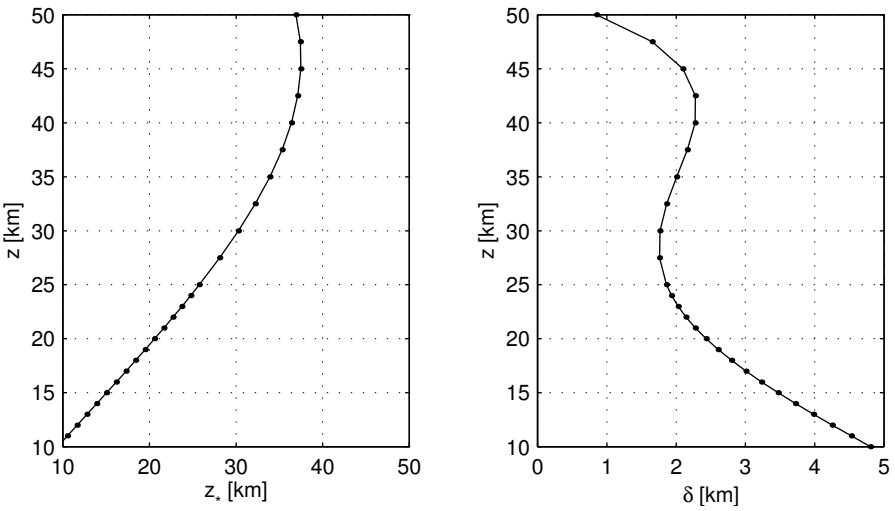

Figure 4.19: Obtained parameters of the resolution kernel at all altitudes. Left: the centroid $z_\star$. Right: the width δ of the Gaussian kernel.

experiment. At high altitudes, extra stabilization was applied, making use of the data statistics: the Direct Method.

Although the *apparent* vertical resolution of the instrument is large (roughly 25 km, as defined by the angular size of the Sun), the excellent signal-to-noise ratio ensures that the actual resolution is much better. We can now proceed to the next chapter, where the separation of the different species contributions will be described.

Chapter 5

The spectral inverse problem.

The previous chapter explained how we retrieved total extinction altitude profiles from the measurements. We can now proceed to the next step, and ask the question: how can we derive the separate contributions of the different species that are present?

The method of the separation is of course based on the extinction cross section wavelength signature of the different species. Hence the word *spectral*. So we will now use all ORA channels at once, which, as will be explained, naturally leads to a new inverse problem.

To find a solution, a specific method is used, based on a well-known (and popular) inverse problem theory. At the end, we will have number density profiles for O_3 and NO_2, and aerosol extinction profiles at six wavelengths. The parameters that characterize the particle size distribution of the aerosols will be derived later, because this requires a separate approach.

The chapter ends with a discussion of the results obtained, and these will (again) be compared with the SAGE II results.

5.1 The formulation

5.1.1 Notation

We start with the mathematical formulation of the inverse problem. To make notation somewhat easier, we will omit the subscripts for extinc-

tion coefficients that were used in previous chapters. The extinction coefficient for constituent j at wavelength λ_i is notated as β_i^{j}. An example for ozone:

$$\beta_{600}^{\mathrm{O_3}} = \mathcal{N}^{\mathrm{O_3}} C_{600}^{O_3} \tag{5.1}$$

with $\mathcal{N}^{\mathrm{O_3}}$ the number density of ozone and $C_{600}^{O_3}$ the extinction cross section of ozone at wavelength λ_{600}. Aerosols on the other hand need a description with a particle size distribution. For example, at 600 nm:

$$\beta_{600}^{\mathrm{aero}} = \int_0^\infty \frac{\mathrm{d}\mathcal{N}^{\mathrm{aero}}(r)}{\mathrm{d}r} C_{600}^{\mathrm{aero}} \,\mathrm{d}r \tag{5.2}$$

5.1.2 The matrix equation

At every altitude, the total extinction at wavelength λ_i is the sum of the separate contributions:

$$\beta_i^{\mathrm{tot}} = \beta_i^{\mathrm{air}} + \beta_i^{\mathrm{O_3}} + \beta_i^{\mathrm{NO_2}} + \beta_i^{\mathrm{aero}} \tag{5.3}$$

We have six wavelengths at our disposal. With the exception of aerosols, every constituent extinction can be expressed as the product of the number density and the cross section. This leads to a linear matrix equation:

$$\begin{bmatrix} C_{340}^{\mathrm{air}} & C_{340}^{\mathrm{O_3}} & C_{340}^{\mathrm{NO_2}} & 1 & 0 & 0 & 0 & 0 & 0 \\ C_{385}^{\mathrm{air}} & C_{385}^{\mathrm{O_3}} & C_{385}^{\mathrm{NO_2}} & 0 & 1 & 0 & 0 & 0 & 0 \\ C_{435}^{\mathrm{air}} & C_{435}^{\mathrm{O_3}} & C_{435}^{\mathrm{NO_2}} & 0 & 0 & 1 & 0 & 0 & 0 \\ C_{442}^{\mathrm{air}} & C_{442}^{\mathrm{O_3}} & C_{442}^{\mathrm{NO_2}} & 0 & 0 & 0 & 1 & 0 & 0 \\ C_{600}^{\mathrm{air}} & C_{600}^{\mathrm{O_3}} & C_{600}^{\mathrm{NO_2}} & 0 & 0 & 0 & 0 & 1 & 0 \\ C_{1013}^{\mathrm{air}} & C_{1013}^{\mathrm{O_3}} & C_{1013}^{\mathrm{NO_2}} & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathcal{N}^{\mathrm{air}} \\ \mathcal{N}^{\mathrm{O_3}} \\ \mathcal{N}^{\mathrm{NO_2}} \\ \beta_{340}^{\mathrm{aero}} \\ \beta_{385}^{\mathrm{aero}} \\ \beta_{435}^{\mathrm{aero}} \\ \beta_{442}^{\mathrm{aero}} \\ \beta_{600}^{\mathrm{aero}} \\ \beta_{1013}^{\mathrm{aero}} \end{bmatrix} = \begin{bmatrix} \beta_{340}^{\mathrm{tot}} \\ \beta_{385}^{\mathrm{tot}} \\ \beta_{435}^{\mathrm{tot}} \\ \beta_{442}^{\mathrm{tot}} \\ \beta_{600}^{\mathrm{tot}} \\ \beta_{1013}^{\mathrm{tot}} \end{bmatrix} \tag{5.4}$$

When we take a look at the values of the cross sections (Figs. 3.5 and 3.6, Chapter 3) we see that they can differ by several orders in magnitude. Therefore, the matrix equation is badly scaled. This is easily cured if we scale a constituent with respect to the wavelength where

its cross section is largest (air scattering: 340 nm; ozone absorption: 600 nm; NO_2 absorption: 435 nm). Using $R_i^{\text{air}} = C_i^{\text{air}}/C_{340}^{\text{air}}$, $R_i^{\text{O}_3} = C_i^{\text{O}_3}/C_{600}^{\text{O3}}$, and $R_i^{\text{NO}_2} = C_i^{\text{NO}_2}/C_{435}^{\text{NO}_2}$, we get:

$$\begin{bmatrix} 1 & R_{340}^{\text{O}_3} & R_{340}^{\text{NO}_2} & 1 & 0 & 0 & 0 & 0 & 0 \\ R_{385}^{\text{air}} & R_{385}^{\text{O}_3} & R_{385}^{\text{NO}_2} & 0 & 1 & 0 & 0 & 0 & 0 \\ R_{435}^{\text{air}} & R_{435}^{\text{O}_3} & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ R_{442}^{\text{air}} & R_{442}^{\text{O}_3} & R_{442}^{\text{NO}_2} & 0 & 0 & 0 & 1 & 0 & 0 \\ R_{600}^{\text{air}} & 1 & R_{600}^{\text{NO}_2} & 0 & 0 & 0 & 0 & 1 & 0 \\ R_{1013}^{\text{air}} & R_{1013}^{\text{O}_3} & R_{1013}^{\text{NO}_2} & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta_{340}^{\text{air}} \\ \beta_{600}^{\text{O}_3} \\ \beta_{435}^{\text{NO}_2} \\ \beta_{340}^{\text{aero}} \\ \beta_{385}^{\text{aero}} \\ \beta_{435}^{\text{aero}} \\ \beta_{442}^{\text{aero}} \\ \beta_{600}^{\text{aero}} \\ \beta_{1013}^{\text{aero}} \end{bmatrix} = \begin{bmatrix} \beta_{340}^{\text{tot}} \\ \beta_{385}^{\text{tot}} \\ \beta_{435}^{\text{tot}} \\ \beta_{442}^{\text{tot}} \\ \beta_{600}^{\text{tot}} \\ \beta_{1013}^{\text{tot}} \end{bmatrix} \tag{5.5}$$

with every component $R_i^{\text{j}} < 1$.

In short notation, this equation becomes:

$$\mathbf{A_z} \cdot \mathbf{x_z} = \mathbf{b_z} \tag{5.6}$$

The subscripts z indicate the altitude that is considered. Indeed, we want to obtain altitude profiles from total extinction profiles, and the matrix $\mathbf{A_z}$ is also dependent on altitude because cross sections are (in general) temperature dependent.

5.1.3 Aerosol spectral dependence

Molecular species have cross sections with a wavelength dependence that is known in advance (from laboratory measurements). Unfortunately, this is not true for the aerosol extinction cross sections. The wavelength dependence for extinction by *one* particle was already discussed in Chapter 3. It depends heavily on the size parameter $x = \text{k}a$ and the relative refraction index m of the fluid/solid of which the particle is made.

Moreover, the extinction wavelength dependence of a *population* of particles depends on the particle size distribution (Eqn. 5.2), that is completely unknown in advance. Several types of analytic formula for aerosol extinction have been proposed in literature, each of them suitable for a specific type of aerosol. The Angstrom law, for example, has

been applied frequently in the past to describe aerosol extinction in the stratosphere.

In usual (non-volcanic) conditions, it is well-known that stratospheric aerosol distributions have a *mean* size that is small (less than 0.1 μm). As we have seen, in these cases, scattering by aerosols much resembles simple molecular scattering (with a λ^{-4} dependency). Also, the actual extinction values are low, and often the aerosol contribution is simply represented by an additional term, to ensure that the residual from the spectral inversion is acceptable.

Our case is quite different. The volcanic aerosol loading is high, and particles are much larger (which has been known for quite some time now from other experimental results). For now, the details will be postponed until Chapter 6. Simulations have shown us that it is inadequate to model aerosol extinction by a monotonic decreasing function of wavelength. The model should at least be able to exhibit some maximum at a certain wavelength. Furthermore, the wavelength dependence should be *smooth*. We decided to use the following expression:

$$\beta_j^{\text{aero}} = c_0 + c_1(\lambda_j - \lambda_{1013}) + c_2(\lambda_j - \lambda_{1013})^2 \tag{5.7}$$

The reference with respect to λ_{1013} is natural because it is the channel that almost exclusively measures aerosol extinction.

What is so special about this form? First of all, if we introduce this equation in our matrix, it will remain linear, so no iterative methods will be necessary. Second, *within* the total extinction measurement error, this form can reasonably approximate a wide variety of extinction wavelength dependencies (constant, λ^{-4}, peaking in the mid-visible, etc.).

Now we can insert Eqn. 5.7 in the matrix problem by a transformation of the state vector $\mathbf{x_z}$:

$$\begin{bmatrix} \beta_{340}^{\text{air}} \\ \beta_{600}^{\text{O}_3} \\ \beta_{435}^{\text{NO}_2} \\ \beta_{340}^{\text{aero}} \\ \beta_{385}^{\text{aero}} \\ \beta_{435}^{\text{aero}} \\ \beta_{442}^{\text{aero}} \\ \beta_{600}^{\text{aero}} \\ \beta_{1013}^{\text{aero}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \lambda_{340} - \lambda_{1013} & (\lambda_{340} - \lambda_{1013})^2 \\ 0 & 0 & 0 & 1 & \lambda_{385} - \lambda_{1013} & (\lambda_{385} - \lambda_{1013})^2 \\ 0 & 0 & 0 & 1 & \lambda_{435} - \lambda_{1013} & (\lambda_{435} - \lambda_{1013})^2 \\ 0 & 0 & 0 & 1 & \lambda_{442} - \lambda_{1013} & (\lambda_{442} - \lambda_{1013})^2 \\ 0 & 0 & 0 & 1 & \lambda_{600} - \lambda_{1013} & (\lambda_{600} - \lambda_{1013})^2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \beta_{340}^{\text{air}} \\ \beta_{600}^{\text{O}_3} \\ \beta_{435}^{\text{NO}_2} \\ c_0 \\ c_1 \\ c_2 \end{bmatrix} \tag{5.8}$$

or, in short notation:

$$\mathbf{x_z} = \mathbf{K} \cdot \mathbf{y_z} \tag{5.9}$$

Using Eqns. 5.5 and 5.8, we finally derive:

$$\begin{bmatrix} 1 & R_{340}^{\text{O}_3} & R_{340}^{\text{NO}_2} & 1 & \lambda_{340} - \lambda_{1013} & (\lambda_{340} - \lambda_{1013})^2 \\ R_{385}^{\text{air}} & R_{385}^{\text{O}_3} & R_{385}^{\text{NO}_2} & 1 & \lambda_{385} - \lambda_{1013} & (\lambda_{385} - \lambda_{1013})^2 \\ R_{435}^{\text{air}} & R_{435}^{\text{O}_3} & 1 & 1 & \lambda_{435} - \lambda_{1013} & (\lambda_{435} - \lambda_{1013})^2 \\ R_{442}^{\text{air}} & R_{442}^{\text{O}_3} & R_{442}^{\text{NO}_2} & 1 & \lambda_{442} - \lambda_{1013} & (\lambda_{442} - \lambda_{1013})^2 \\ R_{600}^{\text{air}} & 1 & R_{600}^{\text{NO}_2} & 1 & \lambda_{600} - \lambda_{1013} & (\lambda_{600} - \lambda_{1013})^2 \\ R_{1013}^{\text{air}} & R_{1013}^{\text{O}_3} & R_{1013}^{\text{NO}_2} & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \beta_{340}^{\text{air}} \\ \beta_{600}^{\text{O}_3} \\ \beta_{435}^{\text{NO}_2} \\ c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \beta_{340}^{\text{tot}} \\ \beta_{385}^{\text{tot}} \\ \beta_{435}^{\text{tot}} \\ \beta_{442}^{\text{tot}} \\ \beta_{600}^{\text{tot}} \\ \beta_{1013}^{\text{tot}} \end{bmatrix} \tag{5.10}$$

In short notation:

$$\mathbf{A_z} \cdot \mathbf{x_z} = \mathbf{A_z} \cdot \mathbf{K} \cdot \mathbf{y_z} = \mathbf{b_z} \tag{5.11}$$

or:

$$\mathbf{B_z} \cdot \mathbf{y_z} = \mathbf{b_z} \tag{5.12}$$

which is a (mathematically!) well-posed problem (6 measurements, 6 unknowns, no linear dependent rows). It should now become clear why

we have chosen an aerosol description (Eqn. 5.7) with 3 parameters. If we had used more, the matrix equation would have been underdetermined.

5.2 Inversion: a Bayesian approach

5.2.1 Probability density functions

Consider a (scalar) quantity $\bar{y}$ with uncertainty σ (for example: experimental error). In fact, this is shorthand for saying that our knowledge of the true value of the measured quantity is described by a *probability density function* (*pdf*)[1] $P(y)$ with a mean $\bar{y}$ and variance σ^2. Very often, one assumes the *pdf* to be *Gaussian*:

$$P(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\bar{y})^2}{2\sigma^2}\right\} \tag{5.13}$$

Now we can generalize to the multi-dimensional case. Let us consider a $(m \times 1)$ vector $\mathbf{y}$. Clearly, if the individual components y_i are independent, then the *pdf* of $\mathbf{y}$ is a product of m partial single-variable *pdf*s such as Eqn. 5.13. If the individual components y_i are not independent, then they are *correlated*. This is expressed by the *covariance* of two components y_i and y_j:

$$S_{ij} = \varepsilon\{(y_i - \bar{y}_i)(y_j - \bar{y}_j)\} = \sigma_{ij}^2 \tag{5.14}$$

with ε the *expectancy* operator. The *correlation* of y_i and y_j is then:

$$\rho_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} = \frac{\sigma_{ij}^2}{\sigma_i \sigma_j} \tag{5.15}$$

with $-1 \leq \rho_{ij} \leq 1$.

The (m × m) covariance matrix $\mathbf{S_y}$ thus fully describes the variances of and correlations between the components y_i.

The *pdf* of $\mathbf{y}$, as a generalization of Eqn. 5.13 thus reads:

$$P(\mathbf{y}) = \frac{1}{(2\pi)^{m/2}\|\mathbf{S_y}\|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\bar{\mathbf{y}})^T \mathbf{S_y}^{-1}(\mathbf{y}-\bar{\mathbf{y}})\right\} \tag{5.16}$$

[1] This is not a probability density function in the statistical view: it is not (necessarily) derived from statistical considerations on a set of repeated measurements.

5.2.2 Probability and inverse problems

Again, consider the (general) inverse problem:

$$\mathbf{y} = F(\mathbf{x}) \tag{5.17}$$

with associated measurement uncertainty, expressed by the covariance matrix $\mathbf{S}_y$. Furthermore, we always have *some* knowledge about the state $\mathbf{x}$ before the measurement is made, although this knowledge can be very minimal. We therefore introduce the a priori mean state $\mathbf{x}_a$ and the a priori covariance matrix $\mathbf{S_a}$. In human words: *before* the measurements are made, we know that the solution $\mathbf{x}$ will be near to some mean state $\mathbf{x_a}$. How near depends on the associated covariance $\mathbf{S_a}$.

We can now introduce a few probability density functions. Let us assume that they are general: no assumptions about their form are needed at this point.

- $P(\mathbf{x})$. This is the *pdf* of the state $\mathbf{x}$ *before* the measurement is taken: the *a priori pdf*.

- $P(\mathbf{y})$. The *pdf* of the measurement $\mathbf{y}$ *before* the measurement is taken. This might seem to appear strange, but we always have some advance information on what the measurement is going to look like.

- $P(\mathbf{y}|\mathbf{x})$. The *pdf* of $\mathbf{y}$ given $\mathbf{x}$. This is the forward model, telling us what $\mathbf{y}$ will be for a given $\mathbf{x}$.

- $P(\mathbf{x}|\mathbf{y})$. The *pdf* of $\mathbf{x}$ given $\mathbf{y}$. This is the *pdf* of $\mathbf{x}$ *after* the measurement has been performed. In other words: the *pdf* of the inverse problem solution.

Now it is clear what a measurement means in the Bayesian view. It is the mapping of some region in the state space onto some region in the measurement space. Both regions are expressed by a *pdf* with associated mean and covariance.

5.2.3 Bayes' theorem

The probability density functions that were defined above can now be linked with the well-known theorem of Bayes:

$$P(\mathbf{y}|\mathbf{x})P(\mathbf{x}) = P(\mathbf{x}|\mathbf{y})P(\mathbf{y}) \tag{5.18}$$

and, rewriting this, we find the *pdf* of the solution:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \tag{5.19}$$

As already said, we have not assumed any specific form of the *pdf*s. Also, no assumptions were made regarding the form of the model. Eqn. 5.19 therefore probably represents the most general expression of an inverse problem solution that can be found in scientific literature.

5.2.4 A linear model with Gaussian statistics

Now, we will be more specific. Suppose that the inverse problem is linear:

$$\mathbf{y} = \mathbf{K} \cdot \mathbf{x} \tag{5.20}$$

where $\mathbf{y}$ is once again the $(m \times 1)$ measurement vector and $\mathbf{x}$ the $(n \times 1)$ state vector.

Furthermore, we assume that all *pdf*s are Gaussian. So we have:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}\|\mathbf{S_y}\|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{K}\mathbf{x})^T \mathbf{S_y}^{-1}(\mathbf{y}-\mathbf{K}\mathbf{x})\right\} \tag{5.21}$$

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\|\mathbf{S_a}\|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x_a})^T \mathbf{S_a}^{-1}(\mathbf{x}-\mathbf{x_a})\right\} \tag{5.22}$$

Applying Bayes' theorem (Eqn. 5.19) we can derive:

$$P(\mathbf{x}|\mathbf{y}) = A \exp\left\{-\frac{1}{2}[(\mathbf{y}-\mathbf{K}\mathbf{x})^T \mathbf{S_y}^{-1}(\mathbf{y}-\mathbf{K}\mathbf{x}) + (\mathbf{x}-\mathbf{x_a})^T \mathbf{S_a}^{-1}(\mathbf{x}-\mathbf{x_a})]\right\} \tag{5.23}$$

with A a normalizing factor. $P(\mathbf{y})$ is not written explicitly, but is included in A.

Evaluating the expression in the exponent, we find that it can be written as the sum of a quadratic form of $\mathbf{x}$ and some extra terms that do not depend on $\mathbf{x}$:

$$\text{exponential terms} = (\mathbf{x} - \tilde{\mathbf{x}})^T \tilde{\mathbf{S}}_{\mathbf{x}}^{-1} (\mathbf{x} - \tilde{\mathbf{x}}) + \text{other terms} \tag{5.24}$$

And therefore:

$$P(\mathbf{x}|\mathbf{y}) = B \exp\left\{ -\frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^T \tilde{\mathbf{S}}_{\mathbf{x}}^{-1} (\mathbf{x} - \tilde{\mathbf{x}}) \right\} \tag{5.25}$$

This is a Gaussian probability density function with mean $\tilde{\mathbf{x}}$ and covariance $\tilde{\mathbf{S}}_{\mathbf{x}}$. Identifying terms between Eqn. 5.23 and 5.25, we obtain:

$$\tilde{\mathbf{x}} = \mathbf{x_a} + (\mathbf{K}^T {\mathbf{S_y}}^{-1} \mathbf{K} + {\mathbf{S_a}}^{-1})^{-1} \mathbf{K}^T {\mathbf{S_y}}^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x_a}) \tag{5.26}$$

and

$$\tilde{\mathbf{S}}_{\mathbf{x}}^{-1} = \mathbf{K}^T {\mathbf{S_y}}^{-1} \mathbf{K} + {\mathbf{S_a}}^{-1} \tag{5.27}$$

Eqns. 5.26 and 5.27 represent the solution of the inverse problem. This solution should be looked upon as the a priori mean and covariance of $\mathbf{x}$, updated with information that has been obtained through measurement. $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{S}}_{\mathbf{x}}$ are also known as the *posterior* state and covariance.

5.2.5 An alternative discussion

From Eqn. 5.23, we see that the posterior *pdf* has a maximum when the expression

$$(\mathbf{y} - \mathbf{K}\mathbf{x})^T {\mathbf{S_y}}^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}) + (\mathbf{x} - \mathbf{x_a})^T {\mathbf{S_a}}^{-1} (\mathbf{x} - \mathbf{x_a}) \tag{5.28}$$

is at its minimum. Therefore, $\tilde{\mathbf{x}}$ is the solution of $\mathbf{y} = \mathbf{K}\mathbf{x}$ that minimizes:

$$\langle \mathbf{y} - \mathbf{K}\mathbf{x} | {\mathbf{S_y}}^{-1} | \mathbf{y} - \mathbf{K}\mathbf{x} \rangle + \langle \mathbf{x} - \mathbf{x}_a | {\mathbf{S_a}}^{-1} | \mathbf{x} - \mathbf{x}_a \rangle \tag{5.29}$$

Without prior information, only the first term would be present. We immediately recognize it as the merit function that is minimized in the

least-squares method. Therefore, Eqns. 5.26 and 5.27 represent the solution to a least-squares problem, but constrained by a priori information.

Since the solution maximizes the posterior *pdf* of $\mathbf{x}$, it is also known as the *maximum posterior solution* [2].

5.2.6 A simple example

Although the equations may seem complicated, the principle is quite simple. Take, for instance, the inverse problem:

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{5.30}$$

with the exact solution $\mathbf{x} = [1\ 1]^T$ (which is unknown!). Now let us add some (uncorrelated) measurement error:

$$\mathbf{S_y} = \begin{bmatrix} 0.04 & 0 \\ 0 & 0.09 \end{bmatrix} \tag{5.31}$$

One specific measurement within the error range could be: $\mathbf{y} = [3.1\ 3.9]^T$ with a normal solution $\mathbf{x} = [1.5\ 0.8]^T$, clearly showing the ill-posedness of the problem.

It is therefore necessary to introduce prior information. Suppose we *know* that the solution should be near the prior state vector $\mathbf{x_a}$ within a range specified by $\mathbf{S_a}$:

$$\mathbf{x_a} = [1.2\ 0.9]^T \tag{5.32}$$

$$\mathbf{S_a} = \begin{bmatrix} 0.09 & 0.072 \\ 0.072 & 0.09 \end{bmatrix} \tag{5.33}$$

Eqns. 5.26 and 5.27 give us the posterior solution:

$$\tilde{\mathbf{x}} = [1.22\ 0.92]^T \tag{5.34}$$

$$\tilde{\mathbf{S}}_\mathbf{x} = \begin{bmatrix} 0.0192 & -0.0049 \\ -0.0049 & 0.0056 \end{bmatrix} \tag{5.35}$$

[2] Better known as the *maximum likelihood solution*, a label that is incorrect however. See [Rodgers, 2000].

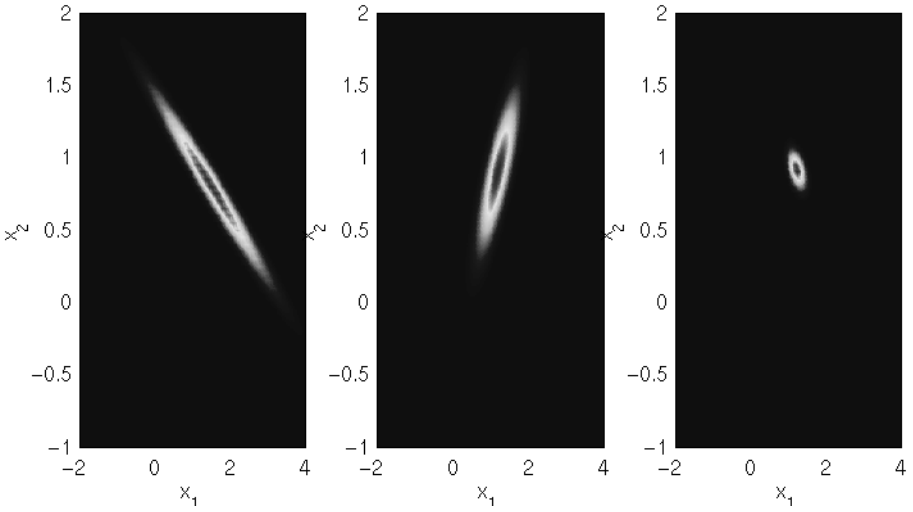

Figure 5.1: Exemplary *pdf*s as function of state vector components x_1 and x_2. Left: $P(\mathbf{y}|\mathbf{x})$. Notice the strong (anti)correlation that is typical for ill-posed problems. Center: $P(\mathbf{x})$, the prior state vector *pdf*. Right: the posterior *pdf*, $P(\mathbf{x}|\mathbf{y})$. The solution has been constrained considerably.

which is not far from what we wanted to achieve. The component x_1 has been resolved with a posterior error of 11 % while x_2 is known with an accuracy of 8 %. Both components are anticorrelated ($\rho = -0.47$).

The entire procedure has been graphically depicted in Fig. 5.1.

5.3 Removal of air scattering

5.3.1 UKMO data

We now have the tools to perform the spectral inversion. Still, since both Rayleigh scattering and aerosol scattering have a smooth wavelength dependence, we need to worry. The worst case will occur when the aerosol particles are small, for then the scattering wavelength dependence is identical to that of air scattering.

We therefore decided to remove the air scattering component, using external information. For years, the *United Kingdom Meteorological Office* (UKMO) has been collecting data from diverse experiments such

as the NOAA satellites and ground-based weather stations, and has inserted these data (temperature, air density, pressure, and wind velocity) into a data assimilation algorithm. Users can provide geolocation and altitude level grids, on which the required data is then interpolated.

For every ORA occultation event, we obtained UKMO temperature and pressure altitude profiles, and calculated air density profiles using the ideal gas law. The extinction profiles were subsequently calculated.

5.3.2 An uninvited guest: diffuse straylight

After comparison of the UKMO Rayleigh extinction profiles with the ORA total extinction profiles, it became clear that something was wrong. At high altitudes, the total extinction was *systematically* lower than the Rayleigh extinction profiles (see Fig. 5.2). This means that we have neglected an additional atmospheric process that '*produces*' light and hereby lowers the measured total extinction.

Inspection of the difference between measured total extinction and meteorological data at high altitudes showed us that the 'lack' of extinction decreases with increasing wavelength. Fitting the mysterious negative contribution with polynomials, we were able to identify a λ^{-4} wavelength dependence. We simply are dealing with straylight!

It is quite easy to see how this is possible. Each ORA module has a $\pm 2°$field of view that is circular in shape. Only about half a degree of this view is occupied by the solar disk. In the remaining part, ORA is able to receive light, scattered from a large part of the atmosphere. Indeed, while the field of view may seem small, ORA is still able to view the entire atmosphere altitude range at once.

We now have an additional term that has to be included in the spectral model. Here, we face a very difficult task. Light, on entering upon an air molecule scatters in all directions according to the phase function. This secondary radiation is attenuated by other species, and will be scattered again at another location in the atmosphere, etc. An analytic model for this complicated behavior does not exist, and it is necessary to simulate the entire process. Needless to say, the computational costs are huge.

On the other hand, the scattered light contribution will not vary a lot during one occultation, since it is formed by partial contributions coming from a wide altitude range. It is therefore reasonable to model

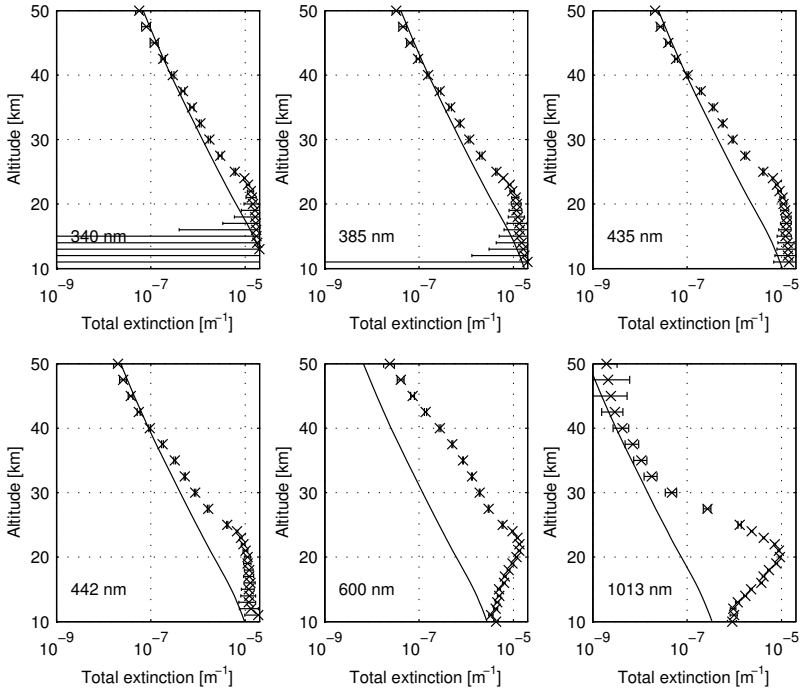

Figure 5.2: ORA total extinction and UKMO meteorological Rayleigh extinction profiles. At short wavelengths and high altitudes, extinction values are clearly too low.

the straylight component as being constant in altitude, with a λ^{-4} wavelength dependency.

5.4 The spectral inversion algorithm

5.4.1 The full matrix equation

After removal of the Rayleigh component, at each altitude, Eqn. 5.5 reads:

$$\begin{bmatrix} R^{O_3}_{340} & R^{NO_2}_{340} & 1 & 0 & 0 & 0 & 0 & 0 \\ R^{O_3}_{385} & R^{NO_2}_{385} & 0 & 1 & 0 & 0 & 0 & 0 \\ R^{O_3}_{435} & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ R^{O_3}_{442} & R^{NO_2}_{442} & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & R^{NO_2}_{600} & 0 & 0 & 0 & 0 & 1 & 0 \\ R^{O_3}_{1013} & R^{NO_2}_{1013} & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \beta^{O_3}_{600} \\ \beta^{NO_2}_{435} \\ \beta^{aero}_{340} \\ \beta^{aero}_{385} \\ \beta^{aero}_{435} \\ \beta^{aero}_{442} \\ \beta^{aero}_{600} \\ \beta^{aero}_{1013} \end{bmatrix} = \begin{bmatrix} \beta^{tot}_{340} \\ \beta^{tot}_{385} \\ \beta^{tot}_{435} \\ \beta^{tot}_{442} \\ \beta^{tot}_{600} \\ \beta^{tot}_{1013} \end{bmatrix} \tag{5.36}$$

and we maintain the short-hand notation:

$$\mathbf{A_z x_z} = \mathbf{b_z} \tag{5.37}$$

The aerosol transformation, Eqn. 5.8 now becomes:

$$\begin{bmatrix} \beta^{O_3}_{600} \\ \beta^{NO_2}_{435} \\ \beta^{aero}_{340} \\ \beta^{aero}_{385} \\ \beta^{aero}_{435} \\ \beta^{aero}_{442} \\ \beta^{aero}_{600} \\ \beta^{aero}_{1013} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda_{340} - \lambda_{1013} & (\lambda_{340} - \lambda_{1013})^2 \\ 0 & 0 & 1 & \lambda_{385} - \lambda_{1013} & (\lambda_{385} - \lambda_{1013})^2 \\ 0 & 0 & 1 & \lambda_{435} - \lambda_{1013} & (\lambda_{435} - \lambda_{1013})^2 \\ 0 & 0 & 1 & \lambda_{442} - \lambda_{1013} & (\lambda_{442} - \lambda_{1013})^2 \\ 0 & 0 & 1 & \lambda_{600} - \lambda_{1013} & (\lambda_{600} - \lambda_{1013})^2 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \beta^{O_3}_{600} \\ \beta^{NO_2}_{435} \\ c_0 \\ c_1 \\ c_2 \end{bmatrix} \tag{5.38}$$

or:

$$\mathbf{x_z} = \mathbf{K y_z} \tag{5.39}$$

We model the straylight as being constant in altitude. First we define a 6-component vector that contains the λ^{-4} scattering behavior:

$$\mathbf{s} = [\lambda^{-4}_{340}\ \lambda^{-4}_{385}\ \lambda^{-4}_{435}\ \lambda^{-4}_{442}\ \lambda^{-4}_{600}\ \lambda^{-4}_{1013}]^T \tag{5.40}$$

Now we can stack the entire ensemble of separate altitude models (Eqn. 5.37) into one large matrix:

$$\begin{bmatrix} \mathbf{A_1} & & & & \mathbf{s} \\ & \mathbf{A_2} & & & \mathbf{s} \\ & & \ddots & & \vdots \\ & & & \mathbf{A_{26}} & \mathbf{s} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_{26}} \\ x_s \end{bmatrix} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \\ \vdots \\ \mathbf{b_{26}} \end{bmatrix} \tag{5.41}$$

In short notation:

$$\mathbf{A_t x_t} = \mathbf{b_t} \tag{5.42}$$

The full aerosol transformation for all altitudes is then:

$$\mathbf{x_t} = \mathbf{K_t y_t} \tag{5.43}$$

with

$$\mathbf{K_t} = \begin{bmatrix} \mathbf{K} & & & \\ & \ddots & & \\ & & \mathbf{K} & \\ & & & 1 \end{bmatrix} \tag{5.44}$$

and we come to the final formulation of the inverse problem:

$$\mathbf{A_t x_t} = \mathbf{A_t K_t y_t} = \mathbf{b_t} \tag{5.45}$$

or:

$$\begin{array}{ccccc} \mathbf{B_t} & & \mathbf{y_t} & = & \mathbf{b_t} \\ (156 \times 131) & & (131 \times 1) & & (156 \times 1) \end{array} \tag{5.46}$$

5.4.2 A priori information

For O_3 and NO_2, we have advance knowledge, provided by the US76 climatological mean density profiles. However, the variance on these profiles is not known. We therefore have to be careful: if the assumed prior variance is too small, we may introduce unwanted bias in the solution. For both gases, we have chosen the prior variance to be 50 % with respect to the climatological profiles. This choice may seem rather arbitrary, but it ensures a positive solution in most cases and, at the same time, provides a constraint that is not too tight. So we have:

$$\begin{array}{ll} \tilde{\beta}^{\mathrm{O_3}}_{600} = C^{\mathrm{O_3}}_{600} \tilde{\mathcal{N}}^{\mathrm{O_3}} & \sigma^{\mathrm{O_3}} = 0.5 \tilde{\beta}^{\mathrm{O_3}}_{600} \\ \tilde{\beta}^{\mathrm{NO_2}}_{435} = C^{\mathrm{NO_2}}_{435} \tilde{\mathcal{N}}^{\mathrm{NO_2}} & \sigma^{\mathrm{NO_2}} = 0.5 \tilde{\beta}^{\mathrm{NO_2}}_{435} \end{array} \tag{5.47}$$

For aerosols, the case is different. Once again, we cannot make use of a climatological profile, because of the abnormal amount of aerosols that are present. We do *know*, however, that after removal of the Rayleigh component, the total extinction in channel λ_{1013} is entirely caused by

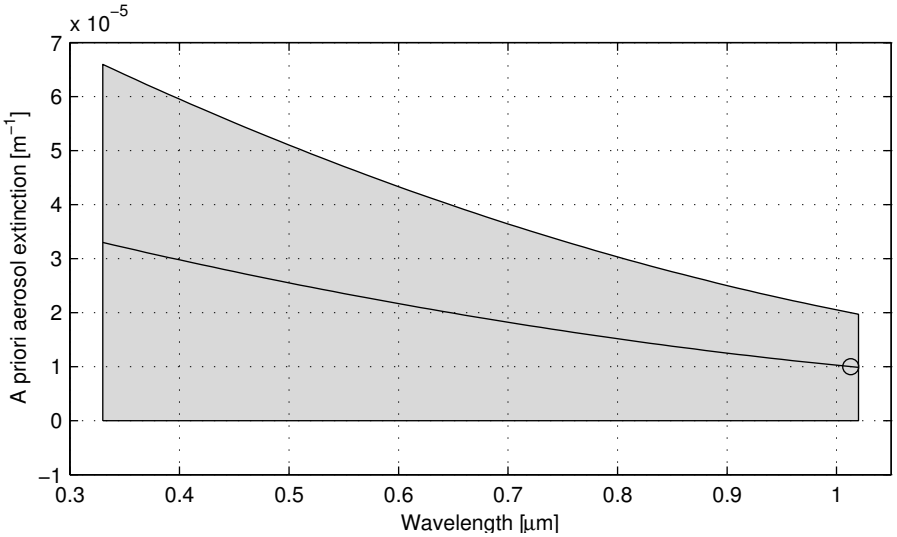

Figure 5.3: The a priori aerosol extinction (solid line) and variance (shaded area), when the measured aerosol extinction at λ_{1013} (circle) equals 10^{-5} m^{-1}.

aerosols. This is an a priori statement. At other wavelengths, the prior aerosol extinction was evaluated as:

$$\tilde{\beta}_{\lambda}^{\text{aero}} = \tilde{c}_0 + \tilde{c}_1(\lambda - \lambda_{1013}) + \tilde{c}_2(\lambda - \lambda_{1013})^2 \tag{5.48}$$

with

$$\tilde{c}_0 = \beta_{1013}^{\text{aero}} \quad \tilde{c}_1 = -2\beta_{1013}^{\text{aero}} \quad \tilde{c}_2 = 2\beta_{1013}^{\text{aero}} \tag{5.49}$$

For this shape, the aerosol extinction gradually increases with decreasing wavelength (Fig. 5.3).

What about the aerosol variance? We have practically no knowledge about the natural variation of Pinatubo aerosols, and therefore we chose a constraint that is not too tight (100 %):

$$\sigma_{\lambda}^{\text{aero}} = \tilde{\beta}_{\lambda}^{\text{aero}} \tag{5.50}$$

And with this, the a priori mean and variance of all constituents have been defined. Now, we can also introduce an artificial covariance. Indeed, one intuitively senses that the profile for a particular constituent has to be *smooth* to a certain extent: it is reasonable to suppose that

the underlying microphysical quantities depend smoothly on altitude (within a range of a few kilometers). This does not mean that fine variations are not possible from one altitude to the next. But they never can be *huge*. For a particular constituent, we can introduce an altitude correlation function that has a Gaussian shape. Ozone, for example, receives this smoothing:

$$(\sigma^{\mathrm{O_3}})^2(z_1, z_2) = \sigma^{\mathrm{O_3}}(z_1)\sigma^{\mathrm{O_3}}(z_2) \exp\left\{-\left(\frac{z_1 - z_2}{L}\right)^2\right\} \tag{5.51}$$

A choice of $L = 5$ km was found to be a good trade-off between efficient regularization and oversmoothing.

All this information can now be stacked in the a priori state vector $\mathbf{x_{t,a}}$ with associated covariance matrix $\mathbf{S_{x,a}}$. We need to apply an extra transformation because we are going to solve Eqn. 5.46 with unknown vector $\mathbf{y_t}$. We know that $\mathbf{x_t} = \mathbf{K_t y_t}$. Thus:

$$\mathbf{K_t}^T \mathbf{x_t} = \mathbf{K_t}^T \mathbf{K_t y_t} \tag{5.52}$$

and

$$\mathbf{y_t} = (\mathbf{K_t}^T \mathbf{K_t})^{-1} \mathbf{K_t}^T \mathbf{x_t} \tag{5.53}$$

The a priori $\mathbf{y_t}$ thus reads:

$$\mathbf{y_{t,a}} = (\mathbf{K_t}^T \mathbf{K_t})^{-1} \mathbf{K_t}^T \mathbf{x_{t,a}} \tag{5.54}$$

with associated covariance matrix[3]:

$$\mathbf{S_{y,a}} = (\mathbf{K_t}^T \mathbf{K_t})^{-1} \mathbf{K_t}^T \mathbf{S_{x,a}} \mathbf{K_t} (\mathbf{K_t}^T \mathbf{K_t})^{-1} \tag{5.55}$$

One last remark concerning the use of prior information. Regularization by the use of covariance matrices is not *strict*. It is not necessarily true that the solution will actually be located in the domain that is defined by the prior mean and covariance. Only in the case where measurement information is lacking will the prior information come into action. Or, otherwise stated: if the *pdf* of the measurement is broader then the prior *pdf*, then the latter will play a role.

[3]For a linear transformation $\mathbf{u} = \mathbf{A v}$, we have $\mathbf{S_u} = \mathbf{A S_v A}^T$.

5.4.3 Solving the problem

If $\mathbf{S_b}$ represents the covariance matrix, associated with the 'measurement' $\mathbf{b_t}$, then Eqns. 5.26 and 5.27 give us the solution of the problem:

$$\tilde{\mathbf{y}}_\mathbf{t} = \mathbf{y}_{\mathbf{t,a}} + (\mathbf{B_t}^T \mathbf{S_b}^{-1} \mathbf{B_t} + \mathbf{S_{y,a}}^{-1}) \mathbf{B_t}^T \mathbf{S_b}^{-1} (\mathbf{b_t} - \mathbf{B_t} \mathbf{y}_{\mathbf{t,a}}) \qquad (5.56)$$

with the posterior covariance matrix:

$$\tilde{\mathbf{S}}_\mathbf{y} = (\mathbf{B_t}^T \mathbf{S_b}^{-1} \mathbf{B_t} + \mathbf{S_{y,a}}^{-1})^{-1} \qquad (5.57)$$

5.4.4 The results

All ORA occultation events were processed using the formalism above. The time required for one event to be processed is moderate, since we are dealing with a linear problem. A few days were sufficient to handle the entire ORA data set. The results (number density profiles for O_3 and NO_2, six aerosol extinction profiles, and associated variances and covariances) were stored on disk.

5.5 Comparison with SAGE II data

In the previous chapter, we have identified 25 ORA occultation events that are coincident with SAGE II occultations, within a window of time and geolocation (criterion 4.63). We can now use the same events to compare the profiles of the individual species.

Table 4.2 indicates the differences in total extinction for the two experiments, from which we can conclude that the comparison is fair, but not spectacular. Therefore, the same conclusion will apply to the obtained O_3 and NO_2 density profiles and aerosol extinction profiles.

Since ORA and SAGE II profiles are calculated on a different altitude grid, SAGE profiles were evaluated at ORA altitude levels using a linear interpolation scheme.

Only three SAGE channels have the same wavelength as those from ORA. In order to compare at other wavelengths, the SAGE aerosol extinction profiles have been fitted with a second order polynomial function (conform with Eqn. 5.7), and have been evaluated at ORA wavelengths.

And finally, the mean and variance of the 25 events were calculated. A complete comparison should include both the variability of the profile set and their individual errors:

$$\epsilon^2 = \frac{\sum_{i=1}^{25}(p_i - \bar{p})^2}{25} + \frac{\sum_{i=1}^{25}\sigma_i^2}{25} \tag{5.58}$$

with p_i and σ_i the i-th profile and its error, respectively. The relative percentual difference between the two experiments was expressed as 100× (ORA-SAGE)/SAGE.

The results for ozone and NO_2 are depicted in Fig. 5.4. Data for SAGE II were only available in the altitude range from 20 to 50 km. For ozone, the difference is at most 40 percent. The NO_2 profiles are also in fair agreement, except at lower altitudes, where the ORA mean profile is abnormally large. The main reason is probably the coupling between the aerosols and NO_2 during the spectral inversion, because the spurious peak of the NO_2 profile occurs at the same altitude as the aerosol peak. Also, this is an altitude region where the relative contribution of NO_2 to the total extinction is *very* low, compared to other constituents.

The results for the aerosol extinction profiles are shown in Fig. 5.5 and 5.6. In most cases, the relative difference between ORA and SAGE does not exceed 100 percent. Note, however, that the difference in extinction profiles for both instruments fits within the range of the error bars.

5.6 Conclusions

At present, inversion methods based on Bayesian grounds are widespread in a variety of scientific disciplines. We have developed a specific form of the method for our needs, that is based on regularization by the application of a priori information on constituent densities. Extra stability is achieved by an altitude smoothing constraint.

In comparison with the SAGE II experiment, ORA profiles are in fair agreement. By *fair*, we mean: not spectacular, but acceptable. If we keep this in mind, we can start making considerations about scientific priorities. It is clear that ozone and NO_2 density profiles do not have the same accuracy as those from more advanced experiments (having a better apparent vertical resolution). We decided to focus our attention to the aerosol extinction results. The reason is clear: the amount of satellite experiments that were space-borne in the period from August 1992 to May 1993 is very limited. But this period is particularly impor-

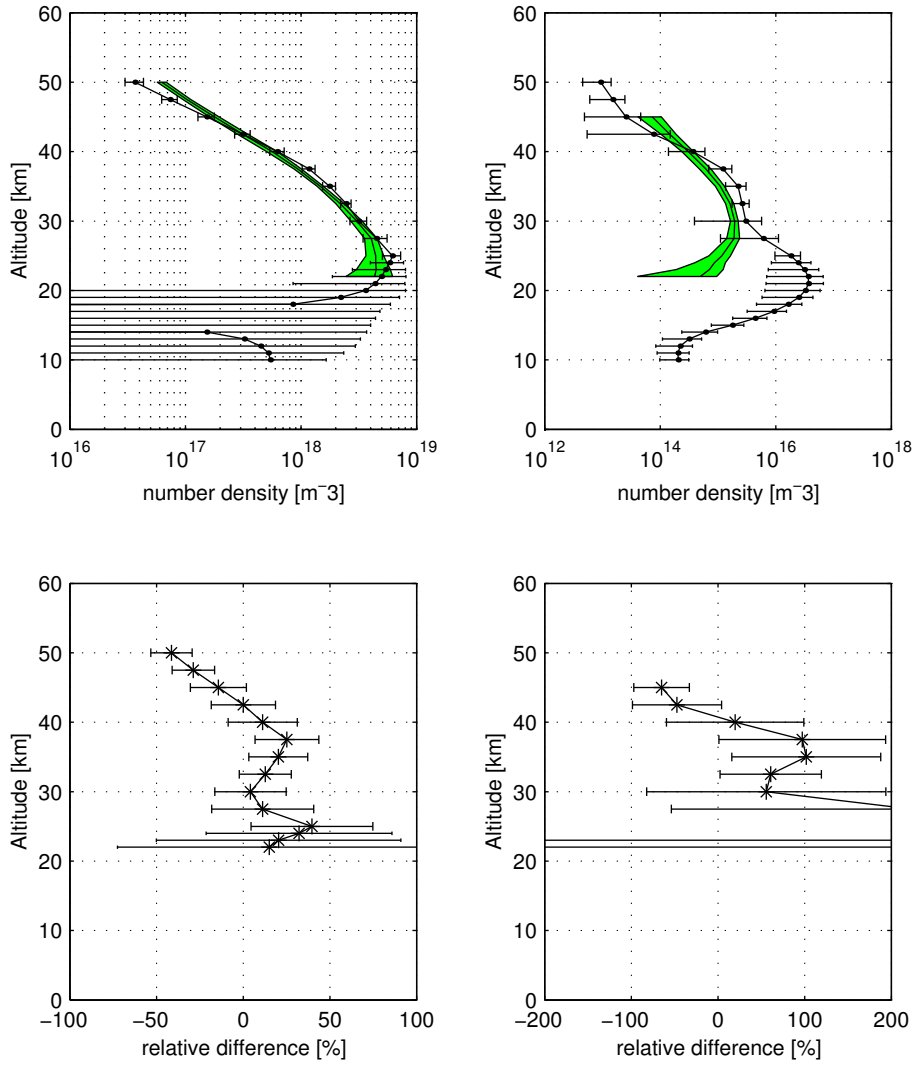

Figure 5.4: Top: Mean number density profiles (25 events) for ozone (left) and NO_2 (right). Shaded regions with solid lines represent SAGE profiles and errors. Solid lines with dots and associated error bars represent ORA data. Bottom: Relative difference between both experiments.

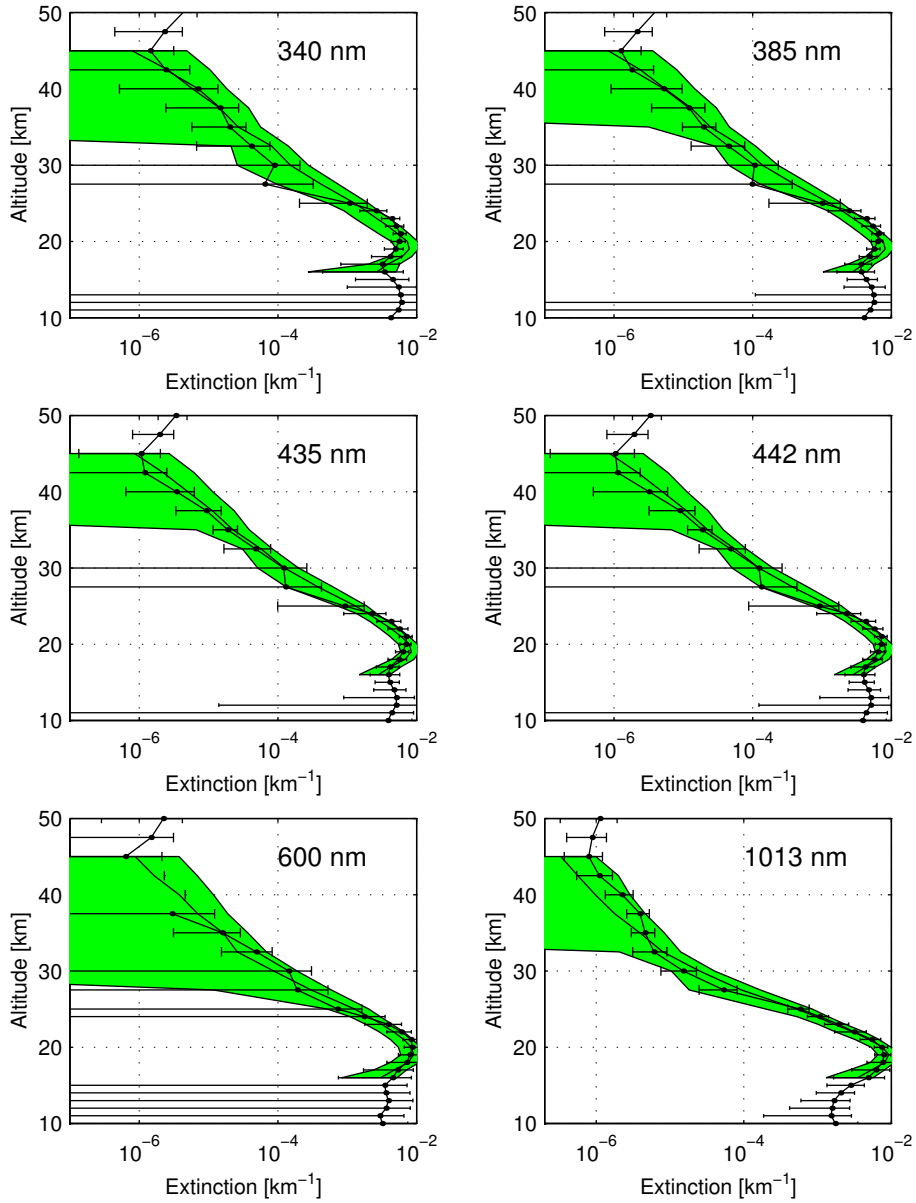

Figure 5.5: Mean aerosol extinction profiles (25 events). Shaded regions and solid lines: SAGE II. Solid lines with dots and error bars: ORA.

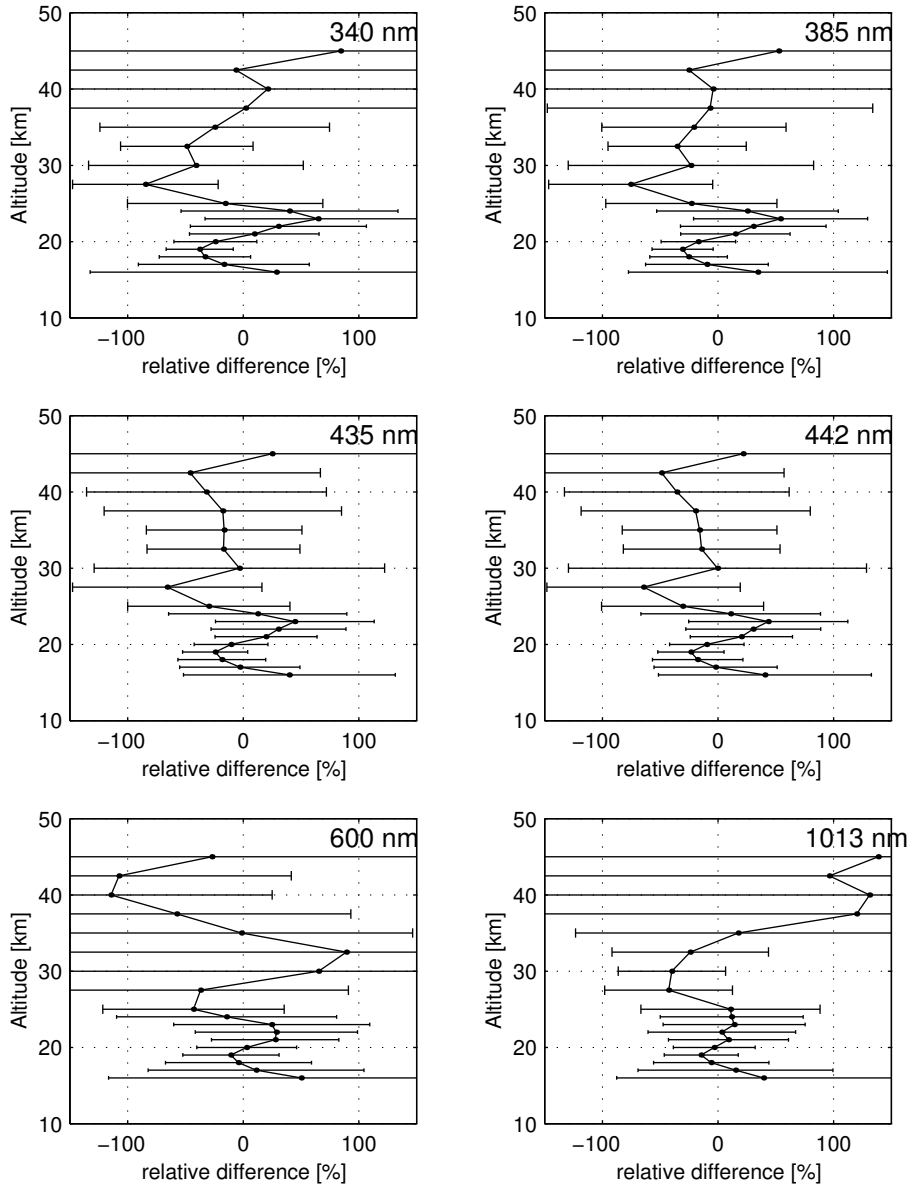

Figure 5.6: Relative differences between ORA and SAGE for mean aerosol extinction profiles (25 events).

tant: the Pinatubo stratospheric aerosols were in full relaxation mode. Investigations of the ORA aerosol results can possibly deliver valuable new information on the behavior of aerosols once they are present in the atmosphere.

Chapter 6

Aerosols: radial inversion and dynamics.

Aerosol extinction coefficients and their spectral dependence result from the interplay between different parameters such as density, particle size, morphology and chemical composition. A model that connects these parameters with extinction is available: the Mie theory (for spherical particles). Once again, a new inverse problem emerges. More specifically, since we want to retrieve information on the *size*, or *radius*, of the aerosol particles, this kind of inverse problem is known as the *radial inversion*.

The results obtained after the inversion can possibly provide an answer to the question: what happens with the aerosols, once they are present in the stratosphere? To investigate these dynamics, the temporal and spatial evolution of the retrieved parameters have to be investigated, and these evolutions have to be connected with theories describing the different dynamic processes that are possibly active.

But before we start with this study, some insight can be gained in the temporal and spatial evolution of aerosols, by investigating the integrated aerosol extinction, or optical thickness.

6.1 Aerosol optical thickness at 1013 nm

6.1.1 Calculation.

The optical thickness offers a quick way to study spatial and temporal variations of the stratospheric aerosols. It is calculated as the upwards integrated aerosol extinction at a certain wavelength, with respect to a reference altitude:

$$\delta = \int_{z_{\text{ref}}}^{+\infty} \beta_{\text{ext}}(z)\,\mathrm{d}z \tag{6.1}$$

The aerosol extinction profiles at 1013 nm are the obvious choice for this calculation. We have seen that the total extinction at this wavelength is mainly influenced by aerosols. Therefore, the aerosol extinction profiles were retrieved with a minimal spectral inversion error.

Which reference altitude should be chosen? Several authors [Trepte et al., 1994, Brogniez and Lenoble, 1987, 1991] take this level to be 2 km above the local tropopause level. This choice is the result of two considerations: (1) it is supposed to reasonably prevent the optical thickness from being affected by possible high-altitude clouds, and (2) the altitude surfaces of equal aerosol extinction roughly 'follow' the global tropopause surface.

It was clear from the start of the data processing that a considerable number of ORA extinction profiles showed abnormally high values at low altitudes. Taking into account that these outliers would distort the final geophysical processing results, we decided to remove them by applying a statistical rejection criterium. After this procedure, 4547 events were left. A first comparison of the 2274 rejected events with monthly mean cloud coverage maps obtained from the ISCCP[1] showed that a high probability of cloud occurrence can be associated with these events. Most of the rejected occultations were located in the equatorial region, where large tower-like *cumulonimbus* clouds can rise to high altitudes, and disturb the extinction measurements.

A further investigation revealed that this rejection criterium was not sufficient. Smaller modulations in the optical thickness and raw transmittance profiles could still be observed, and these are most probably caused by smaller sub-visual clouds that do not produce an abrupt cut-off in the measurements. The reason that ORA measures them, is once

[1] International Satellite Cloud Climatology Project.

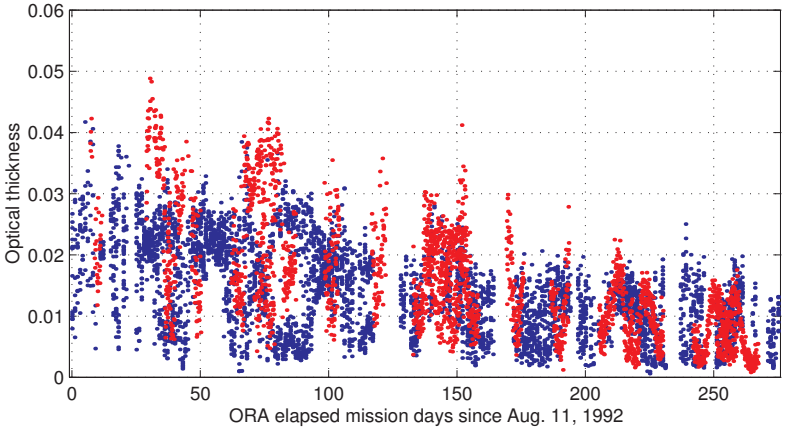

Figure 6.1: Optical thickness at all subsequent occultation events for ORA (blue, 1013 nm) and SAGE II (red, 1020 nm), calculated from $z_T + 6$ km upwards, as function of time. Only the SAGE II events within the [-40°, 40°] latitude band are shown.

again the large field of view. The instrument is still capable of viewing tropospheric clouds two kilometers above the tropopause.

A possible option was to augment the reference altitude. At the end, we chose $z_{ref} = z_T + 6$ km, because the cloud modulation in the optical thickness disappeared at this altitude. For comparison, the ORA optical thickness is shown together with the SAGE II data in Fig. 6.1. Clearly visible are the modulations in time, caused by the latitudinal movement of both satellites. The general agreement is fair, although ORA seems to exhibit a plateau during the first period of the mission. Note however, that such a plateau has also been clearly observed by SAGE II just before this period [Russell et al., 1996]. The plateau is probably not visible in the depicted SAGE II data, due to a low sample density.

6.1.2 Results and interpretation

Using the calculated optical thickness for all 4547 occultation events, we investigated the temporal and latitudinal aerosol variation. We have divided the total ORA latitude range (45°S to 45°N) into 5°bins, with the central bin centered around 0°. The temporal bin size was taken

to be 3 days. For every bin, we calculated the mean optical thickness and the standard deviation. The reason for this binning procedure is the non-uniform sample density in latitude and time, which may lead to unwanted bias.

The temporal variations were fitted by a simple exponential form, to describe the aerosol relaxation:

$$\delta(\phi, d) = A(\phi) \exp(-d/d_0(\phi)) \tag{6.2}$$

where d equals the elapsed days since the start of the ORA mission (Aug. 12, 1992), d_0 a relaxation time parameter and ϕ the latitude. To ensure that A and d_0 have a reasonable smoothness as function of latitude, we used two independent Legendre expansions over the latitude range $\phi =$ (90°S, 90°N):

$$A(\phi) = \sum_{i=0}^{n_1} a_i P_i(\phi) \tag{6.3}$$

$$\frac{1}{d_0} = \sum_{i=0}^{n_2} b_i P_i(\phi) \tag{6.4}$$

These two equations, together with the logarithm of Eqn. 6.2, lead to a linear least squares problem. Investigation of the solution covariance matrix (calculated with $n_1 = n_2 = 10$) indicates that A and d_0 remain quite correlated due to the restricted temporal range. Instead, we present the fitted optical thickness and the time derivative of the optical thickness, evaluated at $d = 100$ and $d = 200$ days, in Fig. 6.2.

On Fig. 6.3, we can observe the optical thickness and its time decay for both ORA and SAGE II as function of latitude. The rather large differences at the beginning of the time range have to be interpreted with caution, because of the low amount of SAGE II measurements. A broad maximum can be seen at all times between 20°N and 20°S, conforming with other measurements and climatologies [Trepte et al., 1994, Long and Stowe, 1994, Hitchman et al., 1994, Mergenthaler et al., 1995]. It is caused by the existence of a tropical aerosol reservoir, maintained by the mean upwards movement of the Brewer-Dobson circulation, periodically enhanced by the Quasi-Biennial Oscillation (QBO) during the easterly shear phase [Trepte et al., 1993, Grant et al., 1996, Choi et al., 1998].

There is a distinct zonal asymmetry at the beginning of the ORA mission, with a peak value at 20°N. This structure was also measured

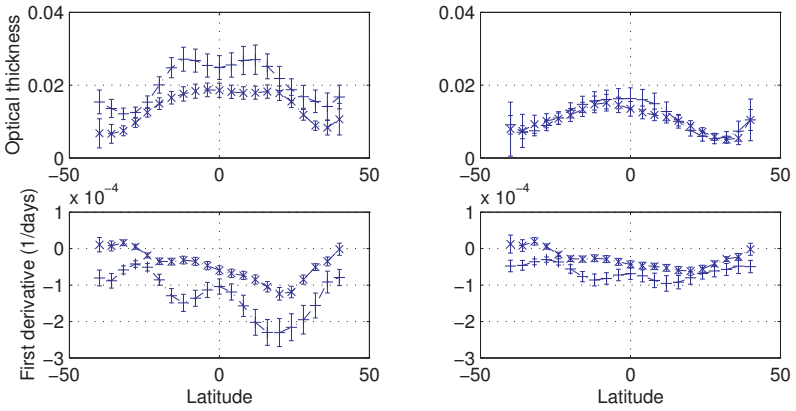

Figure 6.2: Fitted aerosol optical thickness (top) and time derivative (bottom) as function of latitude, evaluated at $d = 100$ days (left, Nov. 19, 1992), and $d = 200$ days (right, Feb. 27, 1993), for ORA (1013 nm, x-signs) and SAGE II (1020 nm, +-signs).

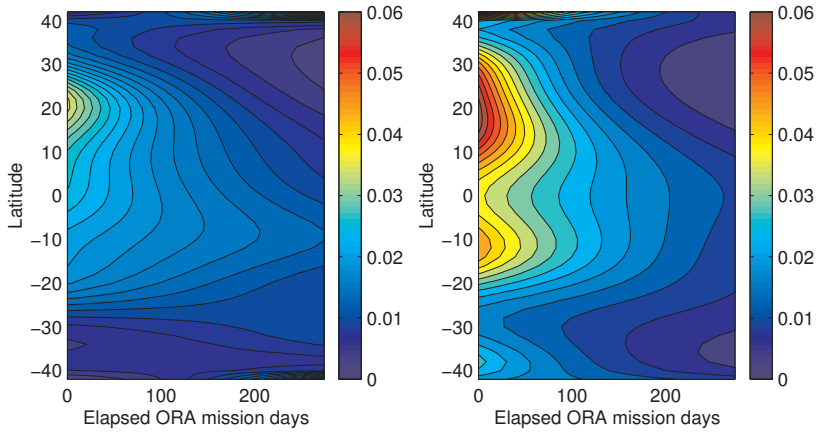

Figure 6.3: Fitted aerosol optical thickness as function of time and latitude. Left: ORA (1013 nm). Right: SAGE II (1020 nm).

by SAGE II, CLAES [Roche et al., 1993, Mergenthaler et al., 1995] and AVHRR [Long and Stowe, 1994]. It is likely to be a seasonal effect caused by enhanced lifting from the tropical reservoir on the summer hemisphere side, or removal of aerosol-rich air and injection of aerosol-poor air from the winter hemisphere extratropics [Hitchman et al., 1994]. The asymmetry reverses towards the end of the mission, resulting in the apparently faster decay at 20°N, as can be seen on the minima in the first time derivative in Fig. 6.2.

6.1.3 Conclusion

This first study of the aerosol 1013 nm optical thickness has given us some insight into *transport* processes. With some relief, we can conclude that the features caused by aerosol dispersion around the globe are similar for both ORA and SAGE II.

But the optical thickness is an integrated quantity. It does not give insight into transport processes that occur in the vertical direction. Furthermore, it is calculated from extinction coefficients of which the values are determined by quantities that are microphysical in nature: the number density, the size, the shape and the composition of the aerosols. Some of these quantities change in time due to a number of different physical processes. Now is a good time to start a discussion on these matters.

6.2 Stratospheric aerosols: microphysical properties

Over the years following the Mount Pinatubo eruption, a wide diversity of measurements have provided a valuable amount of information regarding the characteristics of individual aerosol particles. Some measurements were taken *in situ* (balloons, high-altitude aircraft), other quantities were derived from ground-based and satellite experiments. The quantities of importance, from our point view, are summarized here.

6.2.1 Chemical composition

Investigation of *in situ* captured aerosol samples [Pueschel et al., 1992, 1994, Sheridan et al., 1992], evaporation temperature measurements

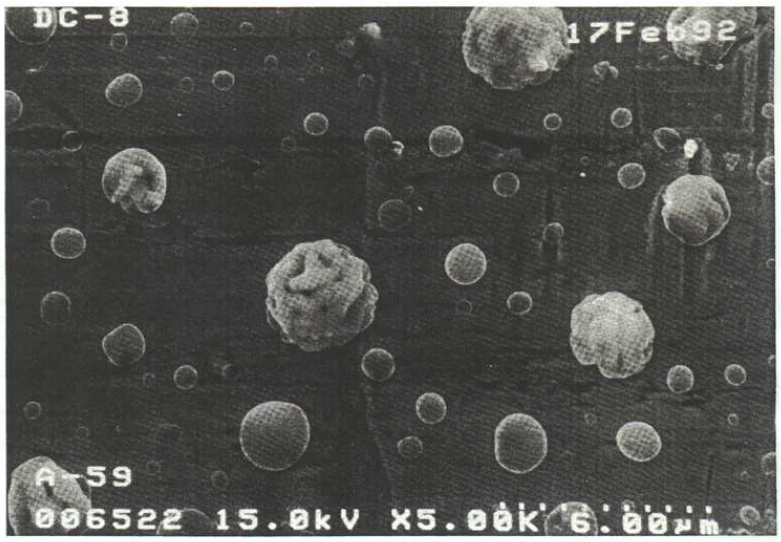

Figure 3. Plane view of stratospheric H_2SO_4/H_2O droplets and sulfuric-acid-coated volcanic ash particles collected on February 17, 1992, near (53°N; 67°W) at 10.7 km msl.

Figure 6.4: Stratospheric H_2SO_4/H_2O droplets (small, spherical) and sulfuric acid coated volcanic ash particles (large), collected on Feb. 17, 1992 near (53°N, 67°W) at an altitude of 10.7 km. The sample was collected by an impactor experiment on a DC-8 airplane [Pueschel et al., 1994].

[Deshler et al., 1992, 1993] and satellite infrared spectroscopy measurements by ISAMS [Grainger et al., 1993] show clearly that the aerosol droplets predominantly consisted of a sulfuric acid and water mixture. More specifically, the mass fraction of the H_2SO_4 component in the mixture was found to be in a range from 65 to 80 %, consistent with theoretical calculations for typical stratospheric temperatures (-50°C) and water vapor mixing ratios of 3 to 5 ppmv [Steele and Hamill, 1981, Russell and Hamill, 1984].

Besides sulfuric acid aerosols, solid particles were also found by impactor experiments [Pueschel et al., 1994], as can be seen on Fig. 6.4. They were identified as volcanic ash of mineral composition. Because these particles are larger and heavier than the droplets, they were for the most part removed from the stratosphere by sedimentation at the time ORA started its mission.

6.2.2 Refractive index

Within the ORA wavelength region, the refractive index imaginary component of H_2SO_4-H_2O mixtures is less than 10^{-6} [Palmer and Williams, 1975]. Absorption of light is therefore *negligible* in our extinction calculations. The real part of the refractive index is dependent on temperature and the mass fraction of H_2SO_4 with respect to H_2O. Values for different situations are available in literature [Russell et al., 1996]. However, since the mass fraction lies in the range from 65 to 80 %, an average value is needed. Many stratospheric aerosol studies therefore assume $N = 1.43 + \mathrm{i}\ 0.00$ (at all wavelengths), and this is also the value that we used.

6.2.3 Shape

Since we are dealing with *droplets* of sulfuric acid solutions, we can assume that they are *spherical*. This most conveniently allows us to use the optical Mie extinction theory. For non-spherical particles, the calculations would be considerably more difficult.

6.3 The particle size distribution

6.3.1 General features

A particle size distribution (*psd*) mathematically expresses the number density of particles with sizes in the range from r to $r + \mathrm{d}r$, and is formally notated as:

$$\frac{\mathrm{d}\mathcal{N}(r)}{\mathrm{d}r} \tag{6.5}$$

After integration for all sizes, we obtain the total number density:

$$\int_0^\infty \frac{\mathrm{d}\mathcal{N}(r)}{\mathrm{d}r}\,\mathrm{d}r = \mathcal{N} \tag{6.6}$$

It is convenient to normalize the *psd*. We then have:

$$\frac{\mathrm{d}\mathcal{N}(r)}{\mathrm{d}r} = \mathcal{N} f(r) \qquad \int_0^\infty f(r)\,\mathrm{d}r = 1 \tag{6.7}$$

From now on, when the particle size distribution is mentioned in this text, we will refer to the normalized function $f(r)$.

Usually, distributions are characterized by quantities such as the mean, variance, skewness, etc. These are easily calculated from the different *moments* of the *psd*. The nth moment of the distribution is defined as:

$$M_f^{(n)} = \int_0^\infty r^n f(r)\,\mathrm{d}r \tag{6.8}$$

6.3.2 The lognormal distribution

Regardless of the specific type of aerosols, one thing is almost certain: particle sizes within one population can cover huge ranges (from a few nanometers to micrometers). This is certainly true for stratospheric aerosols, and the cause can be found in aerosol formation processes. Initially, a particle is formed by the condensation of gases onto a small nucleus. Through a diversity of mechanisms, they can grow to several orders of magnitudes their initial size, before they are removed from the atmosphere.

In the past, several *psd* forms have been proposed for aerosols, such as the *Power-Law* distribution or the *Modified Gamma* Distribution [Seinfeld and Pandis, 1998]. We will use the well-known *lognormal* distribution for three simple reasons: (1) it is able to cover a large size range, (2) it in fact matches *in situ* measured stratospheric aerosols quite well, and (3) lognormals are very flexible in algebraic manipulations.

The huge size domain of the lognormal distribution is obtained by constructing a normal distribution on a logarithmic scale:

$$f(\ln r) = \frac{1}{\sqrt{2\pi}\ln(s)} \exp\left\{-\frac{1}{2}\left(\frac{\ln(r) - \overline{\ln(r)}}{\ln(s)}\right)^2\right\} \tag{6.9}$$

The particle number density in an infinitesimal interval should remain the same, whether the *psd* is expressed on a linear or logarithmic scale:

$$f(r)\,\mathrm{d}r = f(\ln r)\,\mathrm{d}\ln r \tag{6.10}$$

If we also apply the substitution $\ln(r_\mathrm{m}) = \overline{\ln(r)}$, we derive the *lognormal distribution*:

$$f(r) = \frac{1}{\sqrt{2\pi}\ln(s)\,r} \exp\left\{-\frac{1}{2}\left(\frac{\ln^2(r/r_\mathrm{m})}{\ln^2(s)}\right)\right\} \tag{6.11}$$

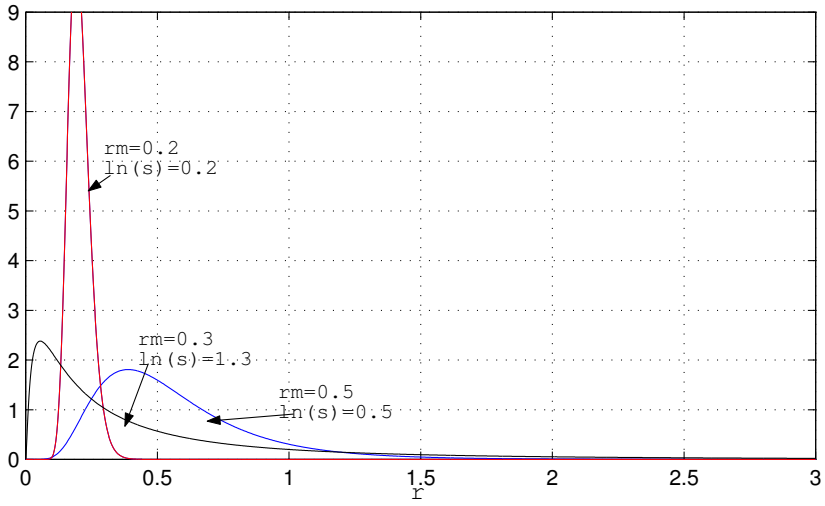

Figure 6.5: The log-normal size distribution for a few different values of (r_m, s).

A few examples have been given in Fig. 6.5, showing the typical form of the function.

The meaning of r_m is exposed by the simple derivation:

$$\int_0^{r_\mathrm{m}} f(r)\,\mathrm{d}r = \frac{1}{2} \tag{6.12}$$

and we see that r_m represents the *median*[2] of the particle size distribution. In a similar way, one can derive that the size range $[r_\mathrm{m}/s, r_\mathrm{m}\,s]$ contains 68 % of the particles in the size distribution. s is known as the *geometric standard deviation*. To simplify the notation, we will use:

$$\sigma = \ln(s) \tag{6.13}$$

Eqn. 6.11 represents a so-called *unimodal* log-normal distribution, because the function is able to exhibit one maximum. Some realistic aerosol size distributions may have multiple maxima. They can be mod-

[2]Exactly 50 % of the particles have a size that is smaller than r_m. The other half of the population is larger in size.

elled by a sum of log-normals:

$$f(r) = \sum_{i=1}^{r} \frac{1}{\sqrt{2\pi}\sigma_i\, r} \exp\left\{ -\frac{1}{2} \left(\frac{\ln^2(r/r_{\mathrm{m},i})}{\sigma_i^2} \right)^2 \right\} \tag{6.14}$$

6.3.3 Moments of the log-normal distribution

The nth moment (Eqn. 6.8) of the log-normal is easily calculated as:

$$M_f^{(n)} = r_{\mathrm{m}}^n \mathrm{e}^{\frac{n^2\sigma^2}{2}} \tag{6.15}$$

In practice, only a few moments are frequently used:

$$\begin{array}{ll} M_f^{(0)} = 1 & M_f^{(1)} = r_{\mathrm{m}} \mathrm{e}^{\frac{1}{2}\sigma^2} \\ M_f^{(2)} = r_{\mathrm{m}}^2 \mathrm{e}^{2\sigma^2} & M_f^{(3)} = r_{\mathrm{m}}^3 \mathrm{e}^{\frac{9}{2}\sigma^2} \end{array} \tag{6.16}$$

6.3.4 Some useful quantities

Several quantities are defined to describe aerosol characteristics, depending on the specific application.

Surface area density. The total surface area of the (spherical) aerosol particles per volume can be calculated by integration of the individual particle surface areas within the distribution, and we obtain the *surface area density* (dimension: area per volume):

$$S = \mathcal{N} \int_0^\infty 4\pi r^2 f(r)\, \mathrm{d}r = \mathcal{N} 4\pi M_f^{(2)} = 4\pi \mathcal{N} r_{\mathrm{m}}^2 \mathrm{e}^{2\sigma^2} \tag{6.17}$$

The interpretation is as follows. Consider a unit volume with $\mathcal{N}$ particles that are distributed according to the log-normal function. The total particle surface area inside the volume then simply equals $\mathcal{N}$ times the surface area of a spherical particle with an equivalent radius $R_{\mathrm{s}} = r_{\mathrm{m}} \mathrm{e}^{\sigma^2}$.

The surface area density can be derived from optical extinction measurements in a straightforward way. Furthermore, in the study of heterogeneous chemistry on aerosol surfaces it becomes an important factor.

Volume density. The total volume of the aerosols per volume (of air that contains the aerosols!) equals:

$$V = \mathcal{N} \int_0^\infty \frac{4}{3}\pi r^3 f(r)\,\mathrm{d}r = \mathcal{N}\frac{4\pi}{3}M_f^{(3)} = \frac{4\pi}{3}\mathcal{N}r_\mathrm{m}^3 \mathrm{e}^{\frac{9}{2}\sigma^2} \tag{6.18}$$

The total particulate volume thus equals the volume of $\mathcal{N}$ particles with equivalent radius $R_\mathrm{v} = r_\mathrm{m}\mathrm{e}^{\frac{3}{2}\sigma^2}$

Effective radius. The radius of a sphere with surface area s and volume v equals $r = 3\frac{v}{s}$. In a similar way, for a *psd* one can define an *effective radius* from the surface area density and the volume density:

$$R_\mathrm{eff} = 3\frac{V}{S} = \frac{M_f^{(3)}}{M_f^{(2)}} \tag{6.19}$$

When different size distributions (that are not necessarily log-normal) must be compared, we need a descriptor that is somewhat independent of the functional form of the distribution. The effective radius meets this requirement. For log-normal distributions, it has an analytical form:

$$R_\mathrm{eff} = r_\mathrm{m}\mathrm{e}^{\frac{5}{2}\sigma^2} \tag{6.20}$$

6.4 Optical extinction by a population of particles

6.4.1 Formulation

In Chapter 3, we discussed the extinction of light by *one* particle. In the case of a spherical particle, the extinction cross section C_ext is successfully given by the exact Mie theory. But in realistic situations, one always has to deal with extinction by a large number of particles, that are not equal in size. In such cases, Eqn. 3.25 has to be applied:

$$\beta_\mathrm{ext}(\lambda) = \int_0^\infty C_\mathrm{ext}(x = \mathrm{k}r)\frac{\mathrm{d}\mathcal{N}(r)}{\mathrm{d}r}\,\mathrm{d}r \tag{6.21}$$

where C_ext is once again written as function of the size parameter $x = \mathrm{k}r$.

Writing this equation in terms of the cross section efficiency, $Q_\mathrm{ext} = \frac{C_\mathrm{ext}}{\pi r^2}$, and using Eqn. 6.7, we obtain:

$$\beta_{\text{ext}}(\lambda) = \mathcal{N} \int_0^\infty \pi r^2 Q_{\text{ext}}(x = \text{k}r) f(r)\, \text{d}r \tag{6.22}$$

By inserting the log-normal distribution for $f(r)$ and applying the substitution $t = \frac{1}{\sqrt{2}\sigma} \ln(r/r_m)$, we derive after some manipulation:

$$\beta_{\text{ext}}(\lambda) = \mathcal{N}\pi \left[r_{\text{m}} e^{\sigma^2}\right]^2 \left[\frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} Q_{\text{ext}}(x = \text{k}r_m e^{\sqrt{2}\sigma t}) e^{-(t-\sqrt{2}\sigma)^2} \text{d}t\right] \tag{6.23}$$

Identifying the first term between brackets as the equivalent radius R_{s}, and the second one as an *equivalent* cross section efficiency $\tilde{Q}_{\text{ext}}$, we get:

$$\beta_{\text{ext}}(\lambda) = \mathcal{N}\pi R_{\text{s}}^2 \tilde{Q}_{\text{ext}}(r_{\text{m}}, \sigma, \text{k}) \tag{6.24}$$

6.4.2 The large particle limit

In the case of one particle, we know that the large particle limit yields:

$$\lim_{x \to +\infty} Q_{\text{ext}}(x) = 2 \tag{6.25}$$

In this limit, the equivalent cross section efficiency transforms to:

$$\begin{aligned} \tilde{Q}_{\text{ext}}(r_{\text{m}}, \sigma, \text{k}) &= \frac{2}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-(t-\sqrt{2}\sigma)^2} \text{d}t && (6.26)\\ &= 2\ \text{erf}(+\infty) && (6.27)\\ &= 2 && (6.28) \end{aligned}$$

and, together with Eqn. 6.24, this leads to:

$$\beta_{\text{ext}} = \mathcal{N}\, 2\pi R_{\text{s}}^2 = \frac{S}{2} \tag{6.29}$$

Extinction for large-particle distributions is therefore independent of wavelength. The conclusions, drawn in Chapter 3 thus still apply: the *flatter* the wavelength curve, the larger the particles.

The consequences are quite important: it is clear from Eqn. 6.29 that a measurement of $\beta_{\text{ext}}(\lambda)$ cannot provide information on $\mathcal{N}$, r_{m} and σ *separately*. This is probably the reason why the retrieval of the surface area density is so popular in aerosol research; it is an integrated quantity, combining the three unknown parameters into one unknown S.

6.4.3 The small particle limit

In a very similar way, one can derive the scattering behaviour for a population of very small particles, using Eqn. 3.41:

$$\beta_{\rm sca}(\lambda) = \mathcal{N}\pi(r_{\rm m}{\rm e}^{3\sigma^2})^2\frac{8}{3}\left|\frac{m^2-1}{m^2+2}\right|^2 {\rm k}^4(r_{\rm m}{\rm e}^{3\sigma^2})^4 \tag{6.30}$$

or, using $R_{\rm a} = r_{\rm m}{\rm e}^{3\sigma^2}$ and $x_{\rm a} = kR_{\rm a}$:

$$\beta_{\rm sca}(\lambda) = \mathcal{N}\pi R_{\rm a}^2\frac{8}{3}\left|\frac{m^2-1}{m^2+2}\right|^2 x_{\rm a}^4 \tag{6.31}$$

Now, the particle population scatters light as if the population consists of identical particles with a size $R_{\rm a}$. We observe a λ^{-4} wavelength dependence, as was expected. But again, the important issue here is: it is impossible to derive the separate parameters $\mathcal{N}$, $r_{\rm m}$ and σ.

6.4.4 Important conclusion

There is obviously a limit to the amount of information about microphysical aerosol properties present in optical extinction measurements. If we want to obtain data on $\mathcal{N}$, $r_{\rm m}$ and σ, we will be in trouble when very large or small particles are present.

For all intermediate sizes, the case is different. Here we see that $\tilde{Q}_{\rm ext}(r_{\rm m}, \sigma, {\rm k})$ has a wavelength dependence, determined by the parameters $r_{\rm m}$ and σ. It is within this region of particle sizes that we should be able to retrieve information. The best contrast is obtained when the mean particle size in the distribution roughly matches the considered wavelength (see also [Echle et al., 1998]). A few examples for $\tilde{Q}_{\rm ext}(r_{\rm m}, \sigma, {\rm k})$ are shown in Fig. 6.6.

6.5 Measured aerosol extinction

At this point already, we can obtain some qualitative information on the size of the aerosol particles, by simply looking at the wavelength dependency of the measured aerosol extinctions. On Fig. 6.7, extinctions are shown versus wavelength, for four different altitudes. In the lower

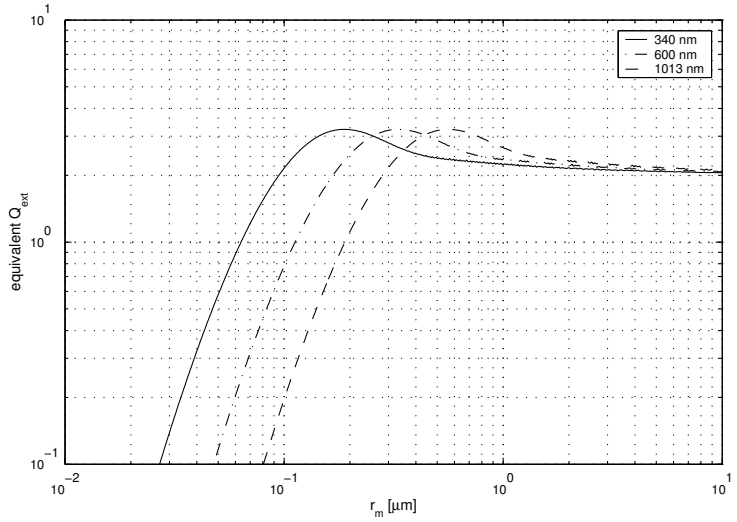

Figure 6.6: The equivalent cross section efficiency $\tilde{Q}_{\rm ext}(r_{\rm m}, \sigma, (k))$ for $\sigma = 0.4$ at 3 different wavelengths.

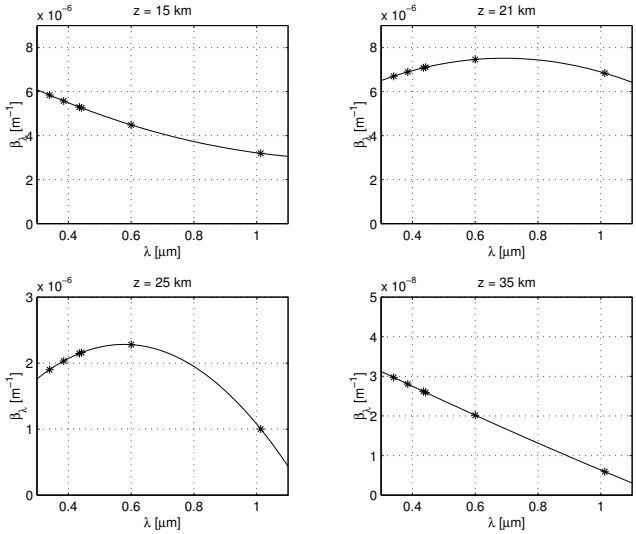

Figure 6.7: A typical example of measured aerosol extinctions, at 4 altitudes. Asterisks represent the actual values at the six ORA wavelengths, while the solid line is the quadratic form that was used in the spectral inversion (Eqn. 5.7).

stratosphere ($z = 15$ km, $z = 21$ km), the curves are rather flat, indicating relatively large particle sizes. At 21 km, the extinction values are also large, suggesting a peak in the aerosol number density. We know from other experiments that this is the altitude region where the peak of the Junge layer is expected [Turco et al., 1982].

At higher altitudes, the extinction curves become steeper, and extinction values are much smaller. The conclusion is now: small particles, with a low number density.

6.6 The inversion algorithm

6.6.1 A first reduction

Since we have aerosol extinctions for 6 wavelengths at our disposal, and three unknowns, the problem is mathematically well-posed. However, because of measurement error, combined with the smooth nature of the model, our inverse problem is highly ill-posed, and (once again) it is necessary to apply some regularization. Several radial inversion methods have been used in the past [Anderson et al., 2000, King et al., 1978, McCormick and Lovill, 1982, Steele and Turco, 1997, Twomey, 1975], but we have developed a specific technique for our purpose.

But first, a very convenient simplification is possible. Inspecting Eqn. 6.24, we see that we can get rid of the unknown number density through a normalization with respect to the aerosol extinction at 1013 nm:

$$\frac{\beta_{\text{ext}}(\lambda)}{\beta_{\text{ext}}(\lambda_{1013})} = \frac{\tilde{Q}_{\text{ext}}(r_{\text{m}}, \sigma, \lambda)}{\tilde{Q}_{\text{ext}}(r_{\text{m}}, \sigma, \lambda_{1013})} \tag{6.32}$$

and we will use the short notation:

$$\beta_{\text{N},\lambda} = \tilde{Q}_{\text{N},\lambda}(r_{\text{m}}, \sigma) \tag{6.33}$$

The reduced problem now consists of 6 measurements (of which 5 are variable and one equals unity) and only 2 unknowns. This new formulation has one very important advantage. For only two unknowns, it is possible to simulate the entire *range* of the inverse problem.

The entire state space, spanned by the two state components r_{m} and σ, can be represented by a two-dimensional fine grid $(r_{\text{m,i}}, \sigma_i)$. For every state, we can now calculate the forward model (Eqn. 6.33), and this has

to be done six times (for each wavelength). The results obtained are then stored on disk.

The big advantage of this method is that most of the calculations have already been done *before* the actual inversion of the measurements starts. No (rather expensive) Mie calculations are necessary *during* the iterations of the inversion algorithm.

Additionally, the topology of the merit function can be investigated *visually*, since it is only dependent on two parameters. This can provide useful information on the pitfalls of the inverse problem, and can lead us to an effective regularization method.

6.6.2 Merit function topology and regularization

The merit function, associated with the inverse problem reads:

$$M(r_{\mathrm{m}}, \sigma) = \left\langle \beta_{\mathrm{M}} - \beta_{\mathrm{N}} | {\mathbf{S}_{\beta_{\mathrm{M}}}}^{-1} | \beta_{\mathrm{M}} - \beta_{\mathrm{N}} \right\rangle \tag{6.34}$$

with β_{M} the (6×1) vector of measured and normalized extinctions, β_{N} the (6×1) vector containing the forward model (Eqn. 6.33), and $\mathbf{S}_{\beta_{\mathrm{M}}}$ the measurement covariance matrix.

A typical example of this merit function is given in Fig. 6.8 for three adjacent altitudes. At $z = 17$ km, we observe a clear minimum, located in a narrow valley. At $z = 18$ km, the situation is worse: a large domain of the state space can be considered as the minimum, and it is far from clear what the actual solution for (r_{m}, σ) should be. At $z = 19$ km, however, the minimum region is again better defined.

It is clear that the solution at $z = 18$ km needs constraining. One way to achieve this is by expressing the unknowns r_{m} and σ as smooth functions of altitude:

$$r_{\mathrm{m}}(z) = r_{\min} + (r_{\max} - r_{\min}) \exp\left\{ - \left(\sum_{i=0}^{6} a_i z^i \right)^2 \right\} \tag{6.35}$$

$$\sigma(z) = \sigma_{\min} + (\sigma_{\max} - \sigma_{\min}) \exp\left\{ - \left(\sum_{i=0}^{6} b_i z^i \right)^2 \right\} \tag{6.36}$$

where $r_{\min}$, $r_{\max}$, $\sigma_{\min}$ and $\sigma_{\max}$ represent the boundaries of the (rectangular) region in the state space where we have calculated the forward model $\beta_{\mathbf{N}}$. We chose $[r_{\min}, r_{\max}] = [0.01, 1]$ μm and $[\sigma_{\min}, \sigma_{\max}] =$

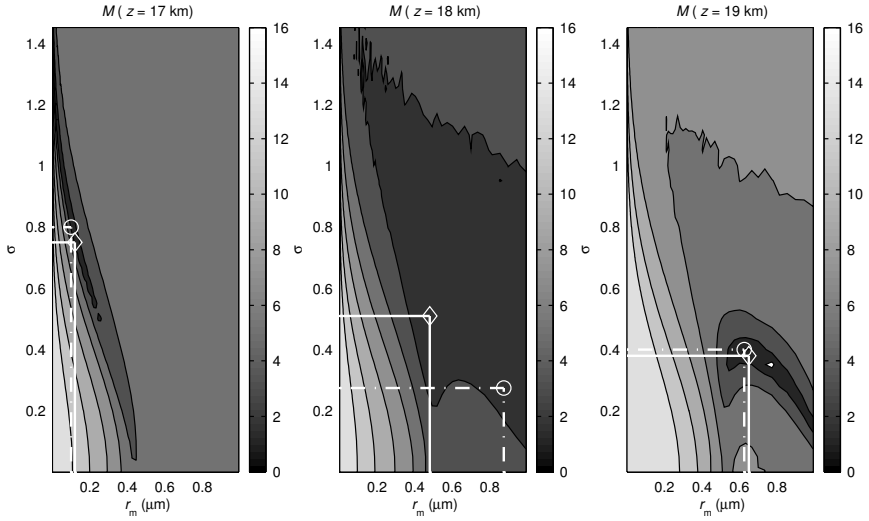

Figure 6.8: The merit function M at three adjacent altitudes may show up quite different topologies that can induce brutal variations on the retrieved parameters. The circles represent the absolute minima and the diamonds are the regularized minima.

$[0.01, 1.5]$, since this is the region where the merit function minima are located.

The functional forms of the $r_{\rm m}$ and σ altitude profiles are characterized by 7 coefficients. This choice was the result of trial and error: 7 coefficients ensure that inversion residuals are acceptable while at the same time the solutions are fairly smooth.

Once again, we have come to a regularization method that consists of altitude smoothing. This of course implies that the radial inversion of one ORA event is comprehensive, i.e. all altitudes at once. Minimization of the merit function is not performed with respect to the $2 \times 26 = 52$ unknowns $r_{\rm m}$ and σ, but to the 14 coefficients (a_i, b_i).

6.6.3 The inversion calculation

The entire inversion process is summarized by the flowchart on Fig. 6.9. First, the entire range $\beta_{\rm N}$ of the inverse problem is calculated on a fine rectangular grid $(r_{\rm m}, \sigma)$, for the six ORA wavelengths, using a Mie

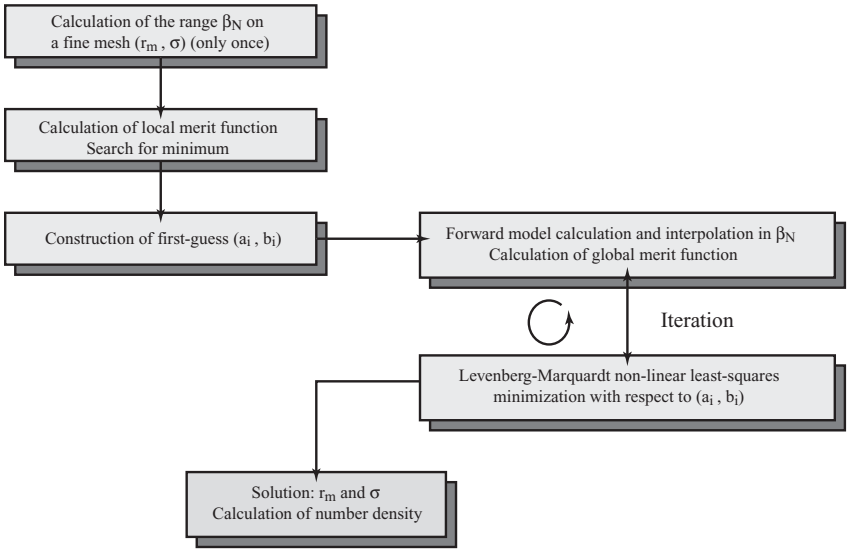

Figure 6.9: Flowchart for the radial inversion

extinction algorithm. This has to be done only once. Then, the actual measurement inversion can start.

Since we are dealing with an iterative process, a first guess is needed. Locally, for every altitude separately, we calculate the merit function, and find the *absolute* minimum by a simple search method. Once finished, we have 26 first-guess r_m and σ coefficients. By the use of a linear least-squares fit on Eqns. 6.35 and 6.36, we then derive the first-guess a_i and b_i coefficients.

Now the iterative procedure can start. First, r_m and σ are calculated from the a_i and b_i coefficients. Using these values, the global merit function is found in the pre-calculated grid by a linear interpolation scheme. It is this merit function that is evaluated by the Levenberg-Marquardt algorithm, which subsequently adapts the a_i and b_i coefficients.

After convergence, we have a solution for r_m and σ as a function of z. The errors are estimated from the *Jacobian* matrix [Press et al., 1992] of the merit function at the final iteration. Using Eqn. 6.24, we also can calculate the number density $\mathcal{N}$, and associated error. And this concludes the inversion.

6.6.4 Results

As already discussed, the quality of the retrieved parameters is strongly determined by the size of the particles. For small particles, it becomes increasingly difficult to infer information on the parameters separately. In the neighborhood of the Junge layer, the structures that can be observed in the r_{m}, σ and $\mathcal{N}$ profiles are quite acceptable, but at larger altitudes (above 35 km) the profiles become increasingly chaotic. However, when we take averages over a certain ensemble of occultation events, clear structures in the profiles begin to show themselves. This is the reason that we will present averaged profiles from now on.

On Fig. 6.10, we present the obtained altitude profiles that were averaged over the entire mission period. The associated uncertainty contains individual profile error and ensemble variability. As can be observed, the size parameter r_{m} reaches a maximum of about 0.57 μm at 17 km, and this result agrees very well with those of other experiments that performed measurements in this time period (see for instance [Russell et al., 1996]). A large plateau in the r_{m} profile is clearly visible between 25 and 35 km above which r_{m} monotonically decreases.

There is no clear correspondence between r_{m} and σ altitude profiles although a broad maximum at 21 km in the latter could suggest the coexistence of two aerosol formation mechanisms inducing a larger spread in the retrieved distribution.

The maximum (Junge layer) in the particle number density culminates at about 23 km with a value of about 4 particles per cm^3, a value that is again very realistic in comparison with other measurements[3]. This maximum seems somehow to be correlated with a slope change in the r_{m} profile.

When comparing aerosol distributions as measured by different experiments, the effective radius R_{eff} (Eqn. 6.20) is the preferred form, since it is a quantity that is somewhat independent of the assumed size distribution form. In Fig. 6.11, the effective radius derived by [Anderson and Saxena, 1996] from SAGE II measurements is plotted, and it agrees reasonably with our data in a common latitude band from 30°N to 40°N. The SAGE II R_{eff} profile (averaged over the year 1993) exhibits a clear maximum of 0.54 μm at about 16 km, slightly shifted with

[3]Measured aerosol number density values are very scarcely found in literature, however, probably due to the large difficulties that are involved in radial inversions.

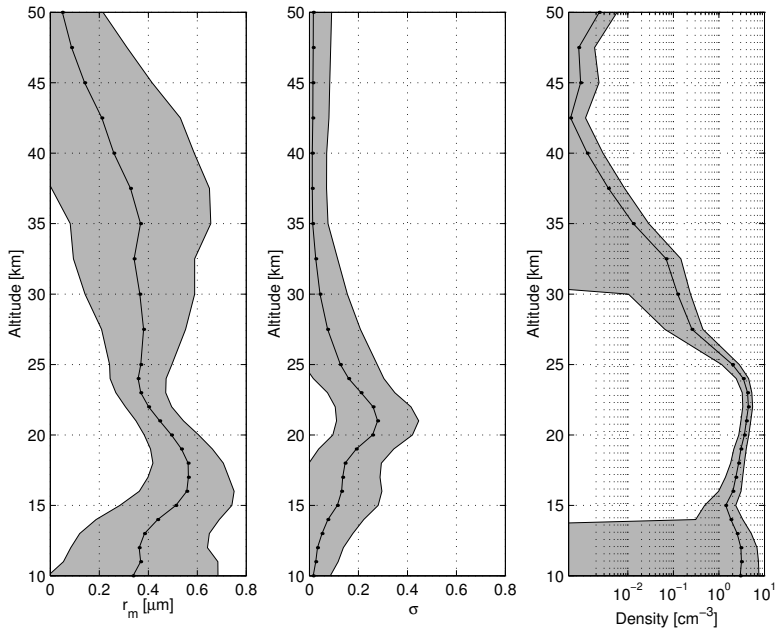

Figure 6.10: Altitude profiles for r_{m}, σ and $\mathcal{N}$, averaged over the entire ORA mission.

respect to the ORA data (averaged over the first four months of 1993) and decreases slowly with increasing altitudes. Again, a broad shoulder (which sometimes becomes a secondary maximum) appears to be present above 23 km and this seems to have been unnoticed by [Anderson and Saxena, 1996] although a close inspection of their Figure 1 reveals the same phenomenon in 1993 and 1994.

Figure 6.11 also shows the effective radius obtained by [Brogniez et al., 1996] from a balloon-borne polarimeter experiment at Gap (44.6°N, 6.1°E, France) in June 1992. Their retrieved values at z=18 and 20 km have been obtained by a similar inversion procedure, assuming a log-normal distribution, and the values seem to be consistent with our averaged data.

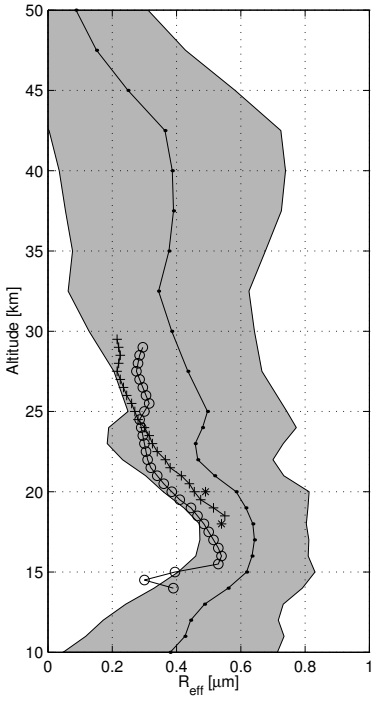

Figure 6.11: Comparison for the effective radius R_{eff} profiles, measured by ORA in the 30°- 40°latitude band (dots) with results obtained by Anderson and Saxena for 1992 (crosses) and 1993 (circles). The results obtained by Brogniez et al. [1996] are also reported (asterisks).

6.7 Temporal evolution of measured r_{m} and $\mathcal{N}$

Fig. 6.12 shows the temporal evolution of the radius parameter r_{m} and aerosol number density $\mathcal{N}$ as function of time for 6 different altitudes. The data have been averaged in monthly bins. At low altitudes, one can clearly observe a positive time gradient for r_{m}. Simultaneously, the number density decreases at lower altitudes. One intuitively senses that this behaviour can be explained by an ongoing *coagulation* process (see further), the merging of several small particles into fewer large particles.

At high altitudes, the situation reverses, with particles becoming smaller while the number density increases. A reverse coagulation process is very unlikely; there is no reason why particles would break up

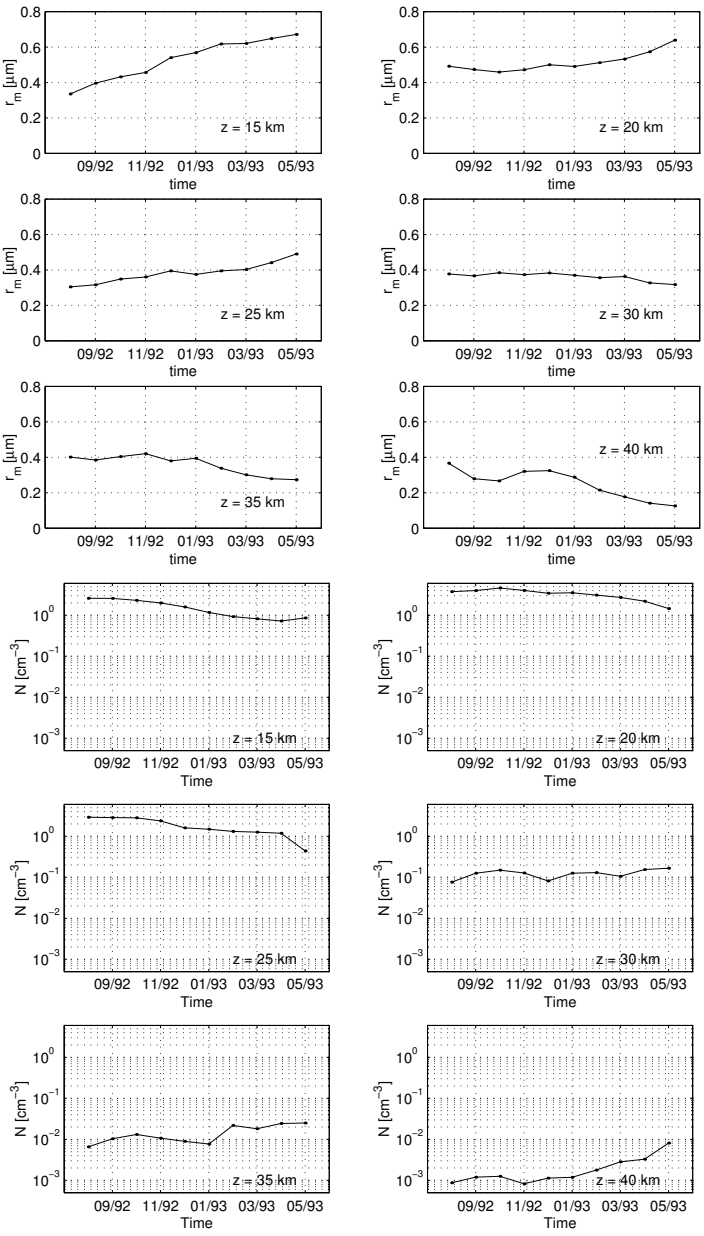

Figure 6.12: Temporal evolution of the retrieved radius parameter r_{m} and the aerosol number density $\mathcal{N}$ (monthly means) at 6 different altitudes.

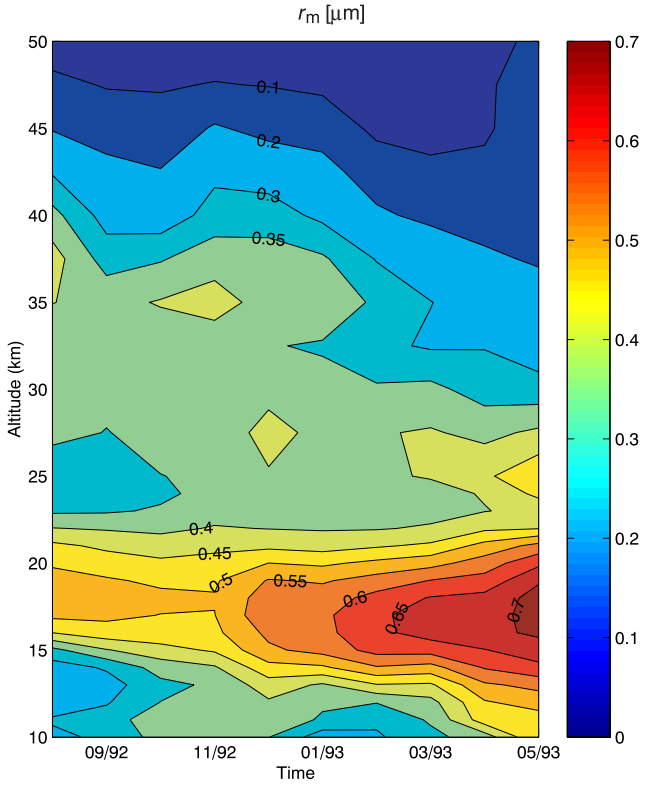

Figure 6.13: Contour plot of $r_{\mathrm{m}}(t, z)$ [μm]. Notice the particle size increase between 15 and 20 km. Gravitational settlement is also suggested.

spontaneously into smaller fragments. Some other process has to be taken into account.

On Fig. 6.13, a contour plot of r_{m} versus time and altitude is represented. The growth of particles is clearly visible. Another feature of importance is the negative slope of the large-particle region. The aerosols are clearly decreasing in altitude, and we immediately suspect the gravitational settlement of particles to be responsible.

We have to be cautious, however. The association of aerosol dynamic processes in this way is based on qualitative observations. Besides that, other processes may have to be taken into account. Therefore, a closer look at aerosol dynamics is necessary.

6.8 Aerosol: dynamic processes

6.8.1 Particle regimes

We are dealing with particles (aerosols) that travel through a fluid (air). A particle 'views' the surrounding fluid in a way that is dependent on its size.

Air molecules can be looked upon as traveling in a straight line in between collisions, and the average distance that is traveled between collisions is the *mean free path* λ_{m}. We can now distinguish two cases. If the particle size is much larger than the mean free path of the surrounding gas molecules, the gas behaves, as far as the particle is concerned, as a continuous fluid. At the other extreme, particles much smaller than the mean free path of the molecules are observed as entities freely moving in between discrete gas molecules. The different regimes can be expressed by the *Knudsen number*:

$$K_{\mathrm{n}} = \frac{\lambda_{\mathrm{m}}}{r} \tag{6.37}$$

with r the radius of the spherical particle. Small Knudsen numbers (< 0.1) represent the *continuum regime*: the gas can be considered as a continuous fluid. Large Knudsen numbers ($\gg 1$) on the other hand represent the *free molecule* or *kinetic regime*, where gas molecules and particles interact as discrete entities. The intermediate case is the *transition regime*. For example, a particle with a radius of 0.1 μm has about $K_{\mathrm{n}} \approx 0.6$ at $z = 0$ km (STP), $K_{\mathrm{n}} \approx 10$ at $z = 20$ km and $K_{\mathrm{n}} > 100$ above $z = 35$ km.

The mean free path of air is critically dependent on density. At an altitude of $z = 10$ km, we have $\lambda_{\mathrm{m}} \approx 2$ μm. At higher altitudes in the stratosphere, the mean free path increases further. For the particle size range that we are dealing with, we can conclude that all regimes are present.

Many theoretical expressions concerning particle dynamics are developed for the continuous regime, and further corrected for non-continuum regimes by the introduction of correction factors. We can conclude by saying that these corrections are important for stratospheric aerosols, and we will have to take them into account.

Another dimensionless number is frequently encountered in aerodynamics. The *Reynolds number* R_{e} expresses the ratio of inertial to vis-

cous forces in a fluid motion. For stratospheric aerosols, R_e is always a very small number ([Turco et al., 1982]), from which we can conclude that viscous forces dominate the motion of a particle in a fluid. In other words: aerosols are not pushed but *dragged* by an air flow.

6.8.2 Gravitational settlement

When a particle is released in the atmosphere, it is attracted by the Earth's gravitational field. From the moment that the particle has a non-zero relative velocity with respect to the surrounding air, it will encounter an opposite directed drag force, generated by the viscous forces that we briefly mentioned in the previous discussion. This drag force is expressed by the *Stokes law*, and it is proportional to the viscosity of air, the particle radius and the relative velocity [Seinfeld and Pandis, 1998]. Initially, the particle is accelerated, but when the downward gravitation force is exactly cancelled by the drag force, the velocity is constant. For the particles of our interest, this *terminal settling velocity* or *sedimentation velocity* is reached after a *very short* time (microseconds), and it is given by:

$$v_{\text{sed}} = \frac{2}{9}\frac{r^2 \rho_{\text{p}} g C_{\text{c}}}{\mu} \tag{6.38}$$

with ρ_{p} the mass density of the particle, g the gravitational acceleration and μ the air viscosity. The velocity depends on the squared particle radius, and this means of course that large particles will be removed faster by gravitation than small ones.

C_{c} represents the dimensionless *slip correction factor*. The Stokes law is derived for the continuum regime (small Knudsen numbers), and is therefore not entirely correct for our purposes. The slip correction factor is given by the empirical expression [Allen and Raabe, 1982]:

$$C_{\text{c}} = 1 + K_{\text{n}}\left[1.257 + 0.4\exp(-\frac{1.1}{K_{\text{n}}})\right] \tag{6.39}$$

Typical values for the terminal settling velocity of sulfuric acid aerosols at different altitudes are given in Fig. 6.14. Notice the large dynamic range of velocities for the different particle sizes. At large altitudes, particles fall relatively fast. This is caused by the large slip correction factor (a large Knudsen number): the mean free path of air molecules is very large due to the low density of the atmosphere at these altitudes.

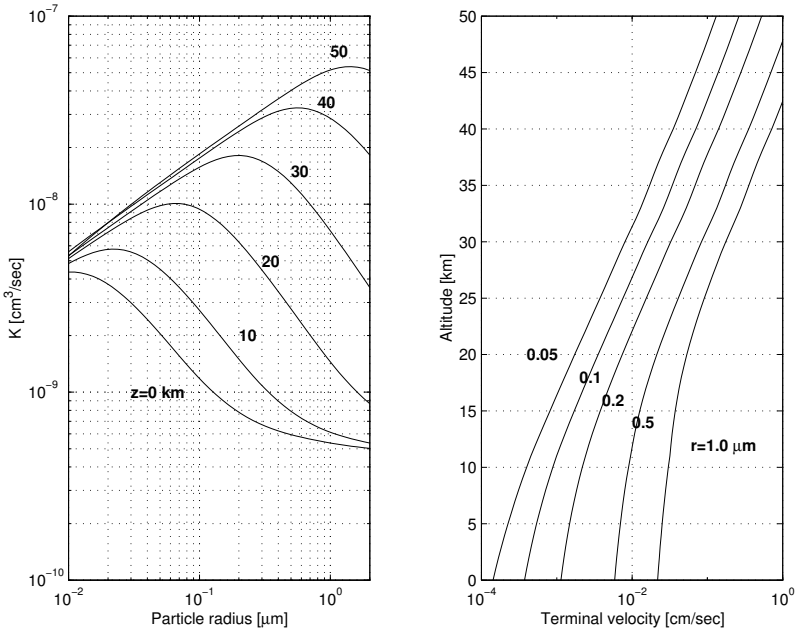

Figure 6.14: Left: Coagulation coefficients at different altitudes for identical particle sizes, versus particle radius. Right: Sedimentation velocity at different altitudes for different particle sizes.

The velocities that we are dealing with are nevertheless quite low. A particle with a radius of $r = 0.2$ μm has a terminal velocity of 0.01 cm s^{-1} at $z = 23$ km. In one year, it will fall a distance of approximately 3 km.

6.8.3 Coagulation

When two moving particles encounter one another, they will physically interact. In the case of fluid particles, cohesive forces ensure that the two separate particles will stick together[4] to form a new, larger particle. Several transport modes can bring two particles into contact, and they provide a means of categorizing coagulation processes.

[4]The probability of this happening is expressed by the *sticking efficiency*, and for fluid particles in the stratosphere it is taken to be unity.

Brownian coagulation

Due to thermal collisions with surrounding gas molecules, a particle will exhibit a random motion. This movement in turn causes collisions between different particles, leading to coagulation. The number of particles that coalesce per second per volume is given by the *coagulation rate* [$\mathrm{cm^{-3}s^{-1}}$]. For a population consisting of particles with 2 different sizes, it is given by [Seinfeld and Pandis, 1998]:

$$J_{12} = \frac{1}{2} K_{12} \mathcal{N}_1 \mathcal{N}_2 \tag{6.40}$$

with K_{12} the *coagulation coefficient*:

$$K_{12} = \frac{2kT}{3\mu} \frac{(r_1 + r_2)^2}{r_1 r_2} \beta \tag{6.41}$$

Here, r_1 and r_2 represent the two particle radii. Without the factor β this equation gives the coagulation coefficient in the continuum regime (small K_n). β is a correction for the transition and free molecule regime that was calculated by [Fuchs, 1964]. It is dependent on temperature, K_n, the particle mass density ρ_p and the gas viscosity μ.

Brownian motion has no directional preference. However, this changes in the presence of temperature gradients (thermophoresis, photophoresis) and gas concentration gradients (diffusiophoresis). In atmospheric conditions, these effects are negligible.

Other mechanisms

Velocity gradients inside an air flow (laminar as well as turbulent) can cause an increased particle collision probability. *Gravitational coagulation* results from small particles being 'trapped' by larger ones that settle faster in a gravitational field.

Additional forces also lead to increased or decreased coagulation rates. The tendency for particles to collide is affected by their electrical charge: *Coulomb forces* induce additional attraction or repulsion. On a molecular scale, *Van der Waals* forces come into play when particles are in close vicinity. *Hydrodynamical forces* also have an influence: air, situated in between two particles, resists being 'squeezed out'.

The relative importance of all these additional processes has been estimated numerically, and it was found that only Van der Waals forces introduce an effect that is not negligible [Seinfeld and Pandis, 1998].

Summarized

From all the processes that govern coagulation dynamics, Brownian coagulation seems to be the main process of importance, but the associated coagulation coefficient K should be corrected for the additional Van der Waals interaction. Also, corrections should be applied for non-continuum effects in our study: at high altitudes, the mean free path of air is comparable and even larger than the aerosol particle sizes.

Fig. 6.14 shows our calculated coagulation coefficients for identical particles at different altitudes (using $\rho_{\mathrm{p}} = 1.66$ g cm^{-3}) as a function of the particle radius. At every altitude, the curve exhibits a maximum for some particle radius. The explanation: small particles have low collision probability, while large particles have little mobility. Both extreme situations are represented by low coagulation coefficients. Somewhere in the middle, a maximum is established. This maximum shifts to larger particle size when altitude increases, because the mobility of the aerosols increases.

The coagulation equation

So far, we have discussed coagulation equations for two particle sizes. We can now generalize to the case of a particle size distribution.

No matter which type of distribution is assumed, it is clear that in an ongoing coagulation process, the total volume of particulate matter is conserved. This is the reason why *the continuous coagulation equation* is described in terms of the particle volume distribution $n(v,t)$:

$$\frac{\partial n(v,t)}{\partial t} = \frac{1}{2}\int_{v_0}^{v-v_0} K(v-q,q)n(v-q,t)n(q,t)\,\mathrm{d}q \qquad (6.42)$$

$$-n(v,t)\int_{v_0}^{\infty} K(q,v)n(q,t)\,\mathrm{d}q \qquad (6.43)$$

with v_0 the smallest volume that is present in the distribution. This equation should be interpreted as follows. (1) a particle with volume v can only be formed through coagulation by two particles with volume q

and volume $v - q$. Hence the positive term[5]. (2) particles with volume v are removed by coagulation with other particles, no matter what size, leading to the negative term.

6.9 A first estimation of process contributions

Modelization of aerosol dynamics is a far from easy task. We have to deal with a 3-dimensional atmosphere, and besides the processes that we have already discussed, the transport of aerosols in the atmosphere by wind plays an important part. We will now try to construct a simple (and very crude!) model that is 1-dimensional: latitudinal and longitudinal transport will be neglected, and two variables remain: the altitude z and the time t. The main argument for such a simplification is of course that atmospheric constituents vary the most along the vertical (altitude) direction.

During the ORA mission, i.e. more than one year after the eruption of Mount Pinatubo, it is reasonable to assume that the full conversion of volcanic H_2SO_4 to stratospheric aerosols is achieved [Turco et al., 1982, Zhao et al., 1995, Weisenstein et al., 1997]. The time variation of the aerosol total number density then consists of three contributions:

$$\frac{\partial \mathcal{N}(z,t)}{\partial t} = T_{\text{sed}} + T_{\text{coa}} + T_{\text{adv}} + T_{\text{snk}} \tag{6.44}$$

respectively representing sedimentation, coagulation, advective[6] vertical transport and an advective meridional sink term (see below).

The temporal variation of the number of particles inside a volume is governed by the *flux* of the particles entering and leaving the volume. For the sedimentation and vertical advection, we can thus write:

$$T_{\text{sed}} = -\frac{\partial}{\partial z}(\mathcal{N} v_{\text{sed}}) \qquad T_{\text{adv}} = -\frac{\partial}{\partial z}(\mathcal{N} v_{\text{adv}}) \tag{6.45}$$

with v_{sed} and v_{adv} the sedimentation and wind vertical velocity, respectively.

In a crude approximation, we estimate the coagulation term by (Eqn. 6.40):

$$T_{\text{coa}} = -\frac{1}{2} K \mathcal{N}^2 \tag{6.46}$$

[5]The factor $\frac{1}{2}$ appears because otherwise all contributions would be counted twice.
[6]The word *advect* means "to be transported by the *mean* wind."

which is the expression for a monodisperse particle distribution (identical particle sizes). The coagulation coefficient K will be evaluated at the measured $r_{\rm m}$ parameters.

The major uncertainty in the model lies in the knowledge of the vertical wind velocity $v_{\rm adv}$ and the meridional sink term $T_{\rm snk}$. The latter can be roughly evaluated by the use of the mass continuity equation, and is given by [Koziol and PudyKiewicz, 1998]:

$$T_{\rm snk} = -2\frac{\mathcal{N}}{\mathcal{N}_{\rm air}}\frac{\partial}{\partial z}(\mathcal{N}_{\rm air} v_{\rm adv}) \tag{6.47}$$

in which $\mathcal{N}_{\rm air}$ is the air number density. The factor 2 enters the expression because horizontal advection occurs towards both the northern and southern hemispheres.

To evaluate the vertical wind velocity, we now rearrange Eqn. 6.44. The air number density is evaluated as $\mathcal{N}_{\rm air}(z) = \mathcal{N}_{\rm air}(z_0)\exp(-\frac{z-z_0}{H})$ where the scale height H has been kept fixed at a (stratospheric) value of 7 km. We obtain:

$$\frac{\partial v_{\rm adv}}{\partial z} = \frac{-1}{3\mathcal{N}}\left[\left(\frac{\partial \mathcal{N}}{\partial z} - \frac{2\mathcal{N}}{H}\right) v_{\rm adv} + \frac{\partial \mathcal{N}}{\partial t} + \frac{\partial}{\partial z}(\mathcal{N} v_{\rm sed}) + \frac{1}{2}K\mathcal{N}^2\right] \tag{6.48}$$

Now the ORA measurements enter the arena. We have calculated monthly mean profiles for $\mathcal{N}$ and $r_{\rm m}$. Except for the unknown $v_{\rm adv}$, all quantities on the right-hand side of Eqn. 6.48 have been evaluated. We are left with a first order differential equation that has been integrated downward by a simple *Euler* method. As boundary condition, we chose $v_{\rm adv}(z = 50{\rm km}) = 0$, a choice that turned out to have little influence on the computed profile. The resulting (mission averaged) vertical velocity profile is shown in Fig. 6.15. Also shown is a climatological profile used by [Koziol and PudyKiewicz, 1998] and a profile computed from the UKMO assimilation model.

One important remark is in order here: the climatological and UKMO profiles should not be taken *too* seriously: the associated uncertainty is *huge*. For instance, UKMO vertical wind profiles are derived as residuals from a 3D transport model, and they represent nothing more than very crude estimates.

Having said this, Fig. 6.15 nevertheless shows that the ORA vertical wind profile is not very different from the climatological profile, and

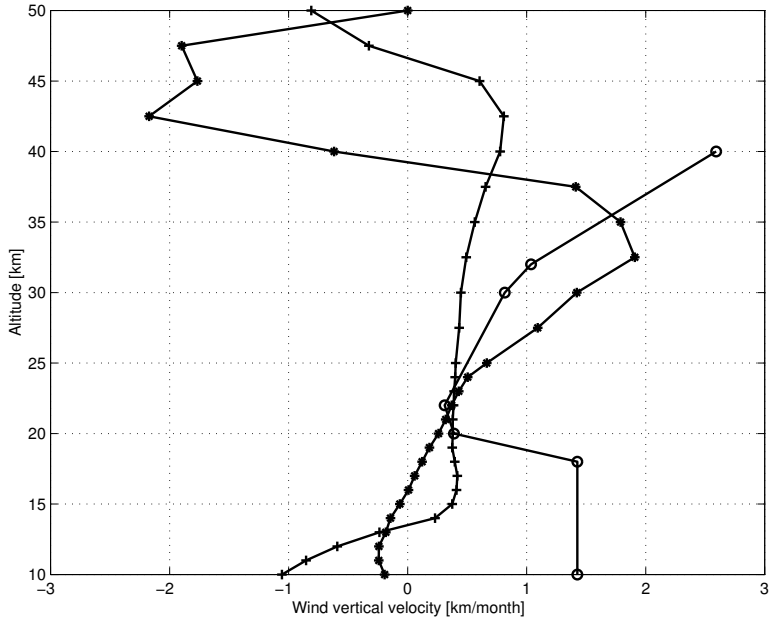

Figure 6.15: Averaged vertical wind velocity profile in a [-40°, 40°] latitude band, obtained from ORA data (filled circles); a climatological profile used by [Koziol and PudyKiewicz, 1998] (open circles); averaged meteorological data from UKMO (crosses).

they both show the same increase with altitude up to 35 km. The ORA profile also has a realistic magnitude in comparison with the UKMO profile, and it seems to exhibit the same sign reversal at high altitudes.

Fig. 6.16 shows the profiles of the relative contributions (Eqn. 6.44) by transport, sedimentation and coagulation to the time variation of the aerosol total number density. A few interesting features can be observed. Between 15 and 23 km, sedimentation results in the *piling up* of particles in time: particles falling from higher altitudes decrease in velocity, resulting in a 'traffic jam'. The effect is however entirely overshadowed by the upwards directed vertical wind, and the net result is a time decrease in the aerosol density. Above roughly 27 km, we see that the time gradient of the total aerosol density becomes positive, conforming with what we already discussed previously (Fig. 6.12). It is caused by the upward transport of particles, partially compensated for

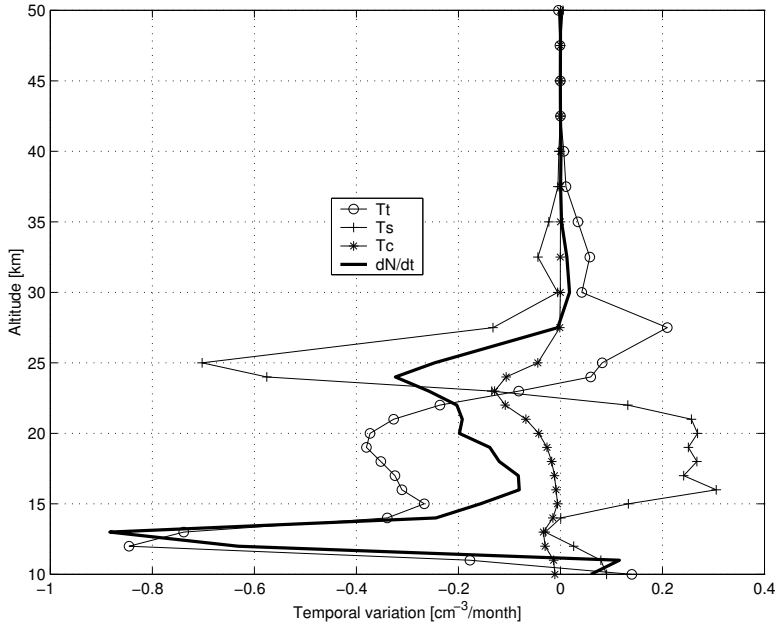

Figure 6.16: Different contributions to the temporal variation of the aerosol total number density $\frac{\partial \mathcal{N}}{\partial t}$ (bold solid line); vertical wind advection (circles), sedimentation (crosses) and coagulation (asterisks).

by downward sedimentation.

Interestingly, where we first intuitively associated the observations with coagulation, it now turns out that this process only plays a secondary role. Nevertheless, its signature in the time evolution of the aerosol density is clearly visible, because sedimentation and vertical advection somewhat compensate for each other over a large altitude range.

6.10 Temporal evolution: a closer look

Until the present, the time evolution of volcanic aerosols in the stratosphere has remained an open and important question. Published relaxation times are about one year [Robock, 2000], an interval that roughly matches the ORA mission period (too short, however, to deliver accurate numbers). It is estimated that the entire stratospheric aerosol loading from a large volcanic eruption is removed after a few years [Robock,

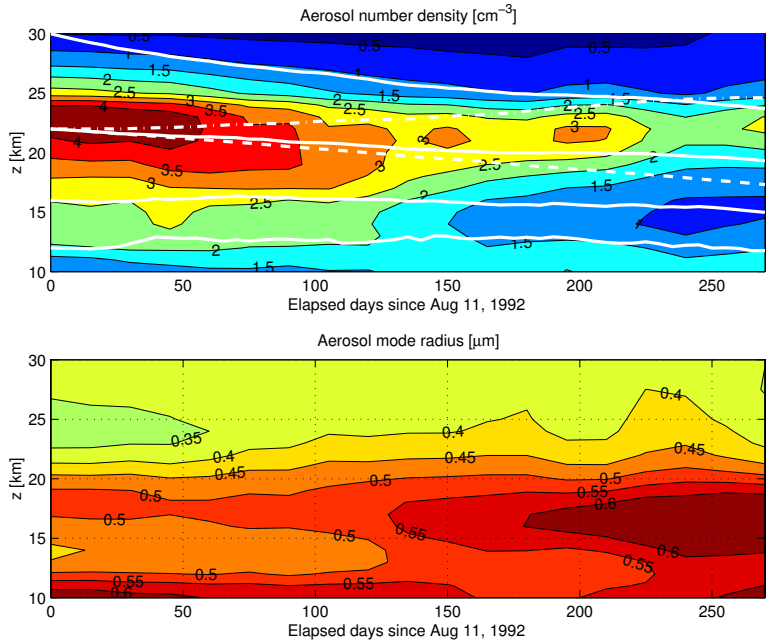

Figure 6.17: Top: latitude averaged time evolution of $\mathcal{N}(t, z)$. Full lines refer to vertical trajectories starting at 12, 16, 22 and 30 km, respectively. The dashed line is the $z = 22$ km trajectory, only accounting for sedimentation, while the dot-dashed line only includes vertical advection (see text). Bottom: time evolution for $r_{\rm m}$.

2000, Turco et al., 1982].

Fig. 6.17 shows a contour plot of the ORA results for $\mathcal{N}(z, t)$ and $r_{\rm m}(z, t)$, averaged along the latitude, and restricted to the upper troposphere - lower stratosphere domain. At the beginning of this chapter, we have seen that the optical thickness decreases in time. We know that this net relaxation results from two opposite effects: the simultaneous decrease of $\mathcal{N}$ and the increase of particle sizes. Although the contributions of sedimentation and vertical transport have greater magnitudes than that of coagulation, they cancel one another out to a certain extent, leaving the coagulation process visible to the naked eye. In one remarkable region on the figure, i.e. from 150 to 270 elapsed mission days at $z = 15$ km, the low density - large particle size association is very clear.

Let us now investigate (more accurately) what is happening. Starting at mission day 1, we calculate the *mean* sedimentation velocity for the associated particle size parameter r_{m}, at a certain altitude z_0. The net vertical velocity v_z is formed by the sum of the sedimentation and vertical transport velocity. The start of a trajectory can now be calculated:

$$z_1 = z_0 + v_z \Delta t \tag{6.49}$$

Repeating this step, every time re-evaluating r_{m} and v_z, we obtain the entire trajectory from the start to the end of the mission. Four trajectories (with initial altitudes at 12, 16, 22 and 30 km) have been superimposed on the density contour plot. Trajectories at high altitudes are steeper, in accordance with what we expect. Trajectories seem to be quite parallel with the aerosol number density isopleths. In the upper troposphere, sedimentation velocities are relatively weak, resulting in the rather flat trajectory.

For the trajectory starting at the aerosol maximum ($z = 22$ km), we have performed the integration with transport only, sedimentation only, and the combined velocity. It is possible that the inaccurate knowledge of the wind vertical velocity may explain the differences between the final integrated altitude (about 19 km), the number density depleted region (about 16 km), and the large particle radius spot (around 17 km).

In order to describe the temporal behavior of aerosol size distributions under the action of coagulation, the integro-differential Eqn. 6.43 should be used. There is however a neat way to avoid this complex task. We know that the total particulate volume inside a volume of air should remain constant under the action of coagulation. In other words, the aerosol volume density (Eqn. 6.18) is constant along a trajectory:

$$V = \mathcal{N} r_{\mathrm{m}}^3 \mathrm{e}^{\frac{9}{2}\sigma^2} = \text{constant} \tag{6.50}$$

or, in terms of the initial values $r_{\mathrm{m}}(t_0)$ and $\sigma(t_0)$:

$$r_{\mathrm{m}}(t) = r_{\mathrm{m}}(t_0) \left[\frac{\mathcal{N}(t_0)}{\mathcal{N}(t)} \right]^{\frac{1}{3}} \mathrm{e}^{\frac{3}{2}(\sigma^2(t_0) - \sigma^2(t))} \tag{6.51}$$

In Fig. 6.18, we present the result of applying the last equation to the evolution of $\mathcal{N}(t)$ and $\sigma(t)$ along the trajectory starting at 22 km. The lower part of the figure shows the comparison of retrieved and

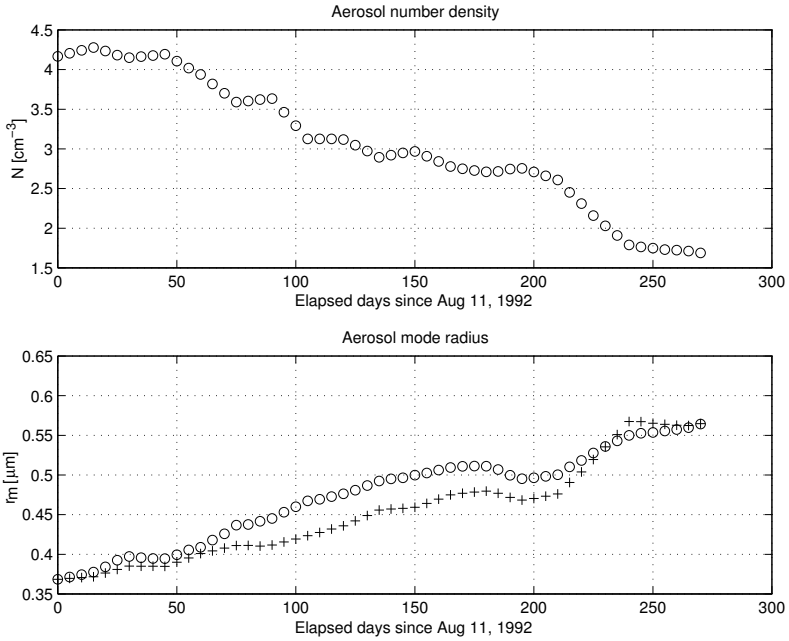

Figure 6.18: Top: retrieved aerosol number density time evolution along the trajectory starting at 22 km. Bottom: retrieved (circles) and computed (crosses) size parameter r_{m} evolution for the same trajectory.

computed temporal evolution of r_{m}. The conclusion is clear: the good agreement strongly supports the coagulation hypothesis.

6.11 Latitudinal variation

Fig. 6.19 represents the mission-averaged aerosol number density in a latitude-altitude contour plot. Besides the well-known presence of an aerosol tropical reservoir at roughly 22 km [Trepte et al., 1994], the evidence of a tongue of aerosols can be seen in the southern hemisphere, which has already been noticed by other investigators [Weisenstein et al., 1997]. It also seems to be a permanent transport pattern because it is clearly visible in the large climatology (from 1984-1994) constructed by [Thomason et al., 1997].

In order to estimate the effect of advective transport, we have super-

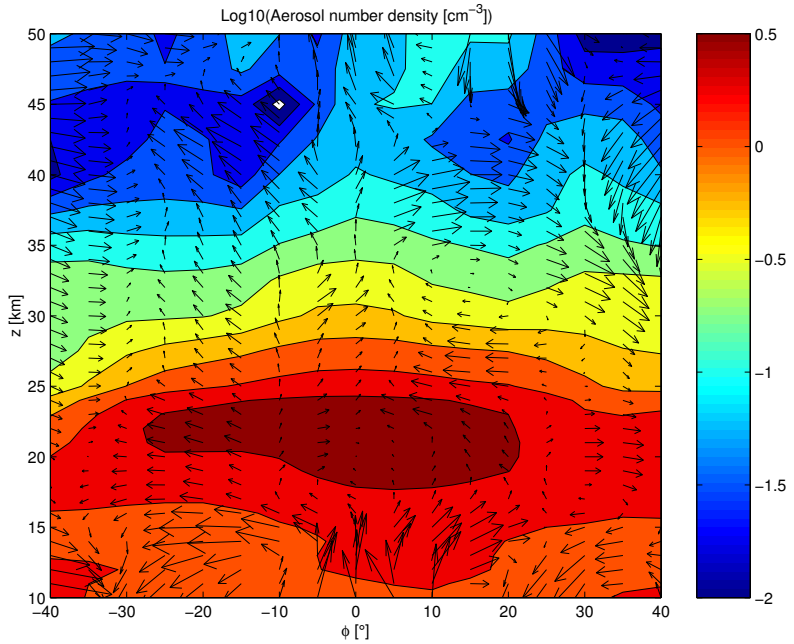

Figure 6.19: Contour plot of the decimal logarithm of the time averaged aerosol total number density $^{10}\log\mathcal{N}$. Superimposed is the UKMO transport vector field [Swinbank and O' Neill, 1994]; arrow lengths are scaled arbitrarily.

imposed UKMO meridional and vertical wind velocities (also mission-averaged) [Swinbank and O' Neill, 1994] on the contour plot. Typical orders of magnitude for meridional and vertical velocities are 500 km month^{-1} (mostly poleward) and 1 km month^{-1} (mostly upward), respectively. It appears that the mean southward transport can be explained by the asymmetry in meridional wind velocities in both hemispheres. While a clear upwelling situation prevails between 10°S and 10°N up to the tropopause, two regions of slightly depleted aerosol levels are visible in the tropics below the tropopause, reflecting the Brewer-Dobson circulation. Similar patterns, including the more intense poleward circulation in the southern hemisphere, have been described elsewhere [Wang et al., 1996].

In the altitude range from 30 to 50 km, there seem to be more particles in the northern hemisphere as a consequence of a large stratospheric

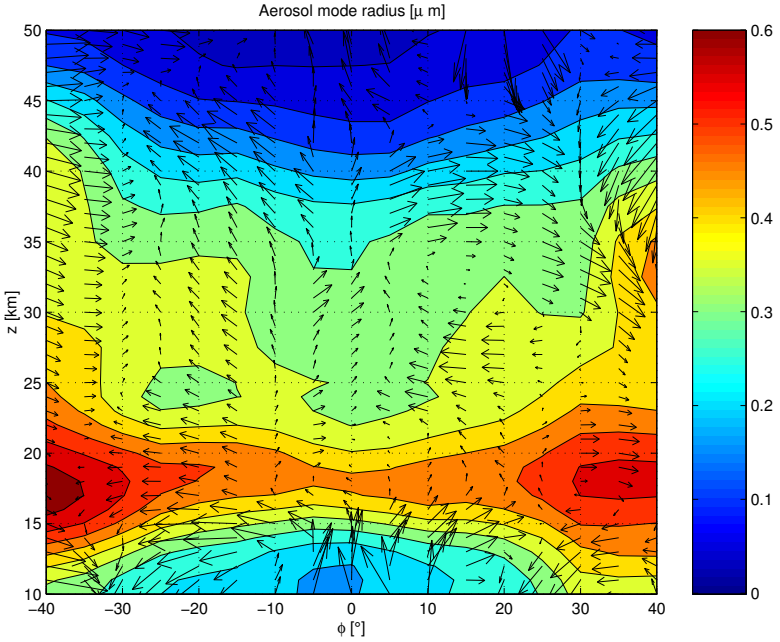

Figure 6.20: Idem as Fig. 6.19, now for the size parameter $r_{\rm m}$.

circulation cell (scc) whose equatorial upwelling branch connects with a downwelling branch located between 20°N and 30°N. As far as we know, this effect has not yet been noticed by other authors.

In Fig. 6.20, we present the isopleths obtained for the aerosol size parameter $r_{\rm m}$ together with the transport velocity fields. The symmetry of the figure is striking, with a general poleward increase of the particle sizes. A very interesting feature is the presence of two symmetrical regions of large particles ($r_{\rm m} \approx 0.6$ μm) located at roughly 17 km in the extratropics. These regions are also characterized by a weak mean meridional circulation or even a circulation cell in the northern hemisphere. This suggests that low transport regions can be associated with large particle sizes. A possible and natural explanation for this relationship could be the fact that, from all particle sizes on the figure, those associated with low transport regions are the *oldest*, which means that they have had enough time to grow under the action of coagulation. At 30°N - 40°N and from 35 to 45km, one can suspect the presence of an-

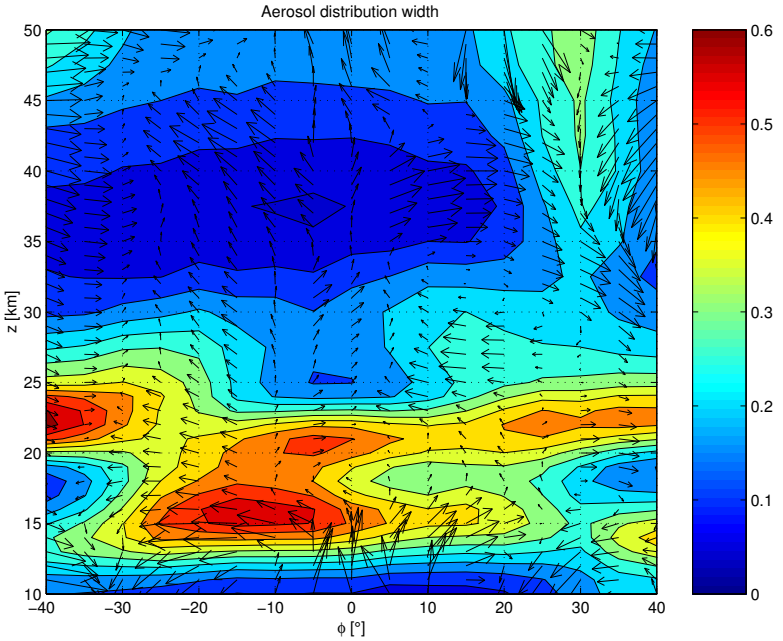

Figure 6.21: Idem as Fig. 6.19, now for the distribution width σ.

other spot of larger particles that might be fed by the above-mentioned circulation cell.

A contour plot of σ is shown in Fig. 6.21. At an altitude of 15 km in the southern tropics, we may observe a large increase of the size distribution width σ, clearly associated with important poleward meridional transport. Also, the downwelling branch of the scc is clearly visible. Furthermore, the above-mentioned stagnation zones are regions of low σ, an observation that is coherent with a coagulation mechanism where the distribution width of particles naturally shrinks with time[7].

The previous contour plots of N, r_{m} and σ show features that have a more or less non-uniform latitudinal dependence, and that are sometimes very local. Whatever the processes may be that cause these observed features, one thing is clear: the total amount of (bulk) aerosol, or its bulk volume should remain constant under the action of transport and

[7]A population of coagulating particles converges asymptotically towards one final particle, or in other words, a size distribution with an infinitely small width.

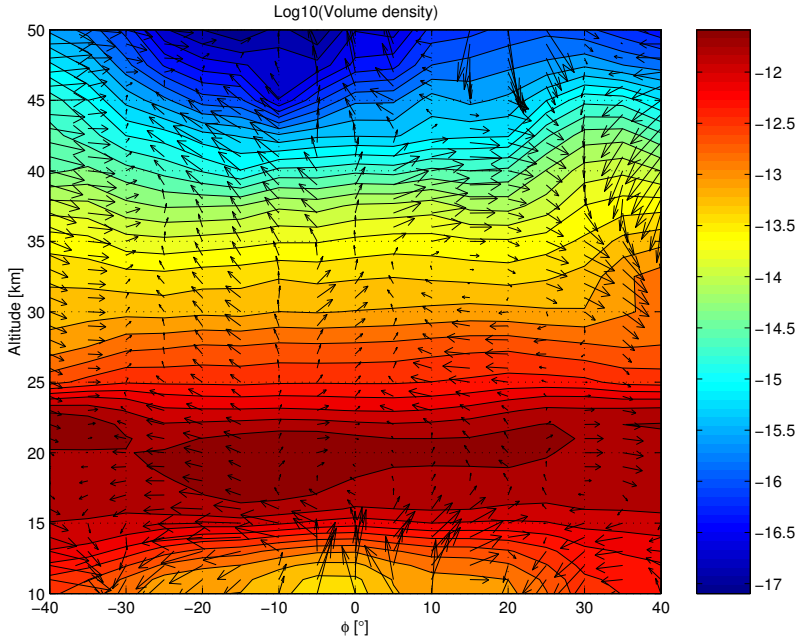

Figure 6.22: Idem as Fig. 6.19, now for the volume density V.

coagulation. Fig. 6.22 shows the aerosol volume density V (calculated with Eqn. 6.18). The features are more uniform now, as expected. Still, the already mentioned southward tongue of aerosols is clearly visible, and the downwelling branch of the scc in the Northern hemisphere is again associated with an increased aerosol loading.

6.12 Conclusion

In general, it is very difficult to extract aerosol microphysical information from optical extinction measurements in the UV-visible spectral region. This is certainly true for ordinary (non-volcanic) background conditions: the particles are too small to induce spectral signatures that reveal the particle sizes. However, we were lucky: in the ORA mission period, the post-Pinatubo aerosols were large enough to induce a scattering behavior that offered sufficient size contrast. The associated inverse problem is however severely ill-posed, and the construction of a regularization

method was imperative. Once again, an altitude smoothing method was used, and the results were found to be satisfactory.

The relative contribution of the active dynamic processes (sedimentation, coagulation, transport) was estimated by the use of a crude 1-dimensional model. The results indicate at least one important feature: vertical transport is not negligible with respect to the other processes. Although the associated velocities are very small, aerosol models should include vertical transport as one of the major governing processes.

At lower altitudes, a simple calculation has shown us that (brownian) coagulation is the main process through which aerosol sizes are altered in time.

A first study of aerosol variation with latitude and altitude revealed some remarkable features. We have proposed transport mechanisms that are possibly responsible for the observed variations. We emphasize here that these are only suggestions: the only way to find the truth is the construction and application of a good 2-dimensional model, and this will be the subject of future work.

Chapter 7

Summary, conclusions and further work.

At the end of this thesis, a summary is in order. All preceding chapters dealt with a specific part of the research, and we will sketch their content in short.

Definitive conclusions can be drawn from the results obtained. Absolute perfection is never achieved in scientific research, and as a consequence there is always room for improvement, both in the experimental and in the data processing stage. Moreover, scientific research is almost never finished, so further work is suggested.

7.1 Summary

7.1.1 The Occultation Radiometer: why and how?

The ORA experiment was designed at the Belgian Institute for Space Aeronomy (BISA) within the scientific framework of atmospheric chemistry and aerosol studies, with a strong focus on the stratosphere. More specifically, it was aimed at the measurement of ozone chemistry-related trace gas densities (O_3, NO_2, H_2O) and aerosols in the stratosphere, while part of the experiment was dedicated to the measurement of mesospheric ozone density. The reason for these measurements is simple. Although current atmospheric chemistry models predict gas densities

reasonably well, they still need lots of fine-tuning, and this can only be achieved by the use of and comparison with new data. Concerning aerosols, lots of issues remain unsolved, and data are needed that can deliver more insight into processes such as aerosol formation, dynamics and transport in the atmosphere. Over the past decade, aerosols have gained lots of attention, because the extremely violent Mount Pinatubo eruption was shown to have had a significant impact on certain trace gas levels and the global climate.

ORA uses the solar occultation method: while in orbit, during subsequent sunsets and sunrises, it measures the solar light, transmitted through the atmosphere, in 8 spectral channels located in the UV, visible, and near-IR wavelength range. The channel wavelengths are specifically chosen to be centered in spectral regions were one of the abovementioned species significantly absorbs or scatters light. Selection of these wavelengths is achieved by the use of filters, and the selected light is detected by the use of a combination of optics and photodiodes. Measured signals can then be digitized and transmitted to Earth.

All in all, ORA is a relatively simple instrument, partly because it is not equipped with a Sun tracking device and telescope. Instead, to ensure that the full solar disk is viewed at all times, while other (disturbing) light sources are effectively blocked out, it has a field of view of $\pm\ 2^{\circ}$. As a result, the *apparent* vertical resolution is low (more on this topic later).

Mounted on the ESA-commissioned EURECA carrier, the experiment was launched on July 31, 1992 onboard the Space Shuttle Atlantis. Over a period of about 10 months, ORA collected data until EURECA was retrieved from space by the Space Shuttle Endeavour on July 24, 1993.

7.1.2 The inverse problem.

A measured transmittance is the result from light rays that are subject to atmospheric absorption, scattering and refraction along their optical paths. Because these rays pass through layers located at a wide range of altitudes, the measured transmittance is in fact an integrated quantity, containing information about light extinction at multiple altitudes. It is the task of the *spatial inversion method* to 'unwrap' this information: from the measurements, the total extinction is calculated at every

altitude.

A total extinction altitude profile itself contains contributions from the different species that absorb or scatter light. In turn, these contributions can be separated by the use of an adequate *spectral inversion method*.

When dealing with inverse problems, one always has to investigate the information content of the measurements in relation to what has to be retrieved from the measurements. More specifically, due to (1) the interdependence of measurements with respect to the experimental noise level, or (2) underdeterminacy (lack of measurements), *null space* problems can emerge: only part of the reality can be retrieved, while the other part resides in a subspace of the solution space, the so-called *null space*. A null space component can never be derived from the measurement, and is therefore meaningless. Basically, there are two remedies: we can decrease the number of solution components, or we can *regularize* the solution by the use of a priori information.

The spatial inversion

In order to derive total extinction altitude profiles from measured transmittance profiles, a model is needed that mathematically represents the different physical processes that are taking place. An exact formulation has been described, based on the Beert-Lambert extinction law, and taking into account the integration over the solar disk source (with solar limb darkening) and the instrument response function. Unfortunately, it was found that the associated inversion would imply a huge computation cost, and so approximations were necessary. The most crucial one consisted of a solar disk discretization in 20 thin horizontal slices, each one acting as a light source with its own optical path through the atmosphere. These paths were calculated, taking refraction into account, from the US76 air density climatology.

The model obtained cannot be inverted analytically, is highly nonlinear and suffers from severe ill-posedness. This last problem is caused by the strong interdependence of subsequent transmittance measurements taken during an occultation. However, we were able to solve the problem with a new method: *Natural Orthogonal Polynomial Expansion* (NOPE). This method combines regularization by the use of a priori information with the reduction of the number of unknowns. The large

amount of post-Pinatubo aerosols lead to an additional problem (a priori extinction is unknown), and a special *mapping method* was required.

The actual spatial inversion of all ORA measurements took about 3 months. The total extinction profiles obtained were realistic, and demonstrated that the actual vertical resolution was much better than expected. However, the constructed model was unable to fit the measurements within the (very small!) digitization error levels. Typically, inversion residuals were found to be about 10 times larger than the digitization errors (roughly 10^{-4}). Most probably, the residuals are caused by model errors (model parameter uncertainties, used approximations, neglected physical processes).

To find out if the results obtained are valid, we compared total extinction profiles obtained from ORA and SAGE II for 25 near-coincident occultations. Over the entire spectral and altitude ranges, a typical relative difference of 20 % was found. Taking into account the imperfect overlap in time, geolocation and viewing geometry, this difference can be considered as acceptable.

At very low and high altitudes, the measured transmittance becomes zero or one, respectively. This means that the information content is very small, and as a result, the associated regions in the total extinction profiles are still somewhat chaotic. In order to stabilize the profiles, a Direct Method was constructed, based on statistical information obtained from the entire ORA data set, combined with the retrieved NOPE total extinction profiles. Also, the results obtained allowed us to estimate the true vertical resolution of the instrument to be about 2 km in altitude.

The spectral inversion

A total extinction profile consists of partial extinctions caused by the different atmospheric species. The separation of these contributions can be achieved when all spectral channels are used at once. The spectral absorption characteristics of trace gases are known in advance: these are given by the absorption-cross sections. Aerosols on the other hand give another story: the scattering spectral behavior is largely determined by the particle sizes, that are not known in advance. We decided to describe aerosol scattering as a quadratic function of wavelength, with the important advantage that the spectral inverse problem became linear.

We have constructed an inversion method, based on *Bayesian* grounds:

using the formalism of probability density functions together with Bayes' theorem, it is possible to incorporate a priori information in a very elegant way.

Before the actual inversion, we decided to remove the Rayleigh scattering contribution by calculating it from meteorological data obtained from the UKMO (United Kingdom Meteorological Office). While doing so, it was discovered that a significant amount of atmospheric straylight had entered the ORA modules, causing an underestimation of the total extinction at high altitudes. It was necessary to take this additional contribution into account.

The spectral inversion was performed by the use of the Bayesian method, in combination with a priori information from climatology and an altitude smoothing constraint. The results: number density profiles for O_3 and NO_2, and aerosol extinction profiles at 6 wavelengths.

These results were compared for the 25 coincident occultation events from ORA and SAGE II, showing that the relative differences were not small, but acceptable. This did not come as a surprise, since the differences were already present in the total extinction comparison.

7.1.3 Aerosols.

The ORA mission period is particularly interesting for the reason that the stratospheric aerosols generated by the Pinatubo eruption are in full relaxation mode. We therefore focused our research mainly on the study of these aerosols.

A first exploration of the latitude-time variation of the 1013 nm optical thickness shows that the results obtained from the ORA and SAGE II experiments are very similar. However, the optical thickness is an integrated quantity (upwards integrated aerosol extinction). Information on the size and number density of aerosols can only be retrieved from the 6 aerosol extinction profiles by the application of yet another inversion: the *radial inverse problem*. The aerosol size and number density can be derived from extinction measurements by the use of a theoretical optical model: the Mie theory. This theory is exact for spherical particles (such as the stratospheric aerosol droplets), and makes use of the refraction index of the aerosols ($N = 1.43 + \mathrm{i}\ 0.01$ for H_2SO_4/H_2O particles with a H_2SO_4 mass fraction of 65 to 80 %).

Actually, within an aerosol population, particle sizes cover a large

range, so we need to introduce a particle size distribution. Usually, a log-normal distribution is assumed. Such a distribution tends to smooth the spectral behavior of aerosol scattering, and in doing so, causes a decrease in the information content of observed aerosol extinction data. This is very unfortunate, because the information that is carried by extinction is already very limited by nature: for small or large particle sizes, no separate data can be derived for the 3 log-normal distribution parameters $\mathcal{N}$ (number density), $r_{\rm m}$ (median size) and σ (distribution width).

A radial inverse problem is therefore severely ill-posed, and (again) an adequate method is needed. Our method involves the simulation of the entire measurement space from the model, an altitude smoothing construction and a non-linear least-squares optimization. The results: altitude profiles for $\mathcal{N}$, $r_{\rm m}$ and σ. Typical values that were obtained for the lower stratosphere (Junge layer): $\mathcal{N} = 4$ particles cm^{-3}, $r_{\rm m} = 0.5$ μm and $\sigma = 0.2$.

Initial investigation suggested that coagulation dominated the aerosol dynamics in the lower stratosphere: as the number density decreased over time, particles clearly grew larger. At high altitudes, the reverse situation was found: particles became smaller as density increased over time.

A simple 1-D model was constructed, taking into account sedimentation, coagulation, vertical and meridional advection. Combined with ORA data, interesting features were found. In the lower stratosphere, the downward sedimentation is entirely overshadowed by upward vertical advection, and the observed density decrease is caused primarily by the coagulation process. At high altitudes ($z > 27$ km), the positive time gradient of number density is caused by the upward transport of particles.

The lower stratospheric coagulation hypothesis has been double-checked in a more accurate way by studying the evolution of aerosol properties along a pre-calculated transport trajectory. Under the action of coagulation, volume density should remain constant, and this was indeed confirmed in a very convincing way.

As far as we know from the literature, ORA has delivered the first latitude-altitude contour plots of the distribution parameters $\mathcal{N}$, $r_{\rm m}$ and σ. Several interesting features have been found. In general, the number

density decreases from equator to pole, while particles become larger, and this again suggests an ongoing coagulation process. Especially 2 regions in the extratropics contain very large particles at an altitude of about 17 km.

Interestingly, the Northern Hemisphere seems to contain more particles as a consequence of a large stratospheric circulation cell (scc) with a downwelling branch located between 20°N and 30 °N. This *scc* might also be responsible for a high-altitude region with larger particles and a very clear signature in the distribution width σ in the upper stratosphere.

Still, a contour plot of the volume density V shows that the total aerosol volume (or mass) is quite homogeneously mixed in the horizontal direction, although elevated aerosol loading is again observed in the Northern Hemisphere.

7.2 Possible improvements

What can be done to improve the results? Let us first consider the point of view of the experiment. Of course, as other instruments demonstrate very clearly, additional technology offers considerable advantages: a small field of view, combined with a Sun tracker decreases the complexity of the spatial inversion algorithm, a CCD array offers much more spectral information , etc. However, let us fix the ORA experimental set-up as it is, and investigate the possibilities.

Usually, when the word *improvement* enters a discussion, one automatically starts thinking in terms of measurement sensitivity. However, increasing the sensitivity is useless since, as we have seen, the spatial inversion model is at present unable to fit the transmittances within the measurement uncertainty. The only way to improve this situation is by making the used model more accurate – with the unfortunate result that computation costs increase significantly.

Is it possible to increase the information content by taking more measurements during an occulation? The answer is again no. It is clear now that the data set from one occultation already shows strong redundancy: two subsequent satellite positions along the orbit deliver data that are strongly dependent, because the wide field of view implies that almost the same atmospheric situation is measured.

It is likely that the results from the spectral inversion would improve if more spectral channels were to be included. However, one has to keep in mind that there is an inherent limit to what can be achieved: the overlapping of extinction cross sections from the different species ensures that there will always be some amount of cross-correlation present in the retrieved parameters. This is not caused by the instrument: it is a simple consequence of the non-orthogonality of extinction cross-sections.

What about the used models? As already mentioned, a more accurate spatial inversion model would have its benefits. More accurate means: less approximations, and the inclusion of other processes such as the reflection of light on the Earth surface (although simulations show this contribution to be very small) and cloud tops. It would also be better to model the incoming straylight more accurately.

In summary, improvements are possible mainly in the domain of the (complex) data processing. The ORA instrument itself (given its limitations) has performed very well — a statement that should not be taken lightly, since the presence of malfunctions in space experiments (caused, for example, by the extreme space environment and launch vibrations) are quite common.

7.3 Further work

In this thesis, we have presented inversion methods and the data exploration of the aerosol results. The remaining part of the retrieved data (O_3 and NO_2 gas densities) has not been fully processed yet, a task that might be performed in the near future. Furthermore, we still need to retrieve the water vapor density profiles from the 943 nm spectral channel measurements, but this requires the development and application of somewhat different inversion methods.

Although important new insights have been gained concerning the aerosols, we have to proceed further, and systematically investigate temporal and spatial variations: we have only presented results that are averaged over rather large data bins (entire mission period, entire latitude range).

Maybe the most important task that lies ahead is the construction of a good 2-D model (altitude versus latitude) that simulates the creation, dynamics and transport of stratospheric aerosols. Only in this way can

we confirm with certainty that the observed aerosol features can be explained by the suggested physical processes. The development of such a model provides us with an important and exciting challenge for the future.

Appendix A

Mesospheric ozone

The ORA spectral channel, centered at 259 nm, was specifically aimed at the measurement of mesospheric ozone, having strong absorption bands (The Hartley bands) in this wavelength region. However, a faulty filter design combined with a strong filter and window degradation forced us to develop a specific retrieval scheme that differs strongly from the one that has been used for the other channels. The description of this method and its results has been published in *Geophysical Research Letters*. We include the article here for completeness.

GEOPHYSICAL RESEARCH LETTERS, VOL. 27, NO. 21, PAGES 3449-3452, NOVEMBER 1, 2000

Ozone profiles from 30 to 110 km measured by the Occultation RAdiometer instrument during the period Aug 1992-Apr 1993.

Didier Fussen, Filip Vanhellemont, Christine Bingen
and Simon Chabrillat
Institut d'Aéronomie Spatiale de Belgique, Brussels

Abstract. We present ozone volume mixing ratio profiles obtained by the ORA instrument during the period Aug 1992-Apr 1993. They have been retrieved by applying a specific inversion algorithm to a radiometric UV channel contaminated by Rayleigh scattering. The results compare reasonably well with other instruments up to the mesopause and are probably the first extended ozone data in the lower thermosphere (90-110 km).

1. Introduction

The Occultation RAdiometer (ORA), a simple UV-visible instrument for atmospheric remote sensing by the solar occultation method, has been developed by the Belgian Institute for Space Aeronomy. It was launched in July 1992 on board the EUropean REtrievable CArrier for a one year mission. The instrument was dedicated to the measurement of stratospheric vertical profiles of O_3, NO_2, H_2O number densities as well as the extinction coefficient of aerosols by using 8 broadband channels ranging from 260 nm to 1013 nm.

ORA recorded about 7000 orbital sunsets and sunrises through the Earth's atmosphere from an altitude of about 500 km. The geolocations of the tangent points lie between 40° S and 40° N due to the low-orbit inclination (28°). In previous publications [*Fussen et al.*, 1997, 1998], we presented the results obtained by the application of a dedicated inversion method. Indeed, the instrument has a large field of view (±2°) suggesting a *apparent* coarse vertical resolution ($\simeq 20\,km$) determined by the Sun's angular size itself. However, this fact is balanced by the quality of the transmittance signal that has a very large signal-to-noise ratio only limited by the digitization (16 bits).

A specific channel was devoted to the study of mesospheric ozone profiles by using the important absorption near the maximum of the Hartley band (260 nm). Unfortunately, this channel was known to be defective before the flight due to a small parasitic transmission window at about 340 nm enhanced by the stronger solar irradiance in that domain. Indeed, the observed transmittance signal turned out to be strongly structured, reflecting the competition between ozone absorption and Rayleigh scattering.

We present the specific retrieving algorithm for this channel and a first analysis of the obtained ozone fields from the stratosphere up to the lower thermosphere.

Paper number 2000GL011575.
0094-8276/00/2000GL011575$05.00

2. Inversion algorithm

A schematic view of the occultation geometry is presented in Fig. 1. The mean tangent altitude h_0 is defined by the straight line joining the satellite and the center of the Sun, assuming that refraction may be neglected above $h_0 \simeq 30\ km$. The Earth's radius R is determined by the sphere tangent to the geoid. The detector records the light emitted by the whole solar disk having an angular size of about 9.3 mrad. This corresponds to a vertical domain of 22 km for the observed atmospheric region at $h_0 = 80\ km$.

The transmittance , defined as the ratio of the occultation signal to the signal measured outside the atmosphere, may be written as

$$T(h_0) = \int_{\lambda_1}^{\lambda_2} \int_{\alpha_1}^{\alpha_2} F(\lambda) G(\alpha, \lambda) \exp(-\tau(\alpha, \lambda))\, d\alpha\, d\lambda \quad (1)$$

where the integral over λ covers the nominal spectral range in the Hartley band as well as the parasite spectral window up to 360 nm. The filter function $F(\lambda)$ contains all the spectral information related with the nominal filter transmittance, the detector sensitivity and the solar irradiance. A second integration runs over the angular domain $[\alpha_1, \alpha_2]$ defining the apparent Sun that is discretized into horizontal slices (a number of 10 slices was found to be a good trade-off between computing time and accuracy). Each slice observed under an angle α can be associated to a tangent height $h(\alpha)$ given by

$$h(\alpha) = \frac{\sqrt{R_o^2 - (h_0 + R)^2} + (h_0 + R)}{\sqrt{1 + \tan(\alpha)^2}} - R \quad (2)$$

The contribution of each slice is weighted by an intensity distribution function $G(\alpha, \lambda)$ across the solar disk that takes into account the relative surface of the slice and the solar limb darkening function [*Allen*, 1985]. According to Bouguer's law, all rays emitted from a particular slice are exponentially attenuated with respect to the slant path optical thickness $\tau(\alpha, \lambda)$ obtained as a third integration along the optical path of the total extinction coefficient $\beta(z, \lambda)$:

$$\tau(\alpha, \lambda) = \int_{s_1}^{s_2} \beta(z(s), \lambda)\, ds \quad (3)$$

The s_1 and s_2 values define the entrance and exit points in the Earth's atmosphere and the local altitude z is related to the optical path length element s by

$$z = \sqrt{s^2 + (h(\alpha) + R)^2} - R \quad (4)$$

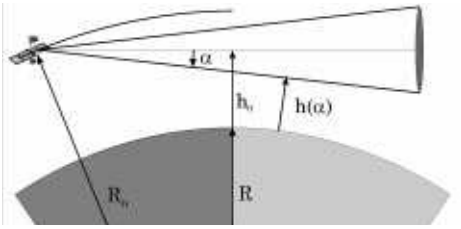

Figure 1. Geometry of a solar occultation

Although a very accurate knowledge of $F(\lambda)$ is usually not critical in an occultation experiment where a monochromatic transmittance is measured (with respect to the intensity recorded outside the atmosphere), this was not the case here where both Rayleigh scattering and ozone absorption may interfere (see Fig. 2).

Some tentatives to re-construct $F(\lambda)$ from pre-flight or post-flight instrument characteristics and published solar irradiance values [*Kurucz et al.*, 1984] were not satisfactory mainly because the temporal evolution of photodiode characteristics was not accurately known. Furthermore, the degradation of the optics was not a linear function of the mission elapsed time. Therefore, we decided to adopt a heuristic procedure where we modelized $F(\lambda)$ as a bi-gaussian function

$$F(\lambda) = c_1 \exp(-(\frac{\lambda - c_2}{c_3})^2) + c_4 \exp(-(\frac{\lambda - c_5}{c_6})^2) \qquad (5)$$

From inspection of the signal, it is clear that the major interference between air and ozone takes place in the $[30 - 70\ km]$ altitude range. For each ORA mission elapsed day, we used US76 [*National Oceanic and Atmospheric Organization*, 1976] standard atmospheric air and ozone profiles to generate a synthetic transmittance. In this low altitude subrange, a non-linear optimization of the set of c_i $\{i = 1..6\}$ was then performed to obtain an accurate fit of one measured event for each day. The ratio of ozone and air cross sections is not the same at different wavelengths and this allows the algorithm to converge toward a solution with well separated centroids. This procedure gives the most probable filter function with respect to a mean atmosphere and it is expected that the temporal evolution of the filter characteristics is smooth, continuous and does not reflect deviations of true ozone profiles with respect to mean values. In the $[30 - 70\ km]$ altitude range, the filter function optimization is strongly constrained to reproduce both ozone and air contributions in both spectral regions. It should be noted that this empirical determination of $F(\lambda)$ remains independent of the transmittance measured between 70 and 120 km.

The retrieved centroids ($c_2 = 272.5 \pm 2.4$ nm and $c_5 = 342.4 \pm 4.3$ nm) and widths ($c_3 \simeq c_6 \simeq 5$ nm) turned out to be very stable during the mission (indicating the absence of wavelength shift) and we present the evolution of $c_{1,2}$ and $c_{4,5}$ in Fig. 3. All coefficients have a well correlated temporal evolution which cannot be attributed to a progressive deviation of ozone with respect to the climatology. Latitudinal effects appear also to be negligible. The relative importance of the nominal filter and of the leak (represented by c_1 and c_4, respectively) showed clearly a strong degradation of the former during the mission. This is probably caused by some opacification of the radiometer optics due to intense UV radiation.

Assuming that the time evolution of the $c_i(t)$ is valid for all events of the same day, the forward model is fully defined by renormalizing $F(\lambda)$ and $G(\alpha, \lambda)$ by

$$\int_{\lambda_1}^{\lambda_2} \int_{\alpha_1}^{\alpha_2} F(\lambda) G(\alpha, \lambda)\, d\alpha\, d\lambda = 1 \qquad (6)$$

The total extinction coefficient $\beta(z, \lambda)$ can be expressed as a sum of products of absorption cross sections times the associated number density profiles as

$$\beta(z, \lambda) = \sigma_{O_3}(\lambda)\, n_{O_3}(z) + \sigma_{Rayleigh}(\lambda)\, n_{Air}(z) \qquad (7)$$

The temperature dependence of the ozone cross section [*Burrows et al.*, 1999] in the Huggins band has been considered. An equivalent cross section was computed by integrating the temperature dependent cross section over the optical path and by using climatological temperature profiles.

The large angular field of view of ORA is responsible for an important overlap between two successive measurements along the orbit, which may cause severe ill-conditioning in the inversion of $T(h_0)$. Therefore we preferred to develop the unknown density profiles on a basis of Chebyshev orthogonal polynomials $P_i(z)$ (defined in the range $[0 - 150\, km]$):

$$n_{O_3}(z) = \widetilde{n}_{O_3}(z) \sum_{i=0}^{N_{O_3}} a_i^{O_3} P_i(z) \qquad (8)$$

$$n_{Air}(z) = \widetilde{n}_{Air}(z) \sum_{i=0}^{N_{Air}} a_i^{Air} P_i(z) \qquad (9)$$

where $a_i^{O_3}$, a_i^{Air} are unknown coefficients defining the relative ozone and air number density profiles ($\widetilde{n}_{O_3}(z)$, $\widetilde{n}_{Air}(z)$ are standard atmospheric profiles used for a correct scaling of the problem). The inversion problem can now be solved

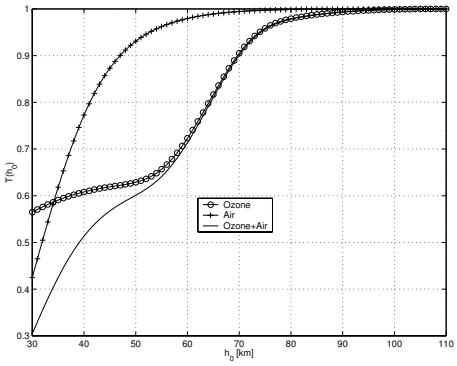

Figure 2. Observed total transmittance signal (full line) and best fit calculated contributions for ozone (circles) and air (crosses)

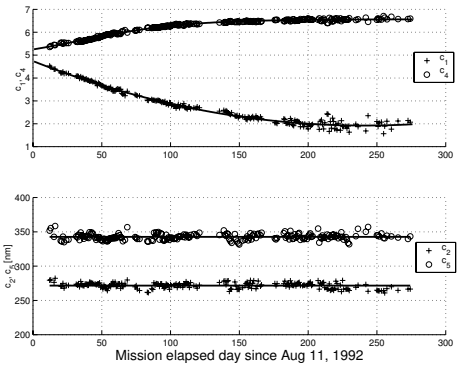

Figure 3. Temporal evolution of the filter coefficients c_1 and c_4

by finding the set of $\{a_i^{O_3},\ a_i^{Air}\}$ coefficients that minimize the merit function M

$$M = \sum_{j=1}^{J} \left(\frac{T_m - T_e}{\sigma(h_0(j))} \right)^2 \tag{10}$$

where $T_m = T_m(h_0(j); a_0^{O_3}, ...a_{N_{O_3}}^{O_3}, a_0^{Air}, ...a_{N_{Air}}^{Air})$ and $T_e = T_e(h_0(j))$ respectively refer to the modelled and experimental transmittances at $J = 81$ tangent altitudes (1 km steps from 30 km to 110 km) during the occultation. The experimental errors $\sigma(h_0(j))$ were estimated by taking into account digitization and offset determination errors. The minimization of M was performed by using a standard Levenberg-Marquardt algorithm applied to 23 unknowns (N_{O_3}=17, N_{Air}=6).

At the minimum, the error on the solution was estimated by means of the Jacobian matrix of M [*Press et al.*, 1992] and quadratically added to errors coming from wavelength and Sun angular size discretization, temperature dependence of the ozone cross section, orbital elements and error estimation of the filter function spectral parameters. We also checked that the solution quality was insensitive to a supplementary optimization of the empirical filter function with respect to the solution profiles. This proves, a posteriori, that the use of climatological profiles was accurate enough for computing the filter function before the vertical inversion.

3. Results and discussion

From the initial 7000 occultations, we rejected all events where the satellite attitude did not have the required stability. Further, we stopped the Levenberg-Marquard optimization after 25 iterations in case of slow convergence and considered the event as not valid. Also, outlier solutions with respect to a 2-sigma level for any altitude were considered as unrealistic and rejected. Finally, a data set of 2530 profiles was obtained.

Figure 4 shows the averaged ozone profiles retrieved from the ORA experiment and the corresponding MLS (composite of 183 GHz and 205 GHz data), HALOE and SME profiles averaged during the period Aug 1992-Apr 1993 between 40S and 40N. Although the instrumental sensitivity limit is reached at about 110 km through the digitization error, we expected the averaged error to be reduced by a factor of $\sqrt{2530} \simeq 50$.

In order to check the consistency of the ORA retrieval, we compared profiles down to 30 km although it was not expected to obtain a better accuracy with our instrument. At 30 km, an ozone deficit of about 25 % for the ORA profile can be attributed to the effect of neglecting refraction. Between 40 and 60 km, ORA seems to exhibit an excedent of about 10-20% compatible however with the estimated error bar. This bias is attributed to the imperfect knowledge of the filter spectral properties and the temperature sensitivity of the ozone cross section in the Huggins bands.

From 60 to 90 km, considerable differences can be observed between the three instruments at the level of the mesospheric ozone minimum around 75-78 km. The HALOE profile reaches a more pronounced minimum while MLS seems to show signs of exaggerated inclusion of a priori information above 85 km. We have also included SME data [*Keating et al.*, 1996] up to about 87 km and these values lie closer to the ORA data at the ozone minimum.

The presence of the ozone maximum around 90 km is quite well understood[*Allen et al.*, 1984; *Fichtelmann and Sonnemann*, 1989; *Sonnemann et al.*, 1998]. It results from a maximal rate of ozone production reactions:

$$O + O_2 + M \ \rightarrow \ O_3 + M \tag{11}$$

due to the opposing gradients of increasing atomic oxygen and decreasing air density when altitude increases. The augmentation in atomic oxygen is itself strongly regulated by a loss reaction with OH and hence by the H_2O concentration.

The diurnal mesospheric ozone variation is very important. Atomic oxygen is converted into a large ozone excess after sunset which is strongly depleted at sunrise by photodissociation. During the day, a slow recovery of ozone occurs due to atomic oxygen production subsequent to O_2 photodissociation. This is the reason for observing a larger mesospheric ozone volume mixing ratio at sunset than at sunrise for a solar occultation experiment. Actually, the photochemical system of the mesosphere represents a para-

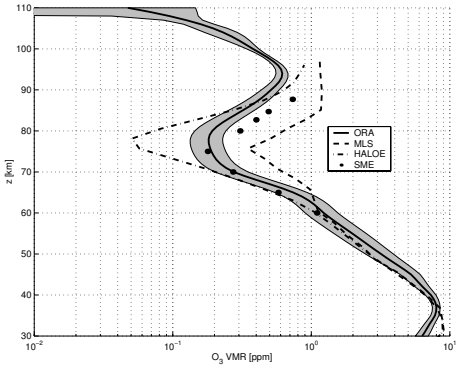

Figure 4. Intercomparison of ozone volume mixing ratio profiles obtained by the SME, MLS, HALOE and ORA instruments

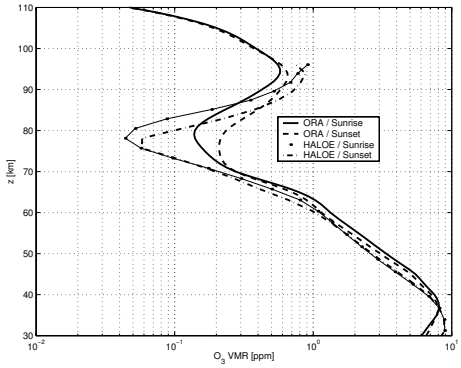

Figure 5. Intercomparison of sunrise/sunset ozone volume mixing ratio profiles for HALOE and ORA

metrically driven photochemical oscillator excited by the diurnal solar radiation leading to a resonance regime in the mesosphere [*Sonnemann and Fichtelmann*, 1997]. In Fig. 5, comparing sunrise and sunset events shows a large diurnal variation effect between 70 and 95 km for ORA and HALOE. Both instruments have a maximal variation around 83 km with a variation factor of about 1.8 for ORA and 2.5 for HALOE. It must be noticed that the retrieval algorithms (onion peeling for HALOE and the method described above for ORA) assume a radial atmosphere with no horizontal nor temporal variations of number densities. This could lead to a smoothing effect in the ORA profiles because the angular integration spans different tangent altitudes associated with different photochemical evolutions of the ozone number density along the line of sight.

Below 70 km, there is an inversion in the sunset-sunrise ratio for both instruments. This slight diurnal effect has been discussed by [*Brühl et al.*, 1996] although no definitive explanation seems to prevail.

To our knowledge, there do not exist other measurements (with extended spatial and temporal coverage) of ozone profiles above 92 km to be compared with ORA values, except for those of [*Riegler et al.*, 1977] which seem to suffer from a possible bias in this altitude range (they report a very large volume mixing ratio of more than 10 ppm at 100 km that could only be explained by critical atomic oxygen and atomic oxygen concentrations).

We conclude that the ORA UV radiometric channel has produced valuable ozone volume mixing ratio profiles in the upper atmosphere, despite a serious handicap due to its non-monochromaticity. In particular, the retrieved profiles agree satisfactorily with results of HALOE, MLS and SME up to the mesopause. Above this altitude, ORA probably supplies the first large dataset of ozone profiles in the lower thermosphere.

In future work, we will investigate the latitudinal and seasonal ozone variations. Some effort will also be devoted to the error budget estimation for the retrieval of an inhomogeneously distributed constituent in a solar occultation experiment. This could require the use of a specific photochemical model.

Acknowledgments. Two of us (F. V. and C. B.) were funded under a Prodex contract granted by the SSTC-DWTC service of the Belgian Government. This work was also supported by the 'Fonds National de la Recherche Scientifique' under grant 1.5.155.98.

References

Allen, C. W. *Astrophysical Quantities, third edition.* The Athlone Press, London and Dover, New Hampshire, 1985.

Allen, M., J. I. Lunine and Y. L. Yung. The Vertical Distribution of Ozone in the Mesosphere and Lower Thermosphere. *Journal of Geophysical Research*, *89*, 4841–4872, 1984.

Brühl, C., S. R. Drayson, J. M. R. III, P. J. Crutzen, J. M. McInerney, P. N. Purcell, H. Claude, H. Gernandt, T. J. McGee, I. S. McDermid and M. R. Gunson. Halogen Occultation Experiment Ozone Channel Validation. *Journal of Geophysical Research*, *101*, 10217–10240, 1996.

Burrows, J. P., A. Richter, A. Dehn, B. Deters and S. Himmelmann. Temperature-dependent cross sections of O3 in the 231-794 nm range recorded with GOME. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *61*, 509–517, 1999.

Fichtelmann, B. and G. Sonnemann. On the Variation of Ozone in the Upper Mesosphere and Lower Thermosphere: A Comparison Between Theory and Observation. *Zeitschrift fur Meteorologie*, *39*, 297–308, 1989.

Fussen, D., E. Arijs, D. Nevejans, F. V. Hellemont, C. Brogniez and J. Lenoble. Validation of the ora spatial inversion algorithm with respect to the stratospheric aerosol and gas experiment II data. *Applied Optics*, *37*, 3121–3127, 1998.

Fussen, D., E. Arijs, D. Nevejans and F. Leclere. Tomography of the Earth's Atmosphere by the Space-Borne ORA Radiometer: Spatial Inversion Algorithm. *Journal of Geophysical Research*, *102*, 4357–4365, 1997.

Keating, G. M., L. S. Chiou and N. C. Hsu. Improved Ozone Reference Models for the COSPAR International Reference Atmosphere. *Advances in Space Resarch*, *18*, (9/10)11–(9/10)58, 1996.

Kurucz, R., I. Furenlid, J. Brault and L. Testerman. *National Solar Observatory No1.* Harvard University, 1984.

National Oceanic and Atmospheric Organization. U. S. Standard Atmosphere. Technical report, National Aeronautics and Space Administration, United States Air Force, Washington DC, 1976.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery. *Numerical Recipes in FORTRAN, Second Edition.* Cambridge University Press, Cambridge, 1992.

Riegler, G. R., S. K. Atreya, T. M. Donahue, S. C. Liu, B. Wasser and J. F. Drake. UV Stellar Occultation Measurements of Nighttime Equatorial Ozone. *Geophysical Research Letters*, *4*, 145–148, 1977.

Sonnemann, G. and B. Fichtelmann. Subharmonics, cascades of period doubling, and chaotic behavior of photochemistry of the mesopause region. *Journal of Geophysical Research*, *102*, 1193–1203, 1997.

Sonnemann, G., C. Kremp, A. Ebel and U. Berger. A Three-Dimensional Dynamic Model of the Minor Constituents of the Mesosphere. *Atmospheric Environment*, *18*, 3157–3172, 1998.

D. Fussen, F. Vanhellemont, C. Bingen, and S. Chabrillat, Institut d'Aéronomie Spatiale de Belgique, 3, avenue Circulaire, B-1180, Brussels, BELGIUM (e-mail: Didier.Fussen@oma.be)

(Received March 3, 2000; revised June 1, 2000;
accepted August 25, 2000.)

Appendix B

Acronyms

ADC	Analogue to Digital Convertor
ATMOS	Atmospheric Trace Molecule Spectroscopy Instrument
AVHRR	Advanced Very High Resolution Radiometer
BISA	Belgian Institute for Space Aeronomy
BST	British Summer Time
CCD	Charge Coupled Device
CIRA	Cospar International Reference Atmosphere
CFC	Chlorofluorocarbon
CLAES	Cryogenic Limb Array Etalon Spectrometer
COSPAR	Commitee on Space Research
DHS	Data Handling System
DM	Direct Method
DU	Dobson Unit
EPROM	Erasable programmable Read-Only Memory
ESA	European Space Agency
ESOC	European Space Operations Centre
ESTEC	European Space Research and Technology Center
EURECA	European Retrievable Carrier
GMT	Greenwich Mean Time
GOMOS	Global Ozone Monitoring by Occultation of Stars

HITRAN	High Resolution Transmission molecular absorption database
IR	Infrared
ISAMS	Improved Stratospheric and Mesospheric Sounder
ISCCP	International Satellite Cloud Climatology Project
LTE	Local Thermal Equilibrium
NASA	National Aeronautics and Space Administration
NAT	Nitric Acid Trihydrate
NOAA	National Oceanic and Atmospheric Administration
NOPE	Natural orthogonal polynomial expansion
ORA	Occultation Radiometer
PDF	Probability Density Function
PIA	Process Interface Adapter
PSC	Polar Stratospheric Cloud
PSD	Particle Size Distribution
QBE	Quasi-Biennial Oscillation
RAM	Random-Access Memory
RAU	Remote Acquisition Unit
SAGE II	Stratospheric Aerosol and Gas Experiment II
SAM II	Stratospheric Aerosol Measurement II
STP	Standard Pressure and Temperature
SVD	Singular Value Decomposition
TTS	Telemetry and Telecommand Subsystem
UKMO	United Kingdom Meteorological Office
UV	Ultraviolet
WMO	World Meteorological Organization

Appendix C

Nederlandse samenvatting

Het ORA experiment: waarom en hoe?

Het *Occultation Radiometer* (ORA) experiment werd ontworpen in het Belgisch Instituut voor Ruimte-Aëronomie (BIRA) binnen het kader van de atmospherische scheikunde en de studie van aërosols, met een sterke focus op de stratosfeer. Meer specifiek, het experiment had als doel het meten van concentraties van spoorgassen (O_3, NO_2, H_2O) en aërosols in de stratosfeer, alhoewel een bepaald onderdeel toegewijd was aan de meting van ozon-concentraties in de mesosfeer. Het motief voor deze metingen is eenvoudig. Alhoewel tegenwoordige atmosfeermodellen gas-concentraties behoorlijk goed kunnen voorspellen blijven er nog steeds onvolmaaktheden aanwezig, wat er op wijst dat heel wat atmosferische processen nog altijd niet perfect begrepen worden. Bijkomende studies en observaties zijn dus nodig. In het geval van aërosols zijn er een aanzienlijke hoeveelheid vragen die nog beantwoord moeten worden, en nieuwe metingen zijn nodig om meer inzicht te verkrijgen in processen zoals de vorming van aërosols, de daaropvolgende dynamische processen en het atmosferisch transport. Het afgelopen decennium is de aandacht voor deze aërosols aanzienlijk gestegen, omdat de extreem gewelddadige uitbarsting van de Pinatubo-vulkaan een significante invloed heeft gehad op bepaalde spoorgas-concentraties en het globale klimaat.

ORA maakt gebruik van de *occultatie-methode*: in een baan rond de aarde, tijdens opeenvolgende zonsop- en ondergangen, wordt het zonlicht

gemeten dat doorgelaten wordt doorheen de atmosfeer van de aarde. Deze meting gebeurt in 8 kanalen met verschillende golflengten in het ultraviolet/zichtbaar/infrarood spectrale gebied. De golflengte van een kanaal werd specifiek gekozen in een spectraal gebied waar één van de voorheen vermelde stoffen aanzienlijk licht verstrooit of absorbeert. De selectie van de golflengte gebeurt door het gebruik van filters, en het geselecteerde licht wordt gedetecteerd met een combinatie van een optisch systeem en een fotodiode. Gemeten signalen kunnen dan in digitale vorm naar de aarde worden gezonden.

Al bij al kunnen we zeggen dat ORA een relatief eenvoudig experiment is, deels omdat het niet uitgerust is met een *Sun tracker* en een telescoop. De volledige zonneschijf wordt gemeten, terwijl eventuele andere storende lichtbronnen (maan, reflecties op de satelliet, etc.) worden afgeschermd door het gezichtsveld te beperken tot $\pm 2°$. Een gevolg hiervan is dat de *schijnbare* verticale resolutie zeer laag is (meer hierover later).

Gemonteerd op het EURECA-platform, gebouwd in opdracht van het Europees ruimte-agentschap ESA, werd het experiment gelanceerd op 31 juli 1992 met de *Space Shuttle Atlantis*. Tijdens het verblijf in de ruimte heeft ORA gegevens verzameld in een periode van 10 maanden, waarna EURECA terug naar aarde werd gebracht door de *Space Shuttle Endeavour* (24 juli 1993).

Het inversie-probleem

Een gemeten transmissie is het resultaat van lichtstralen die worden verstrooid, geabsorbeerd en gebroken langs hun optisch traject doorheen de atmosfeer. Omdat deze stralen passeren doorheen lagen die zijn gelokaliseerd binnen een groot hoogtebereik, is de transmissie in feite een geïntegreerde meting, die informatie bevat over de extinctie van licht op verschillende hoogten ?. Het afleiden van de totale extinctie op elke hoogte afzonderlijk uit de meting gebeurt met behulp van een *ruimtelijke inversie-methode*.

Deze resulterende totale extinctie zelf bevat bijdragen van de verschillende stoffen die het licht absorberen of verstrooien. Met behulp van een *spectrale inversie-methode* worden deze bijdragen gescheiden van elkaar.

Wanneer men een inversie-probleem aanpakt moet men altijd een onderzoek uitvoeren naar de informatie-inhoud van de metingen in relatie tot wat men hoopt af te leiden uit deze metingen. Meer specifiek, vanwege (1) de onderlinge afhankelijkheid van verschillende metingen of (2) het onder-gedetermineerde karakter (gebrek aan metingen), kunnen problemen i.v.m. de zgn. *nul-ruimte* optreden: slechts een deel van de werkelijkheid kan worden afgeleid, terwijl het andere deel van de oplossing zich bevindt in een mathematische deelruimte, de *nul-ruimte*. Deze laatste component kan nooit afgeleid worden uit een meting, en is dan ook zonder enige betekenis. Fundamenteel gezien zijn er twee remedies: het aantal componenten van de oplossing (een vector) verlagen, of de oplossing regularizeren door het gebruik van *a priori* informatie.

De ruimtelijke inversie-methode

Als we totale extinctie-profielen uit de metingen willen afleiden, dan is er een model vereist dat mathematisch uitdrukt hoe de verschillende fysische processen plaats vinden. Een exacte uitdrukking werd afgeleid, gebaseerd op de extinctie-wet van Beert-Lambert, waarbij de integratie over de zonneschijf (met *solar limb darkening*) en de responsfunctie van het instrument in acht werd genomen. Jammer genoeg bleek dat het geassocieerde inversie-probleem aanleiding gaf tot enorme berekeningstijden, en benaderingen in het model waren dan ook noodzakelijk. De meest kritieke benadering bestaat uit de discretizatie van de zonneschijf in 20 dunne horizontale lagen, waarbij elke laag als een aparte lichtbron fungeert, met een eigen optisch traject doorheen de atmosfeer. Deze paden werden berekend, rekening houdend met refractie, uit de US76-klimatologie.

Het uiteindelijke model kan niet analytisch geïnverteerd worden, is niet lineair en is bovendien slecht geconditioneerd. Dit laatste probleem wordt veroorzaakt door de sterke onderlinge afhankelijkheid van opeenvolgende transmissiemetingen tijdens een occultatie. Niettemin waren we in staat om het probleem op te lossen met een nieuwe methode: *Natural Orthogonal Polynomial Expansion* (NOPE). Deze methode combineert regularizatie door middel van a priori informatie met de reductie van het aantal onbekenden. De grote hoeveelheid (Pinatubo) aërosols in de stratosfeer leidde tot een bijkomend probleem (a priori extinctie is niet gekend), en een speciale *mapping*-methode was vereist.

De inversie van alle ORA-metingen nam ongeveer 3 maanden in beslag. De resulterende totale extinctie-profielen waren realistisch, en toonden aan dat de echte verticale resolutie veel beter was dan verwacht. Niettemin was het model niet in staat om de metingen te fitten binnen de (zeer kleine!) onzekerheid, afkomstig van de signaal-digitalisatie. Typisch was het inversie-residu 10 keer groter dan de digitalisatie-onzekerheid (ongeveer 10^{-4}). Het is zeer waarschijnlijk dat deze residu's symptomen zijn van de fouten in het model (onzekerheden op modelparameters, de gebruikte benaderingen, bijkomende fysische processen die verwaarloosd werden).

Om na te gaan of de bekomen resultaten geldig zijn, werden de totale extinctie-profielen van het ORA experiment vergeleken met profielen van het SAGE II experiment voor 25 bijna-coincidente occultaties. Over het volledige golflengte- en hoogtebereik werd een typisch relatief verschil van 20 % gevonden. Rekening houdend met het feit dat de 25 occultaties niet perfect overlappen in tijd en plaats kunnen we stellen dat dit verschil aanvaardbaar is.

Op zeer kleine en grote hoogten wordt de gemeten transmissie respectievelijk nul en één. Dit betekent dat de informatie-inhoud zeer klein is, met als gevolg dat de totale extinctie in de geassocieerde gebieden nog altijd behoorlijk chaotisch is. Met de bedoeling om de profielen te stabiliseren werd een Directe Methode ontwikkeld, die gebaseerd is op het gebruik van statistische informatie, afgeleid van de volledige set van ORA-transmissiemetingen, gecombineerd met de NOPE totale extinctie-profielen. De verkregen resultaten stelden ons bovendien in staat om de echte verticale resolutie van het experiment te schatten op ongeveer 2 km in hoogte.

De spectrale inversie-methode

Een totale extinctie-profiel is opgebouwd uit partiële extincties, veroorzaakt door de verschillende atmosferische stoffen. De scheiding van deze bijdragen kan gebeuren wanneer we alle spectrale kanalen tegelijk gebruiken. De spectrale absorptie-karakteristieken van spoorgassen zijn op voorhand gekend: deze worden gegeven door de werkzame doorsnede voor absorptie. Dit is niet het geval voor aërosols: het spectrale gedrag voor verstrooiing is sterk bepaald door de grootte van de aanwezige deeltjes, die op voorhand niet gekend is. Daarom werd besloten om

verstrooiing door aërosols te beschrijven als een kwadratische functie van golflengte, met het belangrijke voordeel dat de spectrale inversie hierdoor lineair wordt.

Een inversie-methode werd ontworpen die gebaseerd is op *Bayesi-aanse* principes: gebruik makend van het formalisme van waarschijnlijkheidsdichtheid, samen met het theorema van Bayes, is het mogelijk om a priori informatie te incorporeren op een zeer elegante manier.

Voor de eigenlijke inversie van start ging hebben we besloten om de Rayleigh-verstrooiing te verwijderen door gebruik te maken van meteorologische data van UKMO (United Kingdom Meteorological Office). Door dit te doen werd ontdekt dat een behoorlijke hoeveelheid verstrooid licht in de ORA-modules was binnengekomen, wat natuurlijk een onderschatting van de totale extinctie veroorzaakte. Het was dan ook nodig om deze bijkomende contributie in rekening te nemen.

De spectrale inversie werd uitgevoerd door toepassing van de Bayesiaanse methode, gecombineerd met a priori informatie van klimatologieën en een beperking van variabiliteit in hoogte. Resultaat: concentratieprofielen voor O_3 en NO_2, en aërosol extinctie-profielen voor 6 golflengten.

Deze resultaten werden vergeleken voor de 25 coincidente occultaties van ORA en SAGE II, wat aantoonde dat de relatieve verschillen niet klein zijn, maar toch aanvaardbaar. Dit kwam niet als een verrassing, daar deze verschillen alreeds aanwezig waren bij de vergelijking voor de totale extinctie-profielen.

Aërosols

De periode waarin ORA heeft gemeten is uitermate interessant omdat de stratosferische Pinatubo-aërosols nog volop aanwezig zijn. Ons onderzoek was dan ook voornamelijk gericht op de studie van deze aërosols, en de geassocieerde dynamische processen.

Een eerste exploratie van de latitude-tijdsvariatie van de 1013 nm optische dikte toonde dat de resultaten, afkomstig van de ORA en SAGE II experimenten, zeer gelijkend zijn. De optische thickness dikte is niettemin een geïntegreerde grootheid (aërosol-extinctie, geïntegreerd over hoogte). Informatie betreffende de grootte en de concentratie van de aërosols kan alleen verkregen worden uit de 6 aërosol extinctie-profielen door de toepassing van een derde inversie: het *radiale inversie-probleem.*

De grootte en concentratie van aërosols kan uit extinctiemetingen afgeleid worden door het gebruik van een theoretisch optisch model: de *Mie-theorie*. Deze theorie is exact voor sferische deeltjes (zoals de stratosferische aërosoldruppels), en maakt gebruik van de refractie-index van de aërosols ($N = 1.43 + \mathrm{i}\,0.01$ voor H_2SO_4/H_2O-deeltjes met een massafractie voor H_2SO_4 van 65 tot 80 %).

In werkelijkheid bestaat een aërosolpopulatie uit deeltjes waarvan de afmetingen een groot dynamisch bereik innemen. Meestal wordt aangenomen dat de deeltjesdistributie kan benaderd worden met een log-normale functie. Een brede distributie heeft de neiging om het spectrale gedrag van aërosolverstrooiing glad te strijken, en veroorzaakt aldus een daling van de informatie-inhoud van gemeten extincties. Dit is jammer, want deze informatie is van nature al zeer gelimiteerd: voor zeer kleine of grote deeltjes kunnen de drie log-normale parameters ($\mathcal{N}$, r_m en σ) niet apart afgeleid worden.

Een radiaal inversieprobleem is dan ook zeer slecht geconditioneerd, en men moet (alweer) beroep doen op een adequate methode. Onze methode is gebaseerd op de simulatie van de ganse (mathematische) meetruimte vanuit het model, een *constraint* langs de hoogte, en een niet-lineaire optimalisatie. De resultaten: hoogteprofielen voor $\mathcal{N}$, r_m en σ. Typische waarden voor de lagere regionen in de stratosfeer (Jungelaag) zijn: $\mathcal{N} = 4$ deeltjes cm^{-3}, $r_\mathrm{m} = 0.5$ μm en $\sigma = 0.2$.

Een eerste onderzoek van deze parameters suggereerde dat coagulatie het dominante proces was in de dynamiek van aërosols in de onderste stratosferische lagen: terwijl de deeltjesdichtheid kleiner werd met de tijd, werden de deeltjes duidelijk groter. Op grote hoogte werd de omgekeerde evolutie waargenomen: deeltjes die kleiner werden naarmate de dichtheid steeg.

Een eenvoudig 1-D model werd opgebouwd, dat rekening hield met sedimentatie, coagulatie, verticale en meridionale advectie. Gecombineerd met de ORA-data werden interessante eigenschappen gevonden. In de onderste stratosferische lagen wordt de neerwaartse sedimentatie compleet overschaduwd door een opwaartse verticale advectie, en het geobserveerde verval van deeltjesdichtheid is voornamelijk veroorzaakt door coagulatie. Hoger in de stratosfeer ($z > 27$ km) wordt de positieve tijdsgradiënt van de deeltjesdichtheid verklaard door het opwaartse transport van deeltjes.

De coagulatie-hypothese in de lagere stratosfeer werd vervolgens getest op een meer nauwkeurige manier door de evolutie van de aërosol-eigenschappen langs een berekend traject te bestuderen. Onder invloed van coagulatie zou normaal gezien de *volumedichtheid* constant moeten blijven, en dit werd inderdaad bevestigd op een zeer overtuigende wijze.

Voor zover we weten uit de literatuur heeft ORA de eerste zonale contourfiguren (latitude vs. hoogte) opgeleverd voor de distributieparameters $\mathcal{N}$, r_{m} en σ. Verschillende interessante kenmerken werden ook hier weer gevonden. In het algemeen kan gesteld worden dat de deeltjesdichtheid daalt van de evenaar naar de polen toe, terwijl de deeltjes zelf groter worden, wat nogmaals wijst op een coagulatieproces. Vooral 2 zones in de extratropen, op een hoogte van 17 km, bevatten relatief grote deeltjes.

Zeer interessant is het feit dat het noordelijk halfrond meer deeltjes lijkt te bevatten als gevolg van een grote stratosferische circulatie-cel met een neerwaartse tak gelokaliseerd tussen 20°N en 30°N. Deze circulatiecel kan misschien ook verantwoordelijk gesteld worden voor een zone op grote hoogte met relatief grote deeltjes en een zeer duidelijke signatuur in de distributiebreedte σ in de bovenste lagen van de stratosfeer.

Niettemin toont een contour-figuur van de volumedichtheid V aan dat het totale volume (of massa) van de aërosols behoorlijk homogeen gemengd is in de horizontale richting, alhoewel alweer een grotere hoeveelheid aërosols zichtbaar is in het noordelijk halfrond.

Bibliography

M. Abramowitz and I.A. Stegun. *Handbook of Mathematical functions*. Dover Publications, Inc., New York, USA, ninth edition, 1972.

M. Ackerman, C. Lippens, and M. Lechevallier. Volcanic material from Mount St. Helens in the stratosphere over Europe. *Nature*, 287:614–615, 1980.

C.W. Allen. *Astrophysical Quantities*. Athlone, London, UK, third edition, 1985.

M.D. Allen and O.G. Raabe. Re-evaluation of Millikan's oil drop data for the motion of small particles in air. *Journal of Aerosol Science*, 13:537–547, 1982.

J. Anderson and V.K. Saxena. Temporal changes of Mount Pinatubo aerosol characteristics over northern midlatitudes derived from SAGE II extinction measurements. *Journal of Geophysical Research*, 101: 19455–19463, 1996.

J.C. Anderson, C. Brogniez, L. Cazier, V.K. Saxena, and M.P. McCormick. Characterization of aerosols from simulated SAGE III measurements applying two retrieval techniques. *Journal of Geophysical Research*, 105:2013–2027, 2000.

E. Arijs, D. Nevejans, D. Fussen, P. Frederick, and E. Van Ransbeeck. Ora Occulation Radiometer on EURECA - instrument description and preliminary results. In *Space Sciences Part 2*, volume 2 of *Space*

Scientific Research in Belgium, chapter 6, pages 61–78. Federal Office for Scientific, Technical and Cultural Affairs, Wetenschapsstraat 8, B-1040 Brussels, September 1994.

E. Arijs, D. Nevejans, D. Fussen, P. Frederick, E. Van Ransbeeck, F.W. Taylor, S.B. Calcutt, S.T. Werett, C.L. Hepplewhite, T.M. Pritchard, I. Burchell, and C.D. Rodgers. The ORA Occultation Radiometer on EURECA - instrument description and preliminary results. In *Advances in Space Research*, volume 16, pages (8)33–(8)36, 1995.

D.R. Bates and M. Nicolet. The photochemistry of atmospheric water vapour. *Journal of Geophysical Research*, 55:301–327, 1950.

C.F. Bohren and D.R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley & Sons, Inc., New York, USA, 1993.

M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, Oxford, UK, sixth edition, 1980.

G. Brasseur. Ozone depletion: Volcanic aerosols implicated. *Nature*, 359:275–276, 1992.

C. Brogniez and J. Lenoble. Modeling of the stratospheric background aerosols from zonally averaged SAGE profiles. *Journal of Geophysical Research*, 92:3051–3060, 1987.

C. Brogniez and J. Lenoble. Analysis of 5-year aerosol data from the Stratospheric Aerosol and Gas Experiment II. *Journal of Geophysical Research*, 96:15,479–15,497, 1991.

C. Brogniez, J. Lenoble, M. Herman, P. Lecomte, and C. Verwaerde. Analysis of two balloon experiments in coincidence with SAGE II in case of large stratospheric aerosol amount: Post-Pinatubo period. *Journal of Geophysical Research*, 101:1541–1552, 1996.

I.R. Burchell. *Water vapour and temperature retrievals from the EURECA Occultation Radiometer*. PhD thesis, University of Oxford, Department of Physics, 1996.

S.B. Calcutt, T.M. Pritchard, C.L. Heppelwhite, F.W. Taylor, S.T. Werret, E. Arijs, and D. Nevejans. A radiometer for the measurement of

water vapour in the upper atmosphere from space - experiment concept and instrument description. *Applied Optics*, 32:6764–6776, 1993.

M.T. Chahine. Determination of the temperature profile in an atmosphere from its outgoing radiation. *Journal of the Optical Society of America*, 58:1634, 1968.

M.T. Chahine. Inverse problems in radiative transfer: A determination of atmospheric parameters. *Journal of the Atmospheric Sciences*, 27: 960, 1970.

S. Chapman. A theory of upper atmospheric ozone. *Memoirs of the Royal Meteorological Society*, 3:103–125, 1930.

W. Choi, W.B. Grant, J.H. Park, K.M. Lee, H. Lee, and J.M.III Russell. Role of the quasi-biennial oscillation in the transport of aerosols from the tropical stratospheric reservoir to midlatitudes. *Journal of Geophysical Research*, 103:6033–6042, 1998.

J.R. Christy and S.J. Drouilhet. Variability in daily, zonal-mean lower-stratospheric temperatures. *Journal of Climate*, 7:106–120, 1994.

W.P. Chu, E.W. Chiou, J.C. Larsen, L.W. Thomason, D. Rind, J.J. Buglia, S. Oltmans, M.P. McCormick, and L.M. McMaster. Algorithms and sensitivity analyses for Stratospheric Aerosol and Gas Experiment II water vapor retrieval. *Journal of Geophysical Research*, 98(D3):4857–4866, 1993.

W.P. Chu and M.P. McCormick. Inversion of stratospheric aerosol and gaseous constituents from spacecraft solar extinction data in the 0.38-1.0 μm wavelength region. *Applied Optics*, 18:1404–1413, 1979.

W.P. Chu, M.P. McCormick, J. Lenoble, C. Brogniez, and P. Pruvost. SAGE II inversion algorithm. *Journal of Geophysical Research*, 94: 8339–8351, 1989.

P.J. Crutzen. The influence of nitrogen oxides on atmospheric ozone content. *Quarterly Journal of the Royal Meteorological Society*, 96: 320–325, 1970.

P.J. Crutzen. The possible importance of OCS for the sulfate layer of the stratosphere. *Geophysical Research Letters*, 3:73–76, 1976.

Daimler-Benz Aerospace. EURECA user guide. CD-ROM, April 1996.

R. De Beule. *Het Europees Ruimteavontuur*. CODA, Antwerpen, Belgie, 1993.

T. Deshler, D.J. Hofmann, B.J. Johnson, and W.R. Rozier. Balloonborne measurements of the Pinatubo aerosol size distribution and volatility at Laramie, Wyoming during the summer of 1991. *Geophysical Research Letters*, 19:199–202, 1992.

T. Deshler, B.J. Johnson, and W.R. Rozier. Balloonborne measurements of Pinatubo aerosol during 1991 and 1992 at 41° N: Vertical profiles, size distribution and volatility. *Geophysical Research Letters*, 20:1435–1438, 1993.

E.G. Dutton and J.R. Christy. Solar radiative forcing at selected locations and evidence for global lower tropospheric cooling following the eruptions of El Chichón and Pinatubo. *Geophysical Research Letters*, 19:2313–2316, 1992.

G. Echle, T. von Clarmann, and H. Oelhaf. Optical and microphysical parameters of the Mt. Pinatubo aerosol as determined from MIPAS-B mid-IR limb emission spectra. *Journal of Geophysical Research*, 103: 19,193–19,211, 1998.

B. Edlen. The dispersion of standard air. *Journal of the Optical Society of America*, 43:339–345, 1953.

Eureca Users handbook - System Summary. ESA, March 1987.

J.C. Farman, B.G. Gardiner, and J.D. Shanklin. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, 315:207–210, 1985.

O.R. Flöck and M.O. Andreae. Photochemical and non-photochemical formation and destruction of carbonyl sulfide and methyl sulfide in ocean waters. *Marine Chemistry*, 54:11–26, 1996.

N.A. Fuchs. *Mechanics of Aerosols*. Pergamon, New York, USA, 1964.

D. Fussen. NOPE: a new inversion method for the total attenuation profile retrieval in atmospheric tomography from space-borne instruments. In Richard P. Santer, editor, *Atmospheric Sensing and Modelling II*, volume 2582 of *Proceedings of SPIE*, pages 80–87, December 1995.

D. Fussen, E. Arijs, D. Nevejans, and F. Leclere. Tomography of the Earth's atmosphere by the space-borne ORA radiometer: spatial inversion algorithm. *Journal of Geophysical Research*, 102:4357–4365, 1997.

O.K. Garriott. Visual observations from space. *Journal of the Optical Society of America*, 69:1064–1068, 1979.

J.F. Gleason, P.K. Bhartia, J.R. Herman, et al. Record low global ozone in 1992. *Science*, 260:523–526, 1993.

G.H. Golub and C.F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, Maryland, USA, third edition, 1996.

R.G. Grainger, A. Lambert, F.W. Taylor, J.J. Remedios, C.D. Rodgers, M. Corney, and B.J. Kerridge. Infrared absorption by volcanic stratospheric aerosols observed by ISAMS. *Geophysical Research Letters*, 20:1283–1286, 1993.

W.B. Grant, E.V. Browell, C.S. Long, L.L. Stowe, R.G. Grainger, and A. Lambert. Use of volcanic aerosols to study the tropical stratospheric reservoir. *Journal of Geophysical Research*, 101:3973–3988, 1996.

J. Hansen, A. Lacis, R. Ruedy, and M. Sato. Potential climate impact of the Mount Pinatubo eruption. *Geophysical Research Letters*, 19: 215–218, 1992.

J.E. Harries. *Earthwatch - the climate from space*. Praxis Publishing Ltd., Chichester, West Sussex, England, 1994.

J.R. Herman, R. McPeters, R. Stolarski, D. Larko, and R. Hudson. Global average ozone change from November 1978 to May 1990. *Journal of Geophysical Research*, 96:17297–17305, 1991.

M.H. Hitchman, M. McKay, and C.R. Trepte. A climatology of stratospheric aerosol. *Journal of Geophysical Research*, 99:20,689–20,700, 1994.

D.J. Hofmann, S.J. Oltmans, S. Harris, S. Solomon, T. Deshler, and B.J. Johnson. Observations and possible causes of new ozone depletion in Antarctica in 1991. *Nature*, 359:283–287, 1992.

J.R. Holton, P.H. Haynes, M.E. McIntyre, A.R. Douglass, R.B. Rood, and L. Pfister. Stratosphere-troposphere exchange. *Reviews of Geophysics*, 33:403–439, 1995.

B.G. Hunt. Photochemistry of ozone in a moist atmosphere. *Journal of Geophysical Research*, 71:1385–1398, 1966.

L. Innocenti. EURECA-1 science operations. Mission report, ESA, August 1993.

H.S. Johnston. Reduction of stratospheric ozone by nitrogen oxide catalysts from supersonic transport exhaust. *Science*, 173:517–522, 1971.

C.E. Junge, C.W. Chagnon, and J.E. Manson. Stratospheric aerosols. *Journal of Meteorology*, 18:80–108, 1961.

G.S. Kent, P.H. Wang, M.P. McCormick, and K.M. Skeens. Multiyear Stratospheric Aerosol and Gas Experiment II measurements of upper tropospheric aerosol characteristics. *Journal of Geophysical Research*, 100:13,875–13,899, 1995.

M.D. King, D.M. Byrne, B.M. Herman, and J.A. Reagan. Aerosol size distributions obtained by inversion of spectral optical depth measurements. *Journal of the Atmospheric Sciences*, 35:2153–2167, 1978.

D.E. Kinnison, K.E. Grant, P.S. Connell, D.A. Rotman, and D.J. Wuebbles. The chemical and radiative effects of the Mount Pinatubo eruption. *Journal of Geophysical Research*, 99:25,705–25,731, 1994.

T. Koyaguchi and M. Tokuno. Origin of the giant eruption cloud of Pinatubo, June 15, 1991. *Journal of Volcanology and Geothermal Research*, 55:85–96, 1993.

A. S. Koziol and J. PudyKiewicz. High-resolution modeling of size-resolved stratospheric aerosol. *Journal of the Atmospheric Sciences*, 55:3127–3147, 1998.

T.G. Kyle. *Atmospheric Transmission, emission and scattering*. Pergamon Press, Oxford, UK, 1993.

K. Labitzke and M.P. McCormick. Stratospheric temperature increases due to Pinatubo aerosols. *Geophysical Research Letters*, 19:207–210, 1992.

J. Lenoble. Presentation of the European correlative experiment program for SAGE II. *Journal of Geophysical Research*, 94:8395–8398, 1989.

J. Lenoble. *Atmospheric Radiative Transfer - Studies in Geophysical Optics and Remote Sensing*. A. Deepak Publishing, Hampton, Virginia, USA, 1993.

C.S. Long and L.L. Stowe. Using the NOAA/AVHRR to study stratospheric aerosol optical thicknesses following the Mt. Pinatubo eruption. *Geophysical Research Letters*, 21:2215–2218, 1994.

J.D. Lumpe, C.S. Chang, and D.J. Strickland. Atmospheric constituent density profiles from full disk solar occultation experiments. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 46:483–506, 1991.

L.E. III Mauldin, N.H. Zaun, M.P. McCormick, J.H. Guy, and W.R. Vaughn. Stratospheric Aerosol and Gas Experiment II instrument: a functional description. *Opt. Eng.*, 24:307–312, 1985.

M.P. McCormick and J.E. Lovill, editors. *Space Observations of Aerosols and Ozone*, volume 2(5) of *Advances in Space Research*, Oxford (UK), NY (US), 1982. COSPAR, Pergamon Press.

M.P. McCormick and R.E. Veiga. SAGE II measurements of early Pinatubo aerosols. *Geophysical Research Letters*, 19:155–158, 1992.

M.B. McElroy, R.J. Salawitch, S.C. Wofsy, and J.A. Logan. Reductions of Antarctic ozone due to synergestic interactions of chlorine and bromine. *Nature*, 321:759–762, 1986.

J.L. Mergenthaler, J.B. Kumer, and A.E. Roche. CLAES observations of Mt. Pinatubo stratospheric aerosol. *Geophysical Research Letters*, 22:3497–3500, 1995.

O. Minster, L. Innocenti, and D. Mesland. Looking at science on board EURECA. Bulletin BR-80, ESA, Noordwijk, The Netherlands, May 1993.

K.C. Mo and X. Wang. The global climate of June-August 1992: Warm ENSO episode decays and colder than normal conditions dominate the northern hemisphere. *Journal of Climate*, 7:335–357, 1994.

M.J. Molina. Heterogeneous chemistry on polar stratospheric clouds. *Atmospheric Environment*, 25A:2535–2537, 1991.

M.J. Molina and F.S. Rowland. Stratospheric sink for chlorofluoromethanes: chlorine atom-catalyzed destruction of ozone. *Nature*, 249:810–812, 1974.

R. Monastersky. Climate still reeling after Pinatubo blast. *Science News*, 145:70, 1994.

M. Mozurkewich and J.G. Calvert. Reaction probability of N_2O_5 on aqueous aerosols. *Journal of Geophysical Research*, 93:15889–15896, 1988.

D.M. Murphy, D.W. Fahey, M.H. Proffitt, S.C. Liu, K.R. Chan, C.S. Eubank, S.R. Kawa, and K.K. Kelly. Reactive nitrogen and its correlation with ozone in the lower stratosphere and the upper troposphere. *Journal of Geophysical Research*, 98:8751–8773, 1993.

National Oceanic and Atmospheric Organization. U.S. Standard Atmosphere. Technical report, National Aeronautics and Space Administration, Washington D.C., 1976.

D. Nevejans, E. Arijs, and D. Fussen. The electronic and software design of the ORA occultation radiometer. In V.M. Balebanov, editor, *Manufacturing of scientific space instrumentation*, Proceedings of the IVth international seminar, pages 61–73, Frunze, USSR, 1990.

K.F. Palmer and D. Williams. Optical constants of sulfuric acid: Application to the clouds of Venus? *Applied Optics*, 14:208–219, 1975.

M.J. Prather. Catastrophic loss of stratospheric ozone in dense volcanic clouds. *Journal of Geophysical Research*, 97:10187–10191, 1992.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in FORTRAN - The Art of scientific computing*. Cambridge University Press, second edition, 1992.

P. Pruvost, J. Ovarlez, J. Lenoble, and W.P. Chu. Comparison of Stratospheric Aerosol and Gas Experiment II and balloon-borne stratospheric water vapor measurements. *Journal of Geophysical Research*, 98(D3):4889–4896, 1993.

R.F. Pueschel, P.B. Russell, D.A. Allen, G.V. Ferry, K.G. Snetsinger, J.M. Livingston, and S. Verma. Physical and optical properties of the Pinatubo volcanic aerosol: aircraft observations with impactors and a Sun-tracking photometer. *Journal of Geophysical Research*, 99: 12,915–12,922, 1994.

R.F. Pueschel, K.G. Snetsinger, P.B. Russell, S.A. Kinne, and J.M. Livingston. The effects of the 1991 Pinatubo volcanic eruption on the optical and physical properties of stratospheric aerosols. In S. Keevalik and A. Deepak, editors, *IRS92: Current Problems in Atmospheric Radiation*, 1992.

R.G. Roble and P.B. Hays. A technique for recovering the vertical number density profile of atmospheric gases from planetary occultation data. *Planetary and Space Science*, 20:1727–1744, 1972.

A. Robock. Volcanic eruptions and climate. *Reviews of Geophysics*, 38,2: 191–219, 2000.

A.E. Roche, J.B. Kumer, J.L. Mergenthaler, G.A. Ely, W.G. Uplinger, J.F. Potter, T.C. James, and L.W. Sterrit. The Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS: Experiment description and performance. *Journal of Geophysical Research*, 98:10,763–10,775, 1993.

C.D. Rodgers. Characterization and error analysis of profiles retrieved from remote sounding measurements. *Journal of Geophysical Research*, 95:5587–5595, 1990.

C.D. Rodgers. *Inverse Methods for Atmospheric Sounding - Theory and Practice*. World Scientific, first edition, 2000.

F.S. Rowland and M.J. Molina. Chlorofluoromethanes in the environment. *Reviews of Geophysics and Space Physics*, 13:1–35, 1975.

P.B. Russell and P. Hamill. Spatial variation of stratospheric aerosol acidity and model refractive index: Implications of recent results. *Journal of the Atmospheric Sciences*, 41:1781–1790, 1984.

P.B. Russell, J.M. Livingston, R.F. Pueschel, J.J. Bauman, J.B. Pollack, S.L. Brooks, P. Hamill, L.W. Thomason, L.L. Stowe, T. Deshler, E.G. Dutton, and R.W. Bergstrom. Global to microscale evolution of the Pinatubo volcanic aerosol derived from diverse measurements and analyses. *Journal of Geophysical Research*, 101:18,745–18,763, 1996.

R.J. Salawitch. Chemical loss of ozone in the Arctic polar vortex in the winter of 1991-1992. *Science*, 261:1146–1149, 1993.

W.E. Scott, R.C. Torres, S. Self, M.L. Martinez, and T. Jr Nillos. Pyroclastic flows of the June 15, 1991, climactic eruption of Mount Pinatubo. In C.G. Newhall and R.S. Punongbayan, editors, *Fire and Mud: eruptions and lahars of Mount Pinatubo, Philippines*. University of Washington Press, Seattle (US) and London (UK), 1996.

J.H. Seinfeld and S.N. Pandis. *Atmospheric Chemistry and Physics - From Air pollution to climate change*. John Wiley & Sons, Inc., New York, USA, 1998.

J. Sekkate. Rapport de stage effectue a l'Institut d'Aeronomie Spatiale de Belgique. Master's thesis, Institut Superieur Industriel de l' Etat, 158 Rue Royale, B-1000 Brussels, June 1988.

S. Self, J. Zhao, R.E. Holasek, R.C. Torres, and A.J. King. The atmospheric impact of the 1991 Mount Pinatubo eruption. In C.G. Newhall and R.S. Punongbayan, editors, *Fire and Mud: eruptions and lahars of Mount Pinatubo, Philippines*. University of Washington Press, Seattle (US) and London (UK), 1996.

P.J. Sheridan, R.C. Schnell, D.J. Hofmann, and T. Deshler. Electron microscope studies of Mt. Pinatubo aerosol layers over Laramie,

Wyoming during summer 1991. *Geophysical Research Letters*, 19:203–206, 1992.

S. Solomon, R.R. Garcia, F.S. Rowland, and D.J. Wuebbles. On the depletion of Antarctic ozone. *Nature*, 321:755–758, 1986.

S.R. Solomon, R.W. Portmann, R.R. Garcia, L.W. Thomason, L.R. Poole, and M.P. McCormick. Role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes. *Journal of Geophysical Research*, 101:6713–6728, 1996.

H.M. Steele and P. Hamill. Effects of temperature and humidity on the growth and optical properties of sulfuric acid-water droplets in the stratosphere. *Journal of Aerosol Science*, 12:517–528, 1981.

H.M. Steele and R.P. Turco. Retrieval of aerosol distributions from satellite extinction spectra using constrained linear inversion. *Journal of Geophysical Research*, 102:16,737–16,747, 1997.

R. Stolarski, R. Bojkov, L. Bishop, C. Zerefos, J. Staehelin, and J. Zawodny. Measured trends in stratospheric ozone. *Science*, 256:342–349, 1992.

R.S. Stolarski and R.J. Cicerone. Stratospheric chlorine: a possible sink for ozone. *Canadian Journal of Chemistry*, 52:1610–1615, 1974.

L.L. Stowe, R.M. Carey, and P.P. Pellegrino. Monitoring the Mt. Pinatubo aerosol layer with NOAA-11 AVHRR data. *Geophysical Research Letters*, 19:159–162, 1992.

R. Swinbank and A. O' Neill. A stratosphere-troposphere data assimilation system. *Monthly Weather Review*, 122:686–702, 1994.

A. Tarantola. *Inverse Problem Theory*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, second edition, 1987.

L.W. Thomason, L.R. Poole, and T. Deshler. A global climatology of stratospheric aerosol surface area density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984-1994. *Journal of Geophysical Research*, 102:8967–8976, 1997.

O.B. Toon and M.A. Tolbert. Spectroscopic evidence against nitric acid trihydrate in polar stratospheric clouds. *Nature*, 375:218–221, 1995.

C.R. Trepte, L.W. Thomason, and G.S. Kent. Banded structures in stratospheric aerosol distributions. *Geophysical Research Letters*, 21: 2397–2400, 1994.

C.R. Trepte, R.E. Veiga, and M.P. McCormick. The poleward dispersal of Mount Pinatubo volcanic aerosol. *Journal of Geophysical Research*, 98:18,563–18,573, 1993.

R.P. Turco, R.C. Whitten, and O.B. Toon. Stratospheric aerosols: Observation and theory. *Reviews of Geophysics and Space Physics*, 20: 233–279, 1982.

S. Twomey. Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions. *Journal of Computational Physics*, 18:188–200, 1975.

S. Twomey. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. Elsevier Scientific Publishing Company, New York, USA, 1985.

F.P.J. Valero and P. Pilewskie. Latitudinal survey of spectral optical depths of the Pinatubo volcanic cloud-derived particle sizes, columnar mass loadings and effects on planetary albedo. *Geophysical Research Letters*, 19:163–166, 1992.

H.C. van de Hulst. *Light scattering by small particles*. Dover Publications, Inc., New York, USA, 1981.

J.M. Van Doren, L.R. Watson, P. Davidovits, D.R. Worsnop, M.S. Zahniser, and C.E. Kolb. Uptake of N_2O_5 and HNO_3 by aqueous sulfuric acid droplets. *Journal of Physical Chemistry*, 95:1684–1689, 1991.

P. Wang, P. Minnis, M.P. McCormick, G.S. Kent, and K.M. Skeens. A 6-year climatology of cloud occurrence frequency from Stratospheric Aerosol and Gas Experiment II observations (1985-1990). *Journal of Geophysical Research*, 101:29,407–29,429, 1996.

C.R. Webster. Chlorine chemistry on polar stratospheric cloud particles in the Arctic winter. *Science*, 261:1130–1134, 1993.

D.K. Weisenstein, G.K. Yue, M.K.W.K.N.D. Sze, J.M. Rodriguez, and C.J. Scott. A two-dimensional model of sulfur species and aerosols. *Journal of Geophysical Research*, 102:13019–13035, 1997.

S.C. Wofsy and M.B. McElroy. HO_x, NO_x and ClO_x: Their role in atmospheric photochemistry. *Canadian Journal of Chemistry*, 52:1582–?, 1974.

World Meteorological Organization. Scientific assessment of ozone depletion: 1998. Report 44, WMO Global Ozone Research and Monitoring Project, Geneva, 1999.

S. Young. STS-46: Tethered satellite mission ends in disappointment. *Space Flight*, 34:278–279, 1992.

G.K. Yue. Wavelength dependence of aerosol extinction coefficient for stratospheric aerosols. *Journal of Climate and Applied Meteorology*, 25:1775–1779, 1986.

G.K. Yue, M. P. McCormick, W.P. Chu, P. Wang, and M. T. Osborn. Comparative studies of aerosol extinction measurement made by the SAM II and SAGE II satellite experiments. *Journal of Geophysical Research*, 94:8412–8424, 1989.

J. Zhao, R.P. Turco, and O.B. Toon. A model simulation of Pinatubo volcanic aerosols in the stratosphere. *Journal of Geophysical Research*, 100:7315–7328, 1995.

Published works — ORA

Etienne Arijs, Dennis Nevejans, Didier Fussen, Piet Frederick, and Emiel Van Ransbeeck. Ora Occulation Radiometer on EURECA - instrument description and preliminary results. In *Space Sciences Part 2*, volume 2 of *Space Scientific Research in Belgium*, chapter 6, pages 61–78. Federal Office for Scientific, Technical and Cultural Affairs, Wetenschapsstraat 8, B-1040 Brussels, September 1994.

Etienne Arijs, Dennis Nevejans, Didier Fussen, Piet Frederick, Emiel Van Ransbeeck, Fred W. Taylor, Simon B. Calcutt, S.T. Werett, Chris L. Hepplewhite, T.M. Pritchard, Ian Burchell, and Clive D. Rodgers. The ORA Occultation Radiometer on EURECA - instrument description and preliminary results. In *Advances in Space Research*, volume 16(8), pages (8)33–(8)36, 1995.

Didier Fussen. NOPE: a new inversion method for the total attenuation profile retrieval in atmospheric tomography from space-borne instruments. In Richard P. Santer, editor, *Atmospheric Sensing and Modelling II*, volume 2582 of *Proceedings of SPIE*, pages 80–87, December 1995.

Didier Fussen, Etienne Arijs, Dennis Nevejans, and Fabienne Leclere. Tomography of the Earth's atmosphere by the space-borne ORA radiometer: spatial inversion algorithm. *Journal of Geophysical Research*, 102:4357–4365, 1997.

Didier Fussen, Etienne Arijs, Dennis Nevejans, Filip Vanhellemont, Colette Brogniez, and Jacqueline Lenoble. Validation of the ORA spatial

inversion algorithm with respect to the Stratospheric Aerosol and Gas Experiment II data. *Applied Optics*, 37(15):3121–3127, May 1998.

Didier Fussen, Filip Vanhellemont, and Christine Bingen. Evolution of stratospheric aerosols in the post-pinatubo period measured by the Occultation Radiometer ORA. Submitted to Annales Geophysicae, March 2000a.

Didier Fussen, Filip Vanhellemont, and Christine Bingen. Ozone profiles from 30 to 110 km measured by the Occultation Radiometer instrument during the period Aug. 1992 - Apr. 1993. *Geophysical Research Letters*, 27(21):3449–3452, 2000b.

Didier Fussen, Filip Vanhellemont, and Christine Bingen. Remote sensing of the Earth's atmosphere by the spaceborne occultation radiometer, ORA: final inversion algorithm. *Applied Optics*, 40(6):941–948, 2001.

Dennis Nevejans, Etienne Arijs, and Didier Fussen. The electronic and software design of the ORA occultation radiometer. In V.M. Balebanov, editor, *Manufacturing of scientific space instrumentation*, Proceedings of the IVth international seminar, pages 61–73, Frunze, USSR, 1990.

Jaouad Sekkate. Rapport de stage effectue a l'Institut d'Aeronomie Spatiale de Belgique. Master's thesis, Institut Superieur Industriel de l' Etat, 158 Rue Royale, B-1000 Brussels, June 1988.

Filip Vanhellemont and Didier Fussen. Derivation of stratospheric aerosol properties from the ORA experiment. Submitted to and accepted by Advances in Space Research, 2001.

Filip Vanhellemont, Didier Fussen, and Christine Bingen. Temporal and spatial distribution of the stratospheric aerosol optical thickness at 1013 nm derived from the ORA experiment after the Pinatubo eruption. *Geophysical Research Letters*, 27:1647–1650, 2000.