UNIVERSITÀ DEGLI STUDI DI ROMA

TOR VERGATA

Facoltà di Ingegneria
Dipartimento di Informatica, Sistemi e Produzione

Dottorato di Ricerca
in GeoInformazione

USE OF NEURAL NETWORK ALGORITHMS FOR REMOTE SENSING OF ATMOSPHERIC OZONE

DRAFT

Candidato:
Michele Federico Iapaolo

2006

INDEX

CHAPTER 3
NEURAL NETWORKS

3.1 Introduction
3.2 Multi-layer perceptron architecture
3.1.1 Topology
3.1.2 Training
3.1.3 Generalisation
3.1.3.1 Structural stabilisation
3.1.3.2 Regularisation
3.1.3.3 Early stopping

CHAPTER 4
THE GLOBAL OZONE MONITORING EXPERIMENT
4.1 Introduction
4.2 The instrument
4.3 Operation
4.4 Radiometric errors
4.4.1 Noise
4.4.2 Systematic error
3.4.2.1 Peltier cooler interference
3.4.2.2 Diffuser spectral features
3.4.2.3 Scan mirror degradation
3.4.2.4 Etalon effects
3.4.2.5 Polarisation correction
3.4.2.6 Solar calibration unit
3.4.2.7 Pixel scene
3.4.2.8 Temperature variation
4.5 Geophysical products
4.5.1 Overview
4.5.2 GOME ozone profiles retrieval algoritms
3.5.2.1 Optimal Estimation approach
3.5.2.2 Philips-Tikhonov Regularisation approach
3.5.2.3 Data Assimilation approach
4.6 RAL retrieval scheme
4.6.1 Forward model parameters
4.6.2 Retrieval model parameters
4.6.3 Measurement error: noise-floor
4.6.4 Retrieval analysis
4.6.5 Retrieval performance

CHAPTER 5
NEURAL ALGORITHMS FOR THE RETRIEVAL OF OZONE PROFILES

5.1 Selection of input measurements
5.1.1 Methodology based on physical assumption
5.1.2 Sensitivity analysis and Extended Pruning procedure
5.1.3 Feature extraction
5.2 The inversion phase
5.2.1 Optimisation of the hidden layer
5.2.2 Retrieval results
5.2.3 Retrieval error characterization

CHAPTER 6
VALIDATION OF OZONE PROFILES RETRIEVED BY NEURAL NETWORKS

6.1 Comparison with ILAS products
6.1.1 The instrument
6.1.2 Instrument overview
6.1.2.1 Instantaneous field of view (IFOV) tracking
6.1.2.2 Sun-edge sensor
6.1.2.3 Infrared spectrometer
6.1.2.4 Visible spectrometer
6.1.2.5 Instrument performance
6.1.3 Comparison methodology and results
6.2 Validation with lidar measurements
6.2.1 Principle of ozone lidar measurements
6.2.2 Comparison methodology
6.2.2.1 Selection of the representative GOME co-located profile
6.2.2.2 Transformation of the lidar profile into the GOME resolution
6.2.3 Results
6.3 Conclusion

CHAPTER 7
NNs FOR THE RETRIEVAL OF OTHER ATMOSPHERIC PARAMETERS

CHAPTER 8
CONCLUSIONS AND OUTLOOK

List of publications

This thesis is based on the work contained in the scientific papers reported hereafter:

[1] F. Del Frate, M.F. Iapaolo, S. Casadio, S. Godin-Beekmann, M. Petitdidier
"Neural networks for the dimensionality reduction of GOME measurement vector in the estimation of ozone profiles"
Journal of Quantitative Spectroscopy and Radiative Transfer, Volume 92, Issue 3, 15 May 2005, Pages 275-291

[2] F. Del Frate, M.F. Iapaolo, S. Casadio
"Intercomparison between GOME Ozone Profiles Retrieved by Neural Network Inversion Schemes and ILAS Products"
Journal of Atmospheric and Oceanic Technology: Vol. 22, No. 9, pp. 1433–1440, 2005

[3] M.F. Iapaolo, S.Godin-Beekmann, F.Del Frate, S.Casadio, M.Petitdidier, I.S.McDermid, T.Leblanc, D.Swart,
Y.Meijer, G.Hansen, K.Stebel
"Gome Ozone Profiles Retrieved By Neural Network Techniques: A Global Validation With Lidar Mesurements" (Ms# 06-037 - Received 22 March 2006).

The two following papers are also related to the work presented in this thesis:

[4] F. Del Frate, M.F. Iapaolo, S. Casadio, F. Rossi
"A connectionist approach for atmospheric ozone profiling with satellite radiance measurements"
Atti della Fondazione Giorgio Ronchi, Arcetri-Firenze, La Fondazione, 2004.

[5] G. Schiavon, F. Del Frate, M.F. Iapaolo, D. Solimini
"Microwave radiometry and remote sensing applications: Neural network inversion algorithms for satellite temperatures and humidity profiling"
Atti della Fondazione Giorgio Ronchi, Arcetri-Firenze, La Fondazione, 2005.

Preface

Climate change and ozone depletion represent very current scientific issues; understanding the factors causing these changes and assessing the impact of human activities on atmospheric processes is a challenge for modern Earth sciences. Global and systematic monitoring of the atmospheric composition is a key-point for such scientific studies, and a wide variety of observing systems have been designed to this purpose. Among them, satellite-based instruments have been proven to be very effective in last two decades. The advantage of atmospheric sounding performed from space, with respect to ground based techniques, resides in the very high number of available measurements per day and in the global coverage of the Earth surface, allowing for a detailed investigation of the atmospheric status.
A number of different techniques are at hand, using different instruments and viewing geometries. The major problem of remote sensing techniques, with respect with in-situ measurements, is that the atmospheric parameters are not directly sounded, but have to be retrieved from the measured data, which are intrinsically of indirect nature, as they result from the interaction between the electromagnetic radiation and the atmospheric compounds. The retrieval phase requires the solution of the inversion problem, which is never trivial and can be computationally very intensive, especially for non-linear problems. Moreover, satellite-based instruments generally provide huge amounts of data, and very large data sets (increasing in time) have to be processed. Hence, faster algorithms, leading to real-time monitoring, are necessary. Neural Network retrieval schemes represent here a promising technique to achieve this goal, since they can be trained to perform a very fast non-linear inversion of atmospheric parameters from satellite spectral measurements. In this work the application of Neural Networks for the retrieval of a key atmospheric parameter, such as the ozone vertical distribution, from satellite measurements will be analysed, leading to designing a real-time algorithm for very fast data processing.
The work is presented as follows. Chapter 1 introduces the problem of atmospheric ozone, from the point of view of chemical processes involved in its formation and depletion, and monitoring activities aiming at evaluating its distribution. In Chapter 2 the theory of the inversion problem is formalised, and different methodologies aiming at its solution are presented. In Chapter 3 a brief introduction to artificial neural networks is given, and the basic ideas needed to understand the application of neural NNs to the inversion problems are presented. In the context of regression problems, the type of NN called multi-layer perceptron is discussed in details. Chapter 4 describes the Global Ozone Monitoring Experiment (GOME) instrument, and all retrieval techniques aiming at retrieving the information on the ozone vertical distribution from its measurements. In Chapter 5 the design of neural network inversion methods for the retrieval of ozone profiles from GOME data is presented. The analysis and validation of the obtained results is reported in Chapter 6. In Chapter 7 other activities concerning the application of neural networks for the inversion of atmospheric parameters from remote sensing data are presented. Main conclusions of the investigations described in this thesis are summarised, and an outlook for future research activities is presented.
The study presented in this thesis has resulted in three scientific papers, appended at the end of the document. The work has been financed by the GeoInformation PhD Programme of the Università Tor Vergata of Rome, in collaboration also with the European Space Research Institute of the European Space Agency (ESA-ESRIN), and with the Service d'Aéronomie of the Institut Pierre Simon Laplace in Paris.

Chapter 1

Introduction

1.1 Introduction

Atmospheric ozone has a recognized key-role in absorbing the short solar ultraviolet wavelengths, and its vertical distribution affects significantly the chemical and physical processes that dynamically link the troposphere and the stratosphere. Despite the small amount of ozone in the atmosphere, the solar radiation at wavelengths below 310 nm do not penetrates to the lower atmospheric layers, because of the large absorption cross sections characterizing ozone molecules (Figure 1.1). In this way the ozone layer acts as a filter for the harmful UV solar radiation and contributes to the radiative balance of the upper troposphere and lower stratosphere, influencing the radiative forcing of climate. Furthermore, tropospheric ozone is the source of the hydroxyl radical which controls the abundance and distribution of many atmospheric constituents, including greenhouse gases [1].
In the past decades the public concern regarding the impact of human activities on the Earth's atmosphere significantly growed, and a considerable effort has been made by researchers in order to understand the role of anthropogenic gas emissions in the atmospheric chemistry and its relation with the stratospheric ozone depletion, the global climate changing and the increasing pollution of the troposphere [2]. Nevertheless, the relative importance of factors controlling the abundance of O_3 in the lower atmosphere are currently poorly understood. The possibility to improve the scientific knowledge in such a domain relies on monitoring a large set of atmospheric state parameters. This monitoring activity has been performed for a long time with in situ measurements, which include a multitude of ground-based and balloon-borne techniques providing high-quality data with long time series. These techniques generally provide a good spatial and temporal resolution, but a limited geographical coverage. Despite the large number of stations operating worldwide, their distribution is far from being uniform, and not negligible uncovered areas exist, especially in under-developed regions. Remote sensing measurements from satellites can offer an useful alternative for monitoring purposes, allowing a regular and global coverage with an acceptable spatial and temporal resolution.
While in situ techniques are characterized by a physical contact between the sensor and the atmosphere, the rationale of remote sensing measurements relies on the interaction between atmospheric molecules and electromagnetic waves, and on the capabiliy to extract information from such interaction. To this purpose inversion algorithms are needed, which allow to obtain the information of atmospheric parameters from the sensor measurements.
The second European Remote Sensing Satellite (ERS-2), launched in 1995 by the European Space Agency (ESA), carries the Global Ozone Monitoring Experiment (GOME) instrument, whose objective is to measure a range of atmospheric trace constituents with the main interest on global ozone distribution. The GOME sensor is a spectrometer, measuring the solar radiation scattered from the Earth's atmosphere and/or reflected by the Earth's surface in nadir viewing geometry. The objective of this thesis mainly concernes the development of inversion algorithms for the retrieval of atmospheric parameters from satellite measurements, aiming at the same time at demonstrating the potential of observations made by GOME, and future instruments of the same class, to provide valuable global information on the atmospheric state.

1.2 Atmospheric structure and composition

1.2.1 Temperature distribution

The atmosphere is the gas envelope surrounding the Earth. Its major constituents are nitrogen (N_2), oxygen (O_2) and argon (A), but many other species, such as hydrogen (H_2), water vapour (H_2O), ozone (O_3), carbon monoxide (CO) and nitrogen dioxide (NO_2) characterize the chemical composition of the atmosphere and its properties, even if present in much smaller concentration. In Table 1.1 the principal constituents are reported, as well as their abundances.
The distribution of species in the atmosphere is not uniform: different layers, with very distinct chemical and dynamical properties, can be identified on the base of a typical atmospheric temperature profile (Figure 1.2). The lowermost part of the atmosphere, called troposphere, is characterized by a negative temperature gradient with increasing altitude of about 6–10 K/km, mainly determined by the absorption of thermal radiation emitted at the Earth's surface. The extension of the troposphere is approximately 11-km altitude, but large differences occur between high-latitude regions (6-7 km) and tropical regions (18 km). The absorption of Earth's thermal radiation induces a low static stability, leading to convective circulation on timescales of hours to day, which produces a vertically well-mixed region. The atmosphere above the troposphere, up to approximately 50-km altitude, is no longer dominated by the Earth's thermal radiation, but instead the absorption of ultraviolet (UV) solar radiation by ozone determines the vertical temperature structure. In the equilibrium state of the various photo-chemical reaction cycles, the trace gas ozone is more abundant in this region; almost 90% of all atmospheric ozone resides in this region. The resulting positive temperature gradient with increasing altitude implies a large static stability suppressing vertical motions (from here the name of stratosphere), and produces an important horizontal mixing in this region. The region above the stratosphere, up to approximately 85-km altitude, is again characterized by a negative temperature gradient with increasing altitude, as UV absorption by ozone is no longer dominant due to the lower ozone concentrations. The region above the mesosphere is commonly referred to as "upper atmosphere" and is dominated by ionization processes due to the absorption of extreme UV solar radiation by molecular and atomic oxygen. The temperature increases dramatically with altitude and, depending on which process gets the emphasis, the region is called thermosphere or ionosphere. The boundaries at the top of the troposphere, the stratosphere and the mesosphere are called the tropopause, the stratopause and the mesopause, respectively.

1.2.2 Atmospheric composition: the ozone

Despite its small concentration with respect to other gases, the ozone represents a key component in the atmospheric structure. Since almost all the ozone in the atmosphere is present in the stratosphere, this region is usually also referred to as the ozone layer. In Figure 1.3 a tipical ozone vertical distribution is reported. The height of the vertical column above the surface when bringing all the ozone molecules to standard pressure and temperature gives an indication of the thickness of this layer, and is generally measured in Dobson unit (DU); 1 DU corresponds to 0.01 mm of this column. An average layer of about 350–400 DU absorbs all solar radiation with wavelengths below 300 nm and it significantly reduces the UV radiation reaching the Earth's surface in the wavelength range 300–330 nm; in this way it acts as a protection screen for the harmful effects of this radiation, which may cause damage to biological organisms or cause skin cancer to (susceptible) human beings [3]. In the past decades, scientists have revealed significant changes in the extension of the ozone layer, of which the formation of the so-called ozone hole in the springtime above the poles is the most well-known. Small changes in the thickness of the ozone layer are accompanied by large increases in the levels of UV radiation reaching the Earth's surface [4]. In the troposphere the concentrations of ozone are generally very low, and

the main source of ozone is transport from the stratosphere. Nevertheless, near the surface, in the so-called boundary layer, i.e. the turbulent layer of the troposphere reaching from the surface to about 0.2–2 km, ozone is produced in photo-chemical reactions involving some pollutant compounds resulting from human activities, such as nitrogen oxides, hydrocarbons, and carbon monoxide, and is commonly known as photo-chemical smog. High ozone concentrations in the boundary layer are toxic and may cause health problems to human beings [3].

1.2.2.1 Stratospheric ozone

O_3 formation in the stratosphere is initiated by photodissociation of O_2 by solar UV radiation below 242 nm:

$$O_2 + h\nu \rightarrow O + O$$

where $h\nu$ denotes the energy associated with an incident photon. The O atom reacts rapidly with O_2 to form O_3 in the presence of a third body, denoted M, usually N_2 or O_2, which acquires kinetic energy in the process:

$$O + O_2 + M \rightarrow O_3 + M$$

This reaction is the only significant known source of O_3 in atmosphere; if M is not present then O_3 would be formed in unstable vibrational states and rapidly dissociate. O_3 itself dissociates by absorbing UV radiation in the range 240-320 nm, and also reacts with O:

$$O_3 + h\nu \rightarrow O_2 + O$$

$$O + O_3 \rightarrow 2O_2$$

This system is referred to as the Chapman mechanism [5] and leads to a steady-state, i.e. the rate of ozone loss is equal to the rate of ozone formation; the O_3 concentrations characterizing this steady-state are approximately a factor of two higher than observations. The discrepancy is attributed to additional loss cycles, usually driven by an highly reactive (free radical) catalyst, X:

$$X + O_3 \rightarrow XO + O_2$$

$$XO + O \rightarrow X + O_2$$

$$\text{———————}$$

$$O + O_3 \rightarrow 2O_2$$

Important cycles have been identified where X can be HO_x, NO_x [6] or halogen species (in particular Cl and Br) [7]. The relative importance of each cycle depends on the abundance of X and the availability of O; HO_x cycles are particularly important in the upper stratosphere (above 40 km), NO_x in the mid-stratosphere (25-40 km). In the lower stratosphere heterogeneous reactions (i.e. also involving liquid or solid phase) are significant in lowering NO_x, increasing the importance of the HO_x and halogen loss-cycles [8].
The removal of O_x by these cycles is limited by the removal of radicals into relatively stable reservoir species, by reactions such as:

$$ClO + NO_2 + M \rightarrow ClONO_2 + M$$

Under normal circumstances, almost all of the chemically active (stratospheric) chlorine is in the form of HCl or $ClONO_2$. Br compounds are much less abundant in the atmosphere than Cl, but

the related loss cycles are more efficient per molecule due to the much lower fraction in reservoir form [8].
Huge reductions in lower stratospheric O_3 at high latitude are already well known from several decades [4]. The phenomenon have been observed in both hemispheres, more consistently in the Antarctic. The principal cause has been identified in the enhanced release of active chlorine from reservoir species, promoted by heterogeneous multi-phase reaction on polar-stratospheric cloud (PSC) particles [9]. PSC formation requires the very low temperatures (< 200 K) encountered in the polar vortex, when, in the absence of sunlight, the stratospheric air masses cool and descend, developing a circulation mechanism which produces the vortex. The core of the vortex becomes stably isolated from the air outside and the persists until eventual warming and mixing occurs in the spring.
Reactions including the following occur much more readily when the 2nd reactant on the left hand side is absorbed onto the surface of a PSC particle (denoted here by subscript "(s)") [8]:

$$ClONO_2 + HCl_{(s)} \rightarrow Cl_2 + HNO_{3(s)}$$

$$ClONO_2 + H_2O_{(s)} \rightarrow HOCl + HNO_{3(s)}$$

$$HOCl + HCl_{(s)} \rightarrow Cl_2 + H_2O$$

The chlorine species produced in each case all rapidly photolyse, given sun-light in the Spring, releasing Cl which reacts with O_3:

$$Cl + O_3 \rightarrow ClO + O_2$$

Once the ClO concentration reaches a certain level, the following cycle rapidly destroys O_3:

$$ClO + ClO + M \rightarrow ClOOCl + M$$

$$ClOOCl + h\nu \rightarrow Cl + ClOO$$

$$ClOO + M \rightarrow Cl + O_2 + M$$

$$2[\, Cl + O_3 \rightarrow ClO + O_2]$$

$$2O_3 \rightarrow 3O_2$$

The destruction is particularly severe due to the removal of NO_x , sequestered as HNO_3 in PSC particles.

1.2.2.2 Tropospheric ozone
The ozone concentration in the troposphere is not straightforward determined by direct emission from the surface, but by the evolution of the tropospheric chemistry system, modulated by transport within the troposphere, exchange with the stratosphere and deposition at the surface [2]. The chemistry system involving O_3 is mainly controlled by the local distribution of nitrogen oxides (NO_x) and by the oxidant OH, produced from O_3 photolysis and combination with water vapour molecules [8]. In the presence of high concentrations of NO_x, the molecular oxygen (O_2) is oxidated to O_3 by means the following reactions:

$$HO + CO + O_2 \rightarrow HO_2 + CO_2$$

$HO_2 + NO \rightarrow HO + NO_2$

$NO_2 + h\nu \rightarrow NO + O^*$

$O^* + O_2 + M \rightarrow O_3 + M$

In the regions with low concentrations of NO_x the ozone is destroyed by the same radicals which indirectly promote its formation in the presence of high concentration of nitrogen oxides:

$HO + O_3 \rightarrow HO_2 + O_2$

$HO_2 + O_3 \rightarrow HO + 2O_2$

$CO + O_3 \rightarrow CO_2 + O_2$

The tropospheric ozone concentration is therefore determined by the chemical equilibrium:

$[O_3] = k\,[NO_2] / [NO]$

Tropospheric NO_x originates predominantly from fossil fuel combustion [2]. Natural sources are about a factor of 3 lower and include soil emissions and lightning [10]. Although the lifetime of NO_x is short (few hours near the ground, few days at higher altitude), it can can be transported at large distances from localised sources via reservoir species, e.g. peroxyacetylnitrate (PAN). Tropospheric convection is important in transporting NO_x to high altitude and low temperature regions, where both NO_x and PAN are more stable [11].
O_3 production is strongly favoured in urban regions under high-pressure meteorological conditions, since the vertical stability (associated with a temperature inversion) allows the primary pollutants, including NO_x and hydrocarbons, to accumulate. Such conditions typically occur in summer, when photolysis rates are also enhanced. High concentrations of NO_x and VOC are also present in plumes from biomass-burning, which regularly occurs (particularly in tropical regions) during the summer season [2][8]. Halogen compounds have a minor role in tropospheric ozone loss, with the exception of sudden depletion episodes observed in polar spring and attributed to a catalytic cycle involving BrO and heterogeneous chemistry on sea-ice [12].

1.3 Atmospheric ozone measurements

Several methods used to measure atmospheric ozone exist, and some of them are described in this section. This is not intended to be a comprehensive survey but a useful summary to highlight the basic characteristics of the various O_3 measurement techniques.

1.3.1 Ground-based techniques

The total column of O_3 can be measured from the ground using UV-absorption spectroscopy; in particular, the most common instruments are the spectrophotometers developed by Dobson and Brewer respectively. The Dobson instrument is based on the differential absorption method in the UV band, where ozone exhibits strong absorption features. This method exploits the ratio of sunlight intensities at two wavelengths characterized by different ozone absorption properties. The Brewer spectrometer is similar in principle to the Dobson, but the determination of the ozone column amount is obtained from a combination of five wavelengths in the UV spectral range [1]. A analogous instrument is the Systeme d'Analyse par Observation Zenithale (SAOZ), a UV-VIS spectrometer which measures the solar radiation scattered at the zenith angle in the

spectral range 300-600 nm, with a resolution of 0.8 nm [13]. The application of the Umkehr principle to Dobson-Brewer instruments allows to obtain the ozone concentration at 10 different altitude layers [1], giving an estimation of the ozone vertical profile. The height-resolved ozone distribution can be also obtained by mm-wave radiometers [14], which exploit a different spectral range.
Also ozone-sondes can provide in-situ measurements of the ozone profile. This technique is based on the reaction of O_3 with potassium iodide, and provides high precision measures with a good vertical resolution (about 50 m), but is limited as follows [1]:

- the maximum altitude observed is around 30 km
- the stations from which the sondes are launched are limited in number and typically in the Northern-Hemisphere mid-latitudes
- differences in operation between stations and changes in adopted technology limit the long term consistency of the method, especially when data from more than one station have to be compared

In recent years, lidar instruments (which measure the attenuation and delay of a Rayleigh backscattered
laser beam) have provided data of similar quality to the ozone-sondes, spanning an altitude range of approximately 5-45 km. Temporal sampling is potentially better than sonde observations, but generally limited to cloud-free, night-time conditions [15].

1.3.2 Satellite observations

The orbits in which satellite-based ozone-profiling instruments can be successfully employed are limited to Low Earth Orbits (LEOs), with altitudes typically between 600 and 900 km, which generally provide a global coverage of the Earth's surface in few days.
As well as absorbing in the UV, O_3 has vibrational-rotational spectral signature in the mid infra-red (mid-IR) and a pure rotational signature extending into mm-wave spectral range. Remote sensing instruments for measuring atmospheric ozone from space therefore rely on spectroscopic or radiometric techniques, depending on the spectral interval used to perform the measurement.
Active techniques for atmospheric ozone observations, which exploit an actively induced detected signal, are currently restricted to the Differential Absorption Lidar (DIAL) technique [15][16]. Although in recent years several remote sounding satellites have been equipped with lidar instruments, the technology, which is mainly limited by the laser reliability and its power consumption, is still immature for employing ozone DIAL systems from space.
Passive techniques, on the other hand, have been successfully experimented during the last decades; ozone sensing from space can be performed by observing the Earth's natural thermal, solar, lunar and/or stellar radiation in different viewing geometries.
The electromagnetic radiation reflected or emitted from the Earth's atmosphere or at the surface can be measured by looking in the nadir or limb direction (Figure 1.4). Nadir observations generally provide a good horizontal resolution and sampling, but with a limited vertical resolution, which depends on the optical properties of the atmosphere and on the possibility to resolve in altitude the spectral signature. Limb sounders scan the atmosphere in the vertical, viewing tangentially to the surface. Observed radiances are therefore predominantly sensitive to relatively narrow atmospheric layers close to each tangent point, allowing profiles with relatively high vertical resolution to be retrieved. Limitations of the technique are relatively poor horizontal resolution and sampling, and possibly lack of sensitivity to the lower atmosphere, due to obscuration by tropospheric humidity, clouds, aerosols or molecular scattering.
The radiation can also be observed with a direct line of sight (LOS) between the source and the detector, which is done with solar, lunar or stellar occultation techniques (Figure 1.4). The high signal-to-noise ratio allows measurements to be made down to and below the tropopause, with

high vertical resolution. Observations are, however limited by geometrical constraints, such that only a few tens of profiles can be measured per day.

1.3.2.1 Nadir-viewing instruments

The total column of O_3 has been continuosly measured by the series of Total Ozone Mapping Spectrometer (TOMS) instruments since 1978 [17].
The first instrument capable to provide an estimation of the ozone profile, named Backscatter Ultra-Violet (BUV), started its measurements in the early 1970s. It was followed by the Solar Backscatter UV (SBUV, 1978-1984) [18] and by a series of SBUV-2 instruments [19] [20] starting in late 1984 on the operational satellite series of the National Oceanic and Atmospheric Administration (NOAA). This series is now continued as an U.S. national program with SBUV-2 instruments mounted on NOAA's next generation National Polar-orbiting Operational Satellite System (NPOESS). All these instruments perform the spectral measurement of the solar radiation backscattered by the atmosphere at 12 channels in the UV range, and can obtain height resolved ozone profiles by exploiting the large gradient in O_3 optical depths across the Hartley-Huggins bands.
The Global Ozone Monitoring Experiment (GOME), on board the second European Research Satellite (ERS-2) launched by the European Space Agency in 1995, has advanced this technique, providing the continuous spectrum between 240 and 790 nm, i.e. from the UV to the near-infrared (NIR) with relatively-high spectral resolution (~0.2 nm) [21]. The wider spectral range of observation allows to obtain information on the ozone distribution coming from the visible range (Chappuis band) by a Differential Optical Absorption Spectroscopy (DOAS) technique. Also the GOME instrument will be followed by a GOME-2 generation carried on the Meteorological Operational (MetOp) satellite series as part of the European polar orbit satellite system now scheduled to start in 2006 [22]. The Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) on ESA's Environmental Satellite (Envisat) launched in 2002 can be considered as the successor of GOME, providing nadir observations of the solar backscattered radiation from 240 to 2380 nm [23].
In July 2004 the Ozone Monitoring Instrument (OMI) was successfully launched on the Earth Observing System Aura (EOS-Aura) satellite. OMI has similar capabilities for ozone profiling as the GOME and SCIAMACHY instruments, but with a much higher spatial resolution and a daily global coverage [24].

1.3.2.2 Limb-viewing instruments

The Limb Radiance Inversion Radiometer (LRIR) and its successor the Limb Infrared Monitor of the Stratosphere (LIMS) instruments performed thermal emission measurements in the mm-wave and mid-ir spectral range since 1975 [25]. These instruments can provide both day-time and night-time observations, exploiting the O_3 emission lines for the ozone profile retrieval. The same principle has been used by the three instruments launched in 1991 as payload of the NASA Upper Atmosphere Research Satellite (UARS), the Cryogenic Limb Array Etalon Spectrometer (CLAES) [26], the Improved Stratospheric and Mesospheric Sounder (ISAMS) [27] and the Microwave Limb Sounder (MLS) [28]. Also the sub-millimetre radiometer (SMR) [29] on the Swedish experimental satellite Odin, launched in 2001, and the Envisat Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument [30] exploit such a technique to obtain ozone distribution.
In the second half of the 1990s the Improved Limb Atmospheric Spectrometer instruments (ILAS-I and ILAS-II) were launched (on ADEOS-I and –II, respectively), and performed limb spectral observations of the solar scattered radiation [31]. Unfortunately, both only produced data for a short period due to satellite problems. Other sensors have been conceived to perform limb-observations of the solar scattered radiation in the UV, VIS and NIR range, like the Optical Spectrograph and Infrared Imager System (OSIRIS) [29], also carried on the Odin satellite, and

SCIAMACHY on Envisat. Constituent retrieval from such observations is still a challenge, as the radiative transfer problem is much more complicated than that of pure thermal emission.

1.3.2.3 Occultation-viewing instruments

Limb-viewing occultation sensors observations started with the Stratospheric Aerosol and Gas Experiment (SAGE) in 1978 and are still operational [32]. Other instruments involve the second and third Polar Ozone and Aerosol Measurement (POAM-II and POAM-III on the French SPOT-3 and SPOT-4 satellites) [33], the Halogen Occultation Experiment (HALOE on UARS, still operational) [34], but also the previously-mentioned ILAS-I, ILAS-II and SCIAMACHY instruments.
All these instruments view sun-rise and sun-set over the Earth horizon to measure the directly transmitted solar radiation along the line of sight. This technique has the limitation of providing only about 14 sun-rise and 14 sun-set measurement occasions per day, and in a specific latitude band, which changes with season or in time depending on the satellite orbit.
The technique has been recently extended to include lunar (e.g. SCIAMACHY) and stellar occultation, like the Ultraviolet and Visible Imagers and Spectrographic Imagers (UVISI) [35] and the Global Ozone Monitoring by Occultation of Stars (GOMOS) [36].

1.3.2.4 Tropospheric ozone measurements

TOMS observations have been used to deduce the tropospheric ozone column by a number of techniques: e.g. climatological distributions (monthly/seasonal) were obtained by differencing a stratospheric column inferred from SAGE and the total column from TOMS [37]. Another approach is the convective cloud differential (CCD) method, which produces monthly averages by assuming zonal symmetry in the tropical stratospheric ozone column which it derives from measurements over high altitude cloud [38]. Differencing this column from cloud free measurements at the same latitude gives a tropospheric column estimate (limited to the tropics), purely from TOMS data.

1.4 Thesis outline

The main objective of this thesis is to present an inversion technique aiming at the retrieval of ozone profiles from satellite measurements. The research maily focuses on data obtained from the GOME instrument, and relies on a particular inversion methodology based on the use of artificial neural networks (NN). This involves several important aspects, such as the selection of input measurements, the design of the retrieval procedure and the validation of obtained results.
In the following, the content of different chapters will be summarized. In Chapter 2 the theory of the inversion problem is formalized, and different methodologies aiming at its solution are presented. In Chapter 3 a brief introduction to artificial neural networks is given, and the basic ideas needed to understand the application of neural NNs to the inversion problems are presented. In such a context, the type of NN called multi-layer perceptron is discussed in details. In Chapter 4 a detailed overview of the GOME instrument is given, and the principal retrieval methods applied to obtain the information on the ozone distribution are described. In Chapter 5 the design of a neural network inversion algorithms for the retrieval of ozone profiles from GOME data is presented, focusing on all critical issues involved in the design. In Chapter 6, the validation of the retrieval results obtained by the NN algorithms presented in Chapter 5 is reported. The validation activity is performed either with ground-based measurements or satellite observations. In Chapter 7, the main conclusions of the investigations described in this thesis are summarized, and an outlook for future research is given.
The study presented in this thesis was financed by the GeoInformation PhD Programme of the Università Tor Vergata of Rome, in collaboration with the European Space Research Institute of

the European Space Agency (ESA-ESRIN), and with the Service d'Aéronomie of the Institut Pierre Simon Laplace in Paris.

References

[1] Assessment of trends in the vertical distribution of ozone, *WMO SPARC/IO3C/GAW Ozone Research and Monitoring Project Report*, No. 43, 1998.

[2] IPCC Climate Change: The Scientific Basis. Third Assessment Report (TAR), *Cambridge University Press*, Cambridge, UK, 2001.

[3] Atmospheric chemistry and global change

[4] Farman, J.C., Gardiner B.G., Shanklin J.D., Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, *Nature*, 315, 207-210, 1985.

[5] Chapman S., A Theory of upper atmospheric ozone, *Mem. R. Meteorol. Soc*., 3, 103-125, 1930.

[6] Crutzen P.J., The influence of nitrogen oxide on the atmospheric ozone content, *Quart. J. R. Met. Soc*., 96, 320-325, 1970.

[7] Molina M.J. and Rowland F.S., Stratospheric sink for chlorofluoromethanes-chlorine atom catalyzed destruction of ozone, *Nature,*249, 810, 1974.

[8] Seinfeld J.H. and Pandis S.N., Atmospheric chemistry and physics : from air pollution to climate change, *Wiley,* ISBN 0471178160, 1998.

[9] Solomon S., Garcia R.R., Rowland F.S., Wuebbles D.J., On the depletion of Antarctic ozone, *Nature*, 321, 755-758, 1986.

[10] Hauglustaine D.A., Emmons L., Newchurch M., Brasseur G., Takao T., Matsubara K., Johnson J., Ridley B., Stith J., Dye J., On the role of lightning NOx in the formation of tropospheric ozone plumes: a global model perspective, *Journal of Atmospheric Chemistry*, 38 (3), 277-294, 2001.

[11] Singh H.B., Kanakidou M., Crutzen P.J., and Jacob D.J., High-concentrations and photochemical
fate of oxygenated hydrocarbons in the global troposphere, Nature, 378 (6552), 50-54, 1995.

[12] Roscoe H.K., Kreher K., Friess U., Ozone loss episodes in the free Antarctic troposphere, suggesting a possible climate feedback, *Geophysical Research Letters*, 28 (15), 2911-2914, 2001.

[13] Sarkissian A., Vaughan G., Roscoe H.K., Bartlett L.M., O'connor F.M., Drew D.G., Hughes P.A., Moore D.M., Accuracy of measurements of total ozone by a SAOZ groundbased zenith sky visible spectrometer, *Journal of Geophysical Research – Atmospheres*, 102 (d1), 1379-1390, 1997.

[14] D. Maier, N. Kampfer, J. de la Noe,W. Amacher, A. Barcia, P. Baron, B. Barry, G. Beaudin, J. Cernicharo, B. Ellison, J.D. Gallego, M. Gustafsson, A. Karpov, U. Klein, K. Kunzi, J. Louhi,

J. Mallat, D. Matheson, J.R. Pardo, R. Peter, A.V. Raisanen, P. Ricand, R. Siddans, C. Viguerie, and M. Wuthrich., European minor constituent radiometer: a new millimeter wave receiver for atmospheric research, *International Journal of Infrared and Millimeter Waves*, 22 (11), 1555-1575, 2001.

[15] Mc Dermid I.S., Godin S., Walsh T.D., Lidar measurements of stratospheric ozone and intercomparisons and validation, *Appl. Opt.*, 29, 4914-4923, 1990.

[16] Measures, R. M., Laser Remote Sensing, John Wiley, New York, 1984.

[17] McPeters R.D. and Komhyr W.D., Long-term changes in the total ozone mapping spectrometer relative to world primary standard Dobson spectrometer 83, *J. Geophys. Res.*, 96, 2987-2993, 1991.

[18] Mcpeters R.D., Hudson R.D., and Bhartia P.K., The Vertical Ozone Distribution In the Antarctic Ozone Minimum Measured By SBUV, *Geophys. Res. Lett.*, 13 (12), 1213-1216, Suppl. S Nov, 1986.

[19] Miller A.J., Flynn L.E., Hollandsworth S.M., DeLuisi J.J., Petropavlovskikh I.V., Tiao G.C., Reinsel G.C., Wuebbles D.J., Kerr J., Nagatani R.M., Bishop L., and Jackman C.H., Information content of Umkehr and solar backscattered ultraviolet (SBUV-2) satellite data for ozone trends and solar responses in the stratosphere, *Journal of Geophysical Researc-Atmospheres*, 102 (d15), 19257-19263, 1997.

[20] Hilsenrath E., Newman P.A., Cebula R.P., DeCamp P.W., Kelly T.J., Coy L., Ozone change from 1992 to 1993 as observed from SBUV on the ATLAS-1 and ATLAS-2 missions, *Geophysical Research Letters*, 23 (17), 2305-2308, 1996.

[21] J.P. Burrows, M. Weber, M. Buchwitz, V. Rozanov, A. Ladstatter Weissenmayer, A. Richter, R. DeBeek, R. Hoogen, K. Bramstedt, K.U. Eichmann, and M. Eisinger. The global ozone monitoring experiment (GOME): Mission concept and first scientific results. *Journal of the Atmospheric Sciences, Vol 56, Iss 2, pp 151-175*, 1999.

[22] Kerridge B.J.K., Siddans R., Latter B.L., Burrows J.P., Weber M., De Beek R., Aben I., Hartman W., GOME-2 Error Assessment Study, Final Report EUMETSAT Contract, No EUM/CO/01/901/DK, 2002.

[23] Noel S., Bovensmann H., Wuttke M.W., Burrows J.P., Gottwald M., Krieg E., Goede A.P.H., Muller C., Nadir, limb and occultation measurements with SCIAMACHY, *Advances In Space Research,* 29 (11), 1819-1824, 2002.

[24] OMI Ozone Products, Vol. II of OMI Algorithm Theoretical Basis Document, OMI-ATBD-02, V. 2.0, NASA Goddard Space Flight Center, Greenbelt (MD), 2002.

[25] E.E. Remsberg, J.M. Russell, J.C. Gille, L.L. Gordley, P.L. Bailey, W.G. Planet, J.E. Harries, The Validation of NIMBUS-7 LIMS Measurements of Ozone, *Journal of Geophysical Research - Atmospheres,* 89 (nd4), 5161-5178, 1984.

[26] [4] P.L. Bailey, D.P. Edwards, J.C. Gille, L.V. Lyjak, S.T. Massie, A.E. Roche, J.B. Kumer, J.L. Mergenthaler, B.J. Connor, M.R. Gunson, J.J. Margitan, I.S. McDermid, T.J. McGee, Comparison of cryogenic limb array etalon spectrometer (CLAES) ozone observations with

correlative measurements, *Journal of Geophysical Research – Atmospheres*, 101 (d6), 9737-9756, 1996.

[27] B.J. Connor, C.J. Scheuer, D.A. Chu, J.J. Remedios, R.G. Grainger, C.D. Rodgers, F.W. Taylor, Ozone in the middle atmosphere as measured by the improved stratospheric and mesospheric sounder, *Journal of Geophysical Research – Atmospheres*, 101 (d6), 9831-9841, 1996.

[28] L. Froidevaux, W.G. Read, T.A. Lungu, R.E. Cofield, E.F. Fishbein, D.A. Flower, R.F. Jarnot, B.P. Ridenoure, Z. Shippony, J.W. Waters, J.J. Margitan, I.S. McDermid, R.A. Stachnik, G.E. Peckham, G. Braathen, T. Deshler, J. Fishman, D.J. Hofmann, S.J. Oltmans, Validation of UARS microwave limb sounder ozone measurement, *Journal of Geophysical Research – Atmospheres*, 101 (d6), 10017-10060, 1996.

[29] D. Murtagh, U. Frisk, F. Merino, M. Ridal, A. Jonsson, J. Stegman, G. Witt, P. Eriksson, C. Jimenez, G. Megie, J. de la Noe, P. Ricaud, P. Baron, J.R. Pardo, A. Hauchcorne, E.J. Llewellyn, D.A. Degenstein, R.L. Gattinger, N.D. Lloyd, W.F.J. Evans, I.C. McDade, C.S. Haley, C. Sioris, C. von Savigny, B.H. Solheim, J.C. McConnell, K. Strong, E.H. Richardson, G.W. Leppelmeier, E. Kyrola, H. Auvinen, L. Oikarinen, An overview of the Odin atmospheric mission, *Canadian Journal of Physics*, 80 (4), 309-319, 2002.

[30] B.J.K. Kerridge, W.J. Reburn, R. Siddans, S.L. Smith, P.D. Watts, J.J. Remedios, F. Lama, J.J. Barnett, D. Murtagh, J. Stegman, F. Merino, P. Baron, H. Roscoe, D. Hausamann, M. Birk, F. Schreier, B. Schimpf, M. Lopez-Puertas, J.M. Flaud, T. von Clarmann, G.P. Stiller, A. Linden, S. Kellman, M. van Weele, H. Kelder, P. van Velthoven, M. Gauss, I. Isaksen, D. Hauglustaine, C. Clerbaux, and O. Boucher, Definition of Mission Objectives and Observational Requirements for an Atmospheric Chemistry Explorer Mission, ESA Contract 13048/98/NL/GD, Final Report, 2001.

[31] Sasano Y., Suzuki M., Yokota T., Kanzawa H., Improved Limb Atmospheric Spectrometer (ILAS) for stratospheric ozone layer measurements by solar occultation technique, *Geophys. Res. Lett.*, 26, 197-200, 1999.

[32] D.M. Cunnold, W.P. Chu, R.A. Barnes, M.P. Mccormick, R.E. Veiga, Validation of SAGE-II Ozone Measurements, *Journal of Geophysical Research – Atmospheres*, 94 (d6), 8447-8460, 1989.

[33] Lumpe, J. D., R. M. Bevilacqua, K. W. Hoppel, C. E. Randall, POAM III Retrieval Algorithm and Error Analysis, *J. Geophys. Res.*, 107, 4575, 10.1029/2002JD002137, 2002.

[34] C. Bruhl, S.R. Drayson, J.M. Russell, P.J. Crutzen, J.M. McInerney, P.N. Purcell, H. Claude, H. Gernandt, T.J. McGee, I.S. McDermid, M.R. Gunson, Halogen occultation experiment ozone channel validation, *Journal of Geophysical Research – Atmosphere*, 101 (d6), 10217-10240, 1996.

[35] Swartz, W. H., J.-H. Yee, R. J. Vervack Jr., S. A. Lloyd, P. A. Newman, Photochemical ozone loss in the Arctic as determined by MSX/UVISI stellar occultation observations during the 1999/2000 winter, *J. Geophys. Res.*, 107(D20), 8296, doi:10.1029/2001JD000933, 2002.

[36] Bertaux, J. L., G. Megie, T. Widemann, E. Chassefiere, R. Pellinen, E. Kyrölä, S. Korpela, P. Simon, Monitoring of ozone trend by stellar occultations: The GOMOS instrument, *Adv. Space Res.*, Vol. 11(3), 237-242, 1991.

[37] J. Fishman, C.E.Watson, J.C. Larsen, and J.A. Logan, Distribution of Tropospheric Ozone Determined From Satellite Data, *Journal of Geophysical Research – Atmospheres*, 95 (d4), 3599-3617, 1990.

[38] J.R. Ziemke and S. Chandra, Seasonal and interannual variabilities in tropical tropospheric ozone, *Journal of Geophysical Research – Atmospheres*, V104(d17), 21425-21442, 1999.

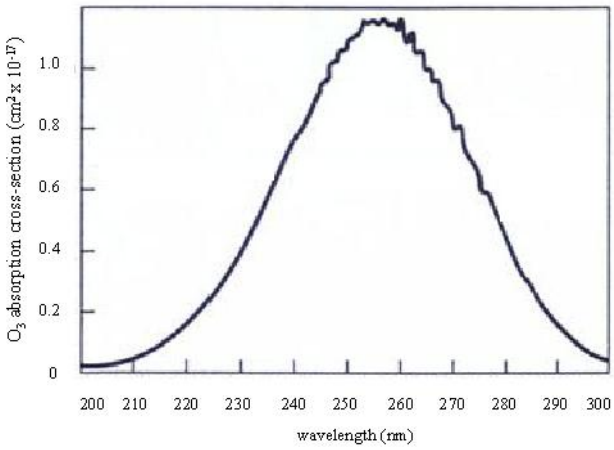

Figure 1.1. Ozone absorption cross section.

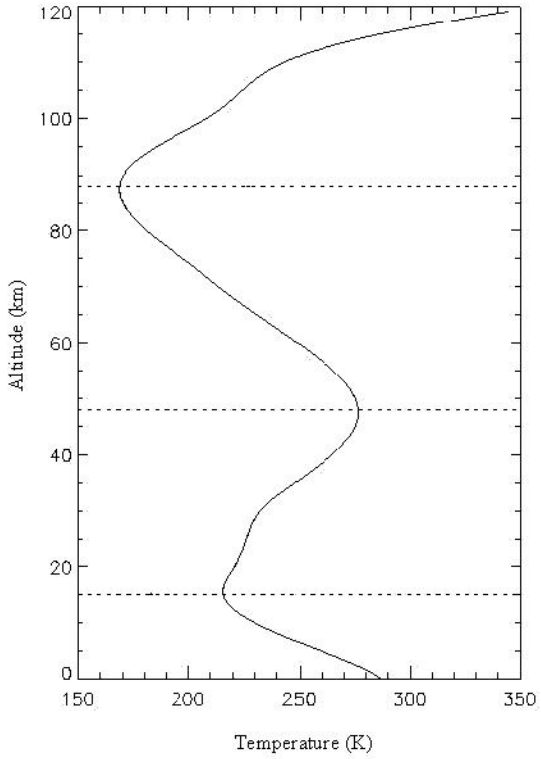

Figure 1.2. Temperature profile.

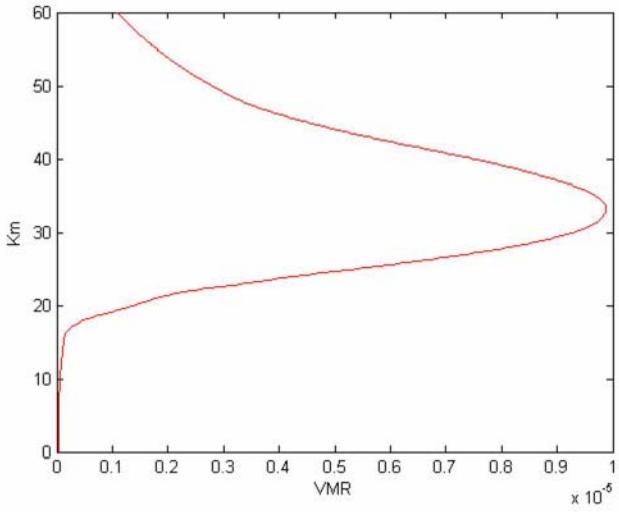

Figure 1.3. Ozone profile.

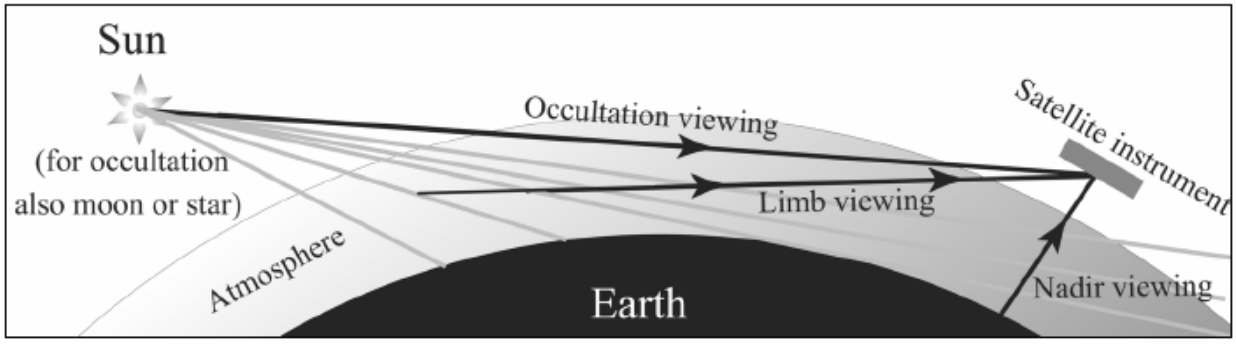

Figure 1.4. Satellite viewing geometries.

N_2	78.08	(%)	Kr	0.00012	(%)
O_2	20.95	(%)	H_2	0.00005	(%)
CO_2	0.03	(%)	CH_4	0.00015	(%)
A	0.93	(%)	NO_2	0.00005	(%)
Ne	0.00182	(%)	H_2O	0 - 4	(ppm)
He	0.00052	(%)	O_3	0 - 7	(ppm)

Table 1.1. Atmospheric composition.

Chapter 2

The inversion problem

The retrieval of atmospheric constituent profiles from spectral measurements is not a well-determined problem, in the sense that more than one atmospheric state may give rise to indistinguishable measurements. This is due to the following reasons: firstly, the experimental errors and the measurement noise are always present, and have to be taken into account; secondly, the atmospheric parameters to be estimated are generally continuous functions, while the set of spectral measurements is discrete. The spatial resolution at which the atmospheric structure can be detected is limited by the observation technique. Therefore, the inversion problem is generally also under-constrained: this class of problem can only be solved by imposing limits (constraints) on the range of allowed atmospheric states, and a number of approaches exist to define these constraints. In this chapter the basic concepts of the inversion theory are given, and different solution methods are presented.

2.1 Inversion problem formalisation

The state of the atmosphere can be represented by a set of parameters (e.g. constituent mixing ratios, temperature, etc.), collected together in a *state vector*, **x**. In defining this vector, continuous variables (e.g. vertical profiles of constituent mixing ratios) are represented in terms of discrete values. Usually these are weights in the linear expansion of the constituent profile in terms of basis functions:

$$x(z) = \sum_{i=0}^{N} x_i B_i(z) \tag{2.1}$$

There are many possible choices for basis functions; among them, the most widely used are the top-hat functions, spanning atmospheric layers (the profile being represented as constant within the layers), and the triangular functions, peaking (with a value equal to 1) at altitude z_i and decaying linearly to 0 at z_{i-1} and z_{i+1}. In Figure 2.1 a tipical triangular function used to represent O_3 vertical profiles is reported.
The set of spectral measurements obtained by a sensor are grouped in a *measurement vector*, **y**. Assuming the existence of a *forward model function*, **F**, which provides an estimate of the measurement vector from a defined state vector and a set of other parameters (grouped in the vector **b**) specifying the radiative transfer calculation and the sensor characteristics, the measurement vector can be expressed as follows:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\varepsilon} \tag{2.2}$$

where **ε** represents the measurement error, i.e. the observed spectral noise.
On the other hand, the *inverse model function*, **I**, permits to calculate an estimation of the state vector, **x`**, from the spectral measurements:

$$\mathbf{x`} = \mathbf{I}(\mathbf{y}, \mathbf{b}, \mathbf{c}, \boldsymbol{\varepsilon}) \tag{2.3}$$

where the vector **c** collects all the auxiliary parameters used to constrain the inversion model. Assuming the measurement error sufficiently small with respect to the measurement vector, the forward model can be linearized around a reference state vector, $\mathbf{x_r}$, and approximated by:

$$\mathbf{F(x, b) = F(x_r, b) + K_x(x\text{-}x_r, b) + K_b(x, b\text{-}b_r)} \tag{2.4}$$

where **K**, known as the *weighting functions matrix*, contains the first derivatives of the forward model **F** with respect to the state vector:

$$\mathbf{K}_x = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \tag{2.5}$$

Also the inverse model can be linearized, defining the *contribution function matrix*, **D**, which contains the partial derivatives of the inverse function with respect to the measurement vector:

$$\mathbf{D}_y = \frac{\partial \mathbf{I}}{\partial \mathbf{y}} \tag{2.6}$$

Using equations (2.2) and (2.4), the problem can now be written in a linear form:

$$\mathbf{y = F(x_r, b) + K_x(x\text{-}x_r, b) + K_b(x, b\text{-}b_r) + \varepsilon} \tag{2.7}$$

or simply:

$$\mathbf{y` = y - F(x_r, b) = K_x\hat{w} + e} \tag{2.8}$$

where $\mathbf{\hat{w} = (x - x_r)}$ represents the optimal estimation of (2.8), and $\mathbf{(\hat{w} + x_r)}$ is the optimal estimate for the original state vector. The measurement error (**ε**) and uncertainties on forward model parameters (**b**) are both taken into account by **e**.
This formalism, taken from [1], [2], [3], is independent from the inversion method actually used. However, the derivation of $\mathbf{D_y}$ and $\mathbf{K_x}$ requires the selection of a particular inversion algorithm, i.e. the forward model parameters, as well as a reference state for the linearization. For non-linear inversion problems, further the reference state is from the true state, the less likely the above mentioned matrices will characterize the solution. The best a priori knowledge of the atmospheric state ($\mathbf{x_a}$) is therefore generally used as the reference state for the linearization.
The difference between the retrieved and the true state vector, i.e. the retrieval error, can be expressed as follows:

$$\mathbf{x`\text{-}x = \delta = (A - I)(x - x_a) + D_y K_b(b - b_a) + D_y \varepsilon} \tag{2.9}$$

where **I** represents the *identity matrix* and **A**, referred to as the *averaging kernels matrix*, is given by the product $\mathbf{D_yK_x}$ and represents the derivatives of the estimated vector **x`** with respect to the true vector **x**. The first term on the right hand side of equation (2.9) is generally known as the *smoothing error*, and originates from the fact that the limited spatial resolution of the inversion method can not reproduce all the features of the true state of the atmosphere, which is naturally not discrete. The second term represents the *forward model error*, and shows how the uncertainties on forward model parameters are mapped into retrieval errors. The last term is called the *measurement error*, and represents the contribution of instrumental noise to the retrieval error.
The smoothing error will always be present, even if we consider an ideal not-noisy sensor and a forward model perfectly reproducing the electromagnetic behaviour of the atmosphere (this error

contribute is intrinsically related to the fact that the inversion problem is ill-posed). Once the measurement and the forward simulation are done, the term

$$\mathbf{e} = \mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b_a}) = \mathbf{K_b}\,(\mathbf{b} - \mathbf{b_a}) + \boldsymbol{\varepsilon} \tag{2.10}$$

can be calculated, and represents the *observation error* [4].
The retrieval error can be therefore expressed as a sum of two terms, the smoothing error and the observation error:

$$\boldsymbol{\delta} = \mathbf{s} + \mathbf{o} = (\mathbf{A} - \mathbf{I})\,(\mathbf{x} - \mathbf{x_a}) + \mathbf{D_y}\,\mathbf{e} \tag{2.11}$$

This error can not be calculated analytically ($\mathbf{x}$ and $\mathbf{b}$ are not known); nevertheless it can be statistically estimated if the covariance matrices characterizing the uncertainties of measurements ($\mathbf{S_\varepsilon}$), forward model parameters ($\mathbf{S_b}$) and retrieved parameters ($\mathbf{S_x}$) are available. The statistical characterization of the observation uncertainties ($\mathbf{S_e}$) can be therefore expressed as:

$$\mathbf{S_e} = \mathbf{K_b}\,\mathbf{S_b}\,\mathbf{K_b}^{\mathbf{T}} + \mathbf{S_\varepsilon} \tag{2.12}$$

the smoothing errors ($\mathbf{S_m}$) and the observation errors ($\mathbf{S_o}$) are then described in a similar way:

$$\mathbf{S_m} = (\mathbf{A} - \mathbf{I})\,\mathbf{S_x}\,(\mathbf{A} - \mathbf{I})^{\mathbf{T}} \tag{2.13}$$

$$\mathbf{S_o} = \mathbf{D_y}\,\mathbf{S_e}\,\mathbf{D_y}^{\mathbf{T}} \tag{2.14}$$

The statistical description of the retrieval error can be finally calculated by:

$$\mathbf{S_\delta} = \mathbf{S_m} + \mathbf{S_o} \tag{2.15}$$

2.2 Solution of the inversion problem

The inversion problem formalised in the above section can be solved using different approaches. As many other estimation problem, the atmospheric parameter retrieval can be placed in the framework of the general Bayesian theory, from which almost all other methods, such as the optimal estimation and regularization, can be derived. In the following section the general concepts of the estimation theory will be given, and the principal estimation methods will be dscribed.

2.2.1 Bayesian estimation

The Bayesian approach solves the inversion problem by determining the *a posteriori* probability distribution function (pdf), $p(\mathbf{x}|\mathbf{y})$, of the state vector $\mathbf{x}$, given the measurement $\mathbf{y}$. Following the Bayes theorem:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \tag{2.16}$$

where $p(\mathbf{x})$ is the a priori probability distribution function of the state $\mathbf{x}$, $p(\mathbf{y}|\mathbf{x})$ is the pdf of the measurement given the state and $p(\mathbf{y})$ is the normalization factor calculated as follows:

$$p(\mathbf{y}) = \int p(\mathbf{y} \mid \mathbf{x}) p(\mathbf{x}) d\mathbf{x} \tag{2.17}$$

Known the a posteriori pdf, the average a posteriori solution can be calculated:

$$\mathbf{x}` = \int \mathbf{x}\, p(\mathbf{x} \mid \mathbf{y}) d\mathbf{x} \tag{2.18}$$

which represents the mean state averaged over the a posteriori pdf. This solution is the one minimising the expected values of the retrieval error variance. A different solution can be found by maximising the a posteriori pdf:

$$\mathbf{x}` = \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\} \tag{2.19}$$

This maximum a posteriori solution is equal to the previous one if the pdf is symmetrical around the mean state; in this case the maximum value of the distribution corresponds to the mean state.

2.2.2 Regularisation

The underlying idea of this approach is to reduce the inversion problem to a set of linear equations [5] [6] [7]:

$$\mathbf{C}\mathbf{x} = \mathbf{y} \tag{2.20}$$

where $\mathbf{x}$ is the n-dimensional state vector, $\mathbf{y}$ the m-dimensional measurement vector and $\mathbf{C}$ the (m x n) kernel matrix. The inversion problem is generally still under-constrained, as the rank of $\mathbf{C}$, p, is smaller than m and n, i.e. the rows of $\mathbf{C}$ are not linearly independent. Moreover, even if a solution is found, the ill-posedness of the problem results in large retrieval errors. The ordinary least square solution, given by:

$$\mathbf{x}` = \min_{\mathbf{x}} \{||\mathbf{C}\mathbf{x} - \mathbf{y}||\} \tag{2.21}$$

where $||\ ||$ is the standard 2-norm, is very sensitive to measurements errors and can not be used in practice. In order to reduce the sensitivity to measurements errors a regularisation function can be used, which results in applying a constrain on the expected solution:

$$\mathbf{x}` = \min_{\mathbf{x}} \{||\mathbf{C}\mathbf{x} - \mathbf{y}|| + \lambda\, ||\mathbf{B}||\} \tag{2.22}$$

where $\mathbf{B}$ represents the imposed constrain and λ is a trade-off factor between the two terms, $||\mathbf{C}\mathbf{x} - \mathbf{y}||$ e $||\mathbf{B}||$. An important group of inversion techniques using regularisation, with a constrain related to the solution magnitude, is usually referred to as constrained least square methods. However, other type of methods, with a constrain of statistical nature, are generally used for atmospheric parameters inversion, and are traditionally known as optimal estimation methods.

2.2.3 Optimal Estimation

The Bayesian solution obtained assuming Gaussian statistics is usually known as the Optimal Estimation solution for the considered inversion problem.
According to [3], maximising the a posteriori solution represented by (2.19) is equivalent to minimise the following expression:

$$[\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b_a})]^T\, \mathbf{S_e}^{-1}\, [\mathbf{y} - \mathbf{F}(\mathbf{x}, \mathbf{b_a})] + (\mathbf{x} - \mathbf{x_a})^T\, \mathbf{S_x}^{-1}\, (\mathbf{x} - \mathbf{x_a}) \tag{2.23}$$

which is analogous to equation (2.22), with the trade-off parameter λ represented here by the matrices that statistically characterize the errors. The state vector minimising the expression is given by:

$$\mathbf{x`} = \mathbf{x_a} + \mathbf{S_x}\,\mathbf{K_x}^T\,\mathbf{S_e}^{-1}\,[\mathbf{y} - \mathbf{F}(\mathbf{x`}, \mathbf{b_a})] \tag{2.24}$$

with the $\mathbf{K_x}$ matrix evaluated at $(\mathbf{x`}, \mathbf{b_a})$. Since the inversion problem is generally not linear, the solution has to be found numerically, and gradient descent techniques are typically used. In the case of steepest descent algorithm the minimisation is pursued by iterating as:

$$\mathbf{x`}_{i+1} = \mathbf{x`}_i - \alpha \mathbf{g_i} \tag{2.25}$$

where $\mathbf{g_i}$ contains the first derivatives of the function expressed by (2.23) evaluated at $(\mathbf{x`}_i)$, and α represents the learning rate. For the Newtonian algorithm:

$$\mathbf{x`}_{i+1} = \mathbf{x`}_i - \mathbf{H_i}^{-1}\mathbf{g_i} \tag{2.26}$$

where $\mathbf{H_i}$ is the Hessian matrix, i.e. contains the second derivatives evaluated at $(\mathbf{x`}_i)$. The steepest descent algorithm is safer during first iterations, avoiding secondary minima, while the Newtonian method converges faster when approaching the global minimum. The Marquardt-Levenberg algorithm merges the two characteristics by means a trade-off parameter, γ:

$$\mathbf{x`}_{i+1} = \mathbf{x`}_i - (\mathbf{H_i} + \gamma \mathbf{I})^{-1}\mathbf{g_i} \tag{2.27}$$

For $\gamma \to \infty$ we obtain the steepest descent algorithm, for $\gamma \to 0$ the Newtonian method. At each iteration step, the solution is expressed by:

$$\mathbf{x`}_{i+1} = \mathbf{x`}_i + (\mathbf{S_x}^{-1} + \mathbf{K_i}^T\mathbf{S_e}^{-1}\mathbf{K_i} + \gamma \mathbf{I})^{-1}\,\{\mathbf{K_i}^T\mathbf{S_e}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{x`}_i)] + \mathbf{S_x}^{-1}[\mathbf{x_a} - \mathbf{x`}_i]\} \tag{2.28}$$

where $\mathbf{K_i}$ is evaluated at each iterated state. In Eq. (2.28) the second derivatives are typically small and have been ignored.
The iterations are stopped when a suitable criterion is satisfied. A common option is to consider the difference $\mathbf{x`}_{i+1} = \mathbf{x`}_i$. As a Gaussian statistic has been assumed for the *n*-dimensional state vector **x**, the variable $(\mathbf{x`} - \mathbf{x})^T\mathbf{S_\delta}^{-1}\,(\mathbf{x`} - \mathbf{x})$ follows a χ^2 distribution of expected value *n*. Approaching the solution the difference $(\mathbf{x`}_{i+1} - \mathbf{x`}_i)$ should be much smaller than $(\mathbf{x`} - \mathbf{x})$, and the criterion to be satisfied can be written as:

$$(\mathbf{x`}_{i+1} - \mathbf{x`}_i)^T\mathbf{S_\delta}^{-1}(\mathbf{x`}_{i+1} - \mathbf{x`}_i) << n \tag{2.29}$$

The obtained solution, $\mathbf{x`} = \mathbf{x`}_{i+1}$, can be tested applying the χ^2 test on the differences $[\mathbf{y} - \mathbf{F}(\mathbf{x`}, \mathbf{b_a})]^T\mathbf{S_e}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{x`}, \mathbf{b_a})]$ and $(\mathbf{x} - \mathbf{x_a})^T\,\mathbf{S_x}^{-1}\,(\mathbf{x} - \mathbf{x_a})$, whose expected values are *m* (measurement vector dimension) and *n* (state vector dimension) respectively.
A retrieval characterization is illustrated in Figure 2.2 and Figure 2.3. The example is taken from the O_3 retrieval algorithm developed at the Rutherford Appleton Laboratory (RAL) for the processing of GOME data. In Figure 2.2 an example of an ozone profile retrieved from GOME spectral measurements (Band 2B) is reported (red line) and compared with the corresponding *a priori* profiles (dashed line) and correlative sonde profiles (black line). In Figure 2.3 the averaging kernels characterizing the retrieval are given; each function represents, for a specific retrieval level, the sensitivity of the retrieved value with respect to a variation of the true state at that level. The Full Width at Half Maximum (FWHM) of the averaging kernel functions gives an indication of the vertical resolution of retrieved profiles.

References

[1] Rodgers C.D., Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys. Space Phys., 14, 609-624, 1976

[2] Rodgers C.D., Characterization and error analysis of profiles retrieved from remote sounding measurements, *Journal of Geophysical Research*, 95, 5587-5595, 1990.

[3] Rodgers C.D., Inverse methods for atmospheric sounding: Theory and practise, *World Scientific Publishing*, 2000

[4] Eriksson P., Analysis and comparison of two linear regularization methods for passive atmospheric observations, *Journal of Geophysical Research*, 105 (D14), 18157-18167, 2000.

[5] Phillips D., A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Math*., 9, 84-97, 1962

[6] Tikhonov A., On the solution of incorrectly stated problems and a method of regularization, *Dokl. Acad. Nauk. SSSR*, 151, 501-504, 1963

[7] Twomey S., On the numerical solution of fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, *J. Assoc. Comput. Math*., 10, 97-101, 1963

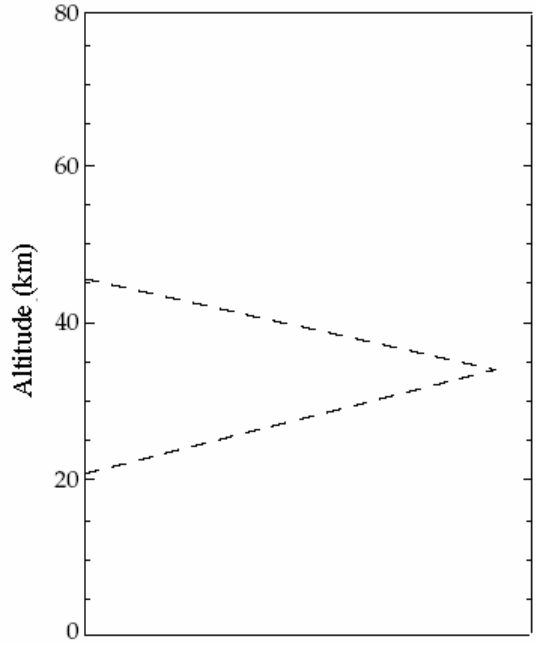

Figure 2.1. Triangular basis function.

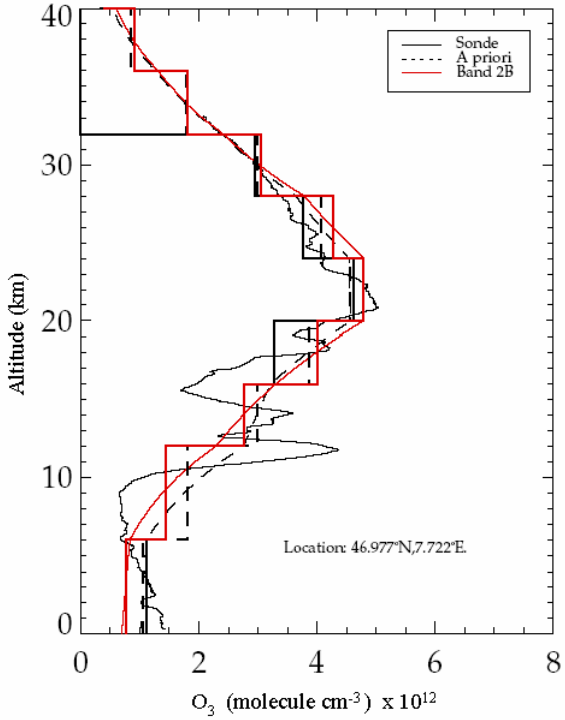

Figure 2.2. Example of ozone profile retrieved by an Optimal Estimation scheme (red line) compared with the corresponding a-priori profile (dashed line) and a coincident profile measured with an ozone-sonde (solid black line).

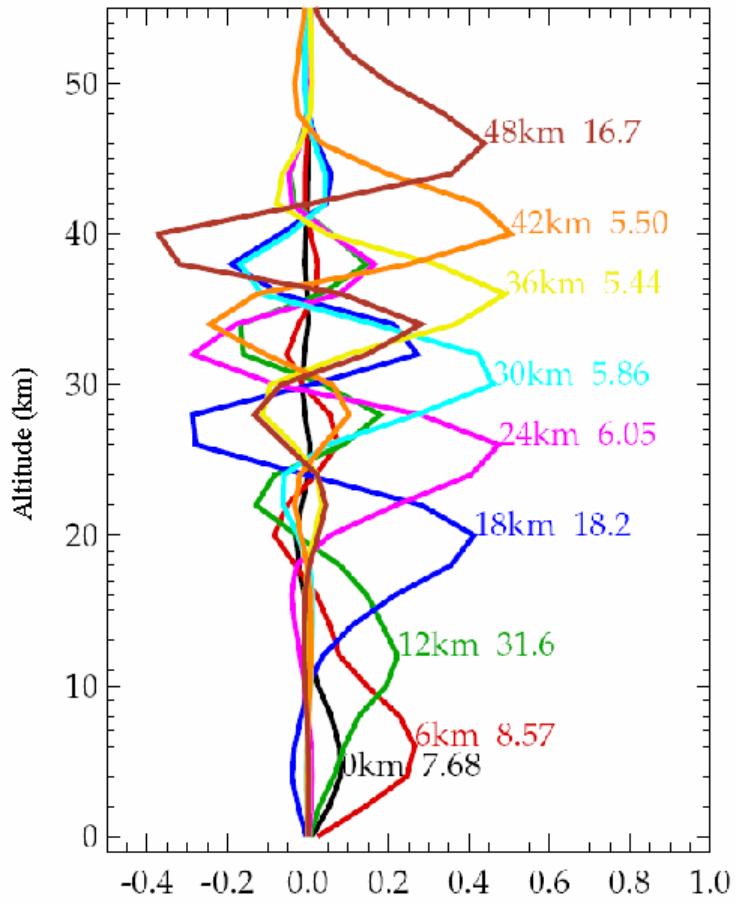

Figure 2.3. Averaging kernels characterizing the retrieved ozone profile reported in Figure 2.2.

Chapter 3

Neural Networks

In this chapter a general overview on artificial neural networks (NNs), aiming at understanding the basic concepts for the application to the solution of inversion problems, will be given. In such a context, a particular network architecture, the multi-layer perceptron, will be considered. An exhaustive description of the topic can be found in [1].

3.1 Introduction

Studies on artificial neural networks started in the 1940s as a result of investigations on the structure and the functionalities of the human brain, which can be viewed as a net made of a large number of nerve cells, called neurons, interconnected by links called synapses. This structure is able to receive and transmitt informations, by means the propagation of electrical signals. When a stimulus (e.g. the electrical signal) is presented to a neuron, the nerve cell elaborates the signal and, if greater than a specific threshold, produces an output. This output is independent from the intensity of the stimulus, and it is transmitted to other cells via the synapses [2]. The behaviour of each single neuron satisfies two basic principles:

- *dynamic polarisaton*: in the nerve cells the electrical information follows a constant and predictable direction;
- *connection specificity*: the connections between nerve cells are not casual, but each cell is specifically linked with some target cells.

Such a structure aims at fulfilling the basic functionalities of the human brain, which can be summarized, according to the modern neurophysiology, in the following:

- *distributed analysis*: the basic functions of human brain are localised in specific areas, and more complicated functions result from the interconnection of such elementary functional regions;
- *parallel processing*: each single function follows more than one connection path, in order to mantain its functionality, even if a connection is damaged;
- *learning process*: represent the capability of a biological system to acquire a knowledge from the external environment, and to adapt its structure on the basis of this acquired knowledge;
- *content addressability*: represent the capability to store the acquired knowledge and to perform queries among stored informations.

The first mathematical model of an artificial neuron was designed by Mc Culloch and Pitts in1943 [3]. This model implements two important characteristics of the biological neuron:

- the artificial neural unit is able to propagate the information if the intensity of the input is greater than a fixed threshold;
- the transmitted information is independent from the intensity of the input.

In the 1960s Rosemblatt designed a new model [4], called *perceptron*, introducing other two fundamental characteristics:

- the artificial neuron receives the inputs via an activation function, which is generally not linear;
- each connection between neurons is characterized by modifiable weights.

According to this model, the neuron processes the information receivesd in input as follows:

$$o = f(\mathbf{W}^{\mathrm{T}}\mathbf{i} + b) \tag{3.1}$$

where o is the output value of the neuron, $\mathbf{i}$ is the vector containing inputs, $\mathbf{W}$ is the vector of weights characterizing the inputs, b an additional bias term and f the activation function. In the following section we will consider a class of artificial neural networks referred to as multi-layer perceptrons (MLPs). In this model the neural units, i.e. the perceptrons, are organised in layers. If the connectivity among neurons in the same layer, or with units of previous layers, is allowed, we obtain recurrent neural networks; otherwise we have feed-forward neural networks, where the information propagates only in the direction of successive layers.
Different activation functions are typically used; among them the linear function:

$$f(a) = a \tag{3.2}$$

or the logistic function:

$$f(a) = \begin{cases} -1 & \text{if } a < 0 \\ 1 & \text{if } a \geq 0 \end{cases} \tag{3.3}$$

This function is not differentiable, and is generally substituted by the sigmoidal function:

$$f(a) = \frac{1}{1 + e^{-a}} \tag{3.4}$$

or by the hyperbolic tangent function:

$$f(a) = tanh(a) \tag{3.5}$$

A neural network based on the multi-layer perceptron model can be viewed as a structure composed of many computational elements, i.e. the neurons, organised in layers, operating in parallel and massively connected by links characterized by different weights. The net necessarely has an input layer, composed by a number of units equal to the number of inputs , an output layer, which produces expected results, and eventually one or more hidden layers, characterized by a variable number of units. A neural network is completely defined by the following characteristics:

- the topology, i.e. the number of input, hidden and output units;
- the properties of each neurons, i.e. the activation function, weights and threshold values;
- the rules for the output calculation;
- the rules for the learning process, i.e. how the weights are modified in order to improve the performance of the net.

Neural networks can be used to solve a wide range of problems, such as pattern recognition, classification, filtering, decision and regression problems and so on [1]. In this work we will consider feed-forward multi-layer perceptrons for the solution of the inversion problem stated in Chapter 2.

3.2 Multi-layer perceptron architecture

MLPs have one or more hidden layers between the input and output layer, and neither recursive nor backward connections. A tipical structure of a MLP is reported in Figure 3.1. The weight characterizing the connection between the node i belonging to the layer l and the node j of the following layer is denoted as w_{ij}^{l}. The weighting matrix for the layer l is:

$$\mathbf{W}^{l} = \begin{bmatrix} w_{11} & \dots & w_{1m} \\ & & \\ w_{n1} & \dots & w_{nm} \end{bmatrix} \tag{3.6}$$

where n is the number of nodes in the layer l and m is the number of nodes of the successive layer. Denoting $\mathbf{I}^{l}$ the input vector of the layer l, and $\mathbf{b}^{l}$ the vector containing the biases, the output vector of the layer is expressed as follows:

$$\mathbf{O}^{l} = f_{l}(\mathbf{W}^{l-1} \mathbf{I}^{l} + \mathbf{b}^{l}) \tag{3.7}$$

where f_l represents the activation function of all nodes of the layer l. MLP networks include at least one layer with not-linear but differentiable activation function; note that, even if it is not mandatory, the output layer is generally characterized by a linear activation function, in order to avoid the output vector of the NN from being limited to a restricted range. Nevertheless, in some cases it can be useful to limit the output range of the MLP, and the typical sigmoidal or hyperbolic tangent functions are used. Designing a multi-layer perceptron involves several actions. First, we need to set-up the topology of the network, i.e. the number of layers, the number of nodes for each layer and the activation functions. Then, we have to build-up a suitable training set, which will be used to fix the parameters of the network, i.e. weights and biases of the nodes. Finally, deciding the training strategy, that will adjust network parameters in order to give a solution for the considered problem.

3.2.1 Topology

In this work MLPs will be used to approximate a mapping function between an input and an output space, i.e. between the vector containing input measurements and the vector containing physical parameters to retrieve. As a consequence, the number of input and output units is automatically given by the dimension of the input and output space. Various theoretical investigations demonstrate that a neural network having one single hidden layer, formed by a suitable number of hidden units, is capable to approximate any not-linear mapping function [1] [5]. In this study only neural networks with one hidden layer will be considered, and the problem of selecting the proper number of hidden units will be addressed (Chapter 5). The sigmoidal function, which has the properties of the logistic function but being continuous and differentiable, is chosen to be the activation function for all the nodes; this choice can be useful in this case, as the output range is constrained to have only positive values.
The internal parameters, i.e. weights and biases determining the estimation properties of the network, will be fixed during the training procedure. However, an initial value is needed. Usually, they are set randomly, but with some care. Assuming sigmoidal activation functions, it is important that the initial values of weights do not drive the neurons into saturation, as the error

surface will be flat and the convergence very slow. On the other hand, they should not be too small, leading otherwise the activation functions to be quasi-linear [1]. In this work symmetric Gaussian distributions, with zero mean and standard deviation set according to the number of input units, has been used.

3.2.2 Training

Given a set of N input-output pairs $\{\mathbf{y},\mathbf{x}\}$, where $\mathbf{y}$ represents the measurement vector and $\mathbf{x}$ the reference state vector corresponding to $\mathbf{y}$, the objective of the MLP is to approximate the mapping function between the input and output space. In statistical terms, the MLP provides an estimation of $p(\mathbf{x}|\mathbf{y})$, i.e. the probability of the output variables conditioned to the input variables. Weights and biases characterizing the MLP will then be set to provide the estimation, that is the approximation of the conditional density function $p(\mathbf{x}|\mathbf{y},\mathbf{w})$, where the vector $\mathbf{w}$ represent all the neural internal parameters. This density estimation can be performed by different classes of methods. The first class, known as parametric methods, assumes a specific function to model the density function, e.g. a Gaussian distribution. In this case the number of parameters does not depend on the size of the data set. The second class, referred to as non-parametric methods, uses a histogram-based approach, without assuming a pre-defined function. Here the number of estimation parameters grows with the size of the dataset. In such a context, MLPs can be considered as a semi-parametric method, since a general class of functions is used, and the number of internal parameters can be increased to build a more flexible model, even if this number does not depend on the size of the dataset.
The training process, also called the learning phase, consists in finding the set of parameters which maximise the estimation properties of the net. This means that, during the training, a function depending on the weights and measuring the estimation capability of the MLP has to be maximised. Such a function is the conditional probability mentioned above, $p(\mathbf{x}|\mathbf{y},\mathbf{w})$, also called the *likelihood of weights*, $\mathcal{L}(\mathbf{w})$. Assuming that inputs and outputs in the training dataset, $\{\mathbf{y}^i,\mathbf{x}^i\}_{i=1,N}$, are independent, the function can be expressed as follows:

$$\mathcal{L}(\mathbf{w}) = \prod_{i=1,N} p(\mathbf{x}^i|\mathbf{y}^i,\mathbf{w})\, p(\mathbf{y}^i) \tag{3.8}$$

Maximising the likelihood function is equivalent to minimise the error function

$$\mathcal{E}(\mathbf{w}) = \sum_{i=1,N} -\ln p(\mathbf{x}^i|\mathbf{y}^i,\mathbf{w}) \tag{3.9}$$

Under the assumption that the target data $\{\mathbf{x}^i\}$ comes from a deterministic function, we can model such a function with a MLP of output $\mathbf{o}(\mathbf{y},\mathbf{w})$; in this case the likelihodd can be written as [1]:

$$p(\mathbf{x}|\mathbf{y},\mathbf{w}) = \frac{1}{Z_a} e^{-\frac{\beta}{2}\|\mathbf{o}(\mathbf{y},\mathbf{w})-\mathbf{x}\|^2} \tag{3.10}$$

where β is a parameter of the distribution and Z_a the normalisation factor.
Combining equation (3.9) and (3.10), the error function to estimate can be written:

$$\mathcal{E}(\mathbf{w}) = \sum_{i=1,N} \left\|\mathbf{o}(\mathbf{y}^i,\mathbf{w}) - \mathbf{x}^i\right\|^2 \tag{3.11}$$

We can therefore see that the set of weights which maximise the conditional density function $p(\mathbf{x}|\mathbf{y},\mathbf{w})$ is obtained by minimising the sum of square errors of the difference between the output vector of the MLP, $\mathbf{o}(\mathbf{y},\mathbf{w})$, and the corresponding target vector, $\mathbf{x}$. As described in Chapter 2, the

expression (3.11) corresponds to the average a posteriori solution, and it is equivalent to the minimum variance solution [1]. The minimisation of square errors can be performed by the gradient descent techniques already presented in Section 2.2.2 and Section 2.2.3.
This solution is known as the least square error solution, and it is generally used to determine the weights of the MLP. Nevertheless, other criteria than maximisation of the likelihood can be taken into account, leading to different error functions. For instances, the least square solution is not optimal in the case of several input units having the same output value, or in the case of outliers having a large contribution to the errors, without being very representative of the dataset. In these cases the use of other norms different from the standard 2-norm can significantly improve the estimation properties of the MLP.

3.2.3 Generalisation

The main objective of neural network inversion algorithms is to approximate the mapping function between two sets of data. It is crucial to underline that the ultimate goal of a NN based retrieval algorithm is to use this function to make estimations on new data; therefore, NNs should be able to catch the general relation between the input and output spaces, not the specific features of the two limited data sets. In general, more complex is the model (e.g. the number of estimation parameters), better is the fit of data. This is risky in terms of generalisations properties, as specific features characterizing the selected dataset can be included in the fit. This phenomenon is called over-fitting, and produces small biases between estimated and true data, but large variances. On the contrary, simpler models are not capable to completely represent the significant relations underlined by the two datasets, producing results characterized by small variances but large biases. This implies that an optimal complexity for the estimation model exists, allowing the best trade-off between the two conflicting requirements [1]. This can be achieved by different strategies.

3.2.3.1 Structural stabilisation

Since the neural network estimation parameters are essentially fixed by the number of biases and weights, and consequently by the number of neurons, the best compromise between the proper fit of data and the generalisation properties essentially depends on the network topology. However, the selection of the optimal number of neural unit is not trivial, and in practise involves testing a relatively large number of MLPs. This process can be very labourious, and other approaches are generally considered.

3.2.3.2 Regularisation

In order to control the model complexity, it is possible to modify the considered error function, measuring the difference between the desired and reference output of the net, with a regularisation function, usually known as *weight decay* [6]. The new error function can be expressed as:

$$\mathcal{E}_D + \lambda\, \mathcal{E}_W = \sum_{i=1,N} \left\| \mathbf{o}(\mathbf{y}^i, \mathbf{w}) - \mathbf{x}^i \right\|^2 + \lambda \left\| \mathbf{w} \right\|^2 \qquad (3.12)$$

where λ is a trade-off multiplier and $\mathbf{w}$ represents the matrix of neural weights. Smaller are the weights, smaller is the possibility for the NN to over-fit the training set.

3.2.3.3 Early stopping

Such a technique, also called cross-validation, essentially consists in monitoring the evolution of the error function during the learning phase, either on the training set or on an independent validation set, which contains input-output patterns different from the training set [7]. The overall retrieval error computed for the training set keeps on decreasing with the number of

training cycles, approaching a value of convergence. Conversely, the error on the validation set reaches a minimum value, after which it will start increasing if the process training is continued, indicating the over-fitting. This is the point to interrupt the learning phase, as the NN has achieved its best generalisation properties.
An example of the application of the early stopping procedure is given in Figure 3.2. The plot is taken from a real experiment carried out during the design of neural network inversion schemes for GOME data (see Chapter 5). The black line represents the total error, computed over all the input-output pairs and for all output units, corresponding to the training set, while the red line shows the total error for the validation set. The minimun of the red curve indicates the point where the trained MLP has achieved its best generalisation properties.

References

[1] Bishop C.M., Neural networks for pattern recognition, *Oxford University Press*, London, UK, 1995

[2] E.R. Kandel, J.H. Schwartz, T.M. Jessel, Principles of Neural Science, 4th Edition, *McGraw Hill*, 2000.

[3] McCulloch W.S. and Pitts W., A logical calculus of ideas imminent in nervous activity, Bulletin of Mathematic Biophysics, 5, 115-133, 1943

[4] Rosemblatt R., Principles of Neurodynamics, Spartan Books, New York, 1959

[5] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[6] Hinton G.E., Connectionist learning procedures, *Artificial Intelligence*, 40,185-234, 1989

[7] C.D. Doan, S.Y. Liong, Generalization for multilayer neural network Bayesian regularization or early stopping, *Proceedings of 2nd Asia Pacific Association of Hidrology and Water Resources Conference (APHW2004)*, Singapore, July 5-8, 2004

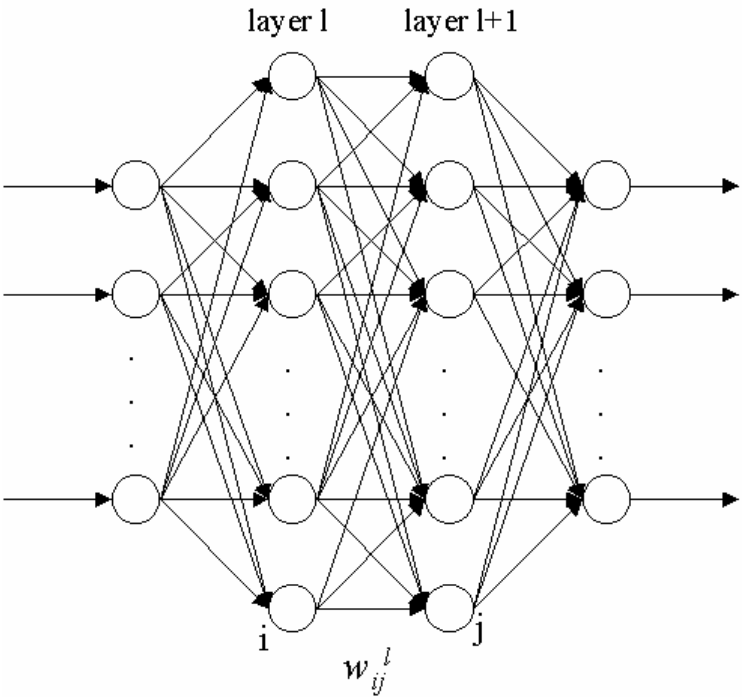

Figure 3.1. General structure of a neural network.

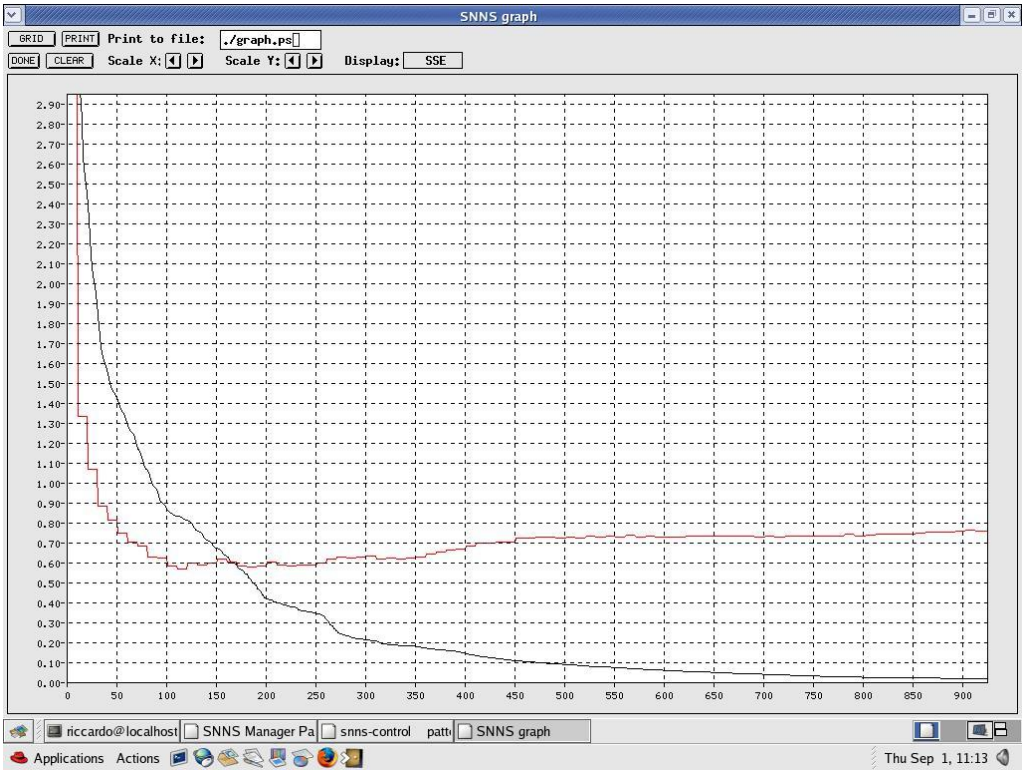

Figure 3.2. Early stopping procedure. The black line represents the NN error evaluated on the training set, while the red line the error assessed on the test set.

Chapter 4

The Global Ozone Monitoring Experiment

4.1 Introduction

The Global Ozone Monitoring Experiment (GOME) is a nadir-viewing, across-track scanning spectrometer, designed to measure solar back-scattered radiation from the atmosphere with a contiguous spectral coverage in the range from 240 to 790 nm.
The instrument is boarded on the second Earth Remote Sensing (ERS-2) satellite, launched on April 1995 by the European Space Agency (ESA). The instrument has been in almost continuous operation since its launch, though recently with somewhat degraded performance. Measured spectra are used for deriving different geophysical parameters, with a special emphasis on the global, height-resolved information on the ozone distribution in the atmosphere. GOME data are conventionally referred to by the processing level:

- Level 0 (L0): raw data transmitted from the satellite;
- Level 1A (L1A): re-formatted data with some trivial unit conversion and calibration parameters appended;
- Level 1B (L1B): calibrated data expressed into scientific units (e.g. radiance values)
- Level 2 (L2): retrieved geophysical products (e.g. total column ozone)

This chapter firstly gives an overview on the GOME instrument, in particular on the operational mode and calibration of measurements. The main sources of error are then identified, followed by a brief survey of the geophysical parameters operationally determined. In the second part of the chapter, a quite comprehensive summary of existing processing algorithms aiming at the ozone profiles retrieval is given. The inversion algorithm designed by the Rutherford Appleton Laboratory is described with more details, as it plays an important role in this work (see Chapter 5).

4.2 The instrument

The instrument optics is reported in Figure 4.1. Light back-scattered from the atmosphere is viewed at nadir via a scan mirror, which allows the instrument's instantaneous field-of-view (IFOV) to be scanned across the satellite ground track. From the scan mirror, light passes through a telescope to the spectrometer entrance slit [1]. From here, light is divided into 4 channels: a pre-disperser prism allows Channel 1 and Channel 2 (below 405 nm) to be separated from Channel 3 and Channel 4; a second prism splits Channel 1 (235-316 nm) and Channel 2 (311-405 nm), while light to Channel 3 (395-611 nm) and Channel 4 (605-793 nm) is divided by a beam-splitter. The spectrum in each channel is generated by a blazed holographic diffraction grating, which disperses the incident radiation onto the detector array [2]. Each array contains 1024 Si diode detector pixels (described in [3]), most of which receive dispersed light from the grating. Some pixels at the ends of the arrays are either physically shielded from incident light (blind pixels) or geometrically positioned so as not to receive direct illumination, but to respond to stray-light. The electronic charge generated by incident photons on the detector pixel is accumulated for a particular integration time, before being discharged and read-out via an

analogue to digital converter (ADC). Level 0 (L0) data are formed by the count, expressed in binary units (BUs), recorded by the ADC. In the absence of light, the detectors still accumulate charge, quantified as dark current, which is a source of spectral noise and has to be accounted for in calibration. Other processes in the detectors and associated electronics give rise to additional noise, independent from the input signal and the integration time, collectively referred to as read-out noise. Both dark current and read-out noise are functions of temperature; therefore, an active Peltier cooler is used to reduce the detector temperature to 200-230 K. Additional detectors are included to measure the polarisation state of the incoming light: the pre-disperser prism directs a fraction of the light from the telescope to three polarisation monitoring detectors (PMDs).
Solar irradiance measurements are obtained switching the scan mirror to the calibration unit. Solar radiation reaches the scan mirror via a diffuser, in order to reduce light intensity and ensure uniform illumination of the slit. The calibration unit also contains a PtNeCr line lamp, which acts as a reference for the wavelength calibration of the instrument.

4.3 Operation

The ERS-2 satellite has a sun-synchronous orbit, with an altitude of 790 km, an equator crossing time at the descending node of 10:30 am local solar time and an inclination of 98.5°. ERS-2 completes 14 orbits a day, with a repeat cycle of 31 days. The principal observing mode of GOME is illustrated in Figure 4.2. The instrument scans across track a 960 km swath, centred on the nadir. The forward-scan from East to West is performed in 4.5 s, while the back-scan takes 1.5 s. Integration periods are synchronised to give 3 pixels (320 km across by 40 km along-track) on the forward scan for all the spectral bands - excepted Band 1A (240-268 nm) - and 1 pixel on the back-scan. Band 1A is integrated over 2 complete scans, and so has a ground pixel size of 960x80 km^2. Given the orbital parameters of ERS-2, GOME gives observations over the whole Earth's surface in 3 days. Occasionally, the instrument scans off-orbit track to view the pole directly, though such data has not been used in this work.
The electronic devices allow reading-out detector pixels with different integration times. This is exploited in Channel 1, where the signal has the greatest dynamic range. The shortest wavelength, referred to as Band 1A, have an integration time of 12 s, while the remainder of Channel 1, Band 1B, as well as Channel 2, Channel 3 and Channel 4 are integrated for 1.5 s. The calibration of GOME observations relies on a database of key-data, mostly generated through pre-flight testing of the instrument, and is performed through several steps [4]:

1. *dark current subtraction*: the dark current BUs for each integration period are subtracted, and the measured signals are then normalised by the integration time;
2. *stray-light subtraction*: stray-light represents the illumination of the detectors originating from anywhere other than the intended grating. This effect, significant for Channel 1 and Channel 2, can be considered spectrally uniform and proportional to the total incident light flux, and it is corrected using pre-flight parameters characterisation;
3. *conversion to radiance units*: leakage and stray-light corrected signals are converted into radiance units, according to the pre-flight characterisation of the radiance sensitivity to un-polarised incident light and the bi-directional scattering distribution function (BSDF) of the calibration unit diffuser;
4. *polarisation correction*: the incident radiance from the Earth is assumed to be un-polarised; the total intensity is estimated by correcting for the polarisation sensitivity of the instrument.

Calibrated radiance spectra are then sun-normalised, assuming that the emission of radiation within the atmosphere is negligible, and neglecting the Ring effect, i.e. the apparent excess intensity in the Fraunhofer lines in scattered sunlight primarily caused by Raman scattering. Sun-

normalised radiances (Level 1B data) are generally used as the basic measurement in UV-visible remote sensing of the atmosphere, since the instrumental gain error, common to both solar irradiances and back-scattered radiances, cancels. Moreover, Fraunhofer spectral structures are removed by the ratio, leaving only the spectral signatures due to the interaction between solar radiation and atmospheric components. Figure 4.3 shows some examples of spectra recorded by the GOME instrument. The two panels on the top represent solar irradiance spectra, expressed in (photons/s/cm^2/nm) as a function of wavelength, from 240 to 350 nm for the panel on the left and from 290 to 370 nm for the panel on the right. The two central panels show the back-scattered radiances spectra (expressed as photons/s/cm^2/nm/sr) for the same wavelength ranges. The two panels on the bottom finally represent the normalised radiances.

4.4 Radiometric errors

In this section a brief overview of the main radiometric errors characterizing GOME observations is given. It is useful to distinguish between [5]:

- Noise, i.e. the purely random (spectrally and temporally) error, which determines the precision of the measurement. These errors are assumed to be Gaussian, i.e. with probability density function given by:

$$P(x) = \frac{1}{\Delta x\sqrt{2\pi}} e^{\frac{(x-\bar{x})^2}{2(\Delta x)^2}} \tag{4.1}$$

- Systematic error, which is correlated on some observable spectral or temporal scale.

Measurement accuracy, i.e. the deviation of the observation from the true quantity being measured, is given by the combination of noise and systematic error.

4.4.1 Noise

The limit of measurement precision depends on the noise affecting the number of detected electrons [3], which follows a Poisson distribution with associated standard deviation:

$$\Delta N_e = \sqrt{N_e} \tag{4.2}$$

The number of detected electrons, N_e, depends on different quantities:

- the incident radiance (photons/s/cm^2/nm/sr)
- the effective telescope area (cm^2)
- the transmission of the optics
- the view solid angle (sr)
- the pixel spectral width (nm)
- the integration time (s)
- the detector quantum efficiency
- the leakage current (A)
- the electron charge (Coulomb)

Two additional noise sources must be considered: the read-out noise, i.e. the random error introduced each time the charge accumulated in a detector pixel is measured, and the digitisation

noise, i.e. the errors introduced by the ADC in reported counts. The signal-to-noise ratio gives a representation of the random error, expressed in radiance units, for each spectral detector element.

4.4.2 Systematic error

4.4.2.1 Peltier Cooler Interference

Early in the mission, it was noted that calibrated Band 1A spectra were subject to spurious, highly significant, quasi-random perturbations correlated with Fraunhofer spectral structures. This was attributed to an increased dark current driven by interference from the Peltier cooler [4]. A scheme aiming at the removal of spurious signals based on correlation with the cooler state was implemented in the GOME Data Processor (GDP); nevertheless some perturbations in Band 1A dark current still persist after the correction.

4.4.2.2 Diffuser spectral features

When comparing GOME spectral measurements with modelled spectral signatures, it has been noticed in various studies that systematic differences exist [6] [7] [8] [9]. Such residual structures have been partially attributed to spectral perturbations introduced by the diffuser BSDF (not corresponding to pre-flight tests), which are quite significant compared to minor trace gas absorption signatures.

4.4.2.3 Scan mirror degradation

Since the beginning of the mission, the scan mirror reflectance has been observed to be degrading, particularly below 300 nm. Under the assumption that this degradation is independent from the scan mirror angle (σ), it can be easily corrected, so as having a little consequence for sun-normalised radiance. However, since January 1999, the degradation of the scan-mirror exhibits dependence from the mirror angle, which has been attributed to the build-up of a contaminant layer [10]. Errors propagate directly into sun-normalised radiances [11], and are expected to have serious implications for ozone profiles retrieval. Empirical correction schemes are generally developed in order to process post-1999 data [12].

4.4.2.4 Etalon effects

The radiance sensitivity to un-polarised light incident on the scan mirror, h, and the polarisation sensitivity of the instrument, η, are determined in part by etalon structures, due to reflection or transmission within layers or between thin surfaces. Spectral features associated with the etalon may shift in wavelength, as a result of small perturbations to the physical etalon in question, giving rise to potentially significant errors in calibrated radiances. Perturbations to h are not significant for sun-normalised radiances, assuming that the etalon has the same impact on both involved optical paths, and that the etalon does not change over a time scale less than a day (i.e. the frequency of solar observations). On the other hand, perturbations to η are more significant, since the polarisation correction only applies to the back-scattered (Earth) radiance.

4.4.2.5 Polarisation correction

The fractional polarisation of the back-scattered light (p) varies rapidly in the spectral range 300-325 nm, which can be therefore particularly affected from radiometric calibration errors arising from the lack of adequately resolved polarisation measurements. Degradation of the UV PMD-1 has been observed from direct-sun observations. The degradation is much faster than the corresponding main channel degradation, leading to further errors in the polarisation correction scheme [7] [13].

4.4.2.6 Solar calibration unit

The observation of GOME solar spectra showed erroneous periodical variations on solar irradiances; the effect is attributed to errors in the diffuser BSDF (only in the direct-sun optical path) and so straightforwardly propagates into sun-normalised radiance errors [7] [14].

4.4.2.7 Pixel scene

The sequential read-out of the detectors during the across-track scan means that each detector pixel sees a slightly different ground scene, which can produce different radiances, e.g. caused by the presence of clouds. These radiance differences are therefore recorded in the measured spectrum. The structure of the scene can produce broad-band spectral errors, giving rise to discontinuities between the overlap regions of the channels (e.g. where the first pixels read-out in Channel 2 overlap in wavelength with the last pixels in Channel 1, but view significantly different ground scenes).

4.4.2.8 Temperature variation

As already mentioned, the instrument temperature can affect the spectral calibration. Significant variations in the instrument temperature have been observed during the GOME life-time [15]: from 1995 to 2002 an increase of 1 K has been observed. Moreover, there is also a seasonal variation correlated with the Sun-Earth distance of 1.5 K (maximum temperatures in December/January) that has to be taken into account.

4.5 Geophysical products

4.5.1 Overview

The Level 2 (L2) geophysical products produced by the GDP [16] include:

- column amounts of O_3, NO_2, BrO, HCHO estimated by the DOAS technique;
- cloud fraction and cloud top height, calculated with the Initial Cloud Flagging Algorithm (ICFA).

In addition, different agencies and research groups have developed their own Level 2 processing schemes, aiming at the retrieval of various atmospheric parameters, such as NO_2, H_2O and SO_2 column densities [17] [18] [19] [20], aerosols optical depth [21] [22] and ozone profiles [23] [24] [25] [26] [27]. In particular for the retrieval of ozone profiles, a number of institutes have proposed a large variety of methods; all these algorithms exploit different spectral ranges and different inversion techniques to extract the information on the ozone vertical distribution. Results from all these schemes have been extensively inter-compared in the framework of the ESA GOME Ozone Profile Working Group. A brief overview on existing retrieval algorithms for GOME ozone profiles is reported in the next section.

4.5.2 GOME ozone profiles retrieval algorithms

Currently, different approaches are used to retrieve the information on the ozone vertical distribution from GOME measurements, and they mainly differ in the use of external information and a-priori constraints. The principal techniques are:

- Optimal Estimation
- Philips-Tikhonov Regularization
- Data Assimilation

The following paragraphs summarise the main characteristics of each of these methods.

4.5.2.1 Optimal Estimation approach

As already stated in Chapter 2, normal least-squares fitting does not work for ill-posed problems (like the ozone profile retrieval from UV-reflectance measurements), since it amplifies measurement noise resulting in profiles with unphysical, large amplitudes. The well-known Optimal Estimation approach solves ill-posed problems by using a-priori information. The use of a-priori profiles tends to stabilize the inversion when relating simulated measurements (from the forward model) to atmospheric parameters. Retrieved ozone profiles from the Optimal Estimation method can be regarded as the a-priori profiles updated with the information contained in the spectral measurements. Using the same formalism already presented in Chapter 2, and neglecting the terms accounting for auxiliary parameters (**b**) and the measurement errors (**ε**), we can write:

$$\mathbf{O}_{3,\,\text{retrieved}} = \mathbf{A}\mathbf{O}_{3,\,\text{true}} + (\mathbf{I} - \mathbf{A})\mathbf{O}_{3,\,\text{a priori}} \tag{4.3}$$

where $\mathbf{O}_{3,\,\text{retrieved}}$, $\mathbf{O}_{3,\,\text{true}}$ and $\mathbf{O}_{3,\,\text{a priori}}$ are vectors of ozone number densities, and correspond to the values of the retrieved, true and a-priori profile, respectively. **A** is the so-called averaging kernel matrix, or model resolution matrix, and it constitutes a mapping between the true anomaly (i.e. the difference between the true and the a priori profile) and the retrieved anomaly (i.e. the difference between the retrieved and a priori profile). According to the theory of the Optimal Estimation, **A** is given by:

$$\mathbf{A} = \mathbf{S}_a \mathbf{K}^T (\mathbf{K}\,\mathbf{S}_a\,\mathbf{K}^T + \mathbf{S}_\varepsilon)^{-1}\,\mathbf{K} \tag{4.4}$$

where $\mathbf{S}_a$ and $\mathbf{S}_\varepsilon$ are the a-priori and the measurement error covariance matrices, respectively, and **K** is the so-called weighting function matrix. **K** describes how the forward model F(**x**), which relates the spectral measurement to the true state vector, is sensitive to changes in the state vector.

When applied to the ozone profile retrieval from nadir UV spectra, the averaging kernels quantify the limited sensitivity of the spectral measurements to fine-scale structures of ozone profile and to the ozone values below the ozone maximum. In addition, the kernels depend on the errors characterizing the inputs to the inversion algorithm. For example, as evident from equations (4.3) and (4.4), for larger a-priori errors the averaging kernel matrix tends towards the identity matrix and hence the OE solution becomes less dependent on the a-priori profile. Conversely, for smaller a-priori errors the averaging kernel elements go to zero, and the solution tends to the a-priori profile. For the measurement errors the situation is reversed. Note that the averaging kernel matrix is influenced not only by the input measurements of the retrieval system, but also by the a-priori selected climatology.

Equation (4.3) quantifies the deviation between the true and the retrieved profile, which is especially important when comparing the retrieved profile to correlative measurements. Note the two extreme cases: when **A** is the identity matrix, the retrieved and the true profiles are equal; when all elements of **A** are zero, the retrieved profile equals the a-priori values.

Different algorithms based on the Optimal Estimation approach have been developed for the retrieval of ozone profiles from GOME measurements. Hereafter most of them are presented, and the main differences between different practical implementations are underlined.

The Institute of Environmental Physics at the University of Bremen in Germany (IUPB) has developed the FURM (Full Retrieval Method) algorithm. This method is essentially based on the Optimal Estimation approach, but coupled with the information matrix method from Kozlov [28], which adapts the number of fit parameters to the information content of the measurement vector. The main purpose of this eigenvector technique consists in using the a-priori information

in the same statistical sense as the original OE approach, but allowing a much faster fit process, as only those parameters for which there is information in the measurement vector are considered. The GOMETRAN radiative transfer model (RTM) [29], specifically designed for GOME retrieval applications, is used for calculations of radiances and weighting functions. GOMETRAN is a monochromatic model which includes a full treatment of multiple-scattering. Rayleigh cross sections and phase functions are based on the work of Bucholtz [30], while aerosol profiles and optical properties are taken from the LOWTRAN-7 aerosol model [31]. GDP spectra are corrected for the Ring effect (including rotational Raman scattering) using a look-up table. Clouds are treated as highly reflecting surfaces at 0 km altitude (i.e. clouds are treated with an "albedo approach"), which means that in the RTM the surface albedo is replaced by the weighted mean of the surface albedo and the cloud albedo, the weight being the fractional cloud cover (the Earth's surface is assumed to be a Lambertian reflector with wavelength dependent albedo). The GDP spectra contain unresolved problems with the radiometric calibration, particularly between 260 and 290 nm [32]. In the retrieval they appear as spectral fit residuals with characteristic structures, but none of the atmospheric fit parameters can account for them [23][33], and the IUPB algorithm was therefore restricted to use only wavelengths longer than 290 nm. In a further development, a new calibration correction scheme has been introduced, that permits adding the wavelength range 275–290 nm to the fit window [34]. The a-priori ozone profiles used in the IUPB algorithm are from the global ozone climatology of Fortuin and Kelder [35], which is based on ozone-sonde and satellite measurements. This climatology provides monthly zonal mean ozone profiles in 10° latitude bands at pressure levels between 1000 and 0.3 hPa. The a-priori variance of these profiles is fixed to 30%, and the complete a-priori covariance matrix is generated assuming an exponential decrease for the off-diagonal elements from the diagonal value, using a correlation length of 5 km, which is the same for all altitudes. The temperature profiles are taken from the UKMO (United Kingdom Meteorological Office) analysis [36].

The Royal Netherlands Meteorological Institute (KNMI) developed the OPERA (Ozone Profile Retrieval Algorithm) scheme. Ozone profiles are derived from GOME data in the spectral range from 270 to 330 nm, co-adding measurements coming from Band 1B and Band 2. The radiometric and wavelength calibration characterizing the GDP Level 1 data were recognized to be too inaccurate for ozone profiles retrieval, and therefore several corrections were applied using the spectral calibration program GomeCAL (available through http://www.knmi.nl/gome_fd/gomecal/) [37] [38] [24].

Sun-normalised radiances are simulated by running the Linearized Discrete Ordinate Radiative Transfer model (LIDORT-A) [39], a simplified and sped-up version of the full LIDORT model [40]. LIDORT-A is only applied for the multiple scattered part of the radiance, and runs with a limited set of 20 layers. The single scattering part is computed with a dedicated, simpler and therefore faster, single scattering model, with the full retrieval grid of 40 layers. LIDORT-A is a scalar model, and it does not treat polarization and the vector nature of the radiation field. This produces errors in the radiance values at the top of the atmosphere that can reach, in the selected spectral range, 10% for high scattering angles. Raman scattering (responsible for the Ring effect) is not treated in the RTM, but accounted for using a high resolution spectrum convolved with the Raman lines [41].

The fractional cloud cover is treated as a Lambertian reflecting layer at the cloud top height, for the fraction of pixels covered with cloud. The cloud fraction and cloud-top height are obtained from the Fast Retrieval Scheme for Cloud Observables (FRESCO), which extracts the information from the Oxygen A-band [42]. Aerosols are not treated, but the FRESCO results incorporate aerosol presence to a certain degree. Ozone cross sections are taken from the temperature-parameterized data set of Bass and Paur [43] and Paur and Bass [44], corrected according to [45], while Rayleigh cross sections are computed using empirical formulae [46]. Other trace gases than ozone are not treated and assumed not to affect the retrieval in the selected spectral range. The a-priori ozone profile information comes from the global ozone climatology

of Fortuin and Kelder [35], with covariance information derived from the same data set (which corresponds to a correlation length of 4–5 km) [47].
The Smithsonian Astrophysical Observatory (SAO) algorithm also uses the Optimal Estimation approach to derive the ozone vertical distribution, and has a particular emphasis on improving tropospheric ozone retrieval [48]. For retrievals in the troposphere, fitting precision is very important and improved wavelength and radiometric calibration is needed. Therefore the algorithm requires a detailed treatment of these aspects, including Ring effect and polarisation correction.
Ozone profiles are retrieved from GDP spectra corrected with GomeCAL. To reduce measurement errors, and because of the relatively broad ozone absorption structure between 289 and 307 nm, 5 neighboring pixels (in wavelengths grid) are co-added and sampled at every 2 pixels. The LIDORT model is used to simulate radiances and weighting functions. The scalar radiances obtained by LIDORT are corrected to incorporate polarization using a look-up table. Surface albedo and parameters for other trace gases (NO_2, SO_2, BrO) are also given as inputs to the model. Auxiliary parameters are provided in order to account for the Ring effect. In the characterization of the atmosphere, the SAO algorithm uses monthly mean stratospheric aerosol data from SAGE-II [49] and tropospheric aerosol model fields from Global Earth Observing System CHEM model (GEOS-CHEM) [50]. Clouds are treated as Lambertian surfaces, and cloud fraction and cloud-top pressure come from the GOME Cloud Retrieval Algorithm (GomeCAT) [51]. Daily temperature profiles from ECMWF (European Centre for Medium-Range Weather Forecast) and surface pressures from NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) are also used by the model. A-priori ozone profiles are initialized with the climatology provided by the Total Ozone Mapping Spectrometer (TOMS-Version 8) [52]. Standard deviations at 40–50 km are assumed as 70%, 60%, 50%, 40%, 30%, 20% for 50, 48, 46, 44, 42, 40 km, respectively. For a-priori standard deviations between 0 and 40 km, whatever is the largest between 40% and the original a-priori standard deviation is used. The off-diagonal elements of the covariance matrix are assumed to exponentially decrease, with a correlation length of 6 km.
The National Oceanic and Atmospheric Administration (NOAA) applied the Version 8 SBUV/2 algorithm, developed for the SBUV instruments, to the GOME data [53]. Unlike the previous four OE algorithms, this algorithm was not specifically designed for GOME data. The SBUV data are measured at the following wavelengths (nm): 251.99, 273.51, 283.27, 287.62, 292.26, 297.54, 301.93, 305.80, 312.50, 317.51, 331.23, and 339.84 with a bandwidth of about 1.1 nm. A triangular filter centered at those values has been used to convert GOME spectral data to the SBUV bandpass. Because the GOME data have large errors below 270 nm, an extrapolation was used to provide the standard input for the SBUV retrieval algorithm at 251.99 nm. The SBUV algorithm uses a single-scattering forward model calculation coupled with adjustments from multiple scattering tables created from the TOMS forward model (TOMRAD) [54]. This solution accounts for all orders of scattering, as well as the effects of polarization. It includes a pseudo-spherical correction, molecular anisotropy [55], and rotational Raman scattering [56]. Clouds are taken into account using climatologies of monthly cloud top pressures and snow/ice coverage. If snow or ice is present, the clouds are treated as though they are at the surface.
The SBUV algorithm normally incorporates its own a-priori ozone profile database. This database was created from 15 years (1988–2002) of ozonesonde measurements and SAGE and/or UARS-MLS data. Over 23.400 sonde profiles from 1988 to 2002 were used in producing this climatology. Average profiles from ozonesondes and SAGE are merged over a range of 4 km. The dataset gives the climatological averages for eighteen 10° latitude bands and 12 months. The a-priori covariance is constructed as follows: the diagonal elements correspond to 50% variance and the non-diagonal covariance elements fall off with a correlation length of approximately two Umkehr layers (about 10 km). The measurement covariance is diagonal and corresponds to radiance errors of 1% in each band. Temperature profiles derive from an interpolation of climatological tables by month and latitude similar to the a-priori ozone profile in size. They are

from a combination of balloon sonde for the lower part, and fields from the National Centre for Environmental Prediction (NCEP) for the upper part.

4.5.2.2 Philips-Tikhonov Regularization Approach

The Philips-Tikhonov Regularization (PTR) approach [57] [58] has been little used for the analysis of atmospheric spectra, e.g., to retrieve ozone profiles. However, it has been extensively studied in the mathematical field of inversion. The analysis of the fundamental problem by Hansen and O'Leary [59] and Hansen [60] provides a basis for the application of PTR to remote sensing problems. In contrast to the OE approach, the PTR approach does not require a-priori ozone profiles and corresponding covariance matrices. The same basic equation of OE (4.3) is applied, but the vector representing the a-priori profile, $\mathbf{x}_{\text{a priori}}$, is set to zero.
Such a technique has been implemented for the ozone profiles retrieval from GOME measurements by the by the Space Research Organization of Netherlands (SRON) [61]. In addition to the least squares minimization between forward model calculations and measurement, this algorithm includes the minimization of the first derivative norm of the profile as a side constraint. The minimization of the least squares condition and the minimization of the first derivative norm are balanced by a regularization parameter. The rationale behind the minimization of the first derivative norm as a side constraint is that the measurement is insensitive to fine scale structures of the ozone profile. These vertical structures do not influence the residual norm but strongly influence the first derivative norm. The regularization parameter should be chosen such that the retrieved profile contains all vertical structures that influence the measurement, while the structures to which the measurement is insensitive should be filtered out. Such a value of the regularization parameter is found from the L-curve [60], which is a parametric plot of the first derivative norm versus the least squares norm (the curve shows an L-shaped corner). The regularization parameter corresponding to the corner of the L-curve yields a good balance between the two minimizations: a decrease of the regularization parameter would not improve the residual norm, but would lead to a strong increase of the first derivative norm; on the other hand, an increase of the regularization parameter would make the residual norm larger. Therefore, when using this value of the regularization parameter, all profile information that is present in the measurement is retrieved, while the part of the profile to which the measurement is insensitive (e.g., fine-scale structures in the profile) is filtered out.
The SRON inversion algorithm stems from calculations performed by a forward model based on the Gauss-Seidel iteration technique, which fully includes multiple-scattering and polarization, described in detail in the paper of Hasekamp and Landgraf [62]. The inclusion of polarization in the radiative transfer calculations reduces errors of up to 10% with respect to the commonly used scalar simulations, which generally neglect the polarization properties of light. The SRON algorithm can be therefore directly applied to the GOME measurements, which are thus not corrected for polarization [63].
The algorithm includes some additional fit parameters, such as the Lambertian-surface albedo and the amplitude of a pre-calculated Ring spectrum. Also the effect of clouds is accounted for, by using the independent pixel approximation, which separates the radiative transfer calculations for cloudy and cloudless scenes. Values for the cloud fraction and cloud-top height are taken from the Fast Retrieval Scheme for Cloud Observables (FRESCO).

4.5.2.3 Data Assimilation Approach

In the operational phase, GOME is primarily used to retrieve total ozone column densities from a spectral window around 330 nm, using the Differential Optical Absorption Spectroscopy (DOAS) technique [64] [16]. In order to derive ozone profiles and a daily global 3 dimensional (3D) ozone analysis, the column observations are assimilated into a 3D chemical-transport model (CTM). While the CTM is driven by meteorological wind and temperature fields, GOME observations are sequentially assimilated into the model using an optimal interpolation scheme [65]. The vertically integrated total column contents of the model are considered as the first-

guess values. The analyzed column values are then vertically distributed, weighted by the corresponding (first-guess) model profile (expressed in ozone mixing ratios). The assimilation scheme accounts for time of observation, for spatial weighting between observation and grid, and for model and observation errors. By applying this method, a global synoptic 3D ozone analysis is available every 6 hours. Note that, unlike methods presented above, this approach does not use the profile information contained in GOME spectra; it can be considered an interesting addition, as it represents the complete a-priori knowledge of the ozone vertical distribution, considering all relevant chemical and physical processes, and the meteorological analyses.
For the assimilation approach, the German Aerospace Center (DLR) used the 3D global CTM called ROSE/DLR, based on the ROSE model described in detail in [66]. The model covers all relevant gas-phase stratospheric chemical processes. Heterogeneous processes on Polar Stratospheric Clouds (PSCs) and on sulfate aerosols are also included in the model. It accounts for about 100 reactions, including oxygen, hydrogen, carbon, nitrogen, chlorine and bromine species. The chemical rate constants and cross sections are taken from [67]. Chemical rate equations are solved by considering an equilibrium state for short-lived species (e.g., ClO, NO, HO, and BrO) and a semi-implicit scheme for the integration of more stable reactants (e.g., HNO_3, N_2O, and CH_4). Short-lived species are grouped and integrated using families (e.g., ClO_x and NO_x), while long-lived species are transported using the semi-Lagrangian scheme [68]. Wind and temperature fields are derived from 24-hour analyses of the UKMO [36]. For the assimilation of GOME total ozone column observations, the model is run with a 5.6° x 5° longitude-latitude spherical discretization. It covers the stratosphere from 8 to 56 km, using 37 levels with a vertical step size of 1.3 km.
The optimal interpolation applied for sequential data assimilation considers the time of observation, the spatial weighting between observation and grid, the model errors, and the observation errors. At each assimilation time step, the model's volume mixing ratios are integrated to total column values, which are then interpolated to the observation grid, i.e. the geo-location of GOME total column observations. Observational increments (i.e., departures from the model) are then determined, and a linear-weight matrix operator (or gain operator) transforms the resulting anomalies back to the model space [69]. Finally, the analyzed total columns are vertically redistributed, weighted by the first-guess model profiles. Model's first-guess and GOME observation errors are set to 18% and 4%, respectively; error covariances are parameterized by hyperbolic functions depending on the horizontal distance between the model grid-point and the observation [70]. The basic time step of the CTM is one hour, and therefore all GOME observations within this interval are binned. A correction for the model bias is applied offline, and it is based on zonal-mean seasonal comparison results with SAGE-II data [71].

4.6 RAL retrieval scheme

The Rutherford Appleton Laboratory (RAL) developed a 3-step scheme to retrieve ozone profiles spanning troposphere and stratosphere [27] [72]. This section will describe in detail the retrieval algorithm designed for GOME flight-data, as the obtained results play a key-role in setting up the neural network retrieval algorithms presented in Chapter 5. The fundamentals of the proposed scheme, i.e. the selection of spectral ranges, the requirements for accurate a-priori knowledge on temperature profiles and the definition of the state vector, stem from some basic analysis of the ozone retrieval from UV-visible back-scattered light measurements in nadir viewing geometry.

4.6.1 Forward model parameters

The RAL inversion scheme is essentially based on the OE approach. Forward model simulations have been carried out with GOMETRAN, which generates monochromatic spectra and

derivatives with respect to ozone (and all known absorbers) in the spectral range 240-740 nm and with a spectral sampling of 0.02 nm. The atmosphere was modelled from 0 km to 100 km, with computational levels every 1 km.
The instrument resolution was assumed to be 0.24 nm, 0.26 nm, 0.44 nm and 0.48 nm in Band 1, Band 2, Band 3 and Band 4 respectively. An integration time of 12 s was assumed below 283 nm and 1.5 s above, as it is currently the case for GOME. Ring effect was neglected, as it showed negligible impact during basic diagnostics performed before the model runs.
Profiles of atmospheric constituents, temperature, pressure and, implicitly, the Rayleigh scattering profile are taken from the Atmospheric Composition Explorer for CHEMistry (ACECHEM) study [73] [74], in order to be representative of mid-latitude conditions.
Accurate profiles of temperature and pressure are required to model the Rayleigh scattering and to extract the information from the Huggins bands temperature dependence; moreover, errors arising from the representation of retrieved profiles in mixing ratio units would be minimised (the measurements are more fundamentally related to absorbed number density).
Temperature and geo-potential height on pressure levels are taken from the United Kingdom Meteorological Office (UKMO) stratospheric analyses [75], obtained from the British Atmospheric Data Centre (BADC). Data are available at a spatial sampling in longitude and latitude of approximately 3.75° x 2.5° respectively, and extends to approximately 55 km in altitude.
For each GOME ground pixel, the nearest UKMO temperature and pressure profiles to the pixel centre are selected and extended to 100 km. The background aerosol profile is assumed from MODTRAN.
Vertical grids are defined for the retrieval state vector and for the radiative transfer model computational levels. To minimise changes in the scheme as it is applied globally, the same sets of levels are always used. Retrieval levels are defined in terms of pressure, so as to follow the variation in tropopause height more closely than geometric altitude. They are referred to in terms of a scale-height in km, referred to as Z*:

$$\frac{Z^*}{km} = 16\left(3.0 - \log_{10}\left(\frac{p}{hPa}\right)\right) \tag{4.5}$$

where p is pressure expressed in hPa. Z* is comparable to geometric altitude at mid-latitudes. Ideally, retrieval levels should be chosen according to:

- widths of averaging kernels;
- atmospheric region of interest (boundary layer, tropopause, stratopause etc.)

In practise, defining levels purely in this way is problematic since the requirements may be contradictory and vary as a function of space and time. Here levels are defined to be:
Z* = 0, 6, 12 km, then at 4 km intervals up to 80 km (corresponding to the pressure levels of approximately 1000, 422, 177, 100, 56, 32, 18, 10, 5.6, 3.2, 1.8, 1.0, 0.56, 0.32, 0.18, 0.10, 0.056, 0.032, 0.018, 0.01 hPa). These levels over-sample the resolution expected on the basis of averaging kernel analysis.

4.6.2 Retrieval model parameters

The measurement vector (**y**) is composed of sun-normalised radiance propagated through an instrument model to units of detected electrons. The weighting functions matrix is treated similarly.

The covariance matrix ($\mathbf{S_y}$) is obtained according to the noise model developed on the base of dedicated analysis [72]. The throughput used here (assumed constant within a channel) is based on [3], and varies from 1 to 3 x 10^{-6} electrons / (photon/s/cm^2/sr/nm).
The state vector ($\mathbf{x}$) is composed of the natural logarithm of O_3 mixing ratios, the surface albedo and the temperature profile. O_3 is represented here on 6 km levels, assuming linear (in log) interpolation between. The square-root of the diagonal elements of the state vector covariance matrix ($\mathbf{S_x}$) gives the expected a-posteriori precision of the estimated state. This vector is referred to here as the retrieval estimated standard deviation (ESD).
Temperature profile, reported on a 2 km grid, is included in the state vector, so that a-priori errors on it can propagate into the O_3 ESD (it is not expected that independent information on temperature exists). The intention here is to account for the importance of the ozone cross-section temperature dependence.
Conservative assumptions are made about a-priori knowledge of surface albedo and its wavelength dependence; the state vector contains a separate entry for albedo at 10 nm intervals. In all cases, triangular basis functions are used to map continuous quantities (profiles and surface albedos) onto the forward model computational grid.
Values of the a-priori profile ($\mathbf{x_a}$) are taken from the Fortuin climatology, interpolated in altitude to the retrieval grid. The covariance matrix ($\mathbf{S_a}$) of the a-priori vector is used to constrain the profile shape and it is assumed to be diagonal. A-priori errors (i.e. the square-root of diagonal elements of $\mathbf{S_a}$) are specified to be 100% (relative to the a-priori profile) for O_3. A-priori errors equal to 1 are assumed for the surface albedos, while the a-priori errors on temperature are varied to investigate their effect on the O_3 ESD. The elements of $\mathbf{S_a}$ are computed assuming a correlation length of Δz_c = 6 km. Some deviations from the Fortuin climatology are imposed, in order to minimise the appearance of spurious spatial/temporal patterns in retrievals at tropospheric altitudes (due to a-priori influence) and to avoid too tight a-priori constraint.

4.6.3 Measurement error: noise-floor

Measurement errors are simulated here assuming photon-noise (Δy_{phot}). In practice, other errors will often dominate. To represent such errors in a simple way, a noise-floor (Δy_{min}) is introduced, in order to define the minimum measurement error represented by $\mathbf{S_y}$:

$$\sqrt{(\mathbf{S_y})_{ii}} = \max\left(\Delta y_{phot,i} \Delta y_{\min,i}\right) \tag{4.6}$$

It is noted that a more natural way to introduce these errors would be to combine $\Delta y_{phot,i}$ and $\Delta y_{min,i}$ in a root-sum-squared fashion; however, the above approach is adopted in the retrieval scheme described in this section.
The RAL retrieval scheme makes use of the estimated random error on measurements provided by the GDP extraction software. In any cases, $\mathbf{S_y}$ is assumed diagonal, and noise-floors (upper limits on the fitting precision) are imposed.

4.6.4 Retrieval analysis

A preliminary analysis aiming at selecting suitable spectral ranges and algorithm parameters was carried out. Several retrieval configurations were tested, in order to illustrate the potential information offered by GOME spectral measurements. The standard configuration assumes a noise-floor of 0.01%, and a temperature a priori error of 1 K. Four different configurations were obtained, respectively, assuming a noise-floor of 0.1% and 1%, a temperature a priori error of 100 K, and a known surface albedo (i.e. not including it in the state vector).
As reported in Figure 4.4, results of the retrieval analysis shows the potentialities of three main ozone-absorption bands:

1. In the Hartley band (between 240 and 310 nm) the vertical optical depth decreases strongly with wavelength. Consequently the penetration depth of solar photons increases and the sensitivity to O_3 perturbations extends to progressively lower altitude. The gradient in penetration depth is sharpened by the λ^{-4} dependence of Rayleigh scattering. Increasing the noise-floor up to 1% has a little impact on the ESD in this region.

2. The Huggins bands (305-335 nm) contribute information extending down to the surface. This information is greatly reduced if the noise-floor is greater than 0.1%. Similarly, if temperature is not constrained, the ESD at low altitudes degrades greatly, indicating the origin of this O_3 information to be predominantly via temperature dependence of the Huggins bands, rather than their absolute opacity. The magnitude of the temperature error (100 K) modelled here is chosen to be consistent with the surface-tropopause temperature difference.

3. The Chappuis bands above 400 nm add some information to that from the Hartley and Huggins bands, particularly on the troposphere, but only if the surface albedo is well known.

Figure 4.4 shows averaging kernels (with respect to true perturbations on a uniform 2 km grid) for the standard state vector configuration as above, for 3 noise-floors and for 3 wavelength coverage options, corresponding approximately to the inclusion of the Hartley, Huggins and Chappuis bands respectively. Kernels for the best case (bottom-left panel), i.e. 0.01% of noise-floor and full spectral coverage, have widths (indicative of vertical resolution) of 5.5 km to approximately 8 km, spanning the range below 40 km. Omission of the Chappuis bands (moving to the centre-left panel) reduces sensitivity of the lowest retrieval level slightly. The retrieval still resolves the lower troposphere at the 0.1% noise-floor (centre panel) but not at 1% (centre-right). If the Huggins bands are not included (top panels) then no resolution within the troposphere is possible at the 0.1% or even 0.01% noise level. At 1% noise level virtually no information exists on the profile below 20 km from wavelengths below 300 nm.
The implemented scheme focuses on the Hartley-Huggins bands below 350 nm. The range is spanned by Band 1A (235-307 nm), Band 1B (307-316 nm), Band 2A (311-312 nm) and Band 2B (312-405 nm). The information content of the Chappuis bands has not been exploited here, because of its strong retrieval sensitivity to the errors in radiometric calibration and to the correct modelling of the wavelength dependencies of atmospheric optical properties [5].
Band 1A spans most of the Hartley Band and contains most of the height-resolved information on the stratosphere. This Band has a 6 times larger Field Of View (FOV) with respect to the other Bands (960x80 km^2 compared to 320x40 km^2 in wide-swath mode, 240x80 km^2 compared to 80x40 in narrow-swath mode). Furthermore, the requirements of the retrieval in terms of fitting precision are different in the Hartley compared to the Huggins bands. To obtain this fitting precision requires a quite different approach to the forward modelling, and it was decided to separate the retrieval process, first fitting the Band 1A spectrum, then using the result to define the a-priori for subsequent retrievals from each of the smaller Band 2 ground-pixels, encompassed by the Band 1A pixel.
The FOV difference between the two bands means, in principle, it may not be possible to fit a consistent single profile to match the observations in both ranges. In practise, one can suggest that the altitude range to which Band 1A is sensitive will be characterised by ozone variability at larger horizontal scales than Band 2, where modulation by lower stratospheric and tropospheric ozone, as well as cloud will be more significant. According to this, the part of the profile measured by Band 1A should be reasonably consistent with that part of the profile over the smaller Band 2 FOV. Nevertheless, it is possible that this inconsistency might arise in practise.

Although this inconsistency problem would be largely removed, it is not desirable to average the Huggins Bands measurements up to the larger pixel size of Band 1A (for reasons regarding cloud coverage), and to maximise the sensitivity to small spatial scale lower-stratospheric and tropospheric ozone perturbations.
In the wavelength range spanned by Band 1B and Band 2A there is a relatively large spectral variation and consequent uncertainty in the fractional polarisation, which can lead to errors of the order of a few percent in sun-normalised radiance. There is a trade-off between the improvement in ESD from including as much of this range as possible and the mapping of polarisation errors. After some experimentation, Band 1B and Band 2A were excluded from the retrieval. Measurements below approximately 265 nm have been shown to improve the retrieval only in upper stratosphere and above. It is also known that this spectral range is particularly prone to errors in dark current, stray-light and is likely to be most subject to instrument degradation over time, and it was discarded. From the remaining Band 1 range, the following sections are ignored to avoid strong Fraunhofer lines (particularly sensitive to errors in modelled leakage current, wavelength calibration and Ring effect) and the NO gamma-bands: 265-269 nm, 278.2-280 nm, 284-286.4 nm and 287.2-288.8 nm.
In order to fit the Huggins bands to the required accuracy (< 0.1%), it is necessary to model the Ring effect and under-sampling. A pre-requisite of such a model is an accurate knowledge of the slit function and the wavelength registration relative to the solar reference spectrum used in the model. Pre-flight spectral calibration of the instrument was insufficient for this purpose and, as previously noted, the on-board calibration lamp does not have adequate lines in the required spectral region. Instead, the scheme developed here attempts to derive the required parameters by fitting the GOME measured solar spectrum to a high-resolution solar reference spectrum [76]. After much experimentation, this problem was overcome and fitting precision better than the required 0.1% in the Huggins bands was achieved, as follows:

- The fitting region was restricted to 322.5 to 334 nm: below this range the fit to the solar reference spectrum shows gross changes in spectral resolution and wavelength calibration. Fit residuals are also larger. The difficulty in fitting sun-normalised radiance below 322 nm is attributed partly to the degradation in spectral performance of the instrument, which cannot be adequately accounted for by the current instrument model.
- Since Band 2 is primarily of interest for the relatively fine-scale temperature dependent structure, the measurements in Band 2B are treated in a manner analogous to DOAS technique [16]. The logarithm of the sun-normalised radiance (**R**) is taken and a polynomial subtracted:

$$\mathbf{R`} = \log_e(\mathbf{R}) - P_2(\lambda, \log_e(\mathbf{R})) \qquad (4.8)$$

  where $P_2(\mathbf{x},\mathbf{y})$ is a 2^{nd} order polinomial fit to a function with values **y** at abscissae **x**.This removes, to a large degree, independent information on the surface reflectance which modulates the mean layer photon-path profile. It is therefore important to specify (not retrieve) an accurate surface albedo as a forward model parameter in this retrieval step. This is obtained from a separate retrieval from measurements in the Huggins absorption minima between 335 and 340 nm. It is assumed that this range is close enough in wavelength to the Band 2 range used for O_3 retrieval that the albedo is appropriate, while being sufficiently insensitive to O_3 absorption that the Band 1 fitted profile can be assumed for the Band 2 albedo fit.
- After restricting the spectral range and adopting the quasi-DOAS approach above, systematic residuals remained at the 0.2% level (in sun-normalised radiance). The retrieval and FM were then modified to allow this pattern to be added to simulated measurements, scaled by a retrieved parameter.

In summary, the proposed retrieval scheme consists of three steps:

1. O_3 profile retrieval from Band 1A
2. Surface albedo retrieval from Band 2B (above 335nm) assuming the Band 1A O_3 result as a-priori
3. O_3 profile retrieval from Band 2B (below 335nm), using the Band 1A O_3 result as a-priori, and assuming the fitted surface albedo

As a pre-cursor to step (3), spectral calibration is performed by fitting the daily GOME solar reference spectrum.

4.6.5 Retrieval performance

4.6.5.1 Ground-pixels processed

The following criteria are applied in selecting Band 1A pixels for processing:

- Standard integration time of 12s
- Line-of-sight zenith angle lower than 30° (i.e. polar scan pixels are ignored)
- Pixels grossly affected by the South Atlantic Anomaly (a systematic error produced by GOME detectors) are discarded.

4.6.5.2 Error analysis

A detailed error analysis has been carried out, in order to assess the main error sources pertaining to the retrieval scheme described in previous sections. The approach adopted was as follows:

- A set of observational and atmospheric scenarios was defined to encompass, so far as practicable, the conditions sampled. These are collectively referred to as *geo-temporal* scenarios.
- These scenarios were used as linearization states about which to compute the sensitivity of the retrieved profile to measurement perturbations.
- A set of fundamental, limiting error sources was identified.
- The “linear mapping” approach was adopted to quantify the propagation of the errors sources into the retrieved state vector for the each observational / atmospheric scenario.
- The mapped errors were combined in a root-sum-squared manner to define a base-line error budget for the profile retrieval.
- By reference to this error budget, the significance of specific aspects of the retrieval scheme and its underlying assumptions were assessed.

The most important findings of the error assessment described above are summarised as follows:

1. The retrieval provides useful information on the O_3 profile below 50 km.
2. Retrieval precision, accounting for measurement noise and other quasi-random errors is expected to be generally in the few-percent range in the stratosphere, increasing to a few 10s of percent in the lowest retrieval levels.
3. Retrieved quantities should be interpreted as estimates of layer-averaged O_3 number density, taking into account the shape of the averaging kernels, and the influence of the a-priori.
4. As intended, the retrieval grid is seen to over-sample the inherent resolution measured by the averaging kernels. Nevertheless, the lower retrieval levels still provide meaningful estimates of tropospheric and lower-stratospheric column densities.
5. The instrumental and RTM errors are generally relatively small, compared to the climatological variance and, in most cases, the Estimated Standard Deviation. Exceptions

are radiometric gain errors including scan-mirror degradation (which has most impact above 40 km) and possibly imperfect knowledge of slit function shape (expected to cause a significant negative bias in the troposphere, though the magnitude is difficult to quantify).

6. High perturbations in aerosol and errors in the assumed temperature profile give rise to retrieval errors in the troposphere of order 10-20%. The temperature error is larger at high solar zenith angle.
7. Radiative transfer model approximations in the retrieval scheme are seen to be adequate, though the number of computational layers used in the tropics could be further optimised. At present a positive bias of 20-30% is predicted between 6 and 12km.

Accounting for a comprehensive range of error sources, information spanning the troposphere and stratosphere is confirmed to be recoverable from GOME measurements using the retrieval scheme described in the previous section.

4.6.6 Validation

Ozone profiles retrieved by the 3-step scheme developed at RAL have been compared to independent, co-located data. The principal focus of this validation work has been the comparison to ozone-sonde measurements, since these provide reliable, high precision observations spanning the altitude range most relevant to the potential application of GOME data [77]. Additionally comparisons have been made with observations from ground-based lidar [78] and the UARS Microwave limb sounder [79].

4.6.6.1 Comparison with sonde/lidar data

Ozone-sondes provide reliable measurements of high vertical resolution from the surface to approximately 25-30 km, depending on balloon burst altitude. Other than the restricted altitude range, the main limitation of the record is its irregular sampling in space and time (large regions having no coverage). To extend the altitude coverage, ground based lidar data, providing information from approximately 10-50 km, are also used. Ozone-sonde and lidar data are treated here in exactly the same way. For the purposes of validation, sonde (and lidar) measurements have been obtained from the following data-centres:

- Nadir data-centre at the Norwegian Institute for Air Research (NILU)
- The WMO World Ozone and UV data-center (WOUDC)
- The Southern Hemisphere ADditional OZonesondes (SHADOZ) archive

The true ozone field is characterised by vertical structure at a resolution higher than that which can be captured by GOME, but which is better represented in the high resolution sonde (and lidar) data (see Figure 2.1). Direct comparison of mixing ratios with sonde/lidar values at the retrieval grid points is therefore likely to lead to a comparison dominated by profile representation errors (i.e. the variability of the true ozone profile over the vertical resolution scale of the retrieval).
The following approaches are adopted to provide a more meaningful comparison:

1. Both sonde and retrieved *number density* profiles are integrated over the layers defined by the retrieval grid. These integrated *sub-columns*, (from retrieval and from sonde) are compared.
2. The sonde profile is used as the “true” profile in equation (4.3), in order to obtain a transformed sonde profile, according to the same observation system used for the retrieval. This version of the sonde profile is then compared to the retrieved profile: this indicates the extent to which the retrieval is consistent with the sonde profile, accounting

for the averaging kernel and a-priori influence, but assuming the retrieval is linear over the range spanned by the solution state and the true profile. The smoothed sonde profile is compared also to the unsmoothed sonde profile, to provide closure by indicating the extent to which the sub-column comparison with the retrieval (see point 1) should behave on the basis of the measurement information, but in the absence of errors (instrumental or otherwise).

To determine the level of agreement between a set of *N* coincident sonde and retrieved sub-columns, the relative difference is formed at each level, for each pair of observations; the mean and standard deviation is then taken over all coincidences. To quantify the improvement of the retrieval over the a-priori, the same statistics have been evaluated, replacing the retrieved profile with the a-priori. Except in the tropics, bias is generally < 10% above 6 km; below 6 km there is a strong negative bias (~ -30%) in the Northern Hemisphere, while the bias at low altitude becomes more positive in the tropics and Southern Hemisphere. The standard deviation is < 10% between 20 and 40 km, except at high latitude; at lower altitude, the standard deviation reaches a maximum of 56% in the tropical troposphere, but it is significantly better in the Northern Hemisphere mid-latitudes. Importantly, there is significant improvement over a-priori standard deviation at all altitudes where the sonde profile itself has greater 5% variability about the climatology.
At Southern Hemisphere high-latitudes it is noted that both standard deviation and bias are relatively poor, compared to other regions, though the improvement over climatology is very significant. Remaining discrepancies are attributable to the limited resolution of the retrieval. Cloud, as measured by ICFA, is shown to influence retrievals in the expected fashion, causing a negative bias which propagates upward as cloud altitude increases.

4.6.6.2 Comparison with UARS MLS data

UARS MLS offers the potential to validate GOME columns to higher altitude and with a more uniform global / temporal coverage than possible with ozone-sondes or lidar observations. The quality of the MLS observations themselves is well known. The main limitations from the point of view of validating GOME are the differences in orbit/viewing geometry, and the fact that the MLS V4 O_3 data used here is expected to be reliable only above 46 hPa ($Z^* = 21$km). Due to the differences in viewing / orbit geometry, the comparison has been performed in terms of zonal mean ozone. In the common altitude range within which both instruments have reasonable precision (25-50 km), the agreement in terms of zonal mean and standard deviation in the zonal mean is good, with a relative difference generally within 10%.

References

[1] J.P. Burrows, M. Weber, M. Buchwitz, V. Rozanov, A. Ladstatter Weissenmayer, A. Richter, R. DeBeek, R. Hoogen, K. Bramstedt, K.U. Eichmann, M. Eisinger, The global ozone monitoring experiment (GOME): Mission concept and first scientific results, *Journal of the Atmospheric Sciences*, 56 (2), 151-175, 1999.

[2] E. Hecht, Optics, Addison-Wesley, 1998.

[3] A. Van der A, SCIAMACHY: Observational noise and errors on atmospheric profiles, *PhD Thesis,* Rijksuniversiteit Leiden, The Netherlands., 1997.

[4] B. Aberle, W. Balzer, A. von Bargen, E. Hegels, D. Loyola, R. Spurr, GOME Level 0 to 1 Algorithms Description, *DLR Technical Report ER TN DLR GO 0022*, Iss./Rev. 5/A, 2000.

[5] B.J.K. Kerridge, R. Siddans, B.L. Latter, J.P. Burrows, M. Weber, R. De Beek, I. Aben, W. Hartman, GOME-2 Error Assessment Study, Final Report EUMETSAT Contract No EUM/CO/01/901/DK, 2002.

[6] I. Aben, GOBELIN - GOME BBM instrument characterization, Final Report, *SRON Tech. Rep., RP-GOB-001SR/99*, 1999.

[7] I. Aben, GDAQI Final Report (GOME Data Quality Improvement), *SRON Tech. Rep., TN-GDAQI-003SR/2000*, 2000.

[8] M. Eisinger, J.P. Burrows, A. Richter, Studies on the Precision of GOME Irradiance and Radiance Products and GOME Measurements of OClO and BrO over Antarctica, GOME Geophysical Validation Campaign, ESA WPP-108, 93-108, 1996.

[9] M. Eisinger, In-orbit monitoring of GOME polarisation sensitivity, ESAMS Workshop in ESTEC, ESA WPP-161 (2), 665-668, 1999.

[10] R. Snel, In-orbit optical path degradation: GOME experience and SCIAMACHY prediction, *Proc. ERS-Envisat Symposium, Gothenburg, ESA SP-461*, 2001.

[11] C.P. Tanzi, R. Snel, I. Aben, Degradation of Observations in the UV of the Global Ozone Monitoring Experiment (GOME), *IRS 2000: Current Problems in Atmospheric Radiation, Deepak*, 2001.

[12] B. Latter, Investigation of UV radiative transfer and its impact on ozone profile retrieval for the Global Ozone Monitoring Experiment, *PhD thesis. University of Reading, in preparation*, 2003.

[13] C.P. Tanzi, I. Aben, S. Slijkhuis, Influence of GOME in-flight degradation on earth radiance measurements, *ESAMS Workshop in ESTEC, ESA WPP-161 (1),* 681–685, 1999.

[14] H. Hegels, S. Slijkhuis, GOME: Correction of Degradation and Update of Key Data, *ESAMS Workshop in ESTEC, ESA WPP-161 (2),* 695-700, 1999.

[15] A. Dehn, Long-term monitoring of the GOME thermal environment, *ESA Technical Note ERSE-SPPA-EOAD-TN-02-0024*, 1998.

[16] R. Spurr, W. Thomas, GOME Level 1 to 2 Algorithms Description, *DLR Technical Report ER-TN-DLR-GO-0025,* Iss./Rev. 2/B, 2000.

[17] A. Richter, J.P. Burrows, Tropospheric NO2 from GOME measurements, *Advances In Space Research*, 29 (11), 1673-1683, 2002.

[18] A.N. Maurellis, R. Lang, W.J. van der Zande, I. Aben, W. Ubachs, Precipitable water column retrieval from GOME data, *Geophysical Research Letters,* 27 (6), 903-906, 2000.

[19] S. Noel, M. Buchwitz, H. Bovensmann, J.P. Burrows, *Retrieval of total water vapour column amounts from GOME/ERS-2 data*, 29, (11), 1697-1702, 2002.

[20] M. Eisinger and J.P. Burrows, Tropospheric sulfur dioxide observed by the ERS-2 GOME instrument, *Geophysical Research Letters,* 25, (22), 4177-4180, 1998.

[21] R. Guzzi, G. Ballista, W. Di Nicolantonio, E. Carboni, Aerosol maps from GOME data, *Atmospheric Environment,* 35 (30), 5079-5091, 2001.

[22] J.P. Veefkind, G. de Leeuw, P. Stammes, R.B.A. Koelemeijer, Regional distribution of aerosol over land, derived from ATSR-2 and GOME, *Remote Sensing of Environment,* 74 (3), 377-386, 2000.

[23] R. Hoogen, V.V. Rozanov, J.P. Burrows, Ozone profiles from GOME satellite data: Algorithm description and first validation, *Journal of Geophysical Research - Atmospheres,* 104 (D7), 1999.

[24] R.J. Van der A, R.F. van Oss, A.J.M. Piters, J.P.F. Fortuin, Y.J. Meijer, H.M. Kelder, Ozone profile retrieval from recalibrated Global Ozone Monitoring Experiment data, *Journal of Geophysical Research - Atmospheres,* 107 (D15), 2002.

[25] O.P. Hasekamp, J. Landgraf, Ozone profile retrieval from backscattered ultraviolet radiances: The inverse problem solved by regularization, *Journal of Geophysical Research - Atmospheres*, 106 (D8), 8077-8088, 2001.

[26] M.D. Muller, A. Kaifel, M. Weber, J.P. Burrows, Neural network scheme for the retrieval of total ozone from Global Ozone Monitoring Experiment data, *Applied Optics,* 41 (24), 5051-5058, 2002.

[27] R. Munro, R. Siddans, W.J. Reburn, B.J.K. Kerridge, Direct measurement of tropospheric ozone distributions from space, *Nature*, 392 (6672), 168-171, 1998.

[28] Kozlov V., Design of experiments related to the inverse problem of mathematical physics, *Mathematical Theory of Experiment Design*, edited by C. M. Ermakov, Nauka, Moscow, 216–246, 1983.

[29] Rozanov V.V., D. Diebel, R. J. D. Spurr, J. P. Burrows, GOMETRAN: A radiative transfer model for the satellite project GOME, the plane-parallel version, *J. Geophys. Res.*, 102, 16683–16695, 1997.

[30] Bucholtz A., Rayleigh scattering calculations for the terrestrial atmosphere, *Appl. Opt.*, 34, 2765–2773, 1995.

[31] Kneiys F.X., E. P. Shettle, L. W. Abreu, J. H. Chetwynd, G. P. Anderson, W. O. Gallery, J. E. A. Selby, S. A. Clough, Users guide to LOWTRAN 7, *Tech. Rep. AFGL-TR-88-0177*, Air Force Geophys. Lab., Bedford, Mass., 1988.

[32] Hilsenrath E., J. Gleason, S. Janz, X. Gu, R. P. Cebula, K. Chance, R. Hoekstra, GOME calibration and validation using backscatter UV techniques, *Proceedings of the GOME Geophysical Validation Campaign Workshop (Frascati, Italy, 24–26 January 1996)*, *ESA WPP-108*, 1996.

[33] McPeters R.D., Climatology of nitric oxide in the upper stratosphere, mesosphere, and thermosphere: 1979 through 1986, *J. Geophys. Res.*, 94, 3461–3472, 1989.

[34] Tellmann S., V.V. Rozanov, M. Weber, J. P. Burrows, Improvements in the tropical ozone profile retrieval from GOME UV/VIS nadir spectra, *Adv. Space Res.*, 34, 739–743, 2004.

[35] Fortuin J.P.F., H. Kelder, An ozone climatology based on ozonesonde and satellite measurements, *J. Geophys. Res.*, 103, 31709 – 31734, 1998.

[36] Swinbank R., A. O'Neill, A stratosphere–troposphere data assimilation system, *Mon. Weather Rev.*, 122, 686–702, 1994.

[37] van Geffen J.H., R. F. van Oss, Wavelength calibration of spectra measured by GOME using a high-resolution reference spectrum, *Appl. Opt.*, 42, 2739–2753, 2003.

[38] Schutgens N.A.J., P. Stammes, A novel approach to the polarisation correction of space-born spectrometers, *J. Geophys. Res.*, 108, 10.1029/2002JD002736, 2003.

[39] van Oss, R. F., and R. J. D. Spurr, Fast and accurate 4 and 6 stream linearised discrete ordinate radiative transfer models for ozone profile remote sensing retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, 75, 177–220, 2002.

[40] Spurr R. J. D., T. P. Kurosu, K. V. Chance, A linearized discrete ordinate radiative transfer model for atmospheric remote sensing retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, 68 (6), 689–735, 2001.

[41] Chance K. V., R. J. D. Spurr, Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering and the Fraunhofer spectrum, *Appl. Opt.*, 36, 5224–5230, 1997.

[42] Koelemeijer R. B. A., P. Stammes, A fast method for retrieval of cloud parameters using oxygen A-band measurements from the Global Ozone Monitoring Experiment, *J. Geophys. Res.*, 106, 3475–3490, 2001.

[43] Bass A. M., R. J. Paur, The ultraviolet cross-sections of ozone: I. The measurements of atmospheric ozone, *Proc. Quadrennial Ozone Symp.*, D. Reibel Publ., Norwell, Mass., 606–610, 1984.

[44] Paur R. J., A. M. Bass, The ultraviolet cross-sections of ozone: II. Results and temperature dependence, *Proc. Quadrennial Ozone Symp.*, D. Reibel Publ., Norwell, Mass., 611–616, 1984.

[45] Orphal J., A critical review of the absorption cross-sections of O3 and NO2 in the 240–790 nm region, *J. Photochemistry and Photobiology A: Chemistry*, 157, 185–209, 2003.

[46] Bates D. R., Rayleigh scattering by air, *Planet. Space Sci.*, 32, 785–790, 1984.

[47] OMI Ozone Products, Vol. II of OMI Algorithm Theoretical Basis Document, OMI-ATBD-02, V. 2.0, NASA Goddard Space Flight Center, Greenbelt (MD), 2002.

[48] Chance K. V., J. P. Burrows, D. Perner, W. Schneider, Satellite measurements of atmospheric ozone profiles, including tropospheric ozone, from UV/Visible measurements in the nadir geometry: a potential method to retrieve tropospheric ozone, *J. Quant. Spectrosc. Radiat. Transfer*, 57, 467–476, 1997.

[49] Bauman J. J., P. B. Russell, M. A. Geller, P. Hamill, A stratospheric aerosol climatology from SAGE II and CLAES measurements: 1. Methodology, *J. Geophys. Res.*, 108, 4382, 10.1029/2002JD002992, 2003.

[50] Martin R. V., D. J. Jacob, K. Chance, T. P. Kurosu, P. I. Palmer, M. J. Evans, Global inventory of nitrogen oxide emissions constrained by space-based observations of NO2 columns, *J. Geophys. Res.*, 108, 4537, 10.1029/ 2003JD003453, 2003.

[51] Kurosu T. P., K. Chance, R. J. D. Spurr, CRAG-Cloud retrieval algorithm for ESA's GOME, *Proc. of the European Symp. on Atmospheric Measurements from Space (ESAMS-'99)*, WPP-161, ESA/ESTEC, Noordwijk, The Netherlands, 513–521, 1999.

[52] McPeters R. D., J. A. Logan, G. J. Labow, Ozone climatological profiles for Version 8 TOMS and SBUV retrievals, *AGU fall meeting*, San Francisco, 2003.

[53] Bhartia P. K., R. D. McPeters, C. L. Mateer, L. E. Flynn, C. Wellemeyer, Algorithm for the estimation of vertical ozone profile from the Backscattered Ultraviolet (BUV) technique, *J. Geophys. Res.*, 101, 18793–18806, 1996.

[54] Dave J. V., Meaning of successive iteration of the auxiliary equation of radiative transfer, *Astrophys. J.*, 140, 1292–1303, 1964.

[55] Ahmad Z. and P. K. Bhartia, Effect of molecular anisotropy on the Backscattered UV Radiance, *Appl. Opt.*, 34, 8309–8314, 1995.

[56] Joiner J., P. K. Bhartia, R. P.Cebula, E. Hilsenrath, R. D. McPeters, H. W. Park, Rotational-Raman Scattering (Ring Effect) in Satellite Backscatter Ultraviolet Measurements, *Appl. Opt.*, 34, 4513–4525, 1995.

[57] Philips P. L., A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Math.*, 9, 84–97, 1962.

[58] Tikhonov A. N., On the solution of incorrectly stated problems and a method of regularization, *Dokl. Akad. Nauk SSSR*, 151, 501–504, 1963.

[59] Hansen P. C., Regularization tools - A Matlab package for analysis and solution of discrete ill-posed problems, *Numer. Algorithms*, 6, 1–35, 1994.

[60] Hansen P. C., D. P. O'Leary, The use of the L-curve in the regularization of discrete ill posed problems, *SIAM J. Sci. Comput.*, 14, 1487–1503, 1993.

[61] Hasekamp O. P., J. Landgraf, Ozone profile retrieval from backscattered ultraviolet radiances: The inverse problem solved by regularization, *J. Geophys. Res.*, 106, 8077–8088, 2001.

[62] Hasekamp O. P., J. Landgraf, A linearized vector radiative transfer model for atmospheric trace gas retrieval, *J. Quant. Spectrosc. Radiat. Transfer*, 75, 221–238, 2002.

[63] Hasekamp O. P., J. Landgraf, R. van Oss, The need of polarization modeling for ozone profile retrieval from backscattered sunlight, *J. Geophys. Res.*, 107, 4692, 2002.

[64] Platt U., Differential optical absorption spectroscopy (DOAS), *Chem. Anal. Series*, 127, 27 - 83, 1994.

[65] Khattatov B. V., J. F. Lamarque, L. V. Lyjak, R. Menard, P. Levelt, X. X. Tie, G. P. Brasseur, J. C. Gille, Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, J. Geophys. Res., 105, 29135-29144, 2000.

[66] Rose K., G. Brasseur, A three-dimensional model of chemically active trace species in the middle atmosphere during disturbed winter conditions, *J. Geophys. Res.*, 94, 16387–16403, 1989.

[67] Sander S. P., A. R. Ravishankara, R. R. Friedl, W. B. De-More, D. M. Golden, C. E. Kolb, M. J. Kurylo, M. J. Molina, R. F. Hampson, R. E. Huie, G. K. Moortgat, Chemical kinetics and photochemical data for use in stratospheric modeling, *Eval. 13, JPL Publ.*, 1281–1298, 2000.

[68] Smolarkiewicz P. K., P. J. Rasch, Monotone advection on the sphere, An Eulerian versus semi-Lagrangian approach, *J. Atmos. Sci.*, 48, 793–810, 1991.

[69] Daily R., Atmospheric Data Analysis, *Cambridge University Press*, 1991.

[70] Riishøjgaard L. P., A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus*, 50A, 42–57, 1998.

[71] Wang H. J., D. M. Cunnold, L. W. Thomason, J. M. Zawodny, G. E. Bodeker, Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res.*, 107, 4691, 2002.

[72] Siddans R., Height resolved ozone retrievals from GOME, *PhD thesis*, Univ. Reading.

[73] B.J.K. Kerridge, W.J. Reburn, R. Siddans, S.L. Smith, P.D. Watts, J.J. Remedios, F. Lama, J.J. Barnett, D. Murtagh, J. Stegman, F. Merino, P. Baron, H. Roscoe, D. Hausamann, M. Birk, F. Schreier, B. Schimpf, M. Lopez-Puertas, J.M. Flaud, T. von Clarmann, G.P. Stiller, A. Linden, S. Kellman, M. van Weele, H. Kelder, P. van Velthoven, M. Gauss, I. Isaksen, D. Hauglustaine, C. Clerbaux, O. Boucher, Definition of Mission Objectives and Observational Requirements for an Atmospheric Chemistry Explorer Mission, *ESA Contract 13048/98/NL/GD, Final Report*, 2001.

[74] J. Langen, J. Fuchs, ACECHEM - Atmospheric Composition Explorer for CHEMistry and climate interation, *ESA Report for Assessment SP-1257(4)*, 2001.

[75] R. Swinbank, A. O'Neill, Quasi-biennial and semi-annual oscillations in equatorial wind fields constructed by data assimilation, *Geophys. Res. Lett.*, 21, 2099-2102, 1994.

[76] C. Caspar, K.C. Chance, GOME Wavelength Calibration using Solar and Atmospheric Spectra, *Proc. 3rd ERS Symp. on Space at the Service of our Environment (ESA SP-414)*, 609-614, 1997.

[77] Assessment of trends in the vertical distribution of ozone, *WMO SPARC/IO3C/GAW Ozone Research and Monitoring Project Report*, No. 43, 1998.

[78] Mc Dermid I.S., Godin S., Walsh T.D., Lidar measurements of stratospheric ozone and intercomparisons and validation, *Appl. Opt.*, 29, 4914-4923, 1990.

[79] L. Froidevaux, W.G. Read, T.A. Lungu, R.E. Cofield, E.F. Fishbein, D.A. Flower, R.F. Jarnot, B.P. Ridenoure, Z. Shippony, J.W. Waters, J.J. Margitan, I.S. McDermid, R.A. Stachnik, G.E. Peckham, G. Braathen, T. Deshler, J. Fishman, D.J. Hofmann, S.J. Oltmans, Validation of UARS microwave limb sounder ozone measurement, *Journal of Geophysical Research – Atmospheres*, 101 (d6), 10017-10060, 1996.

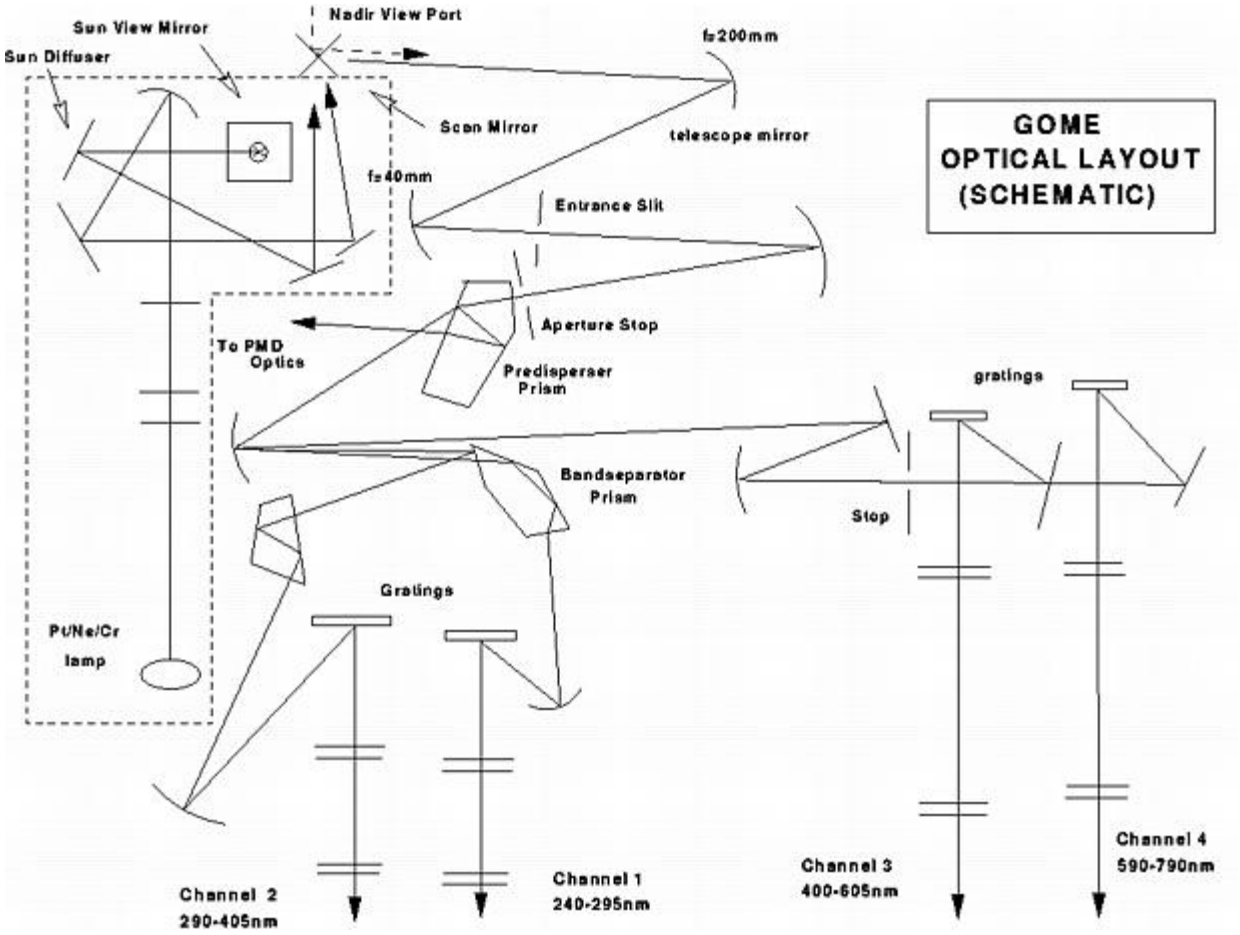

Figure 4.1. Optical layout of the GOME instrument.

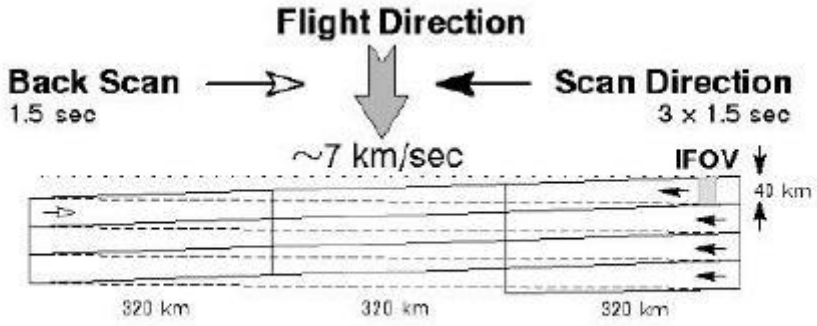

Figure 4.2. GOME principal observing mode.

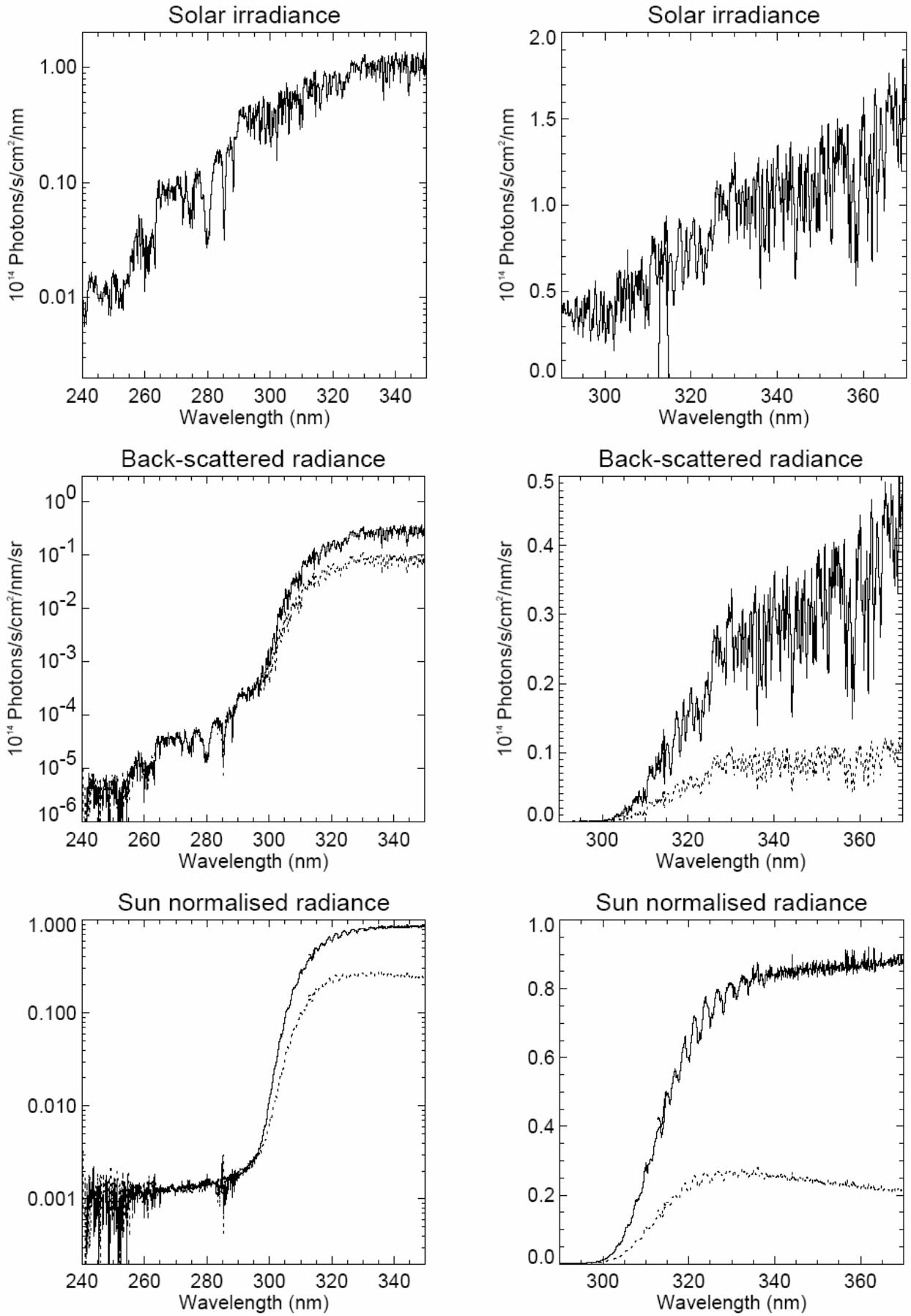

Figure 4.3. Solar irradiance, Earth radiance and sun-normalised radiance spectra measured from the GOME instrument.

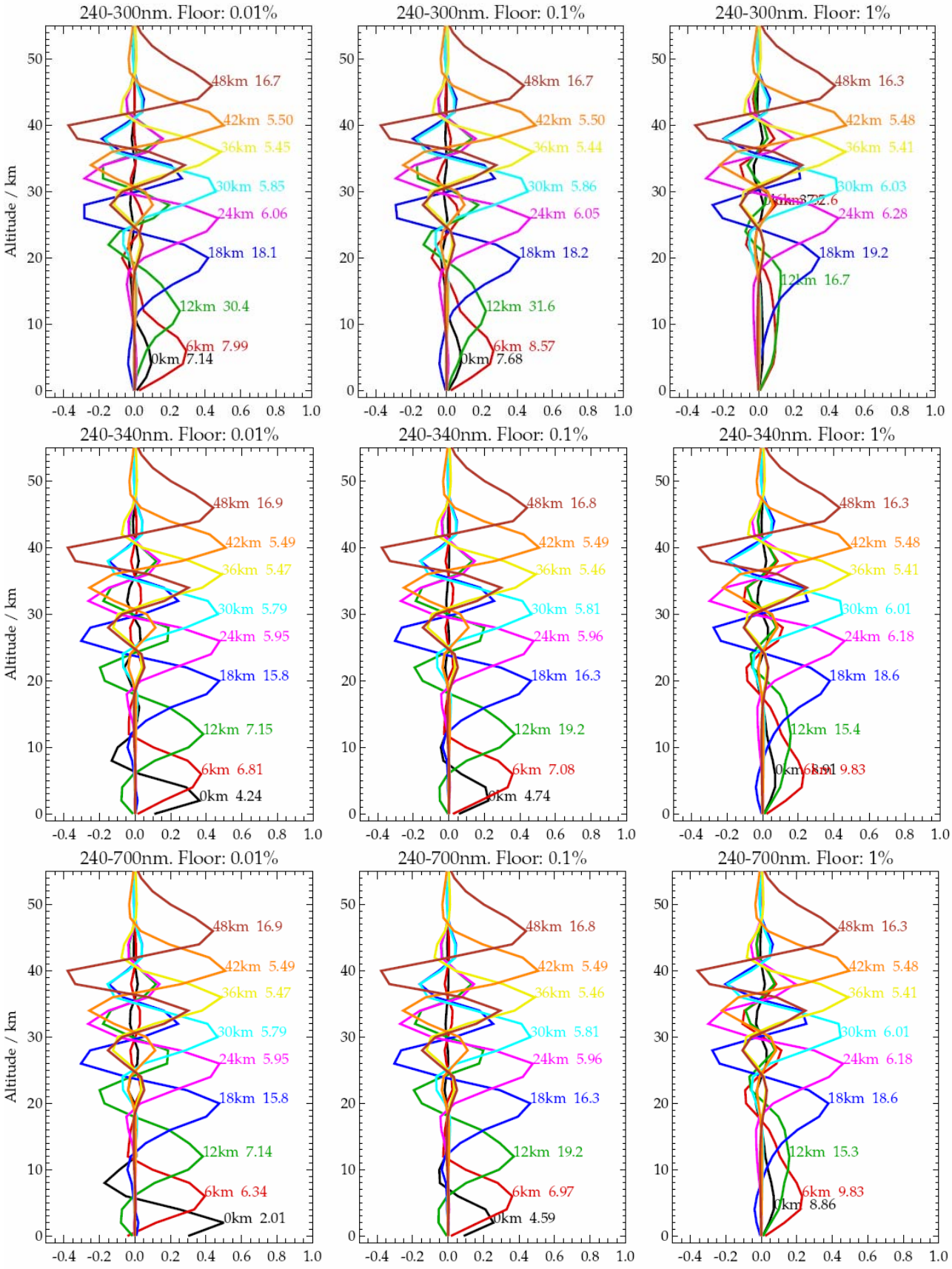

Figure 4.4. Averaging kernels for the RAL retrieval algorithm.

Chapter 5

Neural algorithms for the retrieval of ozone profiles

In order to design effective and reliable NN algorithms, different steps have to be carried out:

1. selection of input measurements
2. generation of a representative training set
3. NN training
4. network topology optimization
5. NN error estimation

In this chapter all these steps needed to design NN algorithms aiming at the retrieval of atmospheric ozone profiles from GOME will presented. In particular we will focuse on the problem of selecting the set of input measurements to be given to the net, as it is crucial for the effectiveness of the neural algorithm itself. Three different approaches have been analysed, and will be described in the following section. Beside traditional techniques, like physical assumption or Principal Components Analysis, a new methodology exploiting NN not only for inversion purposes, but also to automatically extract, from a large set of input measurements, the most suitable subset for the retrieval will be presented, and obtained results assessed.

5.1 Selection of input measurements

Theoretically, the most suitable spectral range to retrieve the information on the vertical distribution of the ozone, from observations of the light back-scattered by the atmosphere, lies between 240 and 340 nm, and includes both Hartley (231–300 nm) and Huggins (300–370 nm) O_3 absorption bands. The reason for this is the strong change of ozone absorption cross-sections in this spectral range, as well as the temperature dependence of ozone absorption in the Huggins band. The interval 240–340 nm is covered by Band 1 (240–315 nm) and Band 2 (313–340 nm) of the GOME instrument. Observations at Chappuis bands (450–800 nm), covered by Band 3 (400–605 nm) and Band 4 (590–790 nm) can be also considered, mainly to improve total column ozone information, but are in practice discarded for its strong sensitivity to radiometric errors. Whatever the inversion scheme might be, the management of hundreds or even thousands of measurements available from GOME Band 1 and 2 is not trivial, and procedures to constrain the information in a lower dimensional space would be recommendable. This is in particular true if the inversion is performed with a neural network approach. Selecting the inputs to the network, on the base of the effectiveness of their information content in estimating the output, eliminates unnecessary or misleading inputs that may confuse the network. Minimising the number of input measurements, without a significant loss of information, affects positively the NN mapping ability and computational efficiency. A network with fewer inputs has fewer adaptive parameters to be determined, which need a smaller training set to be properly constrained. This leads to a network with improved generalization properties and smoother mapping capabilities. In addition, a network with fewer weights may be faster to train. All these benefits make the reduction in the dimension of the input data a normal procedure when designing NN, even for a relatively low dimensional input space.

Dimensionality reduction can be achieved by means of different methods. Among them, the faster and simpler method is essentially based on physical assumptions, and the set of input measurements is selected manually by the algorithm designer. More sophisticated methodologies are based on mathematical analysis of the input dataset, such as the principal component analysis (PCA), also referred to as the eigenvector transformation, the Hotelling transformation or the Karhunen–Loéve (K–L) transformation, which have been recently proven to be useful in the application of NNs for inversion purposes. In this study we investigated the possibility to solve the dimensionality reduction problem by means of an automatic procedure, able to select the most significant wavelengths out of a wide spectral interval. According to this approach, which is rather general and therefore easily applicable to other retrieval problems in remote sensing, NNs can be a suitable tool not only in the inversion phase, but also in the dimensionality reduction phase.

5.1.1 Methodology based on physical assumption

GOME data used in this approach consist of solar irradiance and earth radiance spectra. Solar irradiance spectra are measured daily by GOME and are generally used as the reference light source spectra. The spectral interval selected for estimating ozone profiles is 321–325 nm. This range, characterized by a spectral calibration performed using a fourth-order polynomial, shrinks the frequency interval retrieval down to a subinterval of the GOME Band 2B, where the Huggins ozone absorption band is situated. The considered range has been chosen on the base of some attractive properties:

- higher resolution, with respect to Hartley band (GOME Band 1A, used by most of algorithms presented in Chapter 4), due to a shorter integration time;
- high signal-to-noise ratio (SNR);
- high temperature dependence of the ozone cross sections in the Huggins band below 340 nm [1];
- possibility to compute the ozone slant path content by the Temperature Independent Differential Absorption Spectroscopy (TIDAS) algorithm [2].

Indeed, even if it is true that in the spectral range covered by the GOME Band 1A ozone opacity is large (decreasing monotonically with increasing wavelength from the peak of the Hartley band at 245 nm), measurements in Band 2B still contain information on the ozone profile, as the opacities, though relatively small, may still have significant values. Moreover, the use of neural networks for inversion purposes is often effective because of their capability to simultaneously capture subtle dependencies and complex behaviour of the inversion function.
Once the spectral interval has been selected, earth radiance measurements underpass a normalization procedure, in order to eliminate as much as possible the effects of instrumental parameters and ozone absorption on the spectral shape.
For each wavelength belonging to the selected range, the intensity of the direct sun radiation, I_0, and the intensity of the solar radiation scattered by the atmosphere, I, both measured by GOME have been considered. These two quantities can be put into a relationship using a simplified expression of the Beer–Lambert law:

$$I = I_0 \, e^{-\tau_{O3}} f \tag{5.1}$$

where τ_{O3} is the ozone absorption optical thickness, and f accounts for all the atmospheric broadband radiative phenomena, such as surface albedo, Rayleigh, Mie, Raman, and multiple scattering and absorption by other atoms and molecules. The
logarithm of both terms in (5.1) leads to:

$$\ln \frac{I}{I_0} = -\tau_{O3} + \ln(f) \tag{5.2}$$

The left-hand term in (5.2) has the dimension of an optical thickness, which we can call – τ_{tot}. Moreover, the term ln(*f*) has a very weak wavelength dependence, at least in the range of interest. The mean value of τ_{tot}, $\langle \tau_{tot} \rangle$, calculated over the same range gives:

$$\langle \tau_{tot} \rangle = \langle \tau_{O3} \rangle - \langle \ln(f) \rangle \tag{5.3}$$

Defining $\tau`_{tot}$ as the difference between τ_{tot} and $\langle \tau_{tot} \rangle$, and assuming that the term ln(*f*) is wavelength independent, we can write:

$$\tau`_{tot} = \tau_{O3} - \ln(f) - \langle \tau_{O3} \rangle + \langle \ln(f) \rangle = \tau`_{O3} \tag{5.4}$$

After this transformation, all the quantities involved depend only on the atmospheric temperature and the ozone concentration. Given a certain reference temperature, the ozone slant absorption optical thickness, τ_{sl}, can be expressed as the product of the slant column density, ρ, and the O_3 absorption cross section, σ:

$$\tau_{sl} = \rho\, \sigma \tag{5.4}$$

In our case, the value of 303 K was considered for the reference temperature. A normalized optical thickness can be finally obtained for each wavelength by calculating the following ratio:

$$\tau_{norm} = \frac{\tau`_{tot}}{\tau_{sl}} \tag{5.5}$$

These normalised GOME measurements represent the quantities used as inputs to the neural network in the inversion phase. A number of 26 GOME wavelengths fall within the selected spectral range (321-325 nm), and the corresponding measurements will be used to train a NN to retrieve ozone profiles.

5.1.2 Sensitivity analysis and Extended Pruning procedure

In this section an alternative methodology to analyse which measurements, in the chosen context, are more sensitive to the ozone distribution is presented. Such an approach exploits a Radiative Transfer Model to perform a sensitivity analysis on the forward problem (how radiance measurements change with respect to ozone distribution variations), and a NN-based technique operating on the inverse problem (how estimating ozone vertical distribution on the base of radiance measurements) for the selection of spectral measurements - in the considered GOME spectral range - which might maximise the retrieval accuracy.
To this aim a short summary of NN training technique and pruning algorithms may be useful. As already shown in Chapter 3, the training phase consists in finding the weights and biases minimising the cost function that characterize the learning process. In this work the sum of square errors between the output values calculated by a NN and the corresponding target values has been considered. The minimisation has been performed by means a scaled conjugate gradient (SCG) algorithm.
The ultimate goal, however, is to generalise outside of the training set avoiding over-fitting situations, occurring when the NN learning has been to much polarised on the specific characteristics of the training set. For such a reason it is important to determine when the training procedure has to be stopped; this can be decided by considering the “early stopping” technique

[9]. According to this algorithm, the performance of the net during the training (learning) phase is evaluated either on the training set or on a different independent validation set. In the training set, the overall error in the retrieval of the correct output keeps on decreasing with the training, approaching a value of convergence. Conversely, the error on the validation set reaches a minimum value, after which it will start increasing if the process training is continued. This is the point to interrupt the learning phase.
Normally a pruning procedure is applied to a trained net to get its structural optimization. According to this kind of procedure, a network is examined to assess the relative importance of its weights, and the least important ones are deleted. Typically, this is followed by some further training of the pruned network, and the pruning and training may be repeated for several cycles. The most critical choice in the procedure implementation is how to decide which weights should be removed. To do this, we need some measure of the relative importance, or saliency, of different weights. In this work the simple concept that small weights are less important than large weights has been applied, and the magnitude of a weight value has been used as a measure of its importance. In order to select the most convenient inputs for a specific inversion problem, it is possible to prolong the pruning procedure to the input layer (we will call it "extended pruning procedure"). According to this method, the first inputs to be removed (we remind that an input or hidden unit is removed when it has lost all its connections) coincide with those containing less information; conversely, input units surviving to the extended pruning procedure contain the most of information needed for the retrieval, and are naturally selected for the inversion phase.
A general scheme of the proposed methodology is sketched in Figure 5.1. First of all, the simulation of the forward problem has been set up. This has been performed by using a LInearized Discrete Ordinate Radiative Transfer (LIDORT) model for the generation of earth radiances for a user-defined set of atmospheric variables [3]. The analysis of the complete discrete ordinate solution in a plane parallel multi-layered multiply-scattering atmosphere can be carried out with this model, making it possible the study of how perturbations in layer atmospheric quantities will translate into changes in single-scatter albedos and in the surface albedo, and consequently into changes in the radiance values measured from a satellite at the top of atmosphere. In a second step, a set of NNs have been trained, using as inputs radiance measurements generated by the model, and as reference outputs the corresponding ozone profiles (which in the previous step were the inputs to the forward model). On this set of NNs a first extended pruning procedure has been applied, leading to the selection of spectral sub-intervals more sensitive to the ozone distribution. In the following step, a new set of NNs has been trained using as inputs real GOME measurements belonging to the spectral sub-intervals selected in the previous step, and as reference outputs the corresponding ozone profiles already retrieved by RAL (see Chapter 4 for details). Finally, the extended pruning has been applied again, and a set of specific wavelengths to be used in actual the retrieval phase has been extracted.

Step 1 - Forward problem analysis
In this work the input to the model was determined by a set of atmospheric states generated starting by an assumed standard atmosphere, sub-divided into 201 height levels covering the range of 0-60 km, and changing the ozone vertical distributions and temperature profiles. The generation of atmospheric states, in which hydrostatic equilibrium for pressure and temperature profiles has been considered, mainly focused on the variability of the total ozone content, the vertical position of ozone concentration peak and the vertical thermal gradient, which are the causes most likely to affect radiance variability. In Figure 5.2 we over-plotted several ozone concentration profiles characterized by different vertical positions of the concentration peaks, while in Figure 5.3 we reported profiles characterized by a different total ozone content. In Figure 5.4 and Figure 5.5 the corresponding radiance differences between modified profiles and the reference profile (plotted with a black line) are shown. The examined wavelength range (290-335 nm) was limited by the availability of the ozone cross-sections data base in the LIDORT model [4] [5]. From these figures one can argue that the interval between 290 and 310 nm is the

most sensitive to variations related with the ozone distribution; nevertheless, it has to be noted that the absolute value of radiances within the analysed interval (290-335 nm) is very different, and that the relative variation of radiances – with respect to radiance values corresponding to the reference profile – is mostly important. If we consider the difference between radiances corresponding to modified profiles and radiances of the reference profile, but normalised to the latter, also the spectral range between 310 and 335 nm shows a not-negligible sensitivity, as clearly demonstrated by Figure 5.6 (for ozone profiles reported in Figure 5.2).
A similar analysis has been carried out modifying the temperature profile, which is intrinsically related to the ozone distribution. In Figure 5.7 we report an example where the thermal gradient in the troposphere has been modified (dashed line) with respect to the reference temperature profile (solid line), while in Figure 5.8 we show a profile with an omogeneous reduction of temperature values. In Figure 5.9 and Figure 5.10 the corresponding variation of radiance spectra are reported. From these plots we can clearly observe that the information on the temperature dependence of the ozone cross-section is distributed over all the spectral range between 300 and 335 nm.

Step 2 - Training of neural networks with simulated data
The dataset generated during the first step has been used to train a set of 10 NNs. The input vector contains the radiances calculated by means of LIDORT, while the output vector represents the corresponding ozone profiles, i.e. those profiles that were the input during the run of forward model simulations. LIDORT is intrinsically monochromatic, and calculates the upwelling radiances at the TOA at 960 wavelengths. In order to simplify the procedure, the input measurements have been grouped in 48 spectral intervals, each one containing the average value of radiances belonging to it. On the other hand, ozone profiles contained in the output vector are reported on a grid of 20 pressure levels, which are the same used by RAL estimations (see Chapter 4). This choice stems from the need to have an uniform representation of the ozone distribution during the overall procedure, avoiding the introduction of errors due to different discrete representations of a continuous quantity (note that RAL ozone profiles will be used in Step 4 as references outputs).
We decided to train a set of 10 NNs, which differ only in the initialization of weights and biases before training, in order to obtain statistically consistent results during the extended pruning procedure. The training phase has been performed using the Scaled Conjugate Gradient algorithm and the early stopping procedure (see Chapter 3 for details). The validation set used to determine when the net could be considered “trained” was formed with 10 input-output pairs not included in the trining set.

Step 3 - Extended pruning procedure
During this phase, we applied the extended pruning (as described before) to all NNs trained in Step 2, and a global rank of the relative importance of input units was derived. On the base of obtained results, a set of 5 spectral bands, within the interval 290-335 nm, has been selected, and are summarized in Table 5.1. In Figure 5.11 these 5 bands are plotted over the ozone absorption cross-section spectrum derived from LIDORT. It is interesting to note that the behaviour of the neural networks seem to follow some radiative transfer reasoning. The 5 subintervals are rather regularly distributed throughout both Hartley and Huggins bands. In particular we note the portion of the spectral range with the highest values of the absorption cross-section is included (band B1), as well as the the interval with a good sensitivity to the temperature variations (band B3). On the contrary, the wavelengths characterized by the smallest values ($\lambda > 326$ nm) are discarded. Another comment is on the importance of band B5 (320-325 nm), already selected on the base of physical assumptions, and put in evidence also by the extended pruning procedure.

Step 4 - Training of neural networks with measured data
In the case of retrieval of ozone profiles from GOME data, the use of all the measurements available within the 5 spectral bands selected in Step 3 would lead to an inversion scheme with an input vector of about 200 values. For a NN algorithm to be implemented, this would involve a topology with hundreds of neurons in the input and hidden layers, therefore, a very complex NN, risky in terms of generalization capabilities. The extraction from such many wavelengths of those with more information content and more suitable to perform the inversion is then recommendable. In this step a new set of 10 NNs has been trained; the input vector contains 194 GOME radiance measurements belonging to the bands selected in Step 3, while the reference output vector represents the corresponding ozone profiles provided by the Optimal Estimation scheme developed at the Rutherford Appleton Laboratory (RAL) [6], expressed (as already mentioned before) on a pressure grid of 20 levels, from the Earth surface to a pressure level corresponding to an height of 60 km. In this phase we used a training set with data of 22 different orbits (for an overall number of 3740 patterns) belonging to a time interval of one year, from March 1998 to March 1999, in order to include the maximum number of climatologies and to improve the generalisation properties of the methodology. Also in this case the training phase was performed using the SCG algorithm, with the early stopping procedure. The validation set contains 11 different orbits (i.e. 1670 input-output patterns) from the same time interval.

Step 5 - New extended pruning procedure
We again applied the extended pruning procedure to all trained NNs, in order to single out which wavelengths, from the global set of 194 inputs, could be the most crucial for the retrieval algorithm. In principle, a too little number of inputs would yield a poor accuracy in the estimation. Augmenting the dimensionality of the input vector would be effective up to a certain extent, i.e. until when the addition of further measurements, without provoking a significant improvement of the retrieval, would require a major computational and structural complexity.
A set of 17 wavelengths has been finally selected; removing an additional wavelength from input units would produce a significant increase in the global retrieval error, while adding an additional input would not improve the estimation performance. The final result obtained at the end of Step 5 is reported in Table 5.2, where we listed, for each subinterval extracted during Step 3, the corresponding selected wavelengths.
We can see that the extended pruning procedure has been applied twice: one to the synthetically generated data, the other one to GOME and RAL data. The rationale underlying this hybrid analysis is that, in the selection of the final wavelengths, we should have taken into account either the physical effects embedded by the radiative transfer model or the instrumental effects contained in the spectrometer measurements. All 5 subintervals contain at least 2 wavelengths, and the picked-up lines generally belong to the central part of the band. Subintervals B1 and B5 have the maximum number of wavelengths (5 and 4 respectively), and are then emphasized as the most important ones for ozone retrieval. The smallest number of selected wavelengths is obtained for B3, half-way located between subinterval B1 and subinterval B5. Indeed in this band absorption cross-section values are significantly smaller than those of bands B1 and B2, but, on the other hand, favourable conditions of band B5 are too far to be fully exploited. Nevertheless, it has to be noted that it is not completely discarded, as it carries useful information on temperature distribution.

5.1.3 Feature extraction

Feature extraction is a general name given to the process of generating linear or not-linear combinations of original input (or output) variables in order to reduce the dimensionality of data. PCA (Principal Component Analysis) is a widely used technique for feature extraction or dimensionality reduction. The conventional PCA techniques rely on eigenvector expansions stemming from the variance-covariance matrix describing the variability of the observed

quantities. Mathematically, if **X** is an *n*-dimensional random variable with mean value $\mathbf{X}_M$, the covariance matrix **B** associated to the unknown vector **X** can be evaluated. The generic element of such a matrix will be expressed as:

$$B_{ij} = \langle X_i X_j \rangle \tag{5.6}$$

Then, a new set of variables, i.e. the *n*-dimensional vector **Y**, known as principal components, can be calculated by:

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \ldots + a_{nj}X_n = \mathbf{a_j}^T \mathbf{X} \tag{5.7}$$

Where $\mathbf{a_j}^T$ contains the *n* normalised eigenvectors of the covariance matrix **B**, solution of the eigenvalue problem, stated as follows:

$$[\mathbf{B}][\mathbf{a}] = \lambda[\mathbf{a}] \tag{5.8}$$

The principal components transformation has several interesting characteristics:

- the total variance is preserved in the transformation i.e.

$$\sum_{i=1}^{n} \sigma_i^{\,2} = \sum_{i=1}^{n} \lambda_i \tag{5.9}$$

where σ_i^2 are the variances of the original variables (**X**), λ_i the eigenvalues of **B**, with

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n \tag{5.10}$$

- it minimizes the mean square approximation error
- in a geometrical sense, the transformation may rotate highly correlated features in *n*-dimensions to a more favourable orientation in the feature space, where components are still orthogonal to each other, such that the maximum amount of variance is accounted for in decreasing magnitude along the ordered components

Another common procedure is to base the PCA transformation on the correlation matrix, instead of covariance matrix. The correlation matrix is the standardisation of the covariance matrix obtained by normalising the data with the standard deviations. Such a procedure reduces all the variables to equal importance as measured by scale. In remote sensing of the atmosphere, this can be of importance if the species of interest give emission of very different magnitudes. On the other hand, a non-standardised form may be justified on the basis of possible differences on the signal-to-noise ratio (SNR) characterizing the spectral bands measuring the major magnitudes [7]. In order to concentrate the analysis on the fluctuations of the variables around their mean values, the PCA may be also applied to the new variable **X`**, expressing the deviation of the original variable **X** from its mean and defined as follows [8]:

$$\mathbf{X`} = \mathbf{X} - \langle \mathbf{X} \rangle \tag{5.11}$$

In this work, we decided to apply the PCA on the dataset coming out from Step 3 of the Extended Pruning approach described in Section 5.1.2. This means that feature extraction will be performed on a measurement vector containing 194 radiance values, which belong to the 5 spectral intervals selected during the sensitivity analysis. For applying the PCA we considered the following matrices:

- covariance matrix **B** associated to the variables **X**
- covariance matrix **B`** associated to the variables **X`**
- correlation matrix **R** associated to the variables **X**
- correlation matrix **R`** associated to the variables **X`**

In Table 5.3 we report the ordered list of the first 20 eigenvalues calculated for the 4 cases. We see that most of the variance is in general contained in the first principal component, but a change in the variance distribution can be observed by introducing transformation of (5.11). In this case we have higher variance values carried on by the successive components. As a consequence, dealing with the transformed variables will lead to lower the SNR of the input corresponding to the first principal component and to increase the SNR of the following ones. Another comment can be derived from the analysis of the curves representing the first eigenvector for the 4 cases, as reported in Figure 5.12. We see that for the covariance matrices, band B5 is crucial for determining the value of the first component, while the role of band B1 becomes the most significant when correlation matrices are considered.
Referring to the retrieval performance, the best result was obtained by using the covariance method without applying transformation (5.11). More generally, we have better results with covariance matrices with respect to correlation matrices. Stemming from considerations reported in previous sections, this result puts again in evidence the importance of the subinterval B5 for ozone profiles estimation.
Adopting the same criterium already used for the selection of wavelengths performed with the extended pruning, i.e. minimising the global retrieval error of network estimations, a set of 13 features has been selected, and will be used to design the actual PCA-based retrieval algorithm.

5.2 The inversion phase

Input-output patterns needed for training neural networks for the actual inversion phase have been generated considering, for the input vector, GOME measurements selected with each one of methodologies described above, and the already mentioned offline-retrieved ozone profiles provided by RAL for the reference output vector. Two geophysical quantities have been added as additional inputs to the input vectors, the solar zenith angle characterizing the viewing geometry of GOME measurements, and the ozone slant column, which gives additional information on the total ozone content.
Therefore, the input vector contains:

- 28 input units for the neural algorithm based on the “physical selection”;
- 19 input units for the neural algorithm based on the “Extended Pruning”; selection
- 15 input units for the neural algorithm, based on the PCA feature extraction.

For all the algorithms, the output layer contains 20 units, as RAL ozone profiles are expressed on a pressure grid of 20 levels. In this study we considered a MLP topology with only one hidden layer. In order to decide the number of hidden units, an empyrical approach have been adopted, aiming at optimising the retrieval performance of trained NNs.

5.2.1 Optimisation of the hidden layer

During the training phase, different experimental neural networks have been set up; starting from a hidden layer containing one single neuron, we increased the number of hidden units up to 100. For each network, the retrieval error has been evaluated using a test dataset, which contains different input-output pairs not included in the training dataset. In Figure 5.13 and Figure 5.14 we report, for the first and second neural algorithm respectively, the retrieval error committed on the test set, as a function of the number units contained in the hidden layer. The error is expressed as the Sum of Squared Errors (SSE), evaluated for each output value and for all patterns belonging to the test set.
It can be noted that an optimum number, or at least an optimum range, exists. In fact, if the number of hidden neurons is too small, network input-output associative capabilities are too weak. On the other hand, this number should not be too large, because network capabilities might show a lack of generality (they would be too tailored on the training set). It turns out that a fair compromise between these two contrasting requirements has to be found.
In our study, we used 28 hidden units for both algorithms, as this number was contained in the optimal range in both cases. As far as the PCA implementation is concerned, an hidden layer of 28 hidden units has been also chosen for sake of consistency.

5.2.2 Retrieval results

Once the training process has been completed, the performance of the different neural algorithms on the test set can be evaluated. In Figure 5.15 we show the root mean square error (RMSE), expressed in [molecule/cm^3] x e^{12}, calculated over the test set for the three neural algorithms and reported as a function of altitudes. The solid line represents the standard deviation of the profiles from their means and gives an indication of the a priori profiling accuracy without measurements. We see that first and second neural algorithms (plotted with dashed line and dotted line respectively), for which the input selection was based on physical assumptions (the first one) and on the Extended Pruning procedure (the second one), performs clearly better than the algorithm which used the PCA technique (dash-dotted line). The PCA transformation probably prevents from capturing significant features of the measured spectra. This may be due to the fact that PCA is limited by virtue of being a linear technique, therefore unable to capture more complex nonlinear correlations. On the base of this result, only the first two algorithms will be considered for further analysis. From now, we will refer to GOME_NN1 for the first algorithm and to GOME_NN2 for the second one.
The selection procedure based on the Extended Pruning, if compared with the mothodology based on physical assumption, gives better results up to 40 km, despite the smaller number of input measurements and the simpler pre-processing. At higher altitudes,where ozone concentrations are extremely low, the two algorithms show very similar results. Therefore, it seems then that EP technique provides a very competitive approach to address the problem of dimensionality reduction and selection of input measurements.
In Figure 5.16 we report the results of the comparison between ozone values retrieved by RAL and both neural algorithms, carried out for all the measurements in the test dataset. The scatter plot between RAL and GOME_NN1 ozone concentrations (expressed in [molecule/cm^3] x e^{12}) is represented with blue marks, while red diamonds show the same plot for GOME_NN2. GOME_NN2 shows less scatter around the bisector, and compares better with RAL with respect to GOME_NN1. The correlation coefficient is 0.980 for GOME_NN2 and 0.94 for GOME_NN1, and confirms the effectiveness of both neural algorithms, as well as the better performance of GOME_NN2.
An example on the level of accuracy reachable by means of the Extended Pruning methodology is reported in Figure 5.17, where (lower plot) the contour plots of ozone profiles corresponding to an entire orbit taken as example from the test dataset are shown. The abscissa is the ground

pixel sequence number, corresponding to a latitude range from –75° N to 65° S, while the ordinate is the height above the surface in km. The ozone values are always expressed in [molecule/cm^3] x e^{12} according to the colour scale reported beside the plots. The black areas indicate the surface topology (e.g. mountains). From the comparison between this map and the one provided by RAL (upper plot), derived from measurements of the same orbit, it can been observed that the large scale features of the ozone field are very similar.
In Figure 5.18 we report some examples of ozone profiles retrieved by the GOME_NN2 algorithm (red line), compared with the corresponding RAL profile (black line), taken from the same orbit showed in Figure 5.17. The neural algorithm, if compared with the RAL one, provides a very good representation of the ozone distribution, except for the troposphere, where a little discrepancy can be observed.

5.2.3 Retrieval error characterization

Once an ozone profile has been retrieved by any method, the uncertainty on the retrieved values has to be estimated. According to the formalism used for the Optimal Estimation approach (Chapter 2), the difference between the retrieved value ($\hat{\mathbf{O}}$) and the true state ($\mathbf{O}$) of an ozone profile can be expressed as follows:

$$\hat{\mathrm{O}} - \mathrm{O} = (\mathrm{A} - \mathrm{I})(O - O_a) + G_y \varepsilon \tag{5.12}$$

where $\mathbf{I}$ represent the identity matrix, $\mathbf{O_a}$ the a priori profile, $\mathbf{A}$ the averaging kernel matrix which characterizes the inversion method, $\mathbf{G}$ the matrix containing the gain functions (also called contribution functions) and $\boldsymbol{\varepsilon}$ the errors associated to the measurement vector ($\mathbf{Y}$) used for the inversion. In the equation (5.12), the averaging kernels represent the sensitivity of the retrieved profile to the trace gas vertical distribution, defined as:

$$A = \frac{\partial \hat{\mathrm{O}}}{\partial \mathrm{O}} \tag{5.13}$$

The gain function represents the sensitivity of the retrieved quantities to the input measurements:

$$G_y = \frac{\partial \hat{\mathrm{O}}}{\partial \mathrm{Y}} \tag{5.14}$$

The averaging kernel matrix can be evaluated applying the gain matrix $\mathbf{G_y}$ associated with input measurements (which can be easily determined once the inversion method is determined) to the weighting functions matrix $\mathbf{K}$:

$$A = G_y K \tag{5.15}$$

where $\mathbf{K}$ characterizes the sensitivity of the instrument to the observed species and can be calculated from forward model simulation, by the following expression:

$$K = \frac{\partial \mathrm{Y}}{\partial \mathrm{O}} \tag{5.17}$$

The first term of the right-hand side of equation (5.12) represents the retrieval error associated with the non-ideal sensitivity of the observing system to the real state, and it is generally known

as the smoothing error. The second term, instead, accounts for the influence on the retrieval error of uncertainties on the input measurements.
For the two neural network algorithms no forward simulations have been carried out, and the learning phase has been performed using already retrieved ozone profiles provided by RAL. As a consequence, the averaging kernels can not be determined straightforward from the neural algorithms themselves. Moreover, trained neural networks do not simply reproduce the data used in the training set, but are able to learn the input-output significant relationships and to generalize outside the training set. Therefore, the two neural network inversion schemes cannot be characterized by merely using RAL averaging kernels. Nevertheless, just to have an idea of the accuracy of ozone profiles used for NN training, we report in Figure 5.19 the estimated retrieval error of the RAL OE algorithm, for different measurement conditions, as reported in the legend below the plot. We can see that between 20 km and 50 km the retrieval error is around 5%, while it increases in the upper stratosphere (up to 50%) and in the troposphere (50-70%, depending on the measurement conditions). In the same plot, also the a priori error is reported, and it clearly indicates the capability of the algorithm to provide an improved estimation with respect to the a-priori knowledge.
On the other hand, the Full Width at Half Maximum (FWHM) of the averaging kernel functions (see Chapter 4) represents the extent to which independent information is available in different atmospheric layers, and can provide a measure of the vertical resolution of retrieved profiles. In this study we assumed for the two neural estimation methods the vertical resolution of RAL profiles used in the training phase; hence the FWHM values, which can vary from 10-12 km in the troposphere to 4-5 km in the middle stratosphere (see Chapter 4 for details), can be considered for GOME_NN1 and GOME_NN2.
With regards to the gain function matrix ($\mathbf{G_y}$) and the measurement error ($\mathbf{G_y\varepsilon}$) characterizing the retrieval method, they can be directly determined from the neural parameters. The gain function $\mathbf{G_y}$ is defined from the partial derivatives of the retrievals with respect to the input measurements, and can be calculated by differentiation of the neural network global transfer function. For each retrieval pressure level k the global transfer function is given by:

$$\hat{O}_k = \left[\sum_{j=1}^{Nh} w_{kj}^2 \left(\frac{1}{1+e^{-\left(\sum_{i=1}^{Ni} w_{ji}^1 Y_i + \beta_i \right)}} \right) + \beta_k \right] \tag{5.18}$$

where $\hat{O}_k$ represents the retrieved value of the ozone concentration, Y_i the input measurement at wavelength i, w_{ji}^1 the weights between the input and the hidden layers, w_{kj}^2 the weights between the hidden and the output layers, β_j and β_k the biases of hidden and output units respectively. The inversion errors associated with input measurements uncertainties have been evaluated applying the calculated gain functions to the radiometric noise which characterizes the radiances spectral measurements:

$$E_k = \left(\frac{\partial O_k}{\partial Y_i} \right) \varepsilon_i \tag{5.19}$$

where E_k represents the inversion error at the pressure level k associated with the spectral error ε_i characterizing the input i. The contribution of the radiometric noise affecting GOME radiance measurements to the global inversion error is reported in Table 5.4. The impact of measurements uncertainties is relatively weak at all pressure levels, with values smaller than 3% for GOME_NN1 and 6% for GOME_NN2. These results can be explained by the stability of the spectral measurements performed by the GOME instrument, but also by the robust properties of the neural algorithms, which results poorly affected by the instrumental noise.

An alternative way to evaluate NN retrieval error can be carried out assuming RAL ozone profiles as representative of the true states of the atmosphere. The inversion error is assessed with respect to the validation dataset, whose examples provide a good statistical set of atmospheric conditions and have not been used during the training phase. To this purpose, for all the examples of the validation dataset, the NN retrieved profiles have been compared with the corresponding profiles retrieved by RAL, and the estimated error, expressed as a percentage between the two sets of measurements, has been calculated for both neural algorithms. The results are reported in Figure 5.20. The solid and dotted lines represent the error profiles of GOME_NN1 and GOME_NN2 respectively; the dashed line represents the standard deviation of RAL profiles (i.e. "true" profiles). Globally, the inversion error is expected to be less than 20% between 100 hPa (~ 15 km) and 1 hPa (~ 50 km), with values of 5-10% between 50 hPa (~ 20 km) and 5 hPa (~ 40 km); outside the pressure range between 100 hPa and 1 hPa the retrieval errors are larger, especially at pressure levels smaller than 0.1 hPa, where the concentrations of ozone are extremely low. This analysis shows similar (but not equal) results to those reported for the RAL Optimal Estimation algorithm, and confirms that NNs can provide, on the base of learned rules, independent new estimations with an accuracy comparable to that characterizing OE methods.

References

[1] Burrows J.P., A. Richter, A. Dehn, B. Deters, S. Himmelmann, S. Voigt and J. Orphal, Atmospheric remote-sensing reference data from GOME: 2. Temperature-dependent absorption cross sections of O_3 in the 231–794 nm range, *J. Quant. Spectrosc. Radiat. Transf.*, 61, 509–517, 1999.

[2] S. Casadio, C. Zehner, Temperature independent differential absorption spectroscopy (TIDAS) for the measurement of ozone slant content using GOME data, *Proc. XIX Quadriennial Ozone Symposium*, Sapporo, Japan, 2000.

[3] Spurr R.J., Kurosu T.P., Chance K.V., A linearized discrete ordinate radiative transfer model for atmospheric remote sensing retrieval, J. Quant. Spectrosc. Radiat. Transfer, 68, 689–735, 2001.

[4] Bass A. M., R. J. Paur, The ultraviolet cross-sections of ozone: I. The measurements of atmospheric ozone, *Proc. Quadrennial Ozone Symp.*, D. Reibel Publ., Norwell, Mass., 606–610, 1984.

[5] Paur R. J., A. M. Bass, The ultraviolet cross-sections of ozone: II. Results and temperature dependence, *Proc. Quadrennial Ozone Symp.*, D. Reibel Publ., Norwell, Mass., 611–616, 1984.

[6] Siddans R., Height resolved ozone retrievals from GOME, *PhD thesis*, Univ. Reading.

[7] Singh A., Harrison A., Standardized principal components, *International Journal of Remote Sensing*, 6, 883-896, 1985.

[8] Del Frate F., Schiavon G., A combined natural orthogonal functions/neural network technique
for the radiometric estimation of atmospheric profiles, *Radio Sci*, 33, 405-410, 1998.

[9] Bishop C.M., *Neural networks for pattern recognition*, Oxford University Press, London, UK, 1995

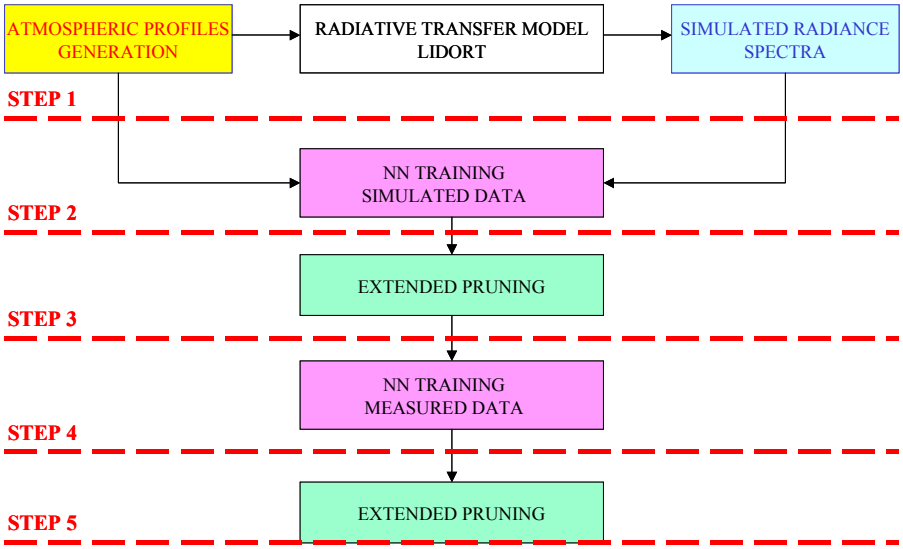

Figure 5.1. General scheme for the input selection methodology based on the Extended Pruning.

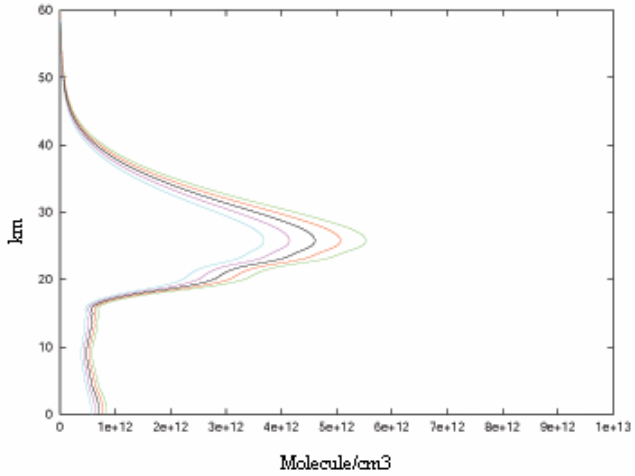

Figure 5.2. Examples of ozone profiles with different total ozone content.

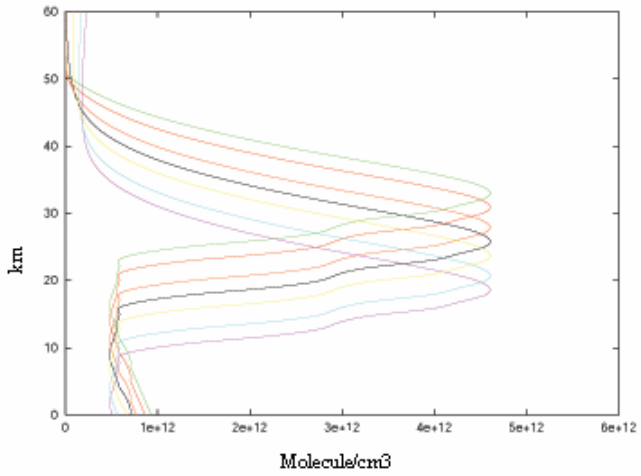

Figure 5.3. Examples of ozone profiles with ozone peak at six different height levels.

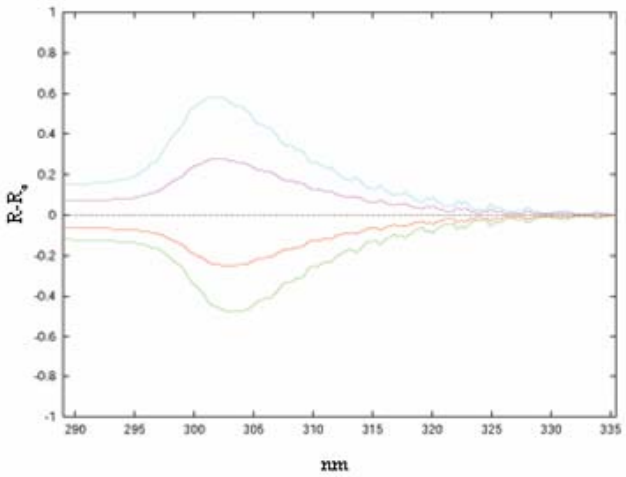

Figure 5.4. Radiance difference spectra corresponding to profiles of Figure 5.2. Each coloured curve represents the difference between the corresponding profile and the standard profile (plotted with a black line in Figure 5.2).

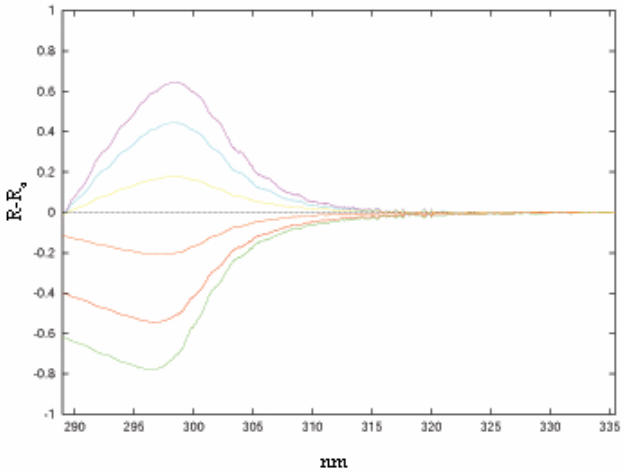

Figure 5.5. Radiance difference spectra corresponding to profiles of Figure 5.3. Each coloured curve represents the difference between the corresponding profile and the standard profile (plotted with a black line in Figure 5.3)

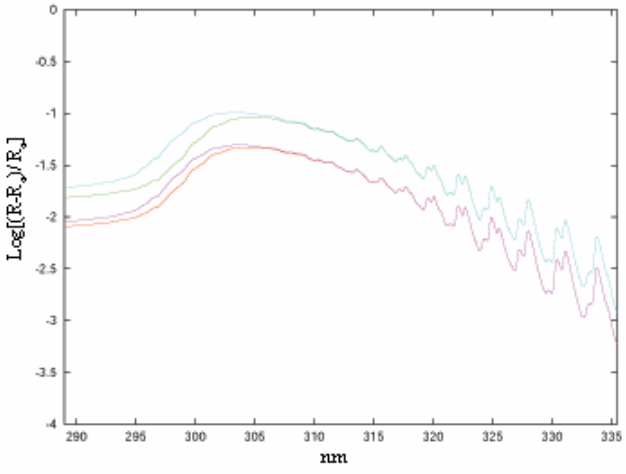

Figure 5.6. Normalised radiance difference spectra corresponding to profiles of Figure 5.2.

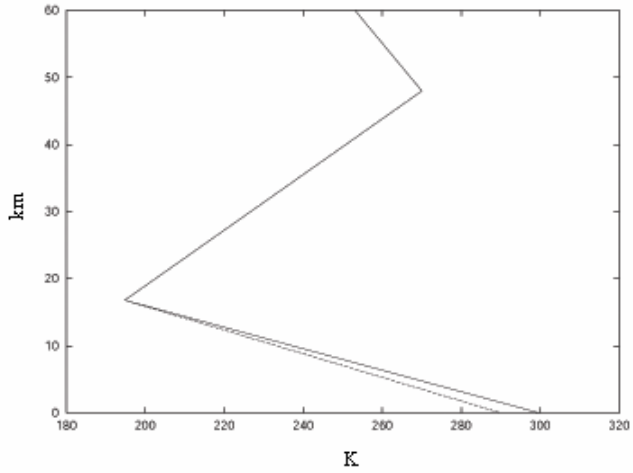

Figure 5.7. Standard temperature profile (solid line) modified by varying the tropospheric thermal gradient (dashed line).

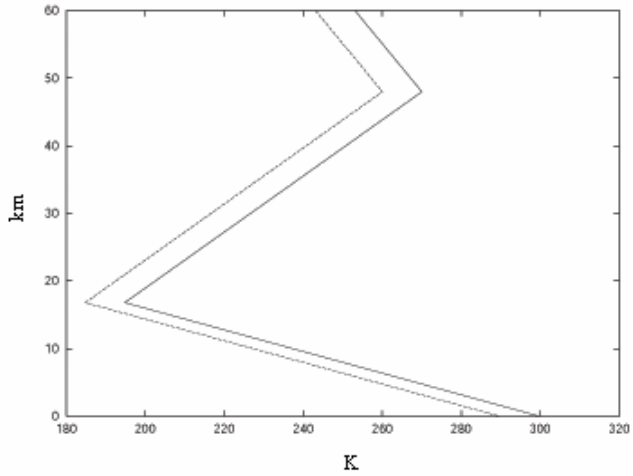

Figure 5.8. Standard temperature profile (solid line) modified with a homogenous variation of temperature values (dashed line).

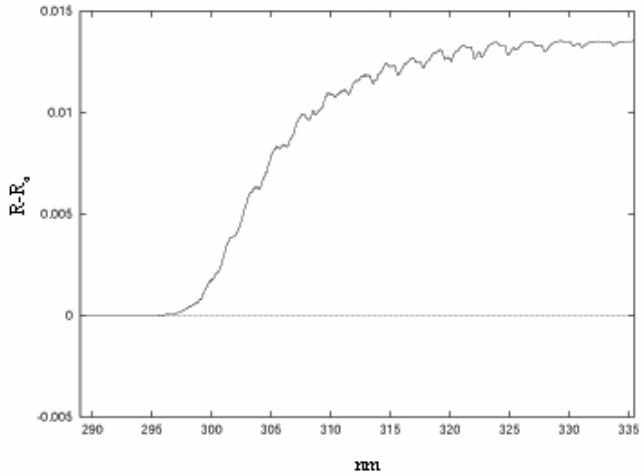

Figure 5.9. Radiance difference spectrum corresponding to profile of Figure 5.7. The curve represents the difference between radiances of the modified profile and the standard profile.

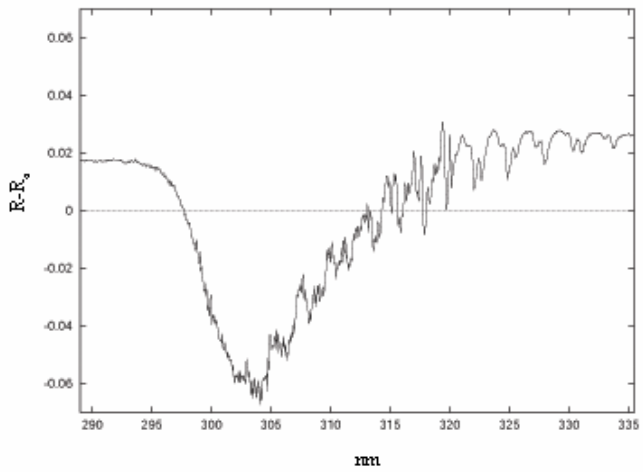

Figure 5.10. Radiance difference spectrum corresponding to profile of Figure 5.7. The curve represents the difference between radiances of the modified profile and the standard profile.

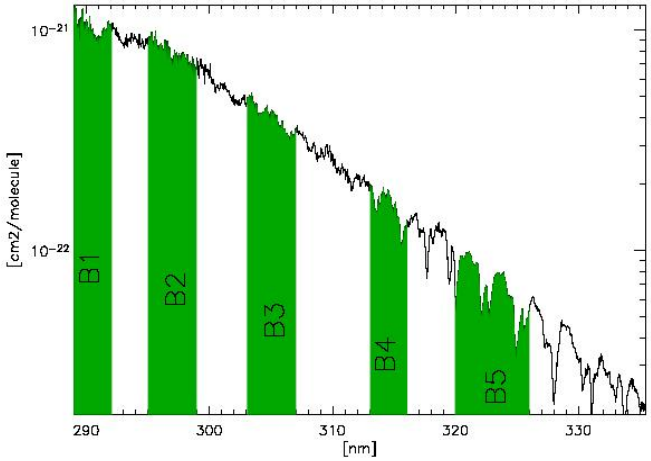

Figure 5.11. Spectral bands selected by means of the first phase of the Extended Pruning procedure.

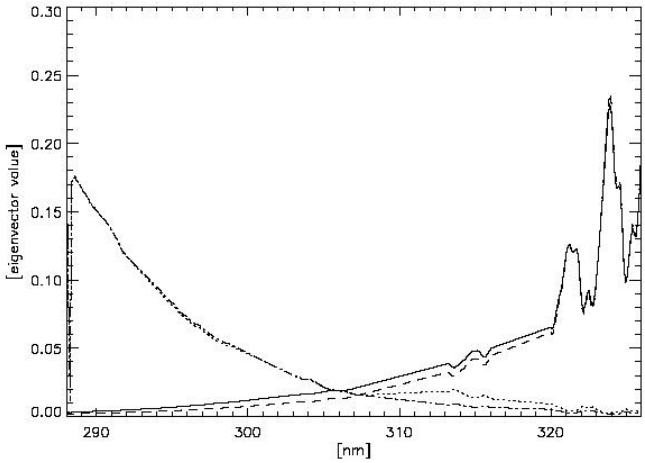

Figure 5.12. Curves representing the first eigenvector for 4 different types of PCA. Solid line: [**B**] matrix, dotted line: [**B`**] matrix, dashed line: [**R**] matrix, dash-dotted line [**R`**] matrix.

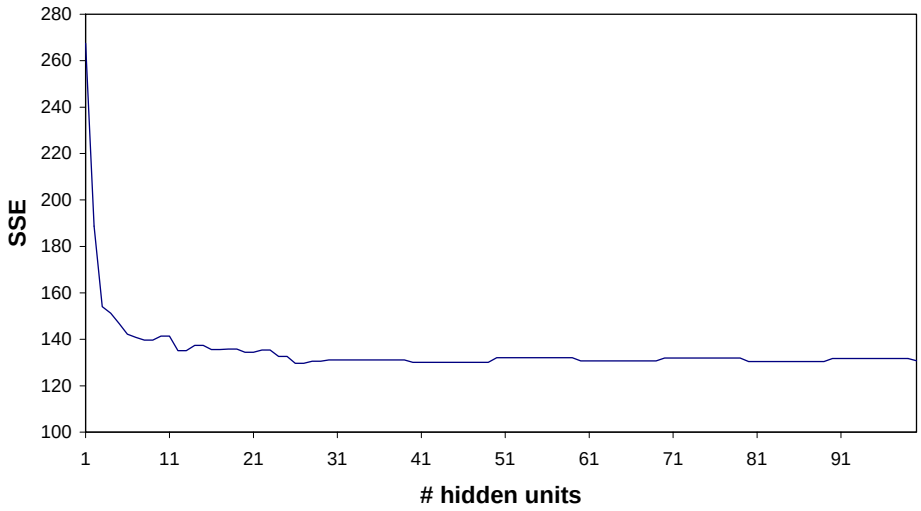

Figure 5.13. Retrieval error (SSE, Sum of Squared Error) evaluated on the test dataset, as a function of the number units contained in the hidden layer, for the first neural algorithm.

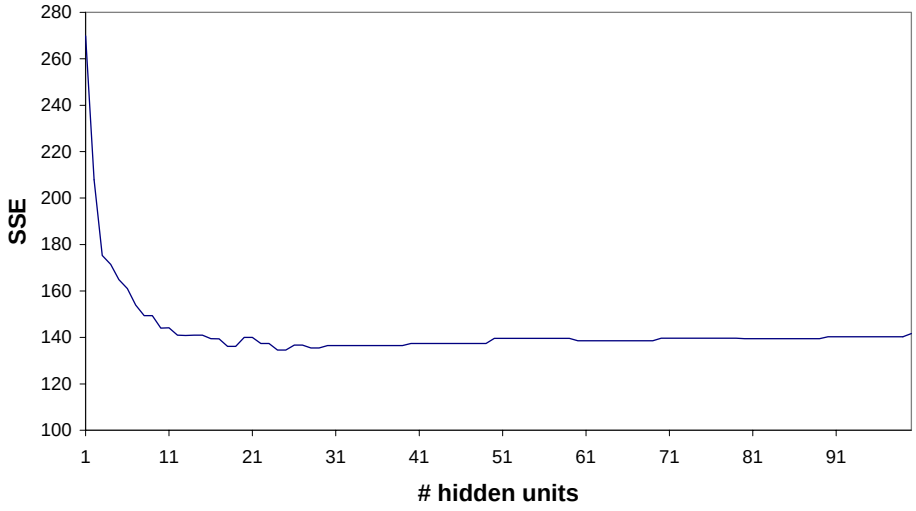

Figure 5.14. Retrieval error (SSE, Sum of Squared Error) evaluated on the test dataset, as a function of the number units contained in the hidden layer, for the second neural algorithm.

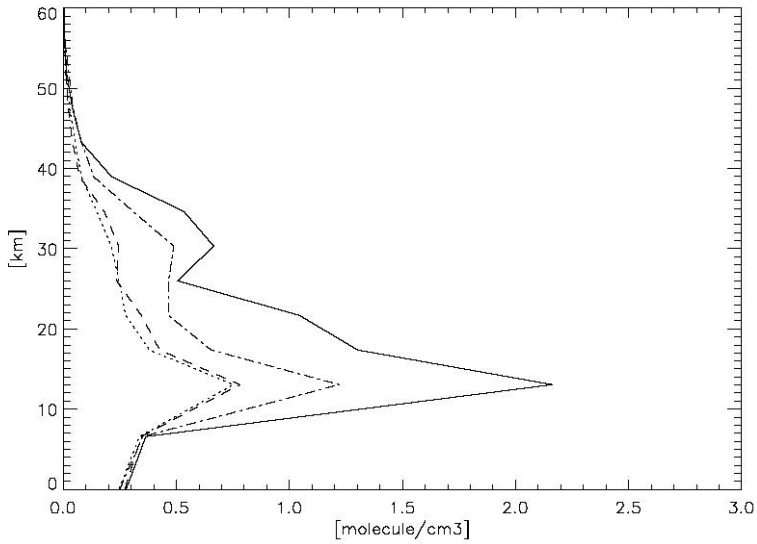

Figure 5.15. Profiles of RMSE of retrieved ozone obtained with three different neural algorithms. Dotted line: EP approach for dimensionality reduction (GOME_NN2); dashed line: dimensionality reduction based on physical assumptions (GOME_NN1); dash-dotted line: PCA approach; solid line: standard deviation of profiles from their means.

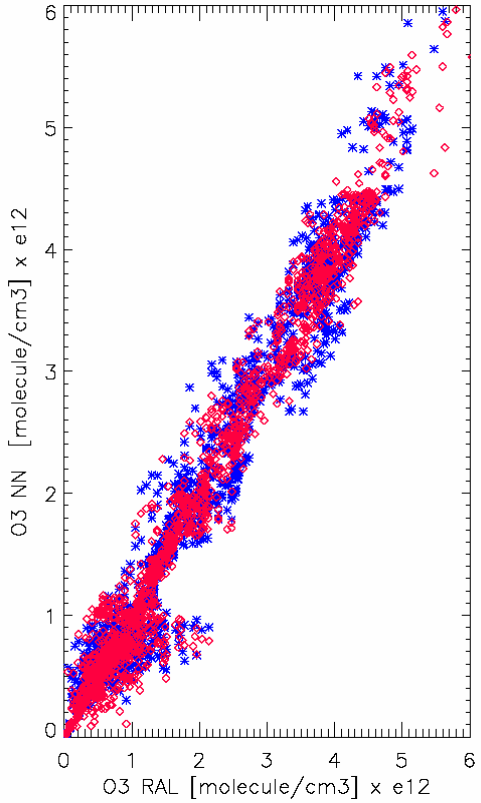

Figure 5.16. Scatter plot between RAL and NN ozone values. Blue marks: algorithm GOME_NN1; red diamonds: algorithm GOME_NN2.

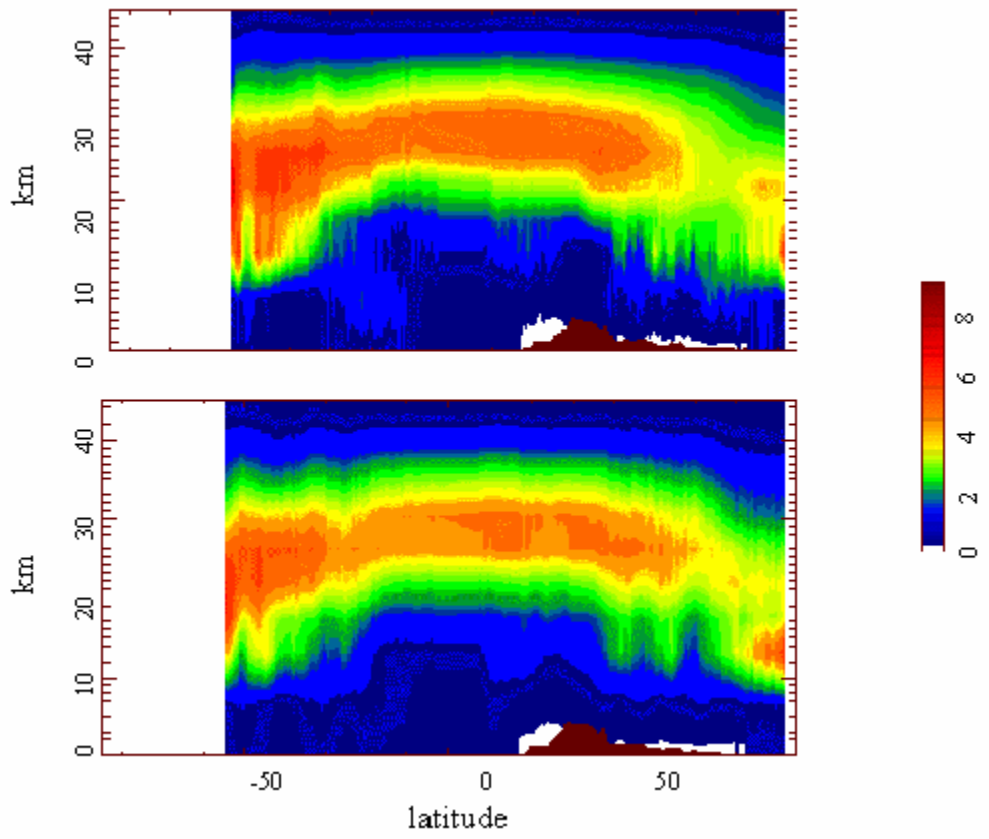

Figure 5.17. Upper plot: ozone map (corresponding to one orbit of the satellite) estimated off-line by RAL. Lower plot: ozone map as estimated by using the real-time GOME_NN2 algorithm. The colour scale represents ozone values reported in [molecule/cm^3] x e12.

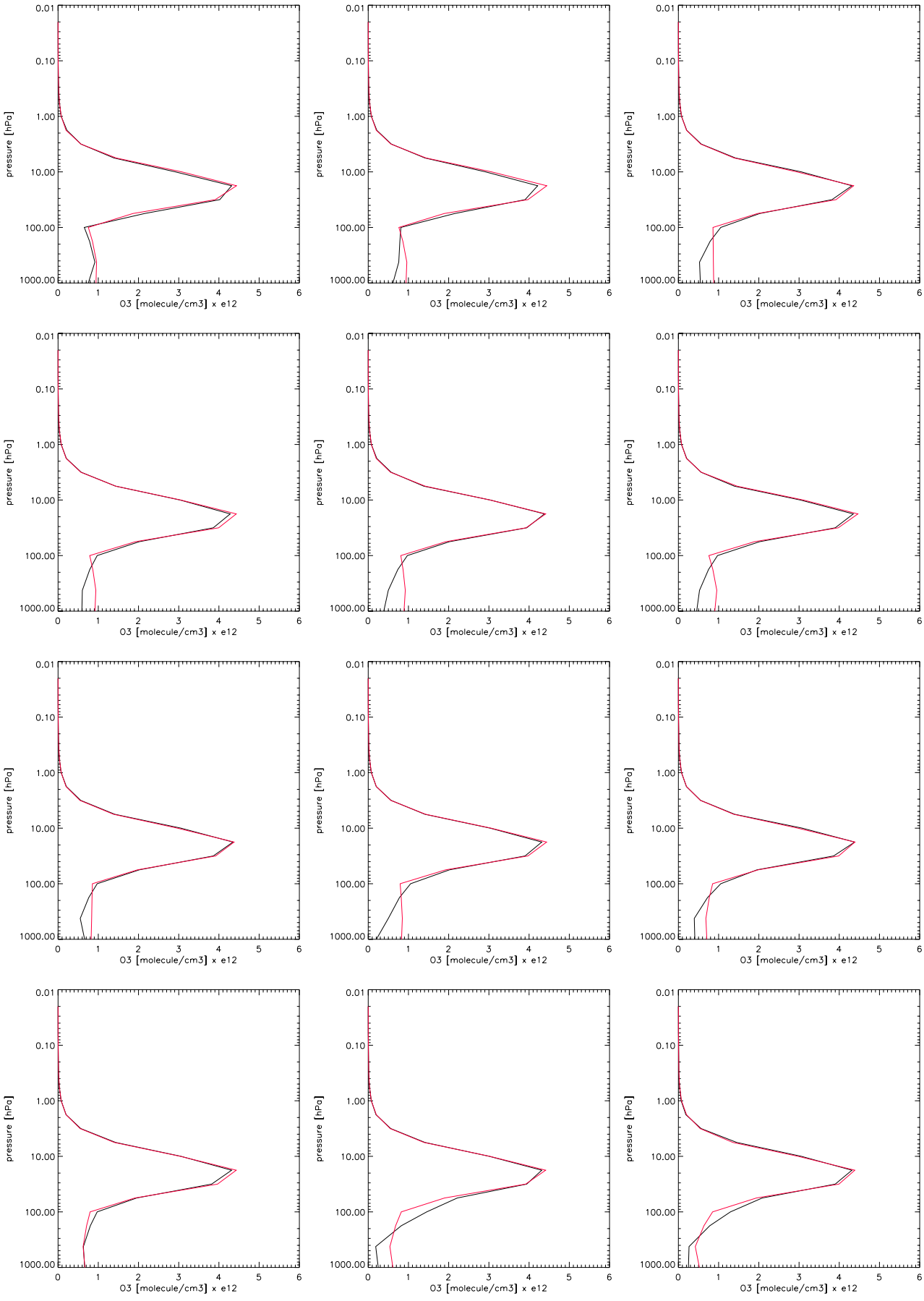

Figure 5.18. Examples of ozone profiles retrieved by GOME_NN2 compared with the corresponding profiles estimated by RAL.

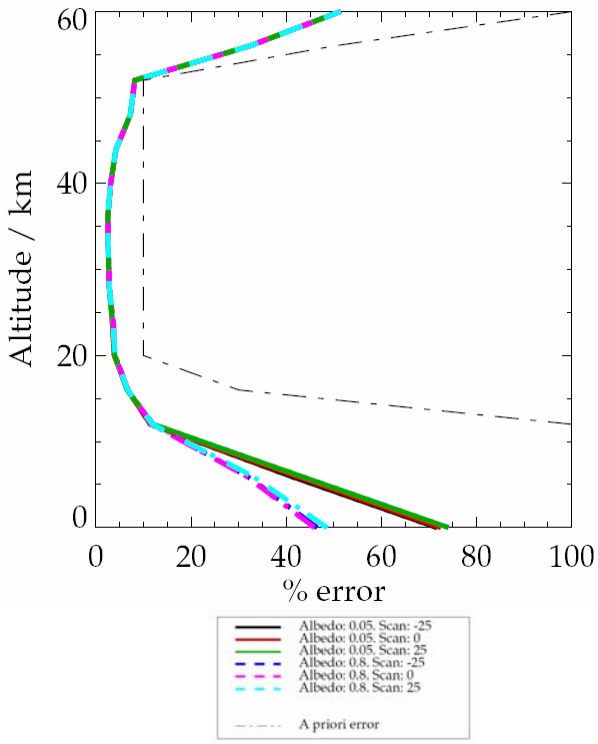

Figure 5.19. Estimated retrieval errors characterizing RAL OE algorithm for different scenarios.

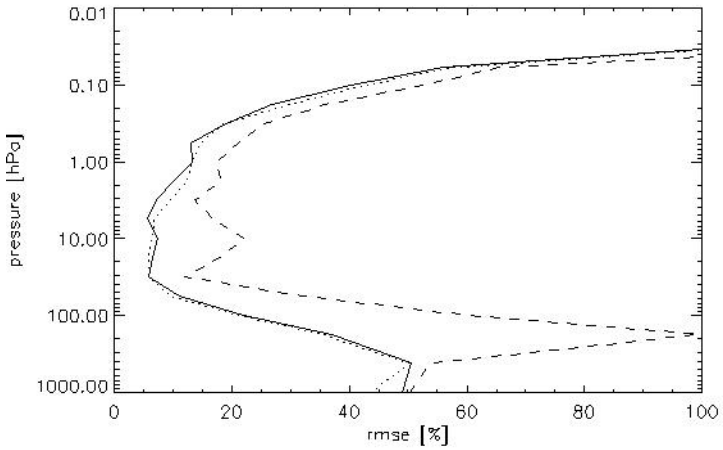

Figure 5.20. Retrieval errors estimated for both neural algorithms. The solid line represents the RMSE for GOME_NN1, while the dotted line for GOME_NN2. The dashed line shows the standard deviation of RAL profiles, assumed to represent the true states.

Band	Spectral range (nm)	Bandwidth (nm)
B1	288.04 - 292.05	4.01
B2	294.97 - 298.96	3.99
B3	303.05 - 307.03	3.98
B4	313.03 - 316.05	3.02
B5	319.98 - 325.96	5.98

Table 5.1. Spectral band selected by means of the first phase of the Extended pruning procedure.

Band	Spectral range (nm)	Wavelengths (nm)
B1	288.04 - 292.05	289.23, 289.99, 290.21, 290.64, 290.75
B2	294.97 - 298.96	297.02, 297.24, 297.35
B3	303.05 - 307.03	305.31, 304.13
B4	313.03 - 316.05	313.49, 314.08, 315.00
B5	319.98 - 325.96	322.06, 322.52, 322.75, 323.21,

Table 5.2. Wavelengths selected by means of the second phase of the Extended pruning procedure.

eigenvalue	$[B]$	$[\tilde{B}]$	$[R]$	$[\tilde{R}]$
λ_1	31820.595	71304.147	1227.623	2617.017
λ_2	19.384	26.904	16.515	24.907
λ_3	6.420	23.534	5.719	23.352
λ_4	0.858	17.833	0.351	15.360
λ_5	0.061	12.810	0.061	10.931
λ_6	0.045	0.293	0.036	0.235
λ_7	0.023	0.176	0.014	0.137
λ_8	0.013	0.115	0.012	0.111
λ_9	0.011	0.093	0.010	0.082
λ_{10}	0.009	0.080	0.006	0.067
λ_{11}	0.008	0.056	0.001	0.048
λ_{12}	0.007	0.040	0.001	0.035
λ_{13}	0.007	0.035	0.000	0.030
λ_{14}	0.006	0.031	0.000	0.027
λ_{15}	0.006	0.030	0.000	0.026
λ_{16}	0.006	0.028	0.000	0.025
λ_{17}	0.005	0.026	0.000	0.021
λ_{18}	0.005	0.023	0.000	0.020
λ_{19}	0.005	0.020	0.000	0.019
λ_{20}	0.005	0.018	0.000	0.016

Table 5.3. First 20 eigenvalues for the 4 different matrices considered for PCA.

pressure (hPa)	σ_{eNN1} (%)	σ_{eNN2} (%)
1000.00	1.4	0.6
421.30	1.0	1.7
177.80	2.0	4.2
100.00	0.9	1.5
56.20	0.15	1.9
31.60	0.9	5.8
17.80	2.3	5.4
10.00	1.8	3.4
5.60	1.7	3.7
3.20	1.6	4.2
1.80	1.2	3.0
1.00	0.6	1.4
0.50	1.1	2.2
0.30	1.0	2.4
0.20	0.2	0.5
0.10	1.0	1.3
0.05	0.7	1.5
0.03	1.1	3.6
0.02	0.3	0.4
0.01	0.3	1.0

Table 5.4. Inversion errors associated with radiometric errors on input radiances for all the pressure levels used for the retrieval.

Chapter 6

Validation of ozone profiles retrieved by neural networks

In this chapter the validation of ozone profiles retrieved by NN algorithms from GOME measurements is presented. The comparison has been conducted either with measurements performed by a different satellite instrument (ILAS) or with ozone profiles obtained by ground-based lidar systems. For both validation exercises an introduction to the observation system to compare with will be given, and the methodology used to perform the comparison will be described. The obtained result will be shown and critically discussed.

6.1 Comparison with ILAS products

The solar occultation sensor ILAS (Improved Limb Atmospheric Spectrometer) on board the ADEOS (Advanced Earth Observing Satellite, launched in August 1996) performs atmospheric layer observations to provide vertical profiles of ozone, methane, water vapour, nitrogen dioxide, nitric acid and nitrous oxide from absorption measurements in the infrared region, and temperature and pressure profiles from measurements of absorption due to oxygen molecules in the visible region. Optical properties of stratospheric aerosol and polar stratospheric clouds (PSCs) are also derived from visible and infrared extinction measurements [1].
The instrument capability to measure trace gases was confirmed from experiments with a gas cell and a black body light source during ILAS pre-flight studies. Moreover, a number of field campaigns using large balloons at Kiruna (Sweden) and ground-based remote sensors at Kiruna, Alaska, Syowa Station, and other locations have been conducted for the validation of ILAS measurements [2].
Next sections briefly describe the instrument, the assessment of performances, data processing algorithms, and validation results.

6.1.1 The instrument

The principle of ILAS measurements is the solar occultation technique, which has been proved to work satisfactorily for stratospheric measurements in previous sensors like the SAGE series [3] and HALOE [4]. In this technique, the sensor detects and disperses spectrally the light from the sun coming through the atmosphere, as shown in Figure 6.1, to give absorption spectra of the atmosphere. Since each gas in the atmosphere has its own characteristic absorption spectral features in the infrared (IR) region, it is possible to identify and quantify the gas components and relative concentrations.
As the satellite orbits the earth, the sensor can continuously measure the changes of light intensity from the sun during the sunrise and sunset events, as seen from the satellite. Vertical profiles of atmospheric components can be derived from the signals, which contain information on different atmospheric layers. The ILAS instrument has an IR and a visible spectrometer. Main targets of the ILAS measurements from IR channel data are vertical profiles of ozone (O_3) and ozone-related species such as nitric acid (HNO_3), nitrogen dioxide (NO_2), nitrous oxide (N_2O), methane (CH_4), and water vapour (H_2O). Profiles of aerosol extinction in IR wavelengths are also derived. With less precision, CFC11 and CFC12 in the troposphere and N_2O_5 in the stratosphere may also possibly be derived. From the visible channel, absorption spectra due to

oxygen molecules can give temperature and pressure profiles estimations [5], as well as aerosol extinction profiles from the signal in the wavelength with no absorption by oxygen molecules can be obtained. The altitude range for data analysis is from the cloud-top to about 60 km and the instantaneous field of view has a 2 km height resolution.
In the stratosphere, heterogeneous reactions on the surface of acid aerosols and/or PSCs are considered to play an important role in ozone destruction processes. Since these particles have characteristic absorption features in the IR region [6], it may be possible to discriminate the particle compositions and phases from the ILAS measurements.
The advantages of the solar occultation technique are the high sensitivity – and then high precision – of measurements, because of the use of the bright sun as a light source, and the self-calibration properties due to the use of the exo-atmospheric measurement as the 100 % reference for each measurement event.
Since the solar occultation technique is employed as the measurement principle and the ADEOS satellite has a sun synchronous polar orbit, with an inclination angle of 98° and an altitude of about 800 km, the measurement region of ILAS is over high latitudes, 55°-70° in the Northern Hemisphere and 63°-87° in the Southern Hemisphere. Figure 6.2 shows the time variation of latitude coverage, which changes according to the season. In the Southern Hemisphere, almost the whole Antarctic continent is covered while, in the Northern Hemisphere, the covered latitude range is narrow. On the contrary, this gives quite unique measurement opportunities, which allow generating daily height-longitude cross sectional maps of the atmosphere.

6.1.2 Instrument overview

The ILAS instrument consists of the following seven major components:

- 2-axis gimbals mirror, controlled to track the sun
- 12 cm diameter Cassegrain telescope
- beam splitter and transfer optics
- IR spectrometer
- visible spectrometer
- sun-edge sensor
- signal processing units

The instrument is mounted on the bottom (the earth side) left (sunny side) of the ADEOS satellite. The sun light passes through two cutaways; one is for sunrise direction and the other is for sunset direction. A plane mirror mounted on the 2-axis gimbals reflects the sun light into the telescope mounted on the slanted optical base. The light passes through the rectangular slit at the primary focal point, and is then split using the dichroic mirror and transfer optics to the IR and visible spectrometers.

6.1.2.1 Instantaneous field of view (IFOV) tracking

ILAS is designed to track the radiometric centre of the sun. The Instantaneous Field Of View (IFOV) of the IR band is of a 13 km width and 2 km height at the tangent point as seen from ADEOS. The IFOV of the visible spectrometer is a 2 km x 2 km square. The Field Of View (FOV) has ± 10° of margin for both the azimuth and the elevation direction. The tracking of the sun is carried out by a digital feedback system (Fine Sun Sensor, FSS). The slant optical base design enables the rectangular IFOV to be parallel to the horizon at the tangent height of 20 km for the two directions, both at sunrise and at sunset, using a common rectangular slit. This configuration of optics minimizes the rotation of IFOV during the tracking from just above the horizon, to 200 km in altitude [7]. This magnitude of rotation is reasonably negligible for the measurements of the stratosphere from the cloud top to 45 km.

6.1.2.2. Sun-edge sensor

Accurate information on tracking in elevation direction, or the absolute direction of the IFOV, is very important in the solar occultation observations. Thus, the ILAS has a sun-edge sensor to measure the direction of IFOV from the top edge of the sun, the direction of which is accurately known, while the tracking itself does not require high elevation angle resolution and accurate spacecraft attitude data. The sun-edge sensor assures that the tangent height for the ILAS measurements can be determined in absolute altitude within 150 m accuracy, provided that the refraction effect can be correctly taken into consideration [7].

6.1.2.3 Infrared spectrometer

The sun light is split by a dichroic mirror and focused on the IR spectrometer entrance slit. The IR bands cover the wavelength region from 11.77 - 6.21 μm, using a 44 element PbTiO3 pyroelectric array detector, which provide a sufficient SNR for solar occultation measurements. The crosstalk between adjacent detectors, which is caused by optical, thermal and electrical interferences, is estimated as less than approximately 3.5 %.

6.1.2.4 Visible spectrometer

The visible sun light is reflected in front of the beam splitter and focused on the entrance slit of the visible spectrometer using 4 lens transfer optics. The visible band uses a part of the 12 cm telescope (3 cm in effective diameter) to reduce the sun light and to obtain better imaging quality. The visible spectrometer uses a holographic concave grating with a 1024 pixel MOS photo diode array detector, which covers the spectral range from 753 nm to 784 nm, with a 0.1 nm FWHM resolution. The visible spectrometer has a slit function approximately identical to the theoretical limit. With this high spectral resolution capability, the visible spectrometer can be well-calibrated for the wavelength in orbit using observation data from the solar Fraunhofer lines.

6.1.2.5 Instrument performance

Data obtained with the IR spectrometer, the visible spectrometer, and the sun-edge sensor, are sampled in 11 bit precision. In Table 6.1 we report the preliminary estimates of ILAS retrievals precision at different altitudes and for various trace gases. These approximate values were derived from the radiometric performance of ILAS, under the assumption that the transmission for a certain optical path is mainly affected by the absorber in the tangent layer. Thus, the precision of ILAS measurements is calculated using the detector noise and the sensitivities of the instrument against changes of the absorber amount at the tangent layers.

6.1.3 Comparison methodology and results

In this study neural network estimated profiles have been compared to the ozone profiles retrieved by the ILAS. From November 1996 to June 1997 all coincident measurements between the two observing systems have been selected. The time interval is limited by the availability of ILAS profiles [1]. In this comparison exercise we defined coincident two measurements if they were taken during the same day and if the centre of the GOME ground pixel is included in the region spanned by a radius of 500 km around the ILAS tangent point. Such a decision stems from the intention of taking into account the air masses displacement that can occur between the time of ILAS and GOME measurements.
A total of 3089 coincident profiles have been found; they are uniformly distributed along all longitudes, but only over the northern and southern high latitude regions. This limitation depends on the particular characteristics of the ILAS instrument, which is expressively designed for the monitoring of atmosphere at those latitudes.
The location of all coincident measurements is shown in Figure 6.3. In the Northern Hemisphere they cover the area between 56° and 72° in latitude, while in the Southern Hemisphere they

cover the area between -63° and -83°. In the first case the northern regions of Europe, Asia and America are monitored, while in the second case measurements only belong to the Antarctic zone. With regards to the temporal location of the measurements, the data set contain different seasonal characteristics, even though not uniformly distributed throughout the data, as reported in Table 6.2. Ozone profiles retrieved from ILAS are reported on a regular grid of 50 equidistant altitude levels, ranging from the Earth's surface to an altitude of 50 km. Nevertheless, limb measurements are not able to provide reliable values of ozone concentration at lower altitudes, and ILAS ozone profiles actually start (values in the first levels are put equal to zero) from the height of cloud top in the lower stratosphere (11-15 km) up to the higher stratosphere (50 km). On the other hand, neural network algorithms used to retrieve ozone profiles from GOME provide the ozone density at 20 fixed pressure levels, ranging from the Earth's surface (1013 hPa) to the mesosphere (0.01hPa), as already described in Chapter 5.
In order to properly compare GOME and ILAS ozone profiles, ILAS vertical resolution has been reduced to the GOME resolution. Taking advantage of the pressure profiles included in the ILAS products and corresponding to the 50 equidistant altitude levels, ILAS ozone values have been interpolated at the GOME pressure levels. The ILAS ozone concentration at one GOME pressure level has been calculated as a mean value of ILAS ozone concentrations that fall into a pressure interval around the fixed GOME pressure level. In this way the ozone values of the two instruments are believed to investigate the same portion of atmosphere and, then, are comparable.
In Figure 6.4 we report the results of the statistical analysis carried out using the coincident profiles belonging to the entire dataset. The relative difference between GOME and ILAS measurements has been defined according to the following expression:

$$RD_i = \frac{O_3\_GOME_i - O_3\_ILAS_i}{O_3\_ILAS_i} \tag{6.1}$$

where *i* denotes the pressure level at which the difference is computed. For each pressure level characterizing the GOME retrieval grid, the mean value and the variance of the relative difference have been evaluated for both GOME algorithms. The solid line represents the profile of the mean value of the relative difference, while the dashed lines indicate the variance. The mean value of the relative difference, averaged over all the pressure levels and for all the available profiles, is 12% for the algorithm GOME_NN1 and 10% for the algorithm GOME_NN2. For levels with pressure values lower than 50 hPa the dispersion around the mean values is around 25%, while it increases at higher pressure values (i.e. down to the troposphere), where ozone concentration variability is larger and the differences between ILAS limb-sounding technique and GOME nadir-viewing geometry becomes stronger.
The variation of ILAS-GOME relative difference as a function of the solar zenith angle (SZA) is shown in Figure 6.5 and Figure 6.6 for GOME_NN1 and GOME_NN2, respectively. No significant trend can be observed until the value of 75°, but for greater angles a decrease of the performance can be noted for both algorithms. This may be due to two main reasons. First, the scarce presence of measurements with SZA larger than 75° in the training set. Second, the longer slant path corresponding to larger zenith angles involves the single-multiple scattering effects to play a not negligible role, decreasing the signal-to-noise ratio and making the retrievals less accurate.
In Figure 6.7 we considered the unfavourable case of a subset of profiles characterized by a SZA greater than 75° and plot the statistical behaviour of the inter-comparison for the two algorithms. We can observe their different robustness, as algorithm GOME_NN1 degrades much more than algorithm GOME_NN2. This may be explained by the fact that the second algorithm, even if using a smaller number of input measurements properly selected by the Extended Pruning procedure, exploits the information from a larger spectral range. On the other hand a smaller

signal-to-noise ratio may attenuate some good physical properties of the GOME_NN1 algorithm such as the sensitivity to temperature.
In Figure 6.8 two examples of ozone profiles retrieved by neural algorithms, compared with the corresponding re-gridded ILAS profiles, are reported. The two profiles refer to different seasonal and geographical condition, as pointed out by the change of position of the ozone peak. The figure shows how GOME_NN2 is more capable (with respect to GOME_NN1) to follow natural variations in the ozone distribution.

6.2 Validation with lidar measurements

The main purpose of this validation work is to assess the capability of the neural network algorithms to generalize out of the temporal window used to build up the training dataset (March 1998-March 1999). Retrieved ozone profiles from GOME measurements from July 1995 to June 2003 have been compared with independent lidar ozone profiles obtained during the same period at 6 stations, and the GOME-lidar profile differences have been evaluated. The lidar was one of the first instruments to be selected in the frame of the Network for Detection of Stratospheric Change (NDSC) to measure the ozone vertical distribution and other atmospheric parameters like temperature, aerosols or water vapour. The NDSC, established in 1991, provides such a consistent, standardized set of long-term measurements of atmospheric ozone, via a network of sites distributed all around the globe. This network consists of more than 70 high-quality, remote-sensing research stations for observing and understanding the physical and chemical state of the stratosphere and upper troposphere, and for assessing the impact of stratosphere changes on the underlying troposphere and on global climate. The measurements priorities include ozone and related parameters, such as temperature, aerosols and various trace gases involved in atmospheric chemical processes. The quality of data is regularly monitored under the NDSC protocol. A detailed description of instruments and measurements sites can be found at the following website: http://www.ndsc.ws.
The validation of the 8 years of ozone profiles was made possible by the resources and the performances of the European Grid project, EGEE (Enabling Grid for E-sciencE) [8]. The retrieval processing was done on EGEE or on a local Grid at ESRIN/ESA. All the 8-years of retrieved data obtained with the two versions of the neural network algorithm (GOME_GOME_NN1 and GOME_NN2) have been stored on EGEE storage elements with the lidar data. This set-up permits to validate in many sites and with different criteria the satellite data in an easy way.

6.2.1 Principle of ozone lidar measurements

The lidar is an active remote sensing instrument based on the interaction between the laser light and the atmosphere. According to the atmospheric parameter to be measured, lidar systems use various light-matter interactions, such as Rayleigh, Mie and Raman scattering, absorption or fluorescence [9]. The DIfferential Absorption Lidar (DIAL) technique used for ozone has the main advantage of providing range-resolved and self-calibrated measurements in the whole stratosphere with a reasonably good vertical resolution [10]. The technique requires the simultaneous emission of two laser beams centred at two wavelengths characterized by a different ozone absorption cross-section; the spectral range is chosen in the UV range, where the ozone absorption is most efficient. The ozone number density is retrieved from the lidar signals according to the following equation [11]:

$$n_{O_3}(z) = \frac{-1}{2\Delta\sigma_{O_3}(z)} \frac{d}{dz} \ln\left(\frac{P(\lambda_1, z) - P_{b1}}{P(\lambda_2, z) - P_{b2}} \right) + \delta n_{O_3}(z) \qquad (6.2)$$

where $n_{O3}(z)$ is the ozone number density at altitude z, $P(\lambda_i,z)$ is the number of detected photons at wavelength λ_i backscattered from altitude z, P_{bi} is the background radiation at wavelength λ_i and $\Delta\sigma_{O3}(z)$ corresponds to the differential absorption cross-section $\sigma_{O3}(\lambda_1,z) - \sigma_{O3}(\lambda_2,z)$ (ozone absorption cross-sections depend on temperature and thus on altitude). δn_{O3} is a correction term depending on absorption by other constituents and Rayleigh and Mie differential extinction and scattering. δn_{O3} is expressed by:

$$\delta n_{O_3}(z) = \frac{1}{\Delta\sigma_{O_3}(z)}\left[\frac{1}{2}\frac{d}{dz}\ln\left(\frac{\beta(\lambda_1,z)}{\beta(\lambda_2,z)}\right) - \Delta\alpha(z) - \sum_e \Delta\sigma_e n_e(z)\right] \qquad (6.3)$$

where $\beta(\lambda_i,z)$ is the total atmospheric backscatter coefficient at wavelength λ_i and altitude z, $\Delta\alpha(z)$ is the differential atmospheric extinction $\alpha(\lambda_1,z) - \alpha(\lambda_2,z)$ linked to Rayleigh and Mie scattering and $\Sigma_e \Delta\sigma_e n_e(z)$ is the differential extinction by other atmospheric compounds.
In the DIAL technique the laser wavelengths are chosen so that the term δn_{O3} represents less than 10% of the term derived from the slope of the lidar signals (the argument of the logarithm in Eq. 6.2) in the altitude range of interest. The derivation of the ozone number density from the laser signals shows thus that the DIAL technique is a self-calibrated technique which does not need the evaluation of instrumental constants. In the altitude range where both laser beams are contained in the telescope field of view, the accuracy of DIAL ozone measurements depends on the accuracy of the ozone absorption cross-sections used in the retrieval, on the laser line width, on the estimation of δn_{O3} and on the linearity of the acquisition device. The absolute value of the ozone absorption cross-sections in the Huggins band below 320 nm are known with an accuracy of ±3% and their relative precision over the same spectral range is reported to be better than 1-2%; the influence of the laser line width was estimated to introduce a small error of the order of 0.8% [11].
The precision of the measurement, defined by the statistical error due to the random character of the detection process, basically follows the Poisson statistics and is the result of a compromise between the experimental system characteristics, as expressed by the following relation [12]:

$$\varepsilon_s(z) \propto (A\Delta z^3 P_0 T_a)^{-\frac{1}{2}} \qquad (6.4)$$

where A represents the telescope area, Δz the vertical resolution, P_0 the emitted laser power and T_a the duration of the measurement. Due to the rapid decrease of the signal-to-noise ratio with the increasing altitude, it is necessary to degrade the vertical resolution of the measurement in order to limit the statistical error. The DIAL stratospheric ozone lidar profiles are thus generally characterized by a vertical resolution varying from several hundred meters in the lower stratosphere to several kilometres around 50 km [11].

6.2.2 Comparison methodology

As already mentioned in previous chapters, neural network algorithms can retrieve new ozone profiles (i.e. not belonging to the training set) by using GOME measurements only. All GOME data from July 1995 to June 2003 have been processed with both neural algorithms. The validation exercise carried out during this study aims at assessing the effectiveness of the two retrieval procedures by the systematic inter-comparison with lidar measurements provided by the NDSC. As already done for the comparison between GOME and ILAS ozone profiles, also for the validation with lidar measurements we need to face with the two following crucial issues: firstly, proper selection criteria, both in space and in time, have to be set in order to compare two measurements concerning approximately the same air masses. Secondly, the profiles to be compared must have an equivalent vertical resolution, in order to represent the ozone

concentrations from the same portions of atmosphere. In what follows a description of the methodology adopted to select the GOME data and transform lidar profiles is given in details.

6.2.2.1 Selection of the representative GOME co-located profile

In this work we assumed that a satellite and a ground-based measurement are co-located when the centre of the GOME ground pixel is within a region of ±2.5° in latitude and ±5.0° in longitude around the lidar station, and when the time interval between the two measurements is less than 12 hours. The chosen criteria are stricter than those adopted by the World Meteorological Organization (WMO) for the assessment of ozone trends [13], which were ±2.5° in latitude and ±12.5° in longitude.
For a single lidar measurement different GOME ozone profiles satisfy the selection criteria and a representative GOME co-located profile must be defined. In this work we considered the mean profile of all selected GOME measurements. This method is not optimal for those lidar stations characterized by a large variability of atmospheric conditions, like the station of Dumont d'Urville (66.7° S, 140.0° E) which is close to the edge of the polar vortex. In this case the probability of considering GOME measurements relative to air masses different from those sounded by the lidar is high. However, the selection of the GOME profile closest to the lidar location is not solving the problem: it should be reminded here that GOME is performing its measurements at around 10:30am local time, while lidar measurements are usually taken at night (10pm to 4am local time). In any case a minimum time difference of about 6-8 hours implies that the air mass probed by GOME is not the same as the one seen by the lidar. The choice of averaging a large number of GOME profiles is assumed to smooth this effect.

6.2.2.2 Transformation of the lidar profile into the GOME resolution

Lidar and GOME ozone profiles do not have the same vertical resolution. On one hand, lidar measurements, obtained with the DIAL technique, are characterized by a vertical resolution ranging from several hundred meters in the lower stratosphere (10-20 km) to several kilometres in the upper part (above 40 km) [11]. On the other hand, GOME ozone profiles, retrieved by means of the neural algorithms, are reported on a grid of 20 fixed pressure levels, from the Earth's surface (1013 hPa) to the mesosphere (0.01 hPa). The vertical resolution, expressed as the FWHM of the averaging kernels characterizing the RAL profiles used in the training phase, is 10-12 km in the troposphere and 4-5 km in the middle stratosphere (see Chapter 5).
In order to compare lidar and GOME ozone profiles, the conversion of the lidar vertical resolution (km) into the GOME profile sampling (hPa) is performed using the pressure profiles included in the lidar products. The lidar pressure profiles result from a combination of nearby radiosounding and the COSPAR (COmmittee on SPAce Research) International Reference Atmosphere 1986 (CIRA-86) model [14]. The mean value of lidar ozone concentrations that fall within a pressure interval centred on each GOME pressure level is taken as the reference lidar value to be compared to the GOME ozone value. This method has been successfully used to compare GOME_GOME_NN1 and GOME_NN2 ozone profiles with the ILAS instrument.

6.2.3 Results

The extended validation has been carried out over the entire period of time covered by available GOME data, from July 1995 to June 2003. We focused on 6 principal lidar stations belonging to the NDSC: Andoya, Norway (69.3° N, 16.0° E), Dumont d'Urville, Antarctic (66.7° S, 140.0° E), Lauder, New Zealand (45° S, 169.7° E), Mauna Loa, Hawaii (19.7° N, 155.1° W), Observatoire de Haute Provence, France (43.9°N, 5.7° E), Table Mountain, USA (34.4° N, 117.7° W). These stations have been chosen to account for different atmospheric conditions. The final dataset consists of 1788 coincident ozone profiles: 738 for the Observatoire de Haute Provence, 323 for Mauna Loa, 243 for Table Mountain, 251 for Lauder, 144 for Andoya and 89 for Dumont d'Urville. We can see that mid-latitude and tropical atmospheric conditions are well-

represented, while high-latitude regions provide a significantly smaller number of coincident profiles. The difference in the number of profiles is explained by the lidar observation constraints, i.e. night-time and clear sky conditions. In Table 6.3 the characteristics of each lidar site are summarized.
The results of the comparison between lidar and GOME ozone values, carried out over all the considered time period, are shown in Figure 6.9 for each lidar station. The scatter plot between lidar and GOME_GOME_NN1 ozone concentrations is represented with blue marks, while red diamonds show the same plot for GOME_NN2; lidar and GOME ozone values are reported, with a logarithmic scale, on the abscissa and on the ordinate respectively. GOME_NN2 shows less scatter around the bisector for all the stations, and seems to compare better with the lidar with respect to GOME_GOME_NN1. In Table 6.4 the coefficients of the linear fit between lidar and GOME measurements are reported, while in Table 6.5 the correlation coefficients, calculated at each pressure level for all the available data, are summarized. The calculated values state that the fit is good for both neural algorithms and for all the lidar stations; nevertheless they confirm that GOME_NN2 performs better than GOME_GOME_NN1, as the linear fit straight line of GOME_NN2 is closer to the bisector than the line of GOME_GOME_NN1, and values of the correlation coefficient (R) for GOME_NN2 is always closer to unity.
In Figure 6.10 we report the results of the statistical analysis for the 6 lidar stations considered in this work. In each panel, the vertical solid line represents, as a function of pressure levels, the mean value of the relative difference between both sets of measurements, calculated as:

$$RD_i = \frac{O_3\_GOME_i - O_3\_LIDAR_i}{O_3\_LIDAR_i} \tag{6.5}$$

while the horizontal bars represent the 2σ error (also known as standard error), which correspond to a confidence interval of 95%, expressed by the following equation:

$$\sigma_i = \sum_{k=1}^{N} \sqrt{\frac{\left(RD_{i,N} - \overline{RD}_i\right)^2}{N-1}} \tag{6.6}$$

Results for GOME_NN1 are reported in blue, and in red those for GOME_NN2; for better readability of plots, the bias and 2σ error profiles are plotted with a small vertical shift between GOME_NN1 and GOME_NN2, in order to avoid overlapping of lines. The statistical analysis shows small biases between GOME and lidar measurements between 100 hPa and 3 hPa for the mid and tropical latitudes stations, even if sometimes significant at 2σ level. The biases values range from 1% to 20% (positive or negative) for both neural algorithms, while the standard error is of about 5%-10%. Above the level of 3 hPa the differences increase, and the biases can reach values of about 30%-50%. More pronounced differences are found, especially for GOME_NN1, at Andoya and Dumont d'Urville stations, where the biases can reach values larger than 20% below the level of 3 hPa and larger than 50% above. Nevertheless, these differences are characterized by higher values of the 2σ errors (30%-50%), which means we can not diagnose the difference between the two measurements. The increasing in biases values can be explained by the fact that these two sites are located in the Arctic and Antarctic region respectively, where the variability of atmospheric conditions is large. Moreover, the solar zenith angles characterizing GOME measurements in these regions are very high (from 75° to 90°). In this case the two neural algorithms can not be sufficiently reliable, as they have been trained with a dataset including only measurements for which the solar zenith angle was smaller than 75°. Such a limitation arises from the RAL algorithm version adopted for our study, which provides ozone profiles only for restricted observation conditions. The necessity to develop a new neural network algorithm expressively designed for poles is unquestionable, and will be considered for future research activities.

The neural algorithm GOME_NN2 shows, in general, smaller biases with respect to GOME_NN1, especially for the polar stations of Andoya and Dumont d'Urville (situated at the edge of the polar vortex, where the variability of atmospheric conditions is large), but also for the Observatoire de Haute Provence and Lauder, which are respectively located in the northern and southern mid-latitude regions. For the tropical sites of Mauna Loa and Table Mountain, where the atmosphere is more stable, the biases of the two neural algorithms are very similar. Such better retrieval performances and robustness properties of GOME_NN2 are due to the wider used spectral range and to the automatic selection procedure characterizing this algorithm, which allows capturing the most significant spectral signatures of ozone absorption. The validation results obtained for the neural algorithms are comparable with those reported for the OE methods in [15], which demonstrated that the ozone profiles retrieved from GOME have a precision of 5-10% and a bias up to 5% or 20% depending on the success of recalibration of the input spectra.
In Figure 6.11 a plot of the average lidar ozone profile measured at all the stations and the corresponding GOME average profiles from both neural algorithms is reported, in order to show how GOME compares in general to the lidar. The black line represents the average lidar profile, the blue and red lines the average GOME_NN1 and GOME_NN2 respectively. The horizontal bars represent the corresponding standard deviations from the mean values. The figure confirms that GOME_NN2 compares better than GOME_NN1 with lidar measurements, especially in the lower part of the stratosphere.
In Figure 6.12 the values of monthly biases between lidar and GOME measurements performed at the Observatoire de Haute Provence are reported, as a function of time, for both neural algorithms (blue marks for GOME_NN1 and red diamonds for GOME_NN2). Each panel is representative of a single pressure level characterizing GOME neural retrievals (only pressure levels from 100 hPa to 1.78 hPa, which contain a statistically significant number of coincident measurements, are considered). For each NN algorithm we also computed the bias linear trend over the entire period of time covered by data, represented by the two solid lines. The amplitude of the seasonal component of the relative differences is rather small although not negligible, and can be ascribed to the dependence of the retrieval performance from SZAs of GOME spectral measurements. Largest values for the biases occur, in general, for the first and/or the last months of the year, when the solar zenith angles are larger (typically between 60° and 67°) with respect to solar angles characterizing the central part of the year (from 25° to 50°). Moreover, it can be noticed that, generally, the seasonal variation of the bias is smaller for GOME_NN2 with respect to GOME_NN1, and also the trend line is smaller for GOME_NN2 at almost all the considered pressure levels. In Table 6.6 we report the values of seasonal biases between lidar and GOME measurements at the Observatoire de Haute Provence for the algorithm GOME_NN2.
This analysis enables to assess the impact of the instrument degradation on the retrieval. In fact, from 2000 on, GOME is suffering from a degradation of the scan diffuser plate and an increase of the internal temperature [16]. The algorithm GOME_NN2 seems to be poorly affected by instrument degradation, while GOME_NN1 exhibits a general increase of the relative difference for almost all pressure levels. Such trends analysis has been carried out for all the stations, and the results are resumed in Table 6.7. For each lidar station, and for both neural algorithms, we report the value of the slope (m) which characterizes the trend line of the relative difference along the considered period (July 1995 – June 2003), expressed in percent per year. The grey-filled boxes indicate the pressure levels where the ozone peak is situated. The difference trends of GOME_NN2 are in general smaller than those of GOME_NN1, in particular for the pressure levels characterized by high ozone concentrations. The better robustness properties of GOME_NN2 with respect to GOME_NN1 are confirmed, for all considered pressure levels.
In Figure 6.13 we report some examples of ozone profiles retrieved from GOME by means of NN algorithms, compared with the corresponding lidar profiles, for the Observatoire de Haute Provence. The re-gridded lidar profile is reported with the black line, GOME_NN1 with the blue line and GOME_NN2 with the red line. Ozone profiles in the upper part of the figure correspond to measurements taken during June 1997, while in the lower part profiles measured during June

2002 are reported. Also from this figure we can observe the better robustness of GOME_NN2, which compares better with the lidar (for both years) and is not affected by the degradation of GOME measurements occured from the year 2000. On the other hand, GOME_NN1 shows a not-negligible worsening of retrieval performance, as clearly emphasized by profiles reported in the lower part of Figure 6.12.

6.3 Conclusion

In this chapter the results of a global validation of the ozone profiles retrieved from GOME radiances using the two independent neural network algorithms are presented. The validation has been carried out either with ozone profiles retrieved from different satellite data (ILAS) or with ground-based lidar measurements belonging to the NDSC database.
In the first case more than 3000 coincident measurements have been selected between November 1996 and June 1997, with different seasonal and geographical characteristics, and the relative difference between ozone profiles retrieved from GOME with both neural algorithms and ILAS profiles has been evaluated. If we consider the entire comparison dataset, the relative difference is generally lower than 30%, with the exception of the very first levels belonging to the lower stratosphere. The results improve with a subset characterized by a more stable atmosphere and a more convenient measurement geometry; in this case the accuracy reaches values around 10%.
The influence of the SZA on the retrieval performance has been also investigated, and a degrade for SZA values larger than 75° has been observed. Besides to a decrease of the S/N ratio, this is also probably due to a poor presence of profiles with such value in the training set so, in this case, the capability of the net to extrapolate from the training dataset is not good. However, mainly due to the wider considered spectral range, algorithm GOME_NN2 shows better robustness properties, with respect to GOME_NN1.
In the second case, the validation has been carried out considering the entire set of available GOME measurements, i.e. from July 1995 to June 2003, and 6 lidar stations, distributed around the world, adhering to NDSC standards. The agreement between GOME and lidar ozone profiles is generally very good for both neural algorithms in the atmospheric pressure range from 56 hPa and 3 hPa, with biases ranging from 1% to 20%, and standard deviations of about 5%-10%. Outside this pressure range, and for GOME measurements performed with a high SZA (> 75°), differences are larger (up to 40-60%), and the ozone number densities retrieved from GOME can not be considered as sufficiently reliable. In this case the influence of the SZA, with the exception of values > 75°, on the performance of neural schemes has been examined, and it results in a weak seasonal variation of the relative difference between lidar and GOME profiles.
This validation exercise emphasizes the better retrieval performance of GOME_NN2 with respect to GOME_NN1, either in terms of retrieval accuracy or robustness, as it is less affected by variations in the atmospheric conditions and by instrument degradation. Hence, the GOME_NN2 algorithm should be selected for a complete GOME mission reprocessing. The retrieved profiles might be used for more general atmospheric studies, such as determination of height resolved ozone trends, mini-holes detection and characterization, and estimate of tropospheric ozone on global scale.

References

[1] Sasano Y., Suzuki M., Yokota T., Kanzawa H., Improved Limb Atmospheric Spectrometer (ILAS) for stratospheric ozone layer measurements by solar occultation technique, *Geophys. Res. Lett.*, 26, 197-200, 1999

[2] Sasano Y., et al., Validation of ILAS Version 3.10 ozone with ozonesonde measurements, *Geophys. Res. Lett.*, 26, 831-834, 1999

[3] McCormick, M. P., R. E. Veiga, W. P. Chu, Stratospheric ozone profile and total ozone trends derived from the SAGE and SAGE II data, *Geophys. Res. Lett.*, 19, 269-272, 1992.

[4] Russell III, J. M., L. L. Gordley, J. H. Park, S. R. Drayson, A. F. Tuck, J. E. Harries, R. J. Cicerone, P. J. Crutzen, J. E. Frederick, The Halogen Occultation Experiment, *J. Geophys. Res.*, 98, 10777-10797, 1993.

[5] Matsuzaki, A., Y. Nakamura, T. Itoh, Rocket observation of the rotational profile of the A-band absorption spectrum of atmospheric oxygen molecule, *Annales Geophys.*, 4, 476-480, 1984.

[6] Toon, O. B., M. A. Tolbert, B. G. Koehler, A. M. Middlebrook, J. Jordan, Infrared optical constants of H2O ice, amorphous nitric acid solutions, and nitric acid hydrates, *J. Geophys. Res.*, 99, 25631-25654, 1994.

[7] Suzuki, M., Y. Sasano, T. Ishigaki, N. Kimura, N. Araki, A. Matsuzaki, Improved Limb Atmospheric Spectrometer, ILAS, *SPIE Proceedings Ser.*, 2268, 103-110, 1991.

[8] EGEE Web Site: http://public.eu-egee.org/

[9] Measures R.M., Laser Remote Sensing, John Wiley, New York, 1984.

[10] Shotland R.M., Journal of Applied Meteorology, Vol. 13, 71, 1974.

[11] Godin-Beekmann S., Porteneuve J., Garnier A., Systematic DIAL lidar monitoring of the stratospheric ozone vertical distribution at Observatoire de Haute-Provence (43.92°N, 5.71°E), *Journal of Environmental Monitoring*, 5, 57-67, 2003.

[12] Godin S., Carswell A., Donovan D., Claude H., Steinbrecht W., Mcdermid S., Mcgee T., Gross M.R., Nakane H., Swart D.P.J., Bergwerff J.B., Uchino O., Von Der Gathen P., Neuber R., Ozone Differential Absorption Lidar Algorithm Intercomparison, *Appied Optics*, 38, 30, 6225-6236, 1999.

[13] WMO, Report 43, Global ozone research and monitoring project, Assessment of trends in the vertical distribution of ozone, 1998

[14] Rees D., Barnett J.J., Labitzke K., CIRA 1986, Middle Atmosphere Models, *Advances in Space Research (COSPAR)*, 10 (12), 1990.

[15] Meijer Y.J., et al., Evaluation of GOME ozone profiles from nine different algorithms, submitted to *Journal of Geophysical Research* in December 2004

[16] Aben,I., Eisinger,M. , Hegels,E., Snel,R., Tanzi,C., GDAQI Final Report, TN-GDAQI 003SR/2000,ESA/ESRIN,2000

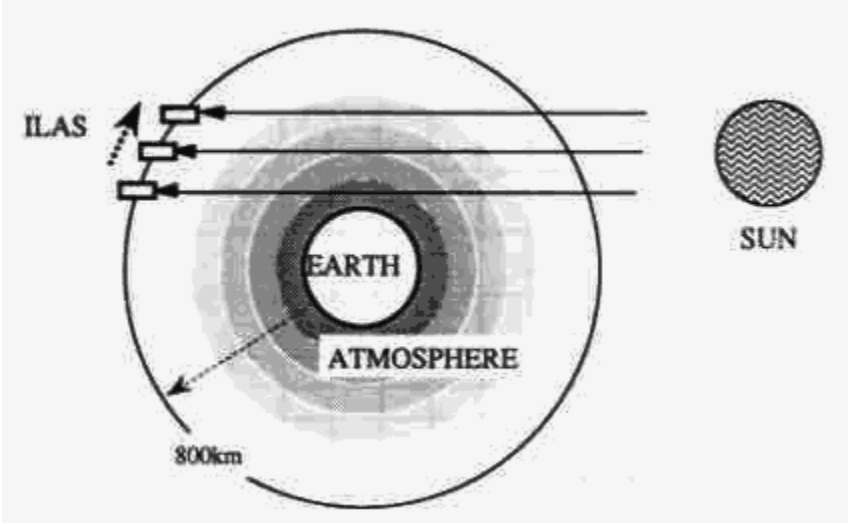

Figure 6.1. Principle of solar occultation measurements.

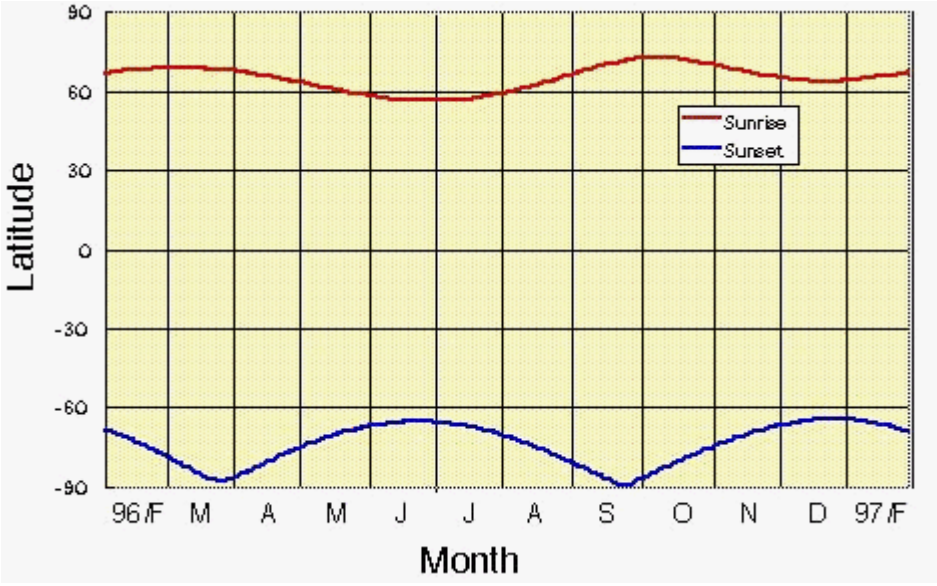

Figure 6.2. Latitude vs. time change of ILAS coverage.

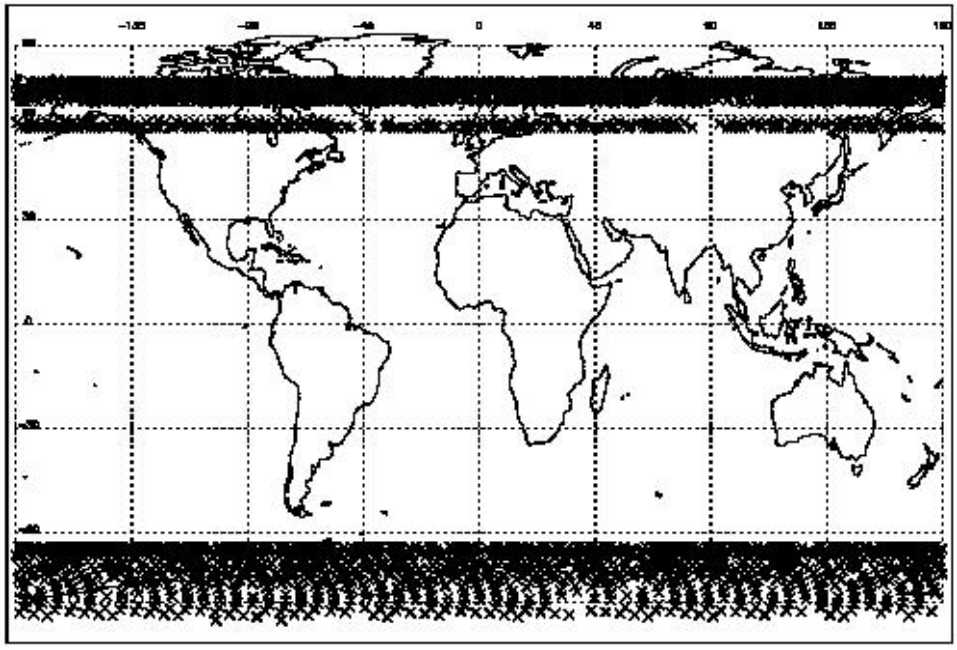

Figure 6.3. Location of coincident measurements of GOME and ILAS ozone profiles between October 1996 and June 1997.

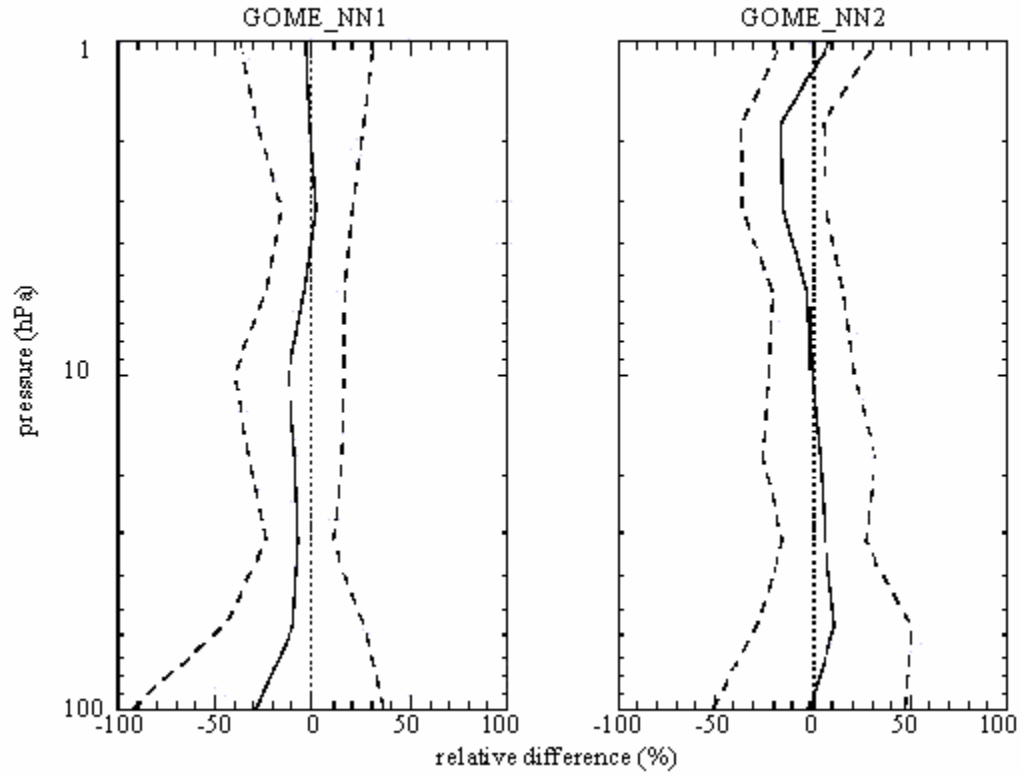

Figure 6.4. Mean (solid line) and variance (dashed lines) values of the relative difference between ILAS and GOME ozone profiles corresponding to the entire dataset.

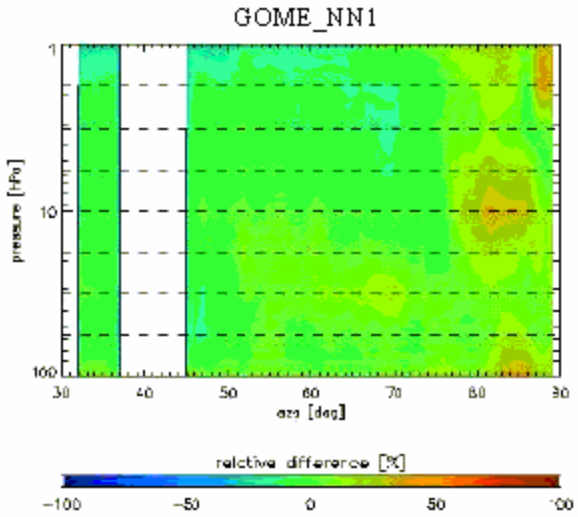

Figure 6.5. Contour plot of ILAS-GOME relative difference for the algorithm GOME_NN1. On the abscissa we report the solar zenith angles, while on the ordinate the pressure levels. Dashed lines represent the pressure levels where ILAS ozone profiles have been interpolated.

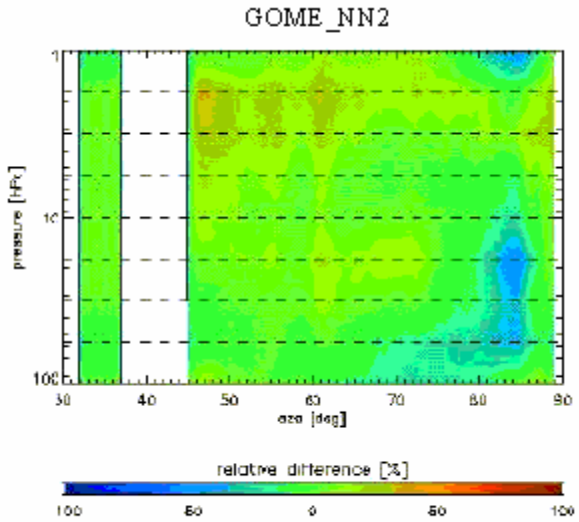

Figure 6.6. Same as Figure 6.6, but for GOME_V2 algorithm.

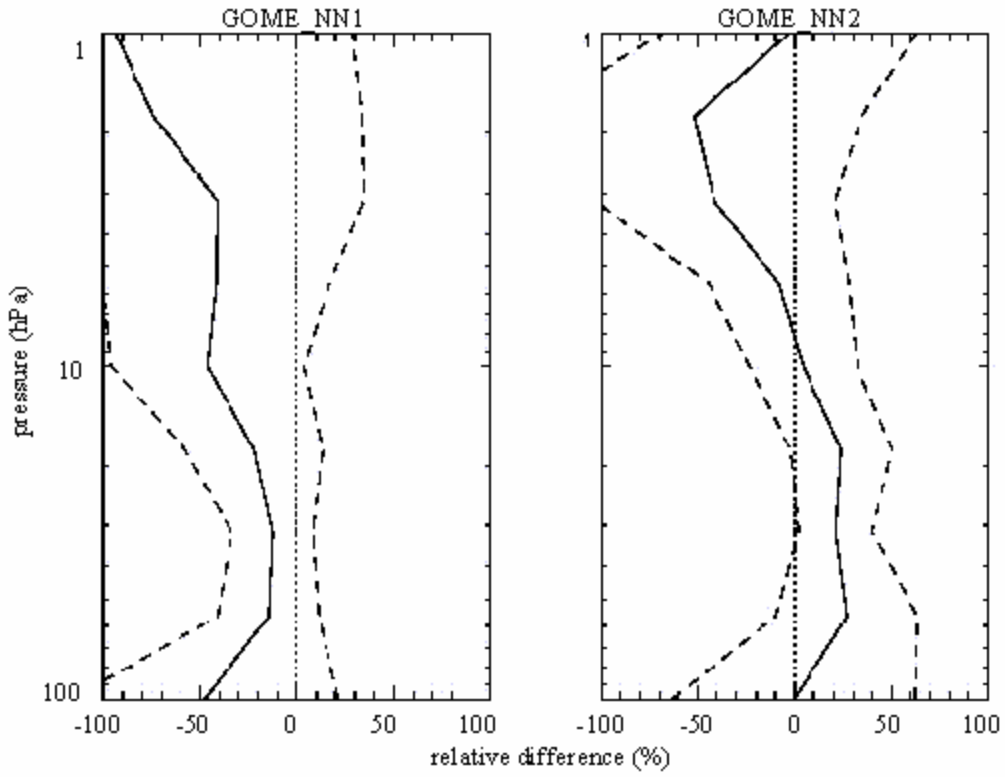

Figure 6.7. Mean (solid line) and variance (dashed lines) values of the relative difference between ILAS and GOME ozone profiles corresponding to a dataset including only profiles for which the solar zenith angle is larger than 75°.

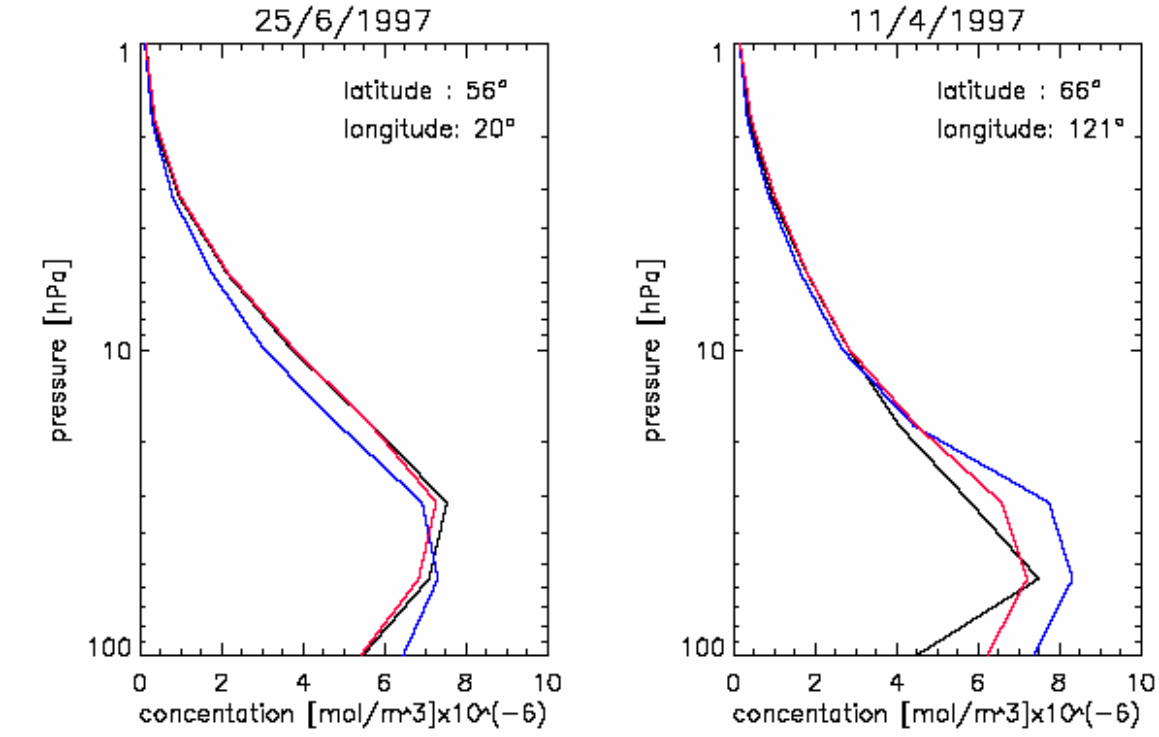

Figure 6.8. Examples of ozone profiles retrieved by NN algorithms (blue line for GOME_NN1 and red line for GOME_NN2) compared with the corresponding ILAS profile.

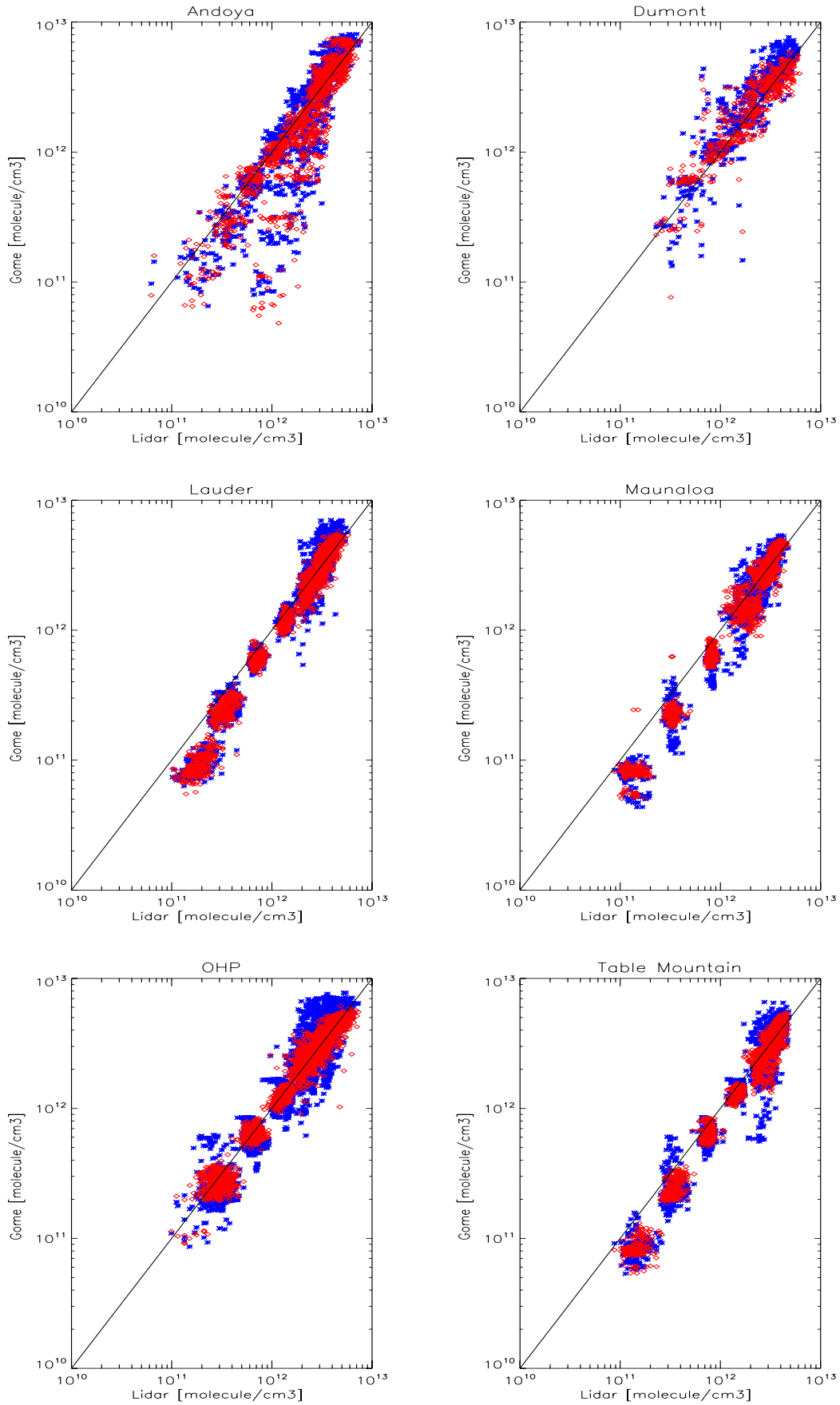

Figure 6.9. Results of the comparison between lidar and GOME ozone values, carried out over the entire dataset. Blue marks represent the scatter plot for GOME_NN1, while red diamonds for GOME_NN2. The solid line represents the bisector, i.e. the ideal linear fit straight line.

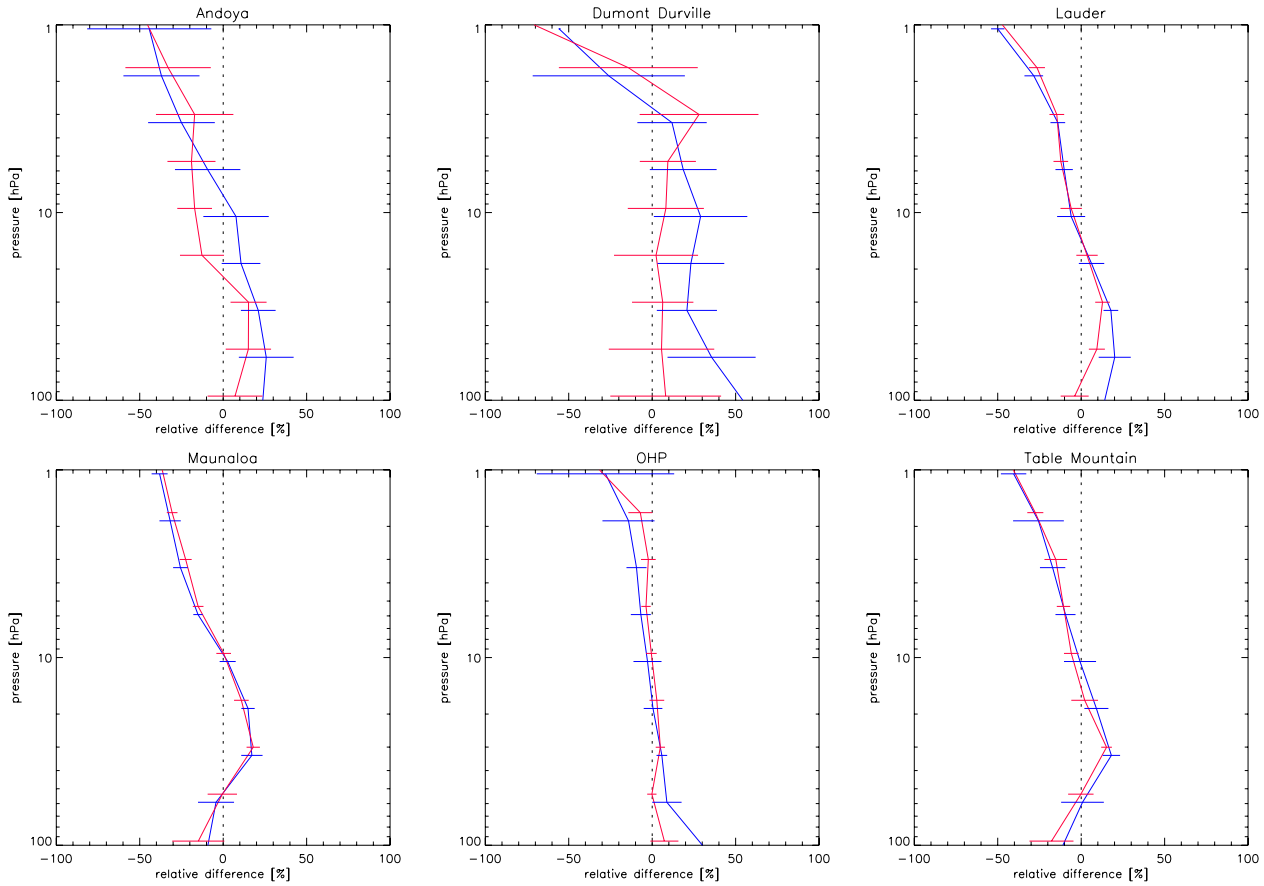

Figure 6.10. Results of the statistical analysis carried out over the entire dataset for the 6 lidar stations considered in this work. The solid line represents, as a function of pressure levels, the mean relative difference between GOME and lidar ozone values, while the horizontal bars indicate the standard deviations (blue lines for GOME_NN1 and red lines for GOME_NN2).

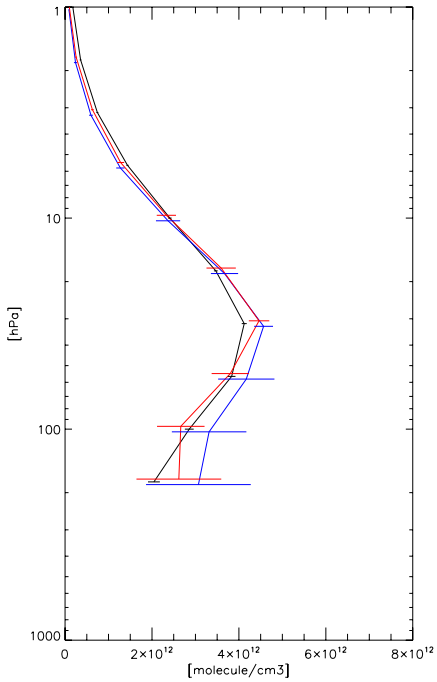

Figure 6.11. Average ozone profiles (lidar in black, GOME-NN1 in blue, GOME-NN2 in red) calculated over all available data. The horizontal bars represent the corresponding standard deviations from the mean values.

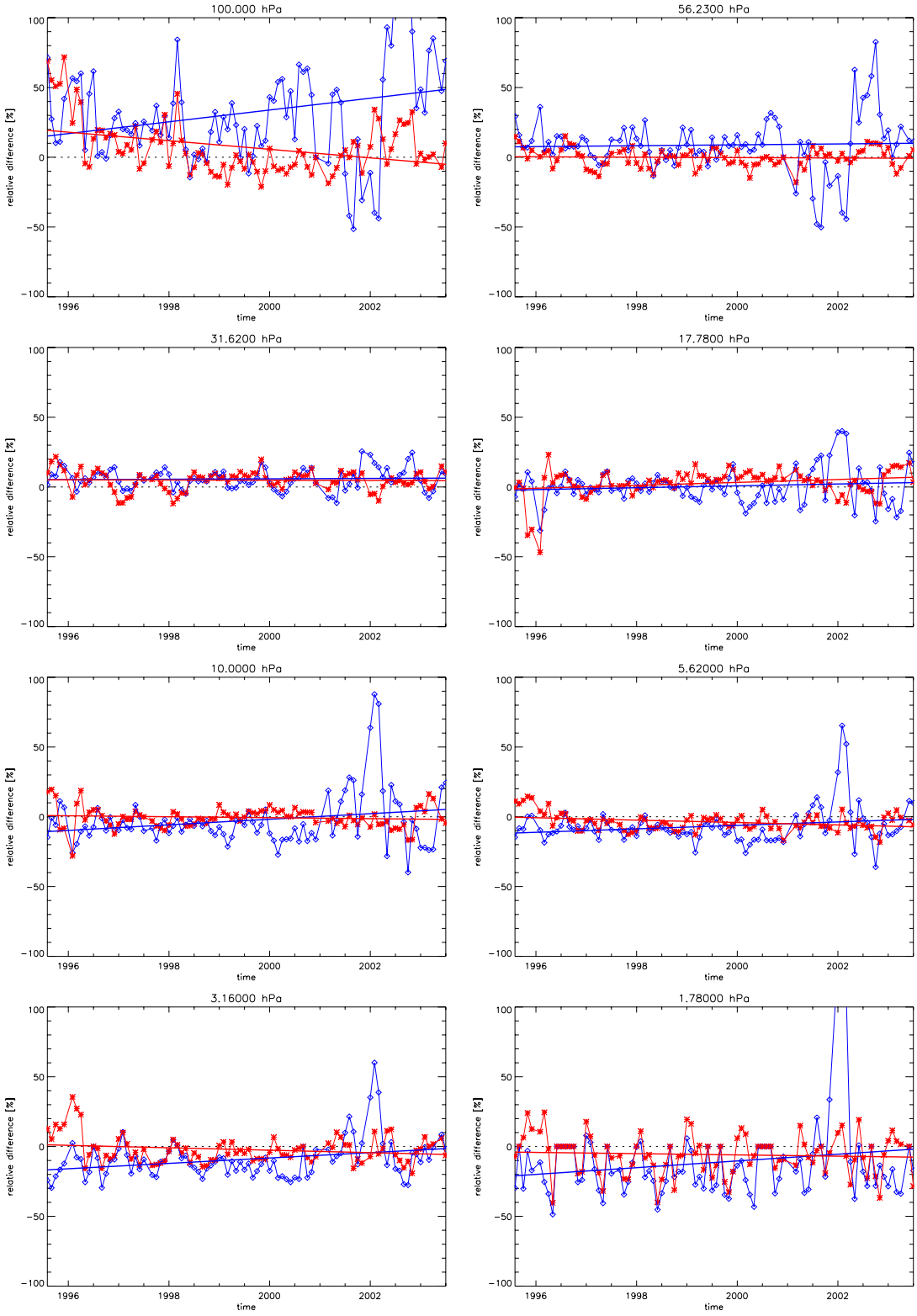

Figure 6.12. Seasonal variation of the relative differences, for both neural algorithms (blue marks for GOME_NN1 and red diamonds for GOME_NN2) and for the OHP co-located data. Each panel is representative of a single pressure level characterizing GOME neural retrievals. The two solid lines represent the trend of the relative difference over the entire period of time covered by data.

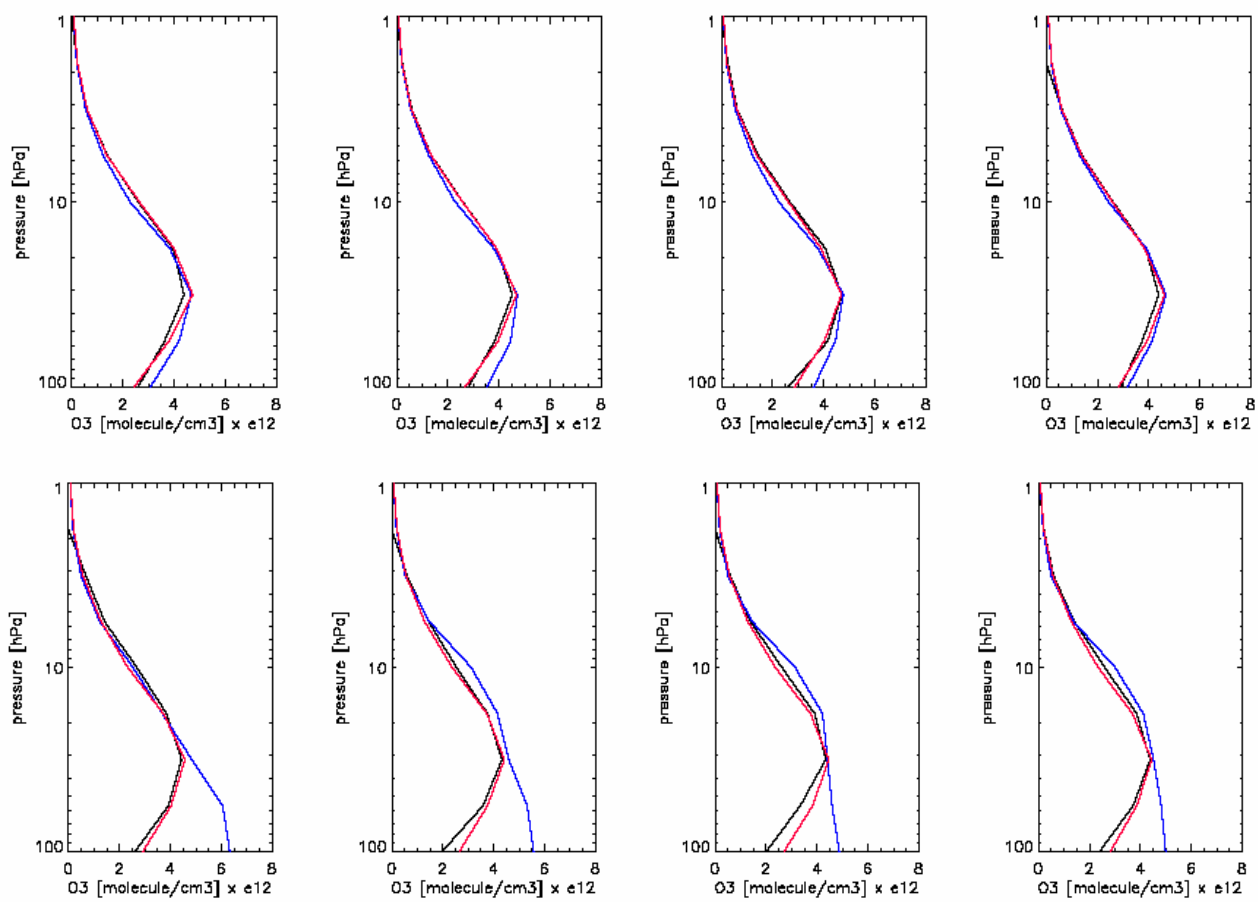

Figure 6.13. Examples of ozone profiles retrieved by NN algorithms (blue line for GOME_NN1 and red line for GOME_NN2) compared with the corresponding lidar profile.

Altitude (km)	10	20	30	40	50
O_3	± 5 %	± 5 %	± 5 %	± 5 %	± 5 %
HNO_3	± 50 %	± 10 %	± 50 %	n.d.	n.d.
NO_2	n.d.	± 5 %	± 5 %	± 20 %	n.d.
N_2O	± 5 %	± 5 %	± 20 %	± 100 %	n.d.
CH_4	± 5 %	± 5 %	± 5 %	± 10 %	± 100 %
H_2O	± 5 %	± 5 %	± 5 %	± 10 %	n.d.
CFC11	± 10 %	n.d.		n.d.	

Table 6.1. Preliminary estimation on the precision of ILAS measurements at different altitudes.

Time window	Hemisphere	Number of profiles
Nov – Dec 1996	North	381
Nov – Dec 1996	South	603
Jan – Mar 1997	North	850
Jan – Mar 1997	South	684
Apr – Jun 1997	North	571

Table 6.2. Characteristics of the entire data set (3089 profiles) grouped into subsets with the same climatologic characteristics of atmosphere.

Station	Latitude	Longitude	Number of profiles
Andoya	69.3° N	16.0° E	144
Dumont d'Urville	66.7° S	140.0° E	89
Lauder	45.0° S	169.7° E	251
Mauna Loa	19.7° N	155.1° W	323
Observatoire de Haute Provence	43.9° N	5.7° E	738
Table Mountain	34.4° N	117.7° W	243

Table 6.3. Characteristics of the 6 lidar stations considered for the validation of GOME ozone profiles.

Station	m (NN1)	m (NN2)	q (NN1)	q (NN2)
Andoya	0.730	0.862	0.455	0.357
Dumont d'Urville	0.756	0.946	0.177	0.033
Lauder	0.826	0.901	0.202	0.159
Mauna Loa	0.850	0.873	0.237	0.187
Observatoire de Haute Provence	0.893	0.957	0.193	0.075
Table Mountain	0.823	0.887	0.232	0.187

Table 6.4. Values of the slope (m) and the y-intercept (q) of the linear fit between lidar and GOME ozone concentrations, for both neural algorithms.

Pressure (hPa)	R (NN1)	R (NN2)
100.00	0.64	0.71
56.23	0.71	0.83
31.62	0.54	0.56
17.78	0.57	0.69
10.00	0.81	0.82
5.62	0.49	0.67
3.16	0.23	0.24
1.78	0.23	0.24
1.00	0.22	0.23

Table 6.5. Values of the correlation coefficient (R) for both neural algorithms, calculated at each pressure level for all the available data.

Pressure (hPa)	Spring		Summer		Autumn		Winter	
	Bias (%)	Standard Error (%)	Bias (%)	Standard Error (%)	Bias (%)	Standard Error (%)	Bias (%)	Standard Error (%)
100.00	0.9	2.0	14.9	4.3	7.3	5.5	8.0	4.3
56.23	0.6	1.2	5.3	1.2	0.3	1.1	5.6	1.4
31.62	6.8	0.9	8.7	1.0	5.1	1.7	1.0	1.5
17.78	5.5	0.9	2.0	1.0	0.4	3.0	4.0	3.1
10.00	1.9	0.9	0.8	1.8	2.6	1.6	1.6	1.9
5.62	3.3	0.7	2.9	1.4	6.0	1.9	2.8	1.2
3.16	2.3	1.7	6.4	1.9	5.2	1.9	4.7	2.4
1.78	15.5	4.0	15.0	3.0	5.3	3.7	1.5	2.7

Table 6.6. Seasonal biases and standard errors between lidar and GOME measurements at the Observatoire de Haute Provence for the algorithm GOME_NN2.

Station	m (% per year)	100.00 hPa	56.23 hPa	31.62 hPa	17.78 hPa	10.00 hPa	5.62 hPa	3.16 hPa	1.78 hPa
Andoya	NN1	0.02	-0.49	-0.36	0.50	1.24	1.50	1.59	1.44
	NN2	0.54	0.44	0.06	0.28	0.46	1.04	1.55	1.59
Dumont d'Urville	NN1	3.26	2.01	0.60	0.65	2.77	0.16	-1.32	2.15
	NN2	-0.23	2.30	1.33	2.07	3.22	0.19	-4.24	0.90
Lauder	NN1	0.24	0.08	0.08	0.11	0.10	0.03	0.08	0.04
	NN2	0.15	0.16	-0.07	-0.03	-0.15	-0.18	-0.32	-0.31
Mauna Loa	NN1	0.10	-0.14	0.30	0.08	-0.01	-0.03	0.05	0.13
	NN2	0.22	0.33	0.27	0.06	-0.31	-0.16	-0.14	-0.08
Observatoire de Haute Provence	NN1	0.52	0.03	0.01	0.08	0.24	0.15	0.24	0.30
	NN2	-0.37	-0.01	-0.00	0.13	-0.04	-0.10	-0.10	-0.05
Table Mountain	NN1	0.36	0.00	0.17	0.33	0.46	0.21	0.21	0.42
	NN2	0.14	0.42	0.03	0.20	0.04	-0.11	-0.21	-0.11

Table 6.7. Values of the slope (m), expressed in percent per year, of the trend line characterizing GOME-lidar relative differences, computed for both algorithms over the entire period of time considered for the validation.

Chapter 7

NNs for the retrieval of other atmospheric parameters

In this chapter the application of the neural network technique to other inversion problems is described. Two cases will be analysed: the first one concerns a retrieval algorithm for the estimation of temperature profiles from the same nadir-viewing radiance spectra already used for the ozone retrieval, while the second one refers to a preliminary study conducted to assess the feasibility of MW/Sub-mm radiometric soundings of the atmosphere from a geostationary orbit, for the estimation of temperature and humidity profiles. Both cases show the large potentialities of neural network in solving non-linear inversion problems, and the possibility to design fast and accurate retrieval algorithms to process satellite data, in order to obtain geo-physical parameters useful for atmospheric monitoring activities.

7.1 Estimation of temperature profiles from GOME

7.1.1 Introduction

The thermal structure of the atmosphere has a relevant influence on the circulation mechanisms of air masses and on the distribution of trace gases. Monitoring continuously over the whole globe the thermal state of the atmosphere is very important in order to understand the processes that control the distribution of trace species therein and to assess the impact of human activities on it. We have already observed that satellites can provide a very powerful platform for quantitative remote-sensing observations on a regular and long-term basis. The GOME instrument (see Chapter 4 for details) is designed to determine the vertical distribution of manifold atmospheric constituents, with a special emphasis on the ozone [1]. In this case, the temperature profile play a key role in determining the distribution of O_3 molecules, as it strongly influences the absorption cross-section of such a trace gas.
Preliminary retrieval of temperature profiles, followed by sequential retrieval of gas components (ozone or others) can be an useful approach to improve estimation results [2]. A drawback of this approach is that the processing time necessary to characterize the atmospheric composition is affected by the time needed to preliminary estimate temperature profiles. As extensively described in previous chapters, neural networks have been proved to have a great potential in approximating any real-valued continuous function [3], and offer the possibility to design very fast retrieval algorithms, allowing reducing the overall processing time.
In this work neural network potentialities have been exploited to design a real-time algorithm for the estimation of temperature profiles from GOME radiance spectra. On the base of good performances obtained for ozone profiles retrieval, the feedforward multi-layer perceptron (MLP) architecture (Chapter 3) has been adopted to estimate temperature profiles also. Before a neural network can be used for the actual retrieval phase, a training phase is necessary, during which the net learns to approximate the input-output significant relation. Such a learning phase consists in finding weights and biases which minimise the error function characterizing the learning process.
The basic idea of this study is to train a neural network using as reference outputs the already existing temperature profiles provided by the UK Met Office, and as inputs to the net the GOME radiance spectra which overpass the sites of temperature measurements. The matching between

GOME data and temperature profiles of the UK Met Office provides the input-output patterns needful to train the neural network. Once the training process is completed, the network will be able to estimate in near real-time new temperature profiles, which can be successively used for the retrieval of trace gases vertical distribution.
In the next section, the methodology followed to design the neural algorithm will be described. This includes the generation of input-output pairs for the training phase and the optimization of the network topology. Once trained, the retrieval performance of the net has been tested on a validation dataset, which contains examples not included in the training dataset.

7.1.2 Methodology

Input-output pairs needed to train the neural net have been generated considering GOME radiance spectra matched with temperature profiles provided by the UK Met Office. The matching between GOME data and temperature profiles was already carried out in the framework of the activities of the GOME Ozone Profiles Retrieval Working Group, coordinated by ESA, where an extensive validation of ozone profiles was accomplished. In the effort of maximizing the retrieval performance of the neural algorithm in terms of robustness and generalization properties, we included in the training set all the available input-output pairs from January 1998 to December 1998. The collected 739 training patterns contain examples representing different seasonal and geographical characteristics.
GOME radiance measurements belonging to the spectral range of 321-325 nm have been used in this study. Such a wavelength interval is attractive because of the strong dependence from the temperature profile of the ozone cross-section, which significantly affect the shape of radiance spectra. Moreover, this range shows an high signal-to-noise ratio, and was already selected during the sensitivity analysis carried out during ozone retrieval preliminary studies (Chapter 5). Taking advantage from previous works, a single input vector contains the 26 radiance values that fall within the selected spectral range, plus the calculated total ozone slant path and the solar zenith angle of GOME measurement. The corresponding output vector represents the temperature profile matched with the GOME measured spectrum, discretized into 14 pressure levels from 1000 mbar to 0.30 mbar.
Also in this case, the minimization of the error function during the learning phase has been pursued by the scaled conjugate gradient (SCG) algorithm [4], coupled with the early stopping procedure. In our case, the validation dataset contains the measurements and the corresponding temperature profiles from January 1997 to December 1997, for a total number of patterns of 606. As already described, this procedure is used in order to avoid overfitting situations and to allow the network to generalize outside of the training dataset. In Figure 7.1 the location of all coincident measurements is reported. The black triangles represent the patterns belonging to the training set, while the blue ones the patterns of the validation set.
During the training phase, the network topology has been then optimized from the point of view of the number of neurons in the hidden layer. Different neural networks have been set up, increasing the number of units in the hidden layer from 1 to 40. The retrieval error committed on the test set by the different networks has been evaluated, as a function of the number of hidden units, and the results are reported in Figure 7.2. It can be noted that an optimum number or at least an optimum range exists. In fact, if the number of hidden neurons is too small, the input-output associative capabilities of the net are too weak. On the other hand, this number should not be too large; otherwise, these capabilities might show a lack of generality (they would be too tailored on the trainin set). It turns out that a fair compromise between these two contrasting requirements has to be found.
In our study, the net having 26 hidden units shows the better performance in terms of the retrieval accuracy assessed over the validation dataset, and has been chosen for the actual retrieval phase.

7.1.3 Results

For the final retrieval algorithm, the neural network with 28 input units, 26 hidden units and 14 output units has been considered. Once trained, the retrieval performance of the net has been tested on the validation dataset. In Figure 7.3 we report some examples of the neural estimation of temperature profiles belonging to this dataset. It can be observed that the designed neural network algorithm is able to catch up the different shapes characterizing different temperature profiles, and in particular is capable to single out the height and the extent of the tropopause, i.e. where the temperature gradient above the troposphere vanishes.
The results of the neural retrieval performed on data not included in the learning set are in good agreement with the corresponding temperature profiles provided by the UK Met Office. The retrieval performance evaluated over the whole validation dataset is reported in Figure 7.4. The root mean square error (dashed line), calculated over the 606 profiles, is significantly smaller (more than 50%) with respect to the standard deviation of the profiles from their means (solid line). In Figure 7.5 we report the profile of the bias between the true profiles and the corresponding neural estimations. The absolute value of the bias is about 1 K in the stratosphere, and 2 K in the troposphere, where the nadir measurements are affected by the presence of the atmospheric layers above. But what is most remarkable is that the neural estimation is real time; once GOME data are made available, few seconds are enough to characterize the atmosphere in terms of temperature values.
The dipendence of retrieval results from pressure levels and solar zenith angles (SZA) which characterize spectral measurements has been also assessed, as reported in Figure 7.6. No significant dependence of the retrieval performance from heights and SZAs can be observed, excepted for lower atmospheric layer (due to the nadir-viewing geometry) and for solar angles larger than 70°, where the relative difference between reference and calculated temperatures can reach values around 10%. Strange larger errors (even if smaller than 5%) can be observed at SZAs between 30° and 35°; a reason for this has not been found at the moment, anf further analysis will be carried out.

7.2. Temperature and humidity profiles from geostationary radiometric soundings

7.2.1 Introduction

The three-dimensional fields of temperature and humidity are most important for weather predictions, as they directly enter the motion equations to be numerically integrated to perform weather forecasting. It is therefore not surprising that, as soon as the space technology enabled it, atmospheric temperature profiling from satellites was energically addressed. Already in 1969 experimental atmospheric sounding missions using IR were introduced (on Nimbus-3, with IRIS and SIRS instruments). MW sounding started some years later (NEMS on Nimbus-5, 1972). The first complete IR/MW sounding system is operating since 1978 (TOVS on TIROS-N), and has been recently upgraded (ATOVS on NOAA-15, 1998). The first experiment of atmospheric sounding from a geostationary orbit, limited to the IR, started in 1980 with the instrument VAS on GOES-4, and has been replaced by SOUNDER on GOES-8. The interest for temperature and humidity profiling from geostationary satellites stems from the requirement for 1-hour observing cycle, which can not be provided by low-orbiting satellites. Moreover, MW observations can provide a significant improvement (with respect to IR) for atmospheric sounding in cloudy areas. In spite of the greatest interest for all-weather temperature/humidity profiling from geostationary orbit, plans to extend MW sounding instruments to geostationary satellites do not exist, since this is considered not feasible with current technology.

This study was conducted in the framework of the Geostationary Observatory for Microwave Atmospheric Sounding (GOMAS) project, partially funded by the Agenzia Spaziale Italiana (ASI). The purpose of GOMAS is to establish the feasibility of MW/Sub-mm sounding from geostationary orbit, to prove the applicability of such observation to frequent and nearly-all-weather atmospheric temperature and humidity sounding, and to estimate very frequently precipitation rate, associated to cloud liquid/ice water discrimination. Such kind of measurements are expected to be very useful for meteorological services and climate research centres, allowing improvements in numerical weather predictions, in the characterization of hydrological regimes and in the description of the water cycle in general circulation models. In the framework of the GOMAS project, this work reports on the development of neural network inversion algorithms for temperature and humidity profiling.

7.2.2 The GOMAS approach

The main problem with MW sounding from geostationary orbit is the diffraction limit which implies, at MW frequencies, very large antenna diameters. Table 7.1 summarizes (for different atmospheric parameters) the relation between the required frequency for sounding, the required spatial resolution and the antenna diameter. In order to alleviate the problem of the antenna dimensionality, the GOMAS project moves to higher frequencies, i.e. in the sub-mm range. The most interesting options are:

- the frequency of 380 GHz for water vapour, implying an antenna of 1.7 m for 20 km resolution
- the frequency of 425 GHz for temperature, implying an antenna of 1.0 m for 30 km resolution

As a matter of fact, temperature/humidity profiling at these frequencies is not attractive, since the strong water vapour continuum in the sub-mm range implies that the lower troposphere is not well-sensed, which is a regrettable problem for geostationary sounding missions addressed to weather forecasting. Therefore, it becomes evident that lower bands at 119 GHz and 183 GHz must be kept. The band centred at 55 GHz must also be included, since it is the less sensitive to clouds; moreover, this band is needed (together with the band at 183 GHz) for intercalibration with MW sounding from polar orbiters.
The GOMAS project envisages a 3-m antenna, which would provide the following capabilities:

- full sounding of the troposphere, from bands at 119 GHz and 183 GHz, with support of the band at 55 GHz, with a spatial resolution of 30 km
- high troposphere sounding, from bands at 380 and 425 GHz, with a resolution of 10 km
- sounding of precipitations, from bands at 380 and 425 GHz, with a resolution of 10 km

It has to be noted that, due to the wide range of frequencies and to the severe radiometric accuracy required for atmospheric sounding, synthetic antenna concepts are ruled out. Receivers at frequencies reported above are not off-the-shelf, but probably do not represent a critical development. For each of the 5 bands, centred at 55, 119, 183, 380 and 425 GHz, 6 to 10 narrow banwidth channels are planned to be implemented by heterodyne spectrometers with quasi-optical bands separation. The total number of channels will be around 40, as reported in Table 7.2.
In Figure 7.7 we report the Incremental Weighting Functions (IWFs) for the 4 bands (with exception of 55 GHz band) selected within the GOMAS project for temperature and humidity profiling [5]. The IWFs represent the relationship between infinitesimal variations in the atmospheric parameter of interest, and the upwelling brightness temperatures measured by the sensor, and provide the theoretical base for the choice of radiometric channels. Numbers reported

beside each function show the frequency offset, expressed in MHz, from the centre of the absorption line. The figure shows that the selected radiometric channels can provide a complete sounding from the Earth surface to 40 km (for temperature profiles) and 20 km (for humidity profiles), as required by the project purpose.

7.2.3 Methodology

In this study the expected performance of a neural network algorithm aiming at the retrieval of temperature and humidity profiles from GOMAS radiometric channels has been assessed. The algorithm has been configured with the choice of the feed-forward multi-layer perceptron architecture (with one hidden layer), and with the choice of network topology, which means the number neurons in each layer. Input data are the brightness temperatures in the proposed channels, while the output ones are temperature and humidity values at different heights in the atmosphere. The study has been carried out by numerical simulations performed with the Millimeter-wave Propagation Model LIEBE-1993 [6] [7].
The analysis includes different steps:

1. generation of a representative dataset
2. training of NNs
3. retrieval exercise
4. analysis of the information content of radiometric channels

In order to generate a representative dataset, radio-soundings collected from 1980 to 1989 at the meteorological station of Brindisi (Italy) have been considered. Temperature and humidity profiles have been given as input to the forward model (LIEBE-1993), which generates the corresponding simulated brightness temperatures as seen from a satellite in the proposed radiometric channels. Clear sky conditions and two different values for the surface emissivity (0.5, 0.9), which accounts for see and land cover respectively, have been assumed. An uncorrelated gaussian noise (with a standard deviation of 0.5K and 1K) has been also considered during model runs.
The global dataset, containing 2578 profiles and the corresponding simulated brightness temperatures, has been then divided in a training set and a validation set, each one composed with 1289 patterns randomly selected. The first one will be used for NNs training (i.e. to update networks internal parameters), while the second one for assessing the retrieval performances achieved by trained networks and stopping the training phase (early stopping procedure, see Chapter 3 for details).
The input vector contains the brightness temperatures generated by the model in the proposed radiometric channels. It is formed by 33 input units: 11 units for the 118 GHz band, 7 for the 183 GHz band, 7 for the 380 GHz band and 14 for the 425 GHz band. On the other hand, the output vector represents the corresponding temperature or humidity profile, from the ground to an altitude of 30 km, with a vertical resolution of 1 km (which leads to a number of 30 output units). On the base of several experiments, 33 neurons have been selected for the hidden layer, either for the network aiming at temperature retrieval or for the net designed for humidity. The error minimization has been performed by means of the scaled conjugate gradient (SCG) algorithm, coupled with the early stopping procedure.

7.2.4 Results

In Figure 7.8 and Figure 7.9 we report the root mean square error (RMSE) obtained for temperature and humidity profiles respectively, for a surface emissivity of 0.5 (plots on the left) and 0.9 (plots on the right). The RMSE (dashed line) is always smaller than the standard deviation of profiles from their mean (solid line) in all cases, and shows the good retrieval

performances achieved by trained NNs. The RMSE for temperature (~ 2 K) is significantly smaller than the standard deviation (~ 8-10 K) at all altitudes. For humidity profiles the good retrieval performances mainly concern the lower part of the atmosphere (below 15 km); nevertheless, it has to be noted that the humidity variability is mostly concentrated in this region, while at upper altitude the atmospheric water content is more stable. In Figure 7.10 we report the RMSE for the unfavorable case of assuming an incorrect value for the surface emissivity, for temperature (plot on the left) and humidity (plot on right) retrievals. In this case the NNs have been trained with brightness temperature corresponding to an emissivity of 0.5, while the retrieval has been performed using brightness temperature generated by the model assuming a surface of emissivity of 0.9. The retrieval performance degrades in both cases, but only in the lower part of the atmosphere (which is consistent with the incorrect assumption on electromagnetic properties of the surface). This analysis points out the importance of having accurate estimations of the surface emissivity in order to obtain good retrieval results. In Figure 7.11 and Figure 7.12 some examples of retrieved profiles are shown, for temperature and humidity respectively, and confirm the capability of NNs to invert radiometric measurements.
A more detailed analysis of the information content of GOMAS radiometric channels has been also carried out. The study has been conducted by means the application of the extended pruning procedure (see Chapter 5) to trained neural networks. We remind here that a pruning procedure aims at removing the weakest, and therefore less significant, connections in a trained neural network, with the purpose of improving its generalization properties. If the procedure is prolonged, it will be a moment where an input unit will have no more connections; at this point, the corresponding input value will not contribute to the output vector estimation, and the input unit can be considered deleted. Such a technique permits to assess the relative importance of input measurements in estimating an output vector. In Table 7.3 we report the results obtained by applying the Extended Pruning to the NN trained for temperature profiles retrieval. The procedure has been stopped when the retrieval error evaluated on the test dataset increases significantly (more than 25%). It is very interesting to observe that input measurements crucial for temperature estimation (i.e. radiometric channels centered at 118 GHz and 425 GHz) have been preserved, while almost all inputs corresponding to humidity estimation (183 GHz and 380 GHz bands) have been deleted. An analogous result has been obtained by applying the same procedure to the NN trained for humidity profiles estimation, as reported in Table 7.4. In this case, the EP technique preserves the radiometric information needed for humidity retrieval (i.e. input units corresponding to 183 GHz and 380 GHz channels), while it eliminates unnecessary input measurements (coming from 118 GHz and 425 GHz bands).
Such a study conducted on the information content of radiometric channels states that NNs can not be merely considered as black boxes, since they clearly show the capability to “understand” the functional relation between the measured electromagnetic quantities and the geo-physical parameters to be retrieved.

References

[1] J.P. Burrows, M. Weber, M. Buchwitz, V. Rozanov, A. Ladstatter Weissenmayer, A. Richter, R. DeBeek, R. Hoogen, K. Bramstedt, K.U. Eichmann, M. Eisinger, The global ozone monitoring experiment (GOME): Mission concept and first scientific results, *Journal of the Atmospheric Sciences*, 56 (2), 151-175, 1999.

[2] Dinelli B.M., et al. Multi-target retrieval (MTR): the simultaneous retrieval of pressure, temperature and volume mixing ratio profiles from limb-scanning atmospheric measurements, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 84, 141–157, 2004.

[3] Bishop C.M., Neural networks for pattern recognition, *Oxford University Press*, London, UK, 1995.

[4] Moller M., A scaled conjugate gradient algorithm for fast supervised learning, *Neural networks*, 6, 525-533, 1993.

[5] M. Klein, A.J. Gasiewski, Nadir sensitivity of passive millimeter and submillimeter wave channels to clear air temperature and water vapour variations, *Journal of Geophysical Research*, 105, D13, 17481-17511, 2000.

[6] H.J. Liebe, G.A. Hufford, and M.G. Cotton, Propagation modelling of moist air and suspended water/ice particles at frequencies below 1000 GHz, AGARD 52nd Specialists' Meeting of the Electromagnetic Wave Propagation Panel, Palma De Mallorca, Spain, 1993.

[7] Liebe H.J., MPM - An atmospheric millimeter-wave propagation model, *Int. J. Infrared and Millimeter Waves*, 10, 631-650, 1989.

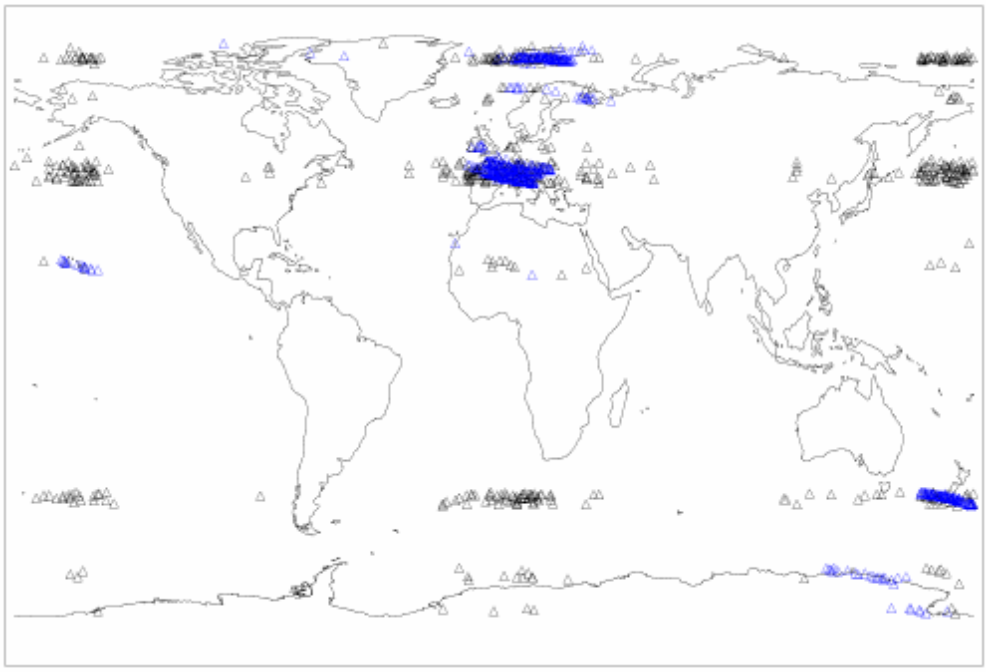

Figure 7.1. Location of temperature profiles provided by the UK Met Office matched with GOME radiance measurements. Black triangles: training dataset. Blue triangles: validation dataset.

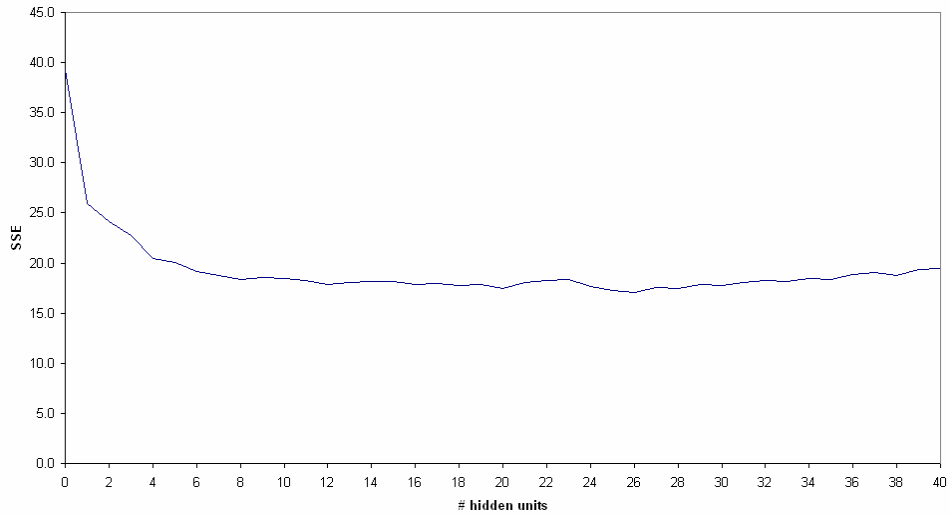

Figure 7.2. Retrieval error committed on the test set by the trained neural network, as a function of the number of units in the hidden layer.

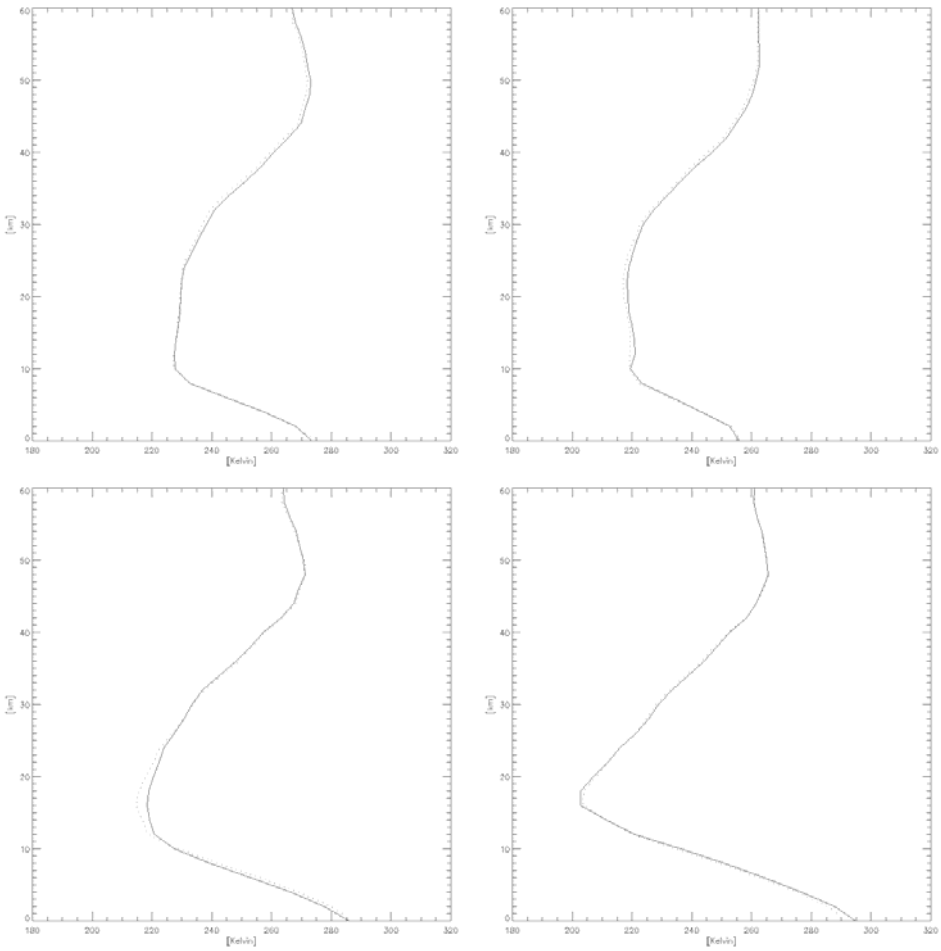

Figure 7.3. Examples of retrieved temperature profiles. The solid line represents the temperature profile estimated by the neural algorithm, while the dotted line indicates the corresponding true profile provided by the UK Met Office.

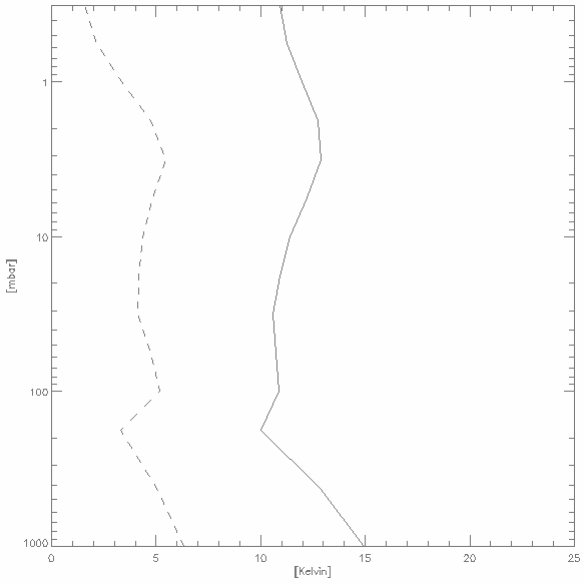

Figure 7.4. Root mean square error (RMSE) of retrieved temperature profiles (dashed line) compared with the standard deviation of the profiles from their means (solid line).

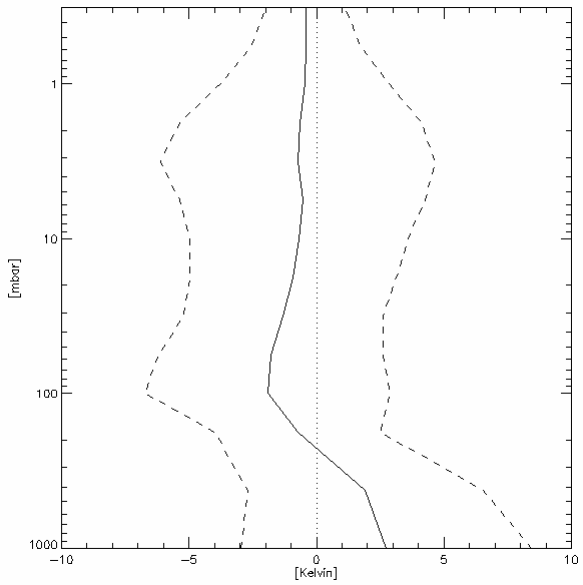

Figure 7.5. Profile of biases between the retrieved temperature profiles and the true profiles belonging to the validation dataset. The solid line represents the mean value of biases, while the dashed lines indicate the standard deviation.

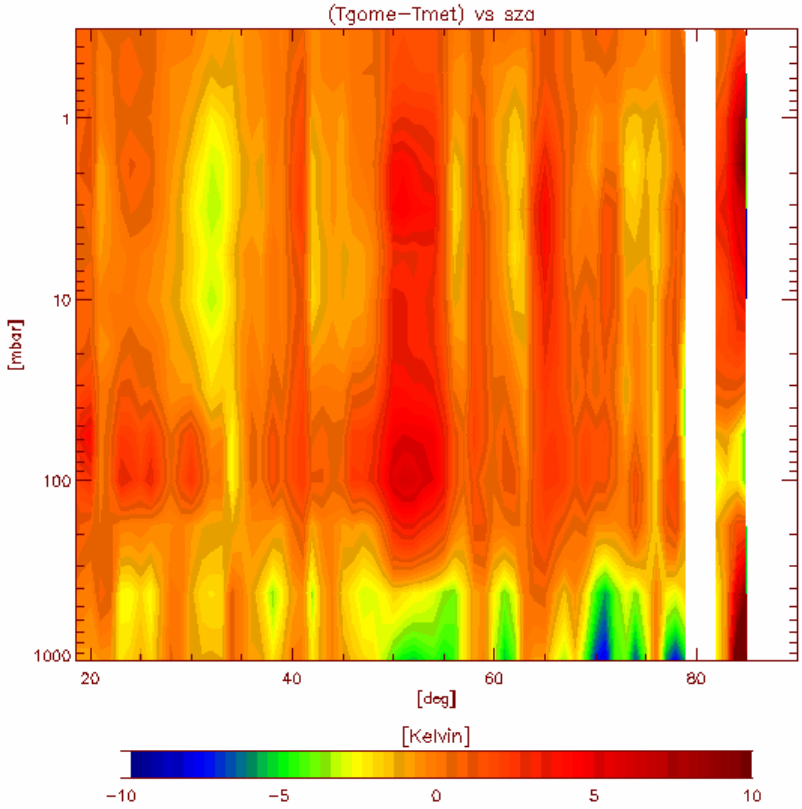

Figure 7.6. Contour plot of differences between retrieved and reference temperature profiles. On the abscissa we report the solar zenith angles, while on the ordinate the pressure levels.

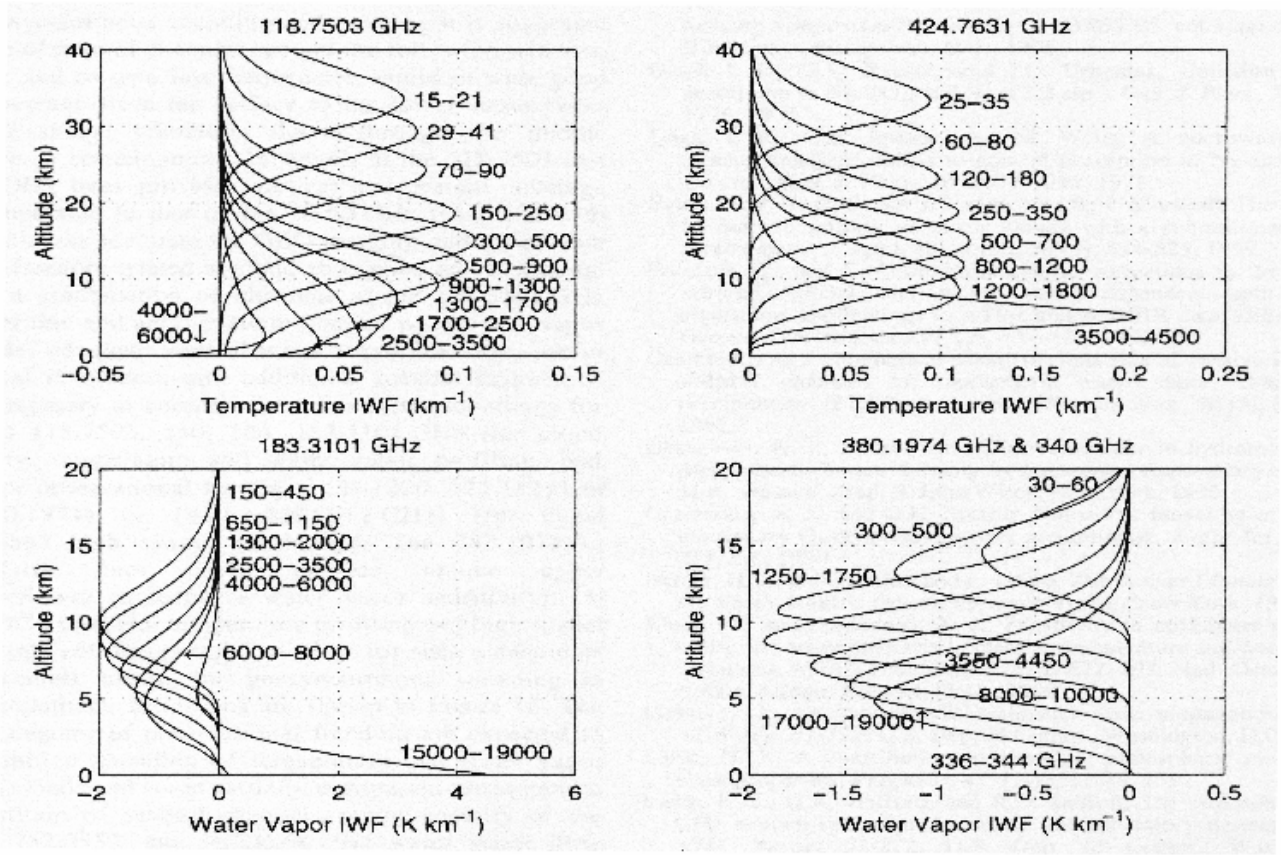

Figure 7.7. Incremental Weighting Functions (IWFs) for the 4 bands (with exception of 55 GHz band) selected within the GOMAS project for temperature and humidity profiling.

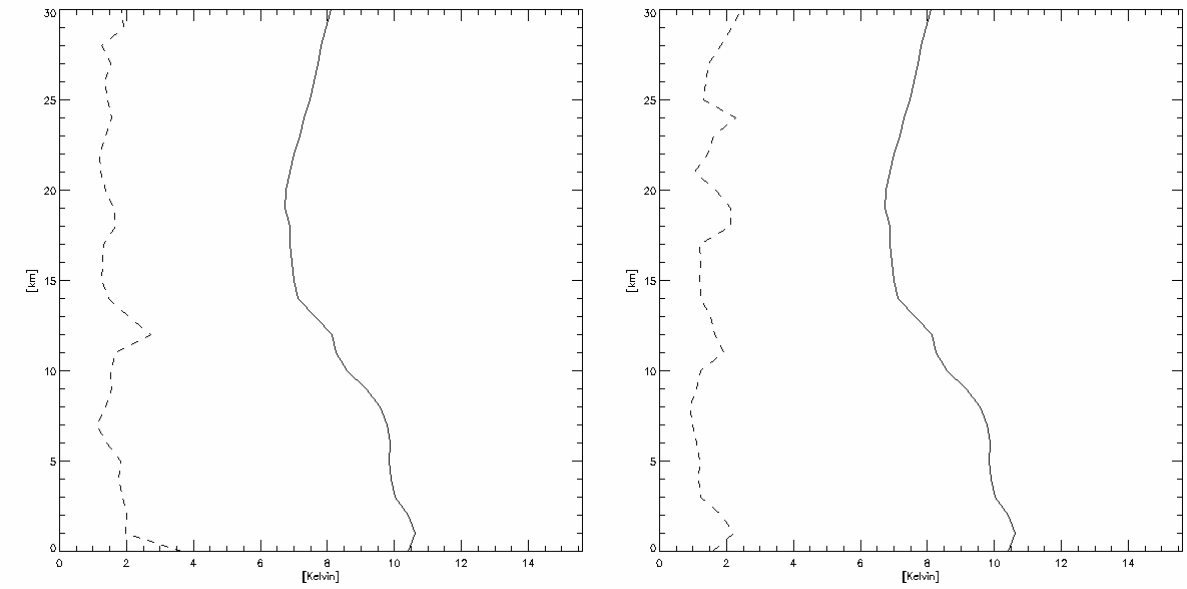

Figure 7.8. Root mean square error (RMSE) obtained for temperature profiles, for a surface emissivity of 0.5 (plot on the left) and 0.9 (plot on the right). The dashed line represents the RMSE, while the solid line the standard deviation of profiles from their mean.

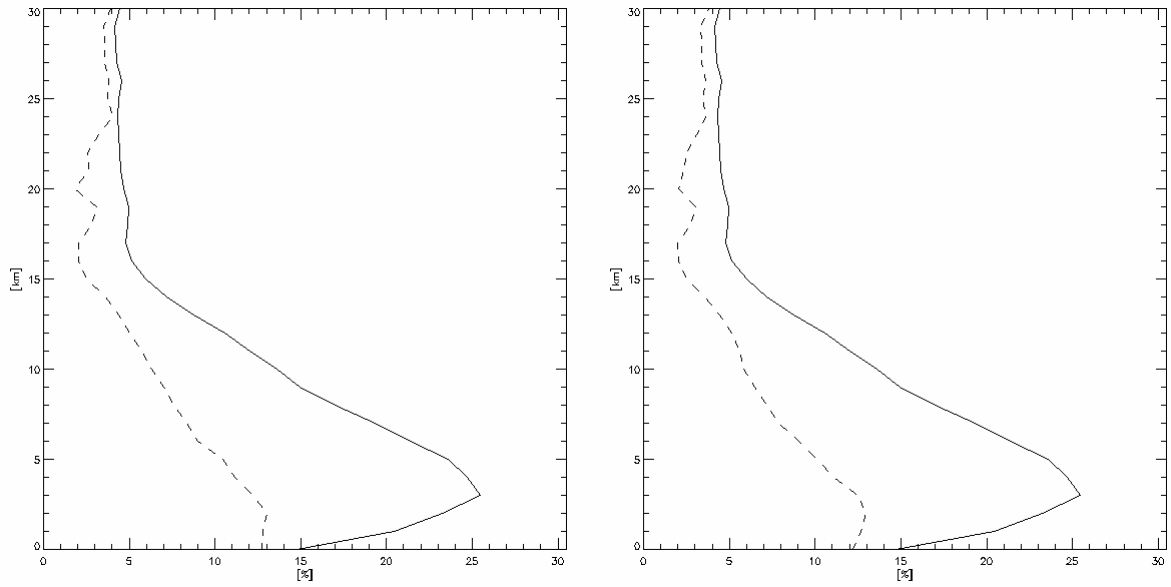

Figure 7.9. Same as Figure 7.8, but for humidity profiles.

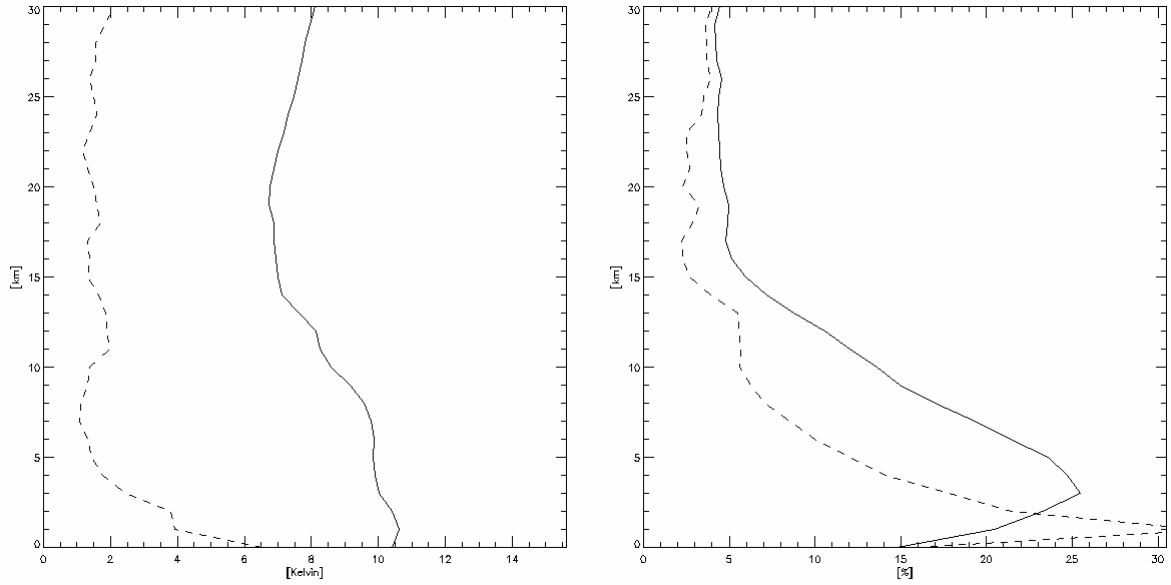

Figure 7.10. RMSE for the unfavorable case of assuming an incorrect value for the surface emissivity, for temperature (plot on the left) and humidity (plot on right) retrievals. NNs have been trained with an emissivity of 0.5, while the retrieval has been performed for an emissivity of 0.9.

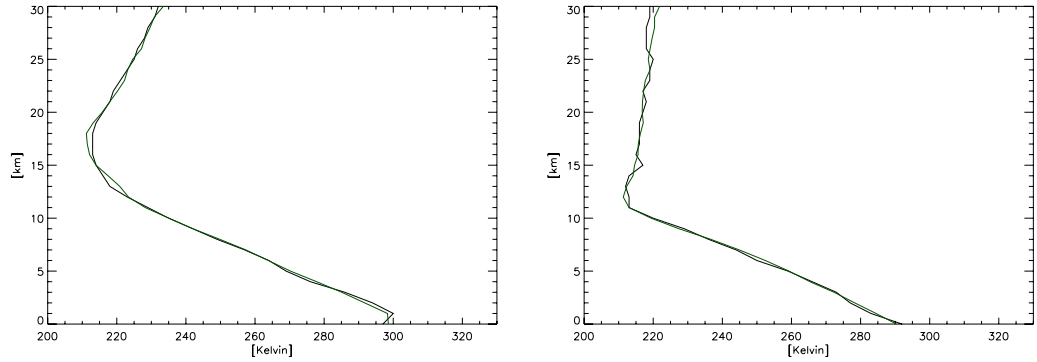

Figure 7.11. Examples of retrieved temperature profiles (green line), compared with corresponding reference profiles.

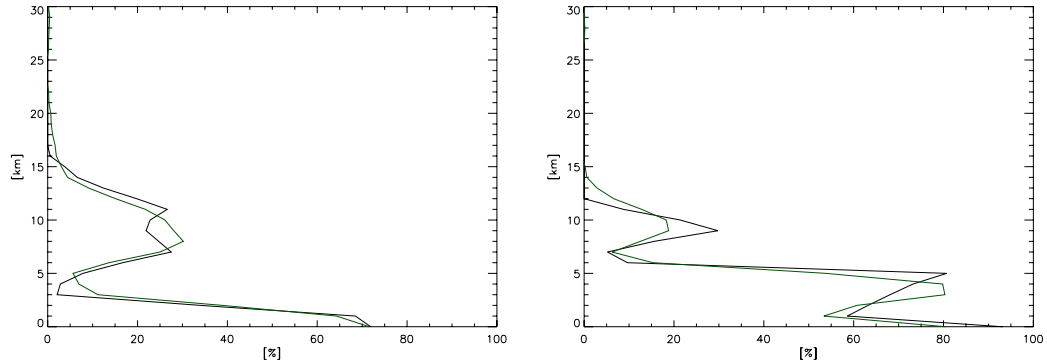

Figure 7.12. Examples of retrieved humidity profiles (green line), compared with corresponding reference profiles.

atmospheric parameter	required frequency (GHz)	required resolution (km)	antenna diameter (m)
temperature profile	55	30	7.8
	119	30	3.6
humidity profile	183	20	3.5
precipitation	90	10	14

Table 7.1. Relation between the required frequency for sounding, the required spatial resolution and the antenna diameter, for different atmospheric parameters.

54 GHz		**118 GHz**		**183 GHz**		**380 GHz**		**425 GHz**	
ν (GHz)	**Δν (MHz)**	**ν (GHz)**	**Δν (MHz)**	**ν (GHz)**	**Δν (MHz)**	**ν (GHz)**	**Δν (MHz)**	**ν (GHz)**	**Δν (MHz)**
56.325	50	118.7503 ± 0.018	6	183.3101 ± 0.300	300	380.1974 ± 0.045	30	424.7631 ± 0.030	10
56.215	50	118.7503 ± 0.035	12	183.3101 ± 0.900	500	380.1974 ± 0.400	200	424.7631 ± 0.070	20
56.025	250	118.7503 ± 0.080	20	183.3101 ± 1.650	700	380.1974 ± 1.500	500	424.7631 ± 0.150	60
55.520	180	118.7503 ± 0.200	100	183.3101 ± 3.000	1000	380.1974 ± 4.000	900	424.7631 ± 0.300	100
54.950	300	118.7503 ± 0.400	200	183.3101 ± 5.000	2000	380.1974 ± 9.000	2000	424.7631 ± 0.600	200
54.400	220	118.7503 ± 0.700	400	183.3101 ± 7.000	2000	380.1974 ± 18.000	2000	424.7631 ± 1.000	400
53.845	190	118.7503 ± 1.100	400	183.3101 ± 17.000	4000	340.0	8000	424.7631 ± 1.500	600
53.290	360	118.7503 ± 1.500	400			optional/auxiliary		424.7631 ± 4.000	1000
52.825	300	118.7503 ± 2.100	800						
51.760	400	118.7503 ± 3.000	1000						
50.300	180	118.7503 ± 5.000	2000						

Table 7.2. Radiometric channels selected within the GOMAS project.

BANDS	# input units before pruning	# input units after pruning
118 GHz	11	10
183 GHz	7	1
380 GHz	7	2
425 GHz	8	8

Table 7.3. Results of the Extended Pruning procedure for temperature.

BANDS	# input units before pruning	# input units after pruning
118 GHz	11	1
183 GHz	7	6
380 GHz	7	6
425 GHz	8	2

Table 7.4. Results of the Extended Pruning procedure for humidity.

Chapter 8

Conclusions and outlook

The study described in this thesis mainly concerns the application of a neural network inversion technique for the retrieval of ozone profiles from a satellite-based nadir-viewing instrument. The effort is justified by the need of very fast retrieval algorithms for the processing of huge amount of data coming from satellite observations. Traditional approaches for solving non-linear inversion problems (as it is the retrieval of the vertical distribution of an atmospheric component from electromagnetic measurements) are based on numerical solutions obtained by means of iterative methods, as explicit expressions can not be derived. This implies large computational burdens and long times for processing. Even if computing technology will provide faster and faster processors, the amount of available satellite data to be processed is also increasing, and with an increasing rate. Moreover, large re-processing activities of archived datasets are often envisaged, in order to improve the information extracted from measurements. It therefore comes out that faster inversion algorithms are always welcome, and neural networks can represent a powerful tool for achieving the goal.

In this work, a neural network algorithm aiming at the retrieval of ozone profiles from radiance measurements of the GOME instrument has been designed. Besides the description of general concepts concerning inversion problems, and the application of neural networks to obtain fast and accurate solutions, the study mainly focuses on the selection of input measurements, i.e. which are the measurements most suitable for addressing the solution. A sensitivity analysis has been carried out, in order to emphasize the relation between atmospheric states and radiances measured by the sensor. Different techniques aiming at the selection of inputs have been explored and discussed. Among them, a new technique, called Extended Pruning (EP), which exploits neural networks to this purpose (and not only for the inversion phase), has been proposed. Obtained results are very promising, and can be straightforward extended to other applications.

Stemming from this preliminary analysis, two neural algorithms (based on two different selection methods) have been designed, and their retrieval performances, as well as their error characterization, have been discussed.

Ozone profiles retrieved by means of the two neural algorithms have been then validated by means of inter-comparisons with ozone measurements performed by other instruments. In a first case, NN retrieved profiles have been compared with the corresponding profiles obtained by another satellite instrument, ILAS. In a second case, a more extended validation has been performed by using ground-based lidar measurements, available from a series of stations belonging to the Network for Detection of Stratospheric Changes (NDSC). Both validation exercises demonstrate the capability of NNs of providing reliable estimations of the ozone vertical distribution, comparable with other techniques in terms of accuracy and sensitivity to observations errors, but with the advantage of doing much faster inversions.

In particular, the algorithm based on the Extended Pruning, has been proven to be more robust with respect the other one, either being less sensitive to instrumental errors, or providing more reliable estimations in correspondance of critical conditions for observations.

Finally, the application of the neural network retrieval technique for the estimation of other atmospheric parameters has been carried out. Two different cases have been considered: the first one concerns the retrieval of temperature profiles from the same data (i.e. radiances measured by GOME) already used for the estimation of the ozone vertical distribution, in order to better characterize the temperture dependence of the ozone cross-section and therefore improve the

ozone retrieval algorithm; the second one refers to a preliminary study conducted in the framework of an innovative project (GOMAS), aiming at assessing the feasibility of radiometric atmospheric soundings (temperature and humidity profiles) from a geostationary orbit.
Results obtained in both cases demonstrate the big versatility of NNs in addressing inversion problems, and confirm their capability to catch the functional relation linking electromagnetic measurements to measured physical parameters. NNs can be therefore considered a very competitive approach for solving this kind of problems.
The research activity presented in this work can be directly extended to other instruments (and in particular to sensors similar to GOME, expressively designed for the ozone monitoring), and the use of NNs can be applied for the inversion of their measurements. In March 2002, the European environmental satellite Envisat was launched with on board the SCIAMACHY instrument, followed in July 2004 by the NASA platform EOS-AURA housing the Dutch-Finnish Ozone Monitoring Instrument (OMI). In addition, for the near future series of GOME-2 (an improved version of GOME) and new SBUV-2 instruments are planned to be launched in orbit. All these instruments are designed to measure the UV solar back-scattered radiation in a nadir-viewing geometry. NNs aiming at the retrieval of ozone profiles and other trace gases can represent an effective approach for processing the huge amount of data that will become available from these instruments. In particular it should be noted that OMI will provide radiance measurements with an improved spatial resolution with respect to previous instruments (13 x 24 km^2 and 13 x 48 km^2 for the global mode, and 13 x 12 km^2 for the zoom-in mode), offering the possibility to monitor tropospheric ozone variability, in particular over urban areas. Strong efforts are needed to design an algorithm for a direct retrieval of tropospheric ozone from nadir-viewing radiance measurements, and NNs can be a powerful tool to address the solution.

List of Figures

List of Tables