J1.1

OZONE PROFILE RETRIEVAL FROM GOME DATA USING A NEURAL NETWORK INVERSE MODEL

Martin D. Müller*, Anton K. Kaifel
Center for Solar Energy and Hydrogen Research (ZSW), Stuttgart, Germany

ABSTRACT

In order to retrieve the atmospheric ozone distribution from the UV-visible satellite spectrometer GOME (Global Ozone Monitoring Experiment), we have modelled inverse radiative transfer directly, using a multilayer perceptron (MLP) type neural network. This network was trained on a data set of measured GOME radiances as input, and collocated ozone profile measurements from ozonesondes, SAGE II, HALOE and POAM III as target values. A partial training method has been devised for dealing with incomplete target data, because neither occultation instruments nor ozone sondes cover the full retrieval height range (1–60 km). Around 70000 collocations from 1996 to 2001 were used for training, and another 12000 for cross-validation in a test data set. Network input consists of a combination of spectral, geolocation and climatological information (latitude and time), the latter making the use of external *a priori* ozone profiles unnecessary. We designate our method as *Neural Network Ozone Retrieval System* (NNORSY). In the stratosphere, NNORSY globally reduces standard deviation (StD) with respect to a well-tried ozone climatology by around 40%. Tropospheric ozone can also be retrieved in many cases, reducing the StD by 10-30% globally. The neural network was found capable of correcting for instrument degradation, pixel cloudiness and scan angle effects on its own, based on the input data provided. Remaining inhomogeneities in the geographical distribution of training data, combined with differing ozone field variability causes retrieval quality to vary with latitude and season. Since retrieval only requires one forward propagation through the network, NNORSY is about 10^3–10^5 times faster than classical, local retrieval techniques like, for instance, Optimal Estimation. Therefore the method is well suited for real-time application and fast data reprocessing. An operational near-real-time prototype of the system is already running successfully at three GOME receiving stations. In order to better characterize single output ozone profiles, a number of local error estimation methods has been investigated. Vertical resolution of the profiles was assessed empirically in comparisons with the high-resolution collocations, and seems to be in the order of 4–6 km. Further developments of NNORSY could involve improvement of training data composition, input parameter optimization, more sophisticated network error functions and training methods, as well as adaptation to other sensors.

1. INTRODUCTION

Although it contributes only about one millionth to the total mass of the atmosphere, the trace gas ozone is one of its most active and important constituents. Not only does it prevent harmful UV radiation from reaching the ground, but it controls the stratospheric temperature distribution, and shares responsibility for summer smog, the oxidation capacity of the atmosphere and global warming. Determining the three-dimensional, global distribution of ozone from the orbit has been a challenge to which a number of different satellite instruments have been assigned in the past. One of these is the *Global Ozone Monitoring Experiment* (GOME) (Burrows *et al.*, 1999) upon the European polar orbiting satellite ERS-2. It measures the backscattered solar radiation from 240 nm to 790 nm at a moderate spectral resolution (0.2–0.4 nm). GOME is a scaled-down version of the SCIAMACHY instrument (*SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY*) (Burrows *et al.*, 1995; Bovesmann *et al.*, 1999), which was recently put into operation on the European ENVISAT satellite.

To exploit the theoretical ozone information content of the backscattered UV-vis radiation, a nonlinear, ill-posed inverse retrieval problem has to be solved for the sun-normalized GOME radiances. To achieve this, various independent retrieval algorithms have been developed, most of which based on the principle of Optimal Estimation (OE) as described by Rodgers (1976, 1990). Simply put, this method tries to find an optimal compromise between information gained from a spectral measurement and an *a priori* ozone profile. The compromise is based on the reliability of either source, i. e. their error characteristics.

Athough this retrieval scheme can be considered

*Corresponding author address: Martin D. Müller, ZSW, Hessbrühlstr. 21c, 70565 Stuttgart, Germany; e-mail: martin.mueller@zsw-bw.de

geophysical, since it relies on the relatively well understood forward calculation of the radiative transfer equation (RTE) used for extracting spectral information, it also possesses a significant statistical component in the form of the optimization framework. In addition, solving the RTE requires certain assumptions about unmeasurable or unavailable parameters of the state of the atmosphere and the instrument. Also, the *a priori* ozone profile together with its variance is usually taken from a climatology, which also consists of a statistical and empirical analysis of a number of measurements.

This paper proposes to let a neural network scheme take care of the heuristics and statistics of the retrieval process, after being provided with a suitable structure and training data set derived from geophysical knowledge and a few assumptions. Such has been done successfully in other fields of atmospheric science (Krasnopolsky, 1997; Chevallier *et al.*, 1998; Jiménez, 2000; Aires *et al.*, 2001) with the network tool of choice being a nonlinear feed-forward neural network, also called a multilayer perceptron (MLP) (Rumelhart *et al.*, 1986). Based upon experience in MLP ozone retrieval from NOAA-TOVS data (Müller and Kaifel, 1999; Kaifel and Müller, 2001), the method has now been applied to GOME data and will be designated in what follows as *Neural Network Ozone Retrieval System* (NNORSY). The motivation behind its development is therefore, that it will complement existing retrieval algorithms by exhibiting different problems and error characteristics, and – even more importantly – that it is faster by several orders of magnitude.

Section 2 of this paper will give a brief overview of the neural networks' relationship to the well-established OE theory, and describe the specific data and training methods needed for retrieval of ozone profiles from GOME. Section 3 presents the retrieval results in the form of error statistics on the training and test data sets, as well as subsets thereof. A brief case study with validation data from Hohenpeissenberg ozone sondes is also included. Section 4 will then elaborate on the advanced topics of estimating local retrieval errors and vertical resolution, while Section 5 will conclude the paper.

2. DATA AND METHODS

2.1. *MLPs in the Retrieval Context*

The idea of retrieving height-resolved ozone information with an orbital instrument by measuring the backscattered UV (BUV) radiation dates back to the 1950's (Singer and Wentworth, 1957) and will be briefly described here. While theoretically possible for other trace gases as well (Rozanov *et al.*, 1993), ozone is particularly suited for this task, because its absorption coefficient rises steeply by about four orders of magnitude in the UV. This alone for solar radiation would already lead to decreasing penetration depth in the atmosphere, but the effect is greatly enhanced by Rayleigh-scattering and increasing air pressure. As a result, incoming radiation observes an almost transparent atmosphere down to a certain depth, at which extinction rises rapidly within a few kilometers. The height of this layer depends on wavelength; it almost acts like a fuzzy mirror, with the reflected photons carrying ozone information mostly from the particular height region. By scanning the UV, a height-resolved ozone profile can thus be constructed.

From radiative transfer theory, it can be shown (e.g. Rodgers, 1990) that the top-of-atmosphere radiance $y(\lambda)$ measured by a satellite is essentially a weighted integral over the atmospheric state $x(z)$

$$y(\lambda) = \int_{z=0}^{\infty} K(z,\lambda)\, x(z)\, dz, \tag{1}$$

where z is the height coordinate, λ the wavelength and $K(z,\lambda)$ the weighting function or kernel of the observation process, which decribes the altitude and fuzziness of the abovementioned 'mirror'. For discrete measurements, Eq. 1 is discretized in such a way that $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is a vector in n-dimensional state space, and $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$ belongs to the m-dimensional measurement space. In retrieval theory, it is furthermore assumed for physical reasons that the relationship between $\mathbf{x}$ and $\mathbf{y}$ can be modeled as

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \varepsilon, \tag{2}$$

where $\mathbf{F}$ is a (radiative transfer) forward model depending on the atmospheric state $\mathbf{x}$ and a number of additional parameters which are for simplicity combined into one vector $\mathbf{b}$. The measurement noise vector ε is assumed to have Gaussian statistics. In general, the model $\mathbf{F}$ will not be linear, but since it represents a continuous function, it can be linearized in the vicinity of a certain state $\mathbf{x}_0$, such that

$$\mathbf{y} - \mathbf{F}(\mathbf{x}_0) = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\Big|_{\mathbf{x}_0} (\mathbf{x} - \mathbf{x}_0) + O(\mathbf{x}^2) \approx \mathbf{K}(\mathbf{x} - \mathbf{x}_0), \tag{3}$$

where $\mathbf{K} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}}$ is called the kernel matrix. However, the underlying Eq. 1 belongs to the class of Fredholm integrals of the first kind, a class containing many ill-posed problems, of which the inverse retrieval problem is no exception. For a comprehensive discussion on this topic, the reader is referred to (Rodgers, 2000). It shall just be stated here that the rank of $\mathbf{K}$ is usually smaller than the state space dimension n, therefore the problem is underconstrained, meaning there exist a multitude of possible solutions $\mathbf{x}$ for any given observation $\mathbf{y}$. It is therefore reasonable to look at the problem from a Bayesian point of view: The most probable solution $\hat{\mathbf{x}}$ is the one that maximizes the likelihood $\mathcal{L} = p(\mathbf{x}|\mathbf{y})$, with $p(\mathbf{x}|\mathbf{y})$ denoting the conditional probability density of $\mathbf{x}$ given the observation $\mathbf{y}$. This is equivalent to minimizing an error function

$$\begin{aligned} E &= -\ln \mathcal{L} = -\ln p(\mathbf{x}|\mathbf{y}) = -\ln \left(\frac{p(\mathbf{y}|\mathbf{x}) p(\mathbf{x})}{p(\mathbf{y})} \right) \\ &= (\mathbf{y} - \mathbf{K}\mathbf{x})^T \mathbf{S}_e^{-1} (\mathbf{y} - \mathbf{K}\mathbf{x}) + \\ &\quad + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) + \text{const.}, \end{aligned} \tag{4}$$

where Bayes' Theorem was used in the first line and the assumption of Gaussian errors in the second. $p(\mathbf{y})$ is assumed to be constant. The covariance matrix $\mathbf{S}_e$ contains forward model and measurement errors, while $\mathbf{S}_a$ tries to model the natural variability of the state, centered around some *a priori* $\mathbf{x}_a$, which in practice mostly represents a climatological mean, i. e. $\mathbf{x}_a = \mathbf{x}_a(\theta, d)$ with θ denoting geographical latitude and d the time of year. Eq. 4 constitutes the core of the OE framework (Rodgers, 1976, 2000). As can be seen, it applies for a single observation $\mathbf{y}$, and requires the knowledge of a forward model, its error covariance, the measurement error covariance, an *a priori* state and an *a priori* covariance.

Since we assumed Gaussian distribution for all errors involved, it follows that $p(\mathbf{x}|\mathbf{y})$ is also a Gaussian. This function can be sampled using a set of observations $\{\mathbf{y}^p\}, \quad p \in \{1, \ldots, P\}$, for which the corresponding states $\mathbf{x}^p$ are known. We further assume the existence of an inverse function $\mathbf{R}$ modelled by

$$\mathbf{x} = \mathbf{R}(\mathbf{y}, \mathbf{c}, \mathbf{w}) + \varepsilon_R, \tag{5}$$

where $\mathbf{c}$ is a vector of additional input parameters, and $\mathbf{w}$ contains the inverse model parameters. Omitting the constant terms, the error function can then be written as

$$\begin{aligned} E &= -\ln \prod_{p=1}^{P} p(\mathbf{x}^p|\mathbf{y}^p) = \\ &= \sum_{p=1}^{P} (\mathbf{R}(\mathbf{y}^p, \mathbf{c}^p, \mathbf{w}) - \mathbf{x}^p)^T \hat{\mathbf{S}}^{-1} (\mathbf{R}(\mathbf{y}^p, \mathbf{c}^p, \mathbf{w}) - \mathbf{x}^p)^2 \end{aligned} \tag{6}$$

In our case, $\mathbf{R}$ is realized by means of a simulated MLP-type neural network. This network consists of a one-dimensional input layer containing enough neurons to receive the input $\mathbf{y}$ and $\mathbf{c}$. Each input neuron is connected to all nodes of a second, hidden layer of neurons via synapses carrying weights $\mathbf{w}$. When presented with an input data vector, the input signals propagate along the synapses whilst being multiplied by the weights. The hidden neurons essentially sum up all incoming signals, and use a fixed nonlinear function – in our case $\tanh$ – to in turn define their outputs, which propagate through another layer of weighted synapses to the output neurons. These neurons again sum up their inputs, and use another $\tanh$ transfer function and a suitable renormalization to define the network output $\mathbf{x}$. Thus, the network is essentially a mapping from measurement space to state space parameterized by means of the weights $\mathbf{w}$.

In principle, more than one hidden layer of neurons could be used, but is has been proven that one layer is enough to make the network model a universal approximator, i. e. it can theoretically approximate any given mapping with arbitrary accuracy (Hornik *et al.*, 1989). However, this proof is not a constructive one, therefore the hidden layer size has to be determined empirically (Tamura and Tateishi, 1997), and the optimal set of weights $\mathbf{w}$ is searched for with a learning algorithm, the choice of which will be described in Section 2.3.. In contrast to conventional approximation schemes with fixed basis functions (e. g. polynoms, trigonometic functions, etc.), where the approximation error increases exponentially with the dimension of the mapping, in approximation tasks carried out by means of neural networks the error was shown to be independent of the dimension (Baron, 1993, 1994).

At this point, we should note the fundamental differences between OE and the neural network approach: While the first term in Eq. 4 represents a distance in measurement space, Eq. 6 corresponds entirely to state space. Also, the neural network method doesn't demand explicit knowledge of a forward model or an *a priori*, but instead it needs a training data set of paired observations and measurements $\{(\mathbf{y}^p, \mathbf{c}^p), \mathbf{x}^p\}$, from which it estimates an

optimal set of parameters $\mathbf{w}$. Once this is done, the resulting mapping $\mathbf{R}$ is optimal in a global sense, and can be applied to all observations, whereas OE determines a new, locally optimal solution for each individual observation. The consequences of these differences will be discussed further below.

In practice, Eq. 6 is often simplified for computational reasons by assuming $\hat{\mathbf{S}} = \sigma^2 \mathbf{I}_n$ for all training pairs.* It then doesn't influence the minimization of E anymore, therefore it can be safely omitted, which leads to the well-established quadratic error function of the form

$$E = \frac{1}{2} \sum_{p=1}^{P} (\mathbf{R}(\mathbf{y}^p, \mathbf{c}^p, \mathbf{w}) - \mathbf{x}^p)^2 =: \sum_{p=1}^{P} E_p, \qquad (7)$$

where the factor $\frac{1}{2}$ is just a convention, and E_p is the error attributed to a single output pattern. This rather coarse assumption is still adequate for ozone retrieval purposes, as will be demonstrated below. However, the description of retrieval error through a single standard deviation σ is unsatisfactory in most cases, thus supplementary methods have to be employed to better characterize the errors (Sec. 4).

Since the ill-posedness of the inverse retrieval problem leads to very unstable solutions with a straightforward, classical approach (Hasekamp *et al.*, 1999; Rodgers, 2000), the second term in Eq. 4 serves as a regularizer. Another way to look a this is by means of the *a priori*, the OE algorithm is drawn toward a physically reasonable solution of Eq. 1, while purely mathematical ones (e. g. involving negative ozone densities), and climatologically improbable ones are being made unattractive. On the other hand, neural networks with quadratic error functions always approximate the conditional mean of the training data with respect to the input Bishop (1995a). In the case of ambiguities, this may lead to bad results, since for instance the mean of two valid branches of a solution is not necessarily itself a solution – it may well lie on neither branch. However it can be argued that since OE yields unambiguous results, the parameters b used in the forward model, in addition to θ and d from the *a priori* serve to remove the ambiguities. Here, these parameters are combined into the vector $\mathbf{c}$, and are thus included in the neural network input to render the network mapping unambiguous.

*Suitable normalization of $\mathbf{y}$ within the preprocessing step ensures that the output variances are all in the same order of magnitude.

2.2. *Training Dataset Assembly*

The NNORSY method relies on collocations with measured ozone profiles from different sources to form the training and test databases. NNORSY was trained on collocations dated from Jan. 1996 to July 2001. The maximum distance for collocations was set to 250 km, with no more than 12 h between ozone and GOME measurements. All data thus collected were interpolated to a common height grid stating average ozone number densities at 1, 2, …, 60 km geopotential height (GPH).

Ozonesonde data were obtained from the World Ozone and Ultraviolet Radiation Data Center (WOUDC) (Wardle *et al.*, 1998) and from the Southern Hemisphere Additional Ozonesondes (SHADOZ) campaign (Thompson *et al.*, 2001). These data generally have a high quality and vertical resolution, but their geographical distribution is uneven, with most stations situated at northern midlatitudes and very few measurements over the oceans. Since sonde data starts becoming unreliable around 25–30 km (SPARC, 1998), all sonde profiles were cut off at a random height in this range. Cutting off all sondes at the same height resulted in artifacs in the retrieved profiles and was therefore avoided. To achieve coverage of the greater heights, the sonde measurements were supplemented by satellite data from solar occultation limb sounders. These were the infrared HALOE instrument (Russell *et al.*, 1993), data version 19 (Lu *et al.*, 1997), the ultraviolet SAGE II sensor (Cunnold *et al.*, 1989), v6.10 (Cunnold *et al.*, 2000), and the POAM III sensor, v3 (Lucke *et al.*, 1999; Lumpe *et al.*, 2002). HALOE and SAGE II sample the entire globe quite evenly between 70°S and 70°N, while the POAM measurement geometry results in sunrise events occurring between 54°N and 71°N, and sunset between 63°S and 88°S. Note that sometimes multiple GOME pixels were collocated with a single ozone profile. This is equivalent to training with noisy input data, and serves to further regularize the retrieval (Bishop, 1995b).

As has been found elsewhere (Lu *et al.*, 1997; Steele and Turco, 1997; Deniel *et al.*, 1997; SPARC, 1998), the aforementioned ozone data sources do not always agree. The neural network will construct a compromise between the different sources, therefore biases stemming from different measurement principles and retrieval algorithms for the limb sounders are likely to cancel out. Yet it is clear that strong biases and variances in the training data set will adversely affect NNORSY retrieval accuracy, therefore some homogenization based on the sta-

tistical properties of the training data set was performed. Since both the distribution of collocations and of ozone in general are largely governed by latitude, the statistics of latitude bands were used for this purpose.

First, the number of collocations over latitude and time had to be checked, for overrepresentation of a certain latitude region might lead to biases in the neural network output, by favoring a certain type of profile (Bishop, 1995a). On the other hand, it is reasonable to include more profiles from latitudes where natural ozone variability is higher – e. g. the fringe of the polar vortex and the northern midlatitudes – such that a greater number of different atmospheric situations is sampled. The availability of data over time should be mostly level, otherwise ozone trends and sensor degradation cannot be accurately modelled. Removal of profiles is done by selectively reducing the allowed maximum collocation distance for ground stations, times, and/or latitude bands. The 2-dimensional homogenization over latitude and time sometimes requires suboptimal compromises to be made. Fig. 1 shows histograms of observation density for NNORSY training data before and after the operation.

It is clearly seen in Fig. 1, that the highest density of collocations is located around 50°– 60°N. POAM data in the NH have therefore been used only poleward of 65°N, to extend data coverage towards the pole, because SAGE and HALOE measurements become sparse in this region. Being of very good quality (Lucke *et al.*, 1999), the withheld POAM data from 54°– 65°N can later be used for validation purposes.

In a second step, the remaining collocations are checked for consistency and submitted to a heuristic screening procedure to remove outliers. Special care has to be taken regarding occultation sounder data below the lowermost stratosphere, because tropospheric aerosols and clouds tend to interfere with their retrieval algorithms and are not always correctly accounted for in the error specifications accompanying the ozone products (Bhatt *et al.*, 1999; Steele and Turco, 1997; SPARC, 1998).

The GOME data used for profile retrieval were processed with the standard GOME Data Processor (GDP), Version 2.1 (Slijkhuis and Loyola, 1999; Bargen *et al.*, 1999), including all standard corrections (straylight, polarization, etc.) except from degradation, which the program applies equally to solar and earth spectra. Thus it cancels out when calculating sun-normalized radiances. Before normalization, all spectra have to be interpolated to a common wavelength grid, since the neural network does not get any wavelength information and therefore assumes each spectral input neuron has a fixed location in wavelength space.

The spectral channels of the GOME instrument and their characteristics are given in Table 1. GOME measures three forward pixels across its track, followed by a single backscan pixel, which is not used here. Apart from that, all valid GOME pixels regardless of cloudiness and ground condition were retained. The GOME forward pixel size of 160 x 40 km^2 being rather large already, retrievals were performed using the full horizontal resolution. This required assigning the same GOME channel 1A spectra to clusters of six forward pixels, since the channel is integrated for a longer timespan to increase the S/N ratio. The channel boundary switch in mid-1998 was found to have no impact on the retrieval with NNORSY, any effects of this are obviously compensated for by the network.

2.3. *Network Training Considerations*

The neural network used for the retrievals has 122 input, 45 hidden and 60 output neurons. Table 2 shows the configuration of the neural network input layer. GOME sun-normalized radiance values are mainly taken from the ozone Hartley- and Hugginsbands, where the absorption coefficient's temperature dependence provides height-resolved ozone information (Chance *et al.*, 1997). The spectral resolution has been decreased to 0.35–1.5 nm by coadding of 4–12 wavelengths, in order to reduce random noise and the number of free parameters (weights) in the network. Since it was seen elsewhere that the ozone Chappuis band can under certain conditions improve retrieval results in the troposphere (de Beek, 1998), some radiances from that band were also included. Additional spectral values provide information on cloud cover and ground albedo. It should be noted that due to the flexibility of neural networks, the method is not very sensitive towards the selection and resolution of wavelength windows, as long as they contain enough physical information altogether. This was found in a number of sensitivity studies (Müller, 2002) which will not be discussed here. Suffice to say that omitting all spectral information, i. e. training climatological networks, increases retrieval errors by over 40% in the stratosphere and above, and by 5–40% in the troposphere (Fig. 2). Therefore, it is clear that the neural network model can indeed be used to extract considerable height-resolved information from the GOME spectra provided.

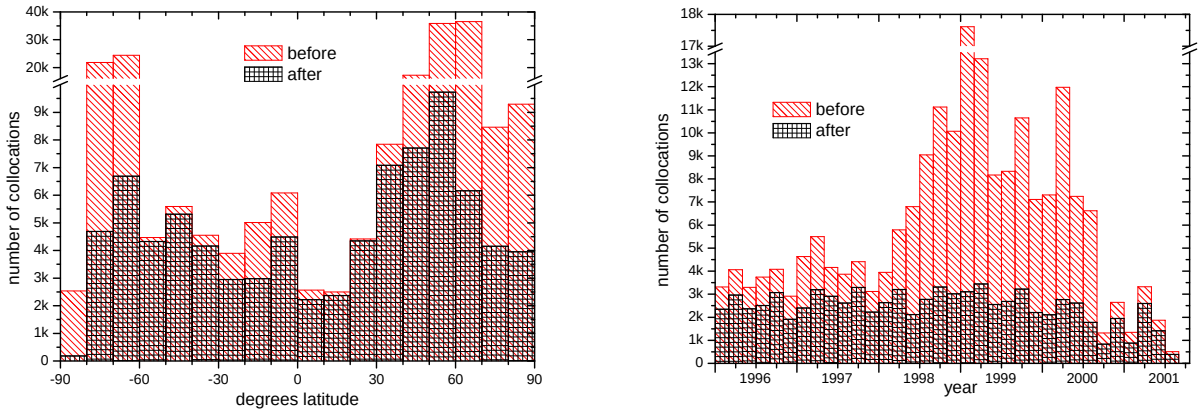

Figure 1: *Training data distribution by latitude and time, before and after the data set homogenization procedure was applied.*

Channel	Wavelenghts λ [nm]	# of photodiodes	δλ [nm]	t_I [s]	SNR
1A	238–283 (307)	400 (625)	0.2	12	10–434
1B	(307) 283–314	295 (70)	0.2	1.5	434–841
2	311–404	841	0.2	1.5	849–5100
3	394–611	1024	0.4	1.5	3500–4200
4	578–790	1024	0.4	1.5	3239–4214

Table 1: *GOME spectral channels and characteristics.* $\delta\lambda$ *is the spectral resolution,* t_I *the integration time, SNR the signal-to-noise ratio, from (Eichmann, 2001). Numbers in brackets were valid before a channel shift on 7th June 1998.*

As mentioned in Section 2.1., apart from the spectral measurements a number of geophysical parameters are provided to the network. These include solar and satellite zenith angles and the scan angle, as well as three separate flags to account for the pixel type (east, nadir, west). Fig. 2 infers that latitude and season allow the implicit construction of some sort of climatology. The continuous in-orbit time provides a means to correct for time-dependent sensor degradation effects to a certain extend. The UKMO temperature profile was provided because of its strong correlation to atmospheric ozone.

Training algorithm selection for MLPs has become a fairly wide field, although many new algorithms are still based on the principle of backpropagation as described by Rumelhart *et al.* (1986). The relative inefficiency of the original algorithm makes it unsuitable for dealing with the vast amount of data present in satellite meteorology, therefore several alternative learning algorithms have been surveyed in the frame of this and previous projects (Kaifel and Müller, 2001). For the given case, the Resilient Propagation algorithm (RPROP) developed by Riedmiller and Braun (1993) has been found advantageous. In RPROP, the magnitude of weight changes does not depend on the derivative of the error function E, as in backpropagation. Instead, each weight w_{ij} is associated with its own weight step Δ_{ij}. These weight steps are modified according to a one-step history of sign changes in the error function derivative. Roughly speaking, the RPROP algorithm accelerates weight adjustment as long as the network is moving towards a minimum, but stops abruptly when stepping over a minimum, and subsequently tries to approach it by interval division. This behaviour was observed to lead to fast and very stable error decreases during training on the data sets provided here (Müller *et al.*, 2001, 2002).

As we mentioned before, a specific problem of training with measured ozone profiles instead of simulations is the incompleteness of the target data. Ozonesondes were used only in a height

Input parameter	# of neurons	Purpose
270–325 nm	74	O_3 Hartley/Huggins band
380–385 nm	13	atmospheric window
598–603 nm	6	O_3 Chappuis band
758–772 nm	9	O_2 band: cloud detection
Satellite & solar zenith angles	4	slant column correction
line-of-sight flags	3	slant column correction
Latitude & season	2	ozone climatology background
In-orbit time	1	instrument degradation correct.
UKMO T-profile	10	atmospheric state info

Table 2: *Neural network input parameters for ozone profile retrievals. The wavelength ranges refer to sun-normalized and logarithmized radiances measured by the GOME instrument.*

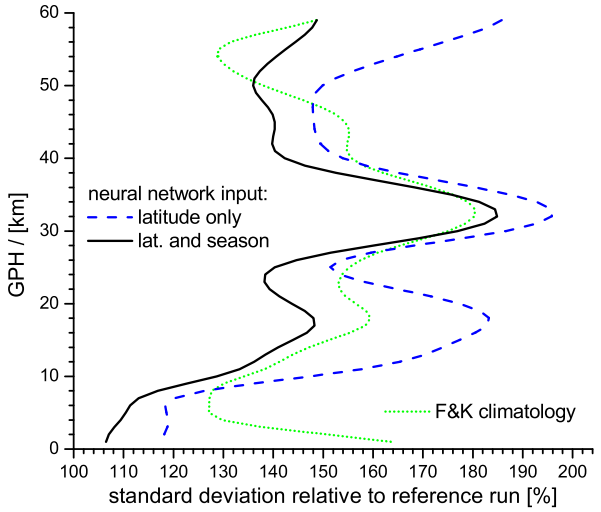

Figure 2: *Comparison of observed test data set errors for climatological networks, i. e. ones without GOME spectral input data. The reference run used the configuration from Tab. 2, but without T-profile, to clarify the impact of spectral information. F&K refers to using the ozone climatology from (Fortuin and Kelder, 1998) instead of NNORSY retrievals.*

range reaching up to 30 km, while data from limb-sounding satellite instruments become unreliable below 10–20 km (SPARC, 1998). These different measurements do not normally coincide in space and time, and hence cannot just be joined together. Therefore, for each target profile $\mathbf{t}^p = (t_1^p, t_2^p, \ldots, t_n^p)^T$, data were either flagged as missing ($m_k^p = 0$) or as available ($m_k^p = 1$). With $o^p = \mathbf{R}(\mathbf{y}^p, \mathbf{c}^p, \mathbf{w})$ being the network output, the pattern output error calculation (cf. Eq. 7) was modified according to

$$E_p = \frac{1}{2}\left(\sum_{k=1}^{n} m_k^p\right)^{-1} \sum_{k=1}^{n} (t_k^p - o_k^p)^2 \cdot m_k^p. \qquad (8)$$

This way, only available data contribute to the weight adjustment. Note that this modification does not noticably change the well-behavedness of the RPROP algorithm, as has been shown in (Müller *et al.*, 2002).

3. RETRIEVAL ERROR STATISTICS

3.1. *Global Error Statistics*

The standard method for evaluating the quality of neural network output is to assess the error statistics on a test collocation data set independent of the training data. This is already being done during training, which is considered completed once the test data set error decreases slower than a given threshold rate. In this sense, the test data set is not completely independent of the training results in that it determines the best configuration to use and when to stop training (Plutovski, 1996). However, it turns out that for the kind of problem faced here, statistics do not chance significantly when using a full-sized third, truly independent evaluation data set for validation purposes only. Since good collocations are a valuable asset, the validation data was composed solely of collocations with ozonsondes from Hohenpeissenberg (47.8°N, 11.0°E) and Syowa (69.0°S, 39.6°E). However, care has been taken to prevent any ozone profiles from appearing in both training and test data, which can happen with multiple collocations and might disrupt statistical independence. The test data set constructed consists of 12281 collocations with a distribution similar to the training data (70048 colloc.).

Figure 3 shows the global error statistics for both training and test data. With $\mathbf{d}^p = \mathbf{o}^p - \mathbf{t}^p$, and $\bar{\mathbf{d}}$ being the average of $\mathbf{d}$ of the entire data set, note that relative standard deviation is defined as

$$s = \frac{\sum_p^P \sqrt{\frac{1}{P-1}(d_k^p - k)^2}}{\frac{1}{P}\sum_p^P t_k^p}, \qquad (9)$$

not as $\sqrt{\frac{1}{T-1}\sum(d_p - \bar{d})^2/t_p}$. Since we do not use averaging kernels or other smoothing methods, sonde profiles used in the collocations were only integrated to 1 km layers. Retrieval vertical resolution is probably considerably lower (cf. Sec. 4.2.). This would give rise to extreme relative errors due to unresolved fine structure whenever the latter calculation method is applied.

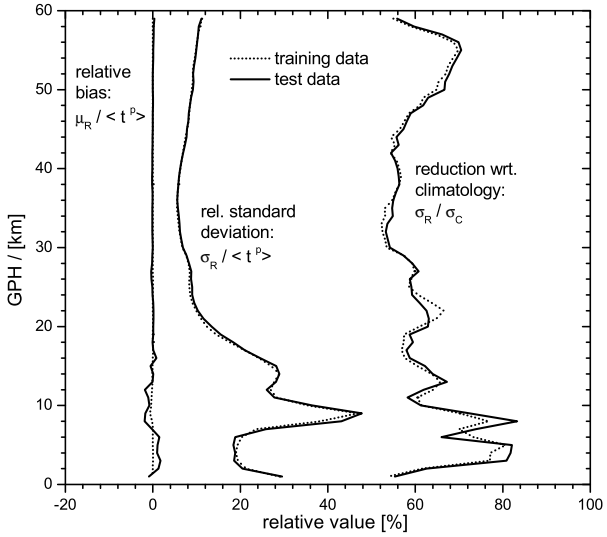

Figure 3: *Global relative statistics of NNORSY ozone profile retrievals compared to training and test data set collocations.* σ *denotes standard deviation, whereby* σ_C *is calculated by using profiles from the Fortuin and Kelder (1998) climatology instead of retrievals.*

As can be seen in Fig. 3, there is almost nonexistent bias between the collocations and NNORSY output. This is a property of the quadratic neural network error function (Bishop, 1995a), and shows the importance of constructing a representative training data set. Overall, NNORSY-GOME reduces the standard deviation of the monthly mean Fortuin and Kelder climatology (Fortuin and Kelder, 1998, referred to as F&K in the following) with respect to the collocated sonde and limb-sounder data by around 20–45%, depending on height.

Test data set errors are only marginally higher than the ones from the training data, which is a sign of the good generalization ability of the network. When further training is carried out, we observe slight improvements above the ozone peak, but the network starts overfitting the sonde data, which leads to higher test data errors in the troposphere. We are still investigating the optimal stopping point for training, and whether the results can be further improved by trimming the number of values per height level and/or by combining the output of several different or differently trained networks.

Some other features which can be recognized in the figure are the sharp peak in the relative error around 9 km GPH. Since the retrievals still reduce standard deviation wrt. the climatology, it can be concluded that this peak is mostly due to the high temporal and spatial variability of ozone in this layer, which also cannot be captured in monthly means. The errors in the lower troposphere are also surprisingly low, considering the fact that gaining ozone information from GOME data in this height range is severely limited by clouds and the low S/N ratio (Hoogen *et al.*, 1999). However, the collocated tropospheric ozone profiles are only representative for the station locations, which means there is not much information on the errors e. g over the oceanic troposphere contained in the training and test data. There are hints that the system systematically underestimates tropospheric ozone on a global scale, but this is still under investigation.

3.2. *Sensitivity Analyses*

Sensitivity studies can performed by dividing the test data set into subsets according to input parameter values. Of these, the latitudinal dependency of the errors will be discussed below. No significant dependency on pixel line-of-sight (LOS) was detected, therefore we conclude that the network is well capable of correcting for scan angle effects. With time, the errors do not follow a significant trend, although in the stratosphere errors for 2001 tend to be slightly higher than in the previous years. This can be explained in part by the underrepresentation of training data from 2001, and will be further looked at in a case study (Sec. 3.3.).

Figure 4 displays the relative statistics for the test data set divided into latitude strips. The variations observed result from a combination of the latitudinal changes of ozone natural variability with the latitudinally uneven distribution of training data (especially ozonesonde data). For instance, in the southernmost strip 1245 of the 1704 collocations stem from

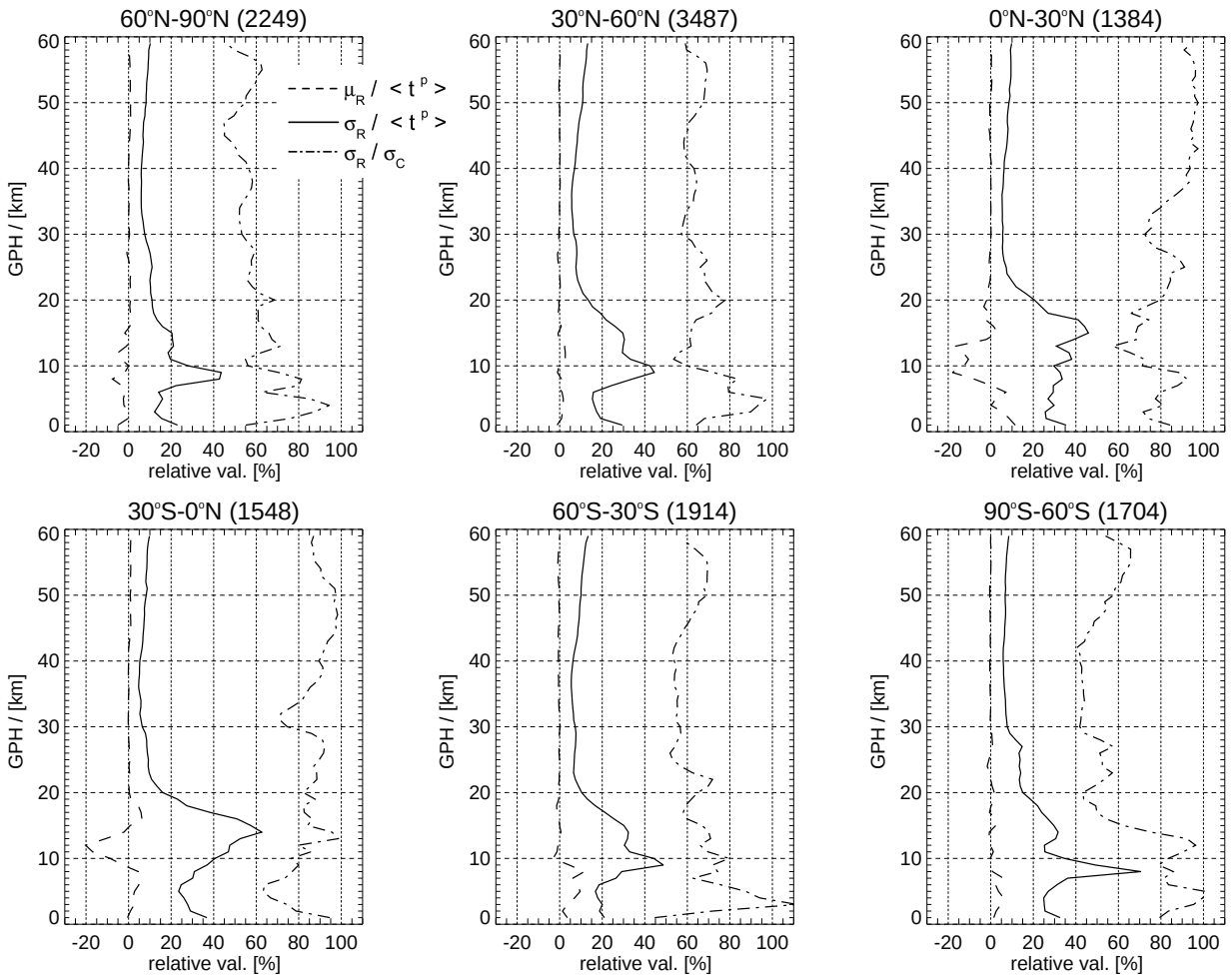

Figure 4: *Same as Fig. 3, but only test data set, divided into different latitude strips.*

the limb sounders, which obviously contribute considerable knowledge in the stratosphere, but sonde data are rare, therefore the standard deviation in the troposphere is fairly high. It is clear that the training algorithm yields lower errors in regions were data are abundant, because it receives more error feedback from these areas. The generally large solar zenith angles (SZAs) in the Antarctic region also contribute to the retrieval problems.

In the north polar region there is ample ozonesonde data (49%), but no satellite data north of 71°N. Thus, the network has enough data to learn a suitable SZA correction for the lower heights, but the relatively low errors above 20 km are somewhat misleading, since they can only be computed from satellite profiles southward of 71°N. Therefore errors closer to the pole might be somewhat higher, but cannot be estimated from the test data.

In the tropics, the climatology obviously reproduces the ozone concentration above 20 km rather well, with NNORSY yielding a less significant reduction in standard deviation. This is a result of the low ozone variability in this part of the atmosphere, where ozone distribution is mostly governed by photolytic ozone production, not by short-term transport and chemistry processes. We also note a low bias of up to 20% in the region around 12 km, the reason of which is not completely understood. A combination of different effects are suspected to contribute, like frequent cloud occurrence, fewer number of collocations, and model errors resulting from the distincively non-Gaussian distribution of target values at these heights. However, due to the low ozone molecule number density in the tropical troposphere, in absolute terms this bias amounts to only $\sim 0.1 \cdot 10^{18}$ m^{-3}, and is thus easily offset globally, with the average extratropical ozone number

density at 12 km ranging around $2 \cdot 10^{18}$ m^{-3}.

3.3. *Ozonesonde Case Study*

Fig. 5 depicts the timeline of both NNORSY and Hohenpeissenberg ozone sondes. For this comparison, layers of 10 km GPH thickness have been calculated to reduce errors resulting from the different vertical resolution of the ozone profiles.

The timeline plots reveal a good agreement of NNORSY with the sondes in all height ranges, but especially between 10 and 20 km. Note that for these plots, all GOME collocations within 250 km of a given ozonesonde were averaged into a single profile. As expected, largest deviations arise in Spring, where there temporal and spatial variability of ozone in the northern midlatitudes is high. There is however no significant bias or drift of the NNORSY partial columns with respect to the sondes, even after beginning of the year 2001, when the ERS-2 satellite was switched to gyroless mode and the operational GOME Level 1 data started developing a number of problems (Aben *et al.*, 2000). It seems the neural network is sophisticated enough to implement a correction for radiance data degradation effects based on the input data. However, it must be noted that on a test dataset average, the errors increase slightly after January 2000, so the correction is by no means perfect. The F&K climatology, also shown in the figure, cannot reproduce the short term changes in the ozone columns, and seems to be somewhat biased to higher values in the 20 to 30 km height range, which might hint at noticable ozone losses in this height range (Reid *et al.*, 2000), since the climatology is compiled from data in the time frame 1980–1991.

4. PROFILE CHARACTERIZATION

In Section 2.1., it was shown that the neural network method is based on a quadratic error function, giving rise to a single standard deviation parameter, which is unsatisfactory in most cases. Instead, the test data set statistics presented in the previous sections may in principle be used as a crude accuracy estimate for the ozone profiles. They encompass GOME spectral measurement noise, collocation errors, smoothing errors, and biases between the different collocated profiles. The uncertainties of collocated target profiles exhibit their influence in two different ways: In the training data, they serve as a noise source and help regularize the retrieval system during the training phase, while in the test data they add directly to the observed error statistics. Not included in these statistics are errors due to the spatial and temporal distribution of training data, i. e. errors connected to the representativeness of the training collocations. These come into play especially in places where collocations are sparse, at the north pole for instance, and over the southern hemisphere oceanic troposphere.

4.1. *Local Error Estimation*

On the other hand, a precision estimate for individual retrieved profiles would be desirable for a number of applications. In a classical system, these estimates can be achieved by calculating the Jacobian of the inversion,

$$\mathbf{D_y} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}}, \tag{10}$$

and propagating the input data errors according to

$$\mathbf{S}_\mathrm{D} = \mathbf{D_y}^T \mathbf{S}_y \mathbf{D_y}. \tag{11}$$

Here, $\mathbf{S}_y$ is the measurement error covariance matrix, and $\mathbf{S}_\mathrm{D}$ the retrieval error covariance matrix. In our case, the diagonal elements of $\mathbf{S}_\mathrm{D}$ yield variance estimates for each profile height level. Usually, the input noise is assumed to be uncorrelated, hence $\mathbf{S}_y$ is diagonal and the profile errors effectively result from a quadratic propagation of the input errors using Gauss' error propagation rule.

Although the NNORSY Jacobian can be easily calculated (Bishop, 1995a; Aires *et al.*, 2001), the above procedure seems to grossly overestimate the observed errors, leading to values of several thousand percent in some instances. Looking at the Jacobian column vectors – the so-called contribution functions (Rodgers, 2000) – which each describe the reaction of the output profile to a change of one input parameter, reveals very similar patterns between neighboring spectral values, sometimes with opposing sign. It would thus seem that the network determines the output largely from groups of input parameters, thereby cancelling their errors in a nonlinear, sophisticated way which cannot be reproduced by the linear approximations in Eq. 11. These findings are generally in line with what Aires *et al.* (2001) report on the nonlinear properties of retrieval networks.

Alternative local error estimation methods for neural networks have been extensively investigated in current literature (e. g. Papadopoulos *et al.*, 2001,

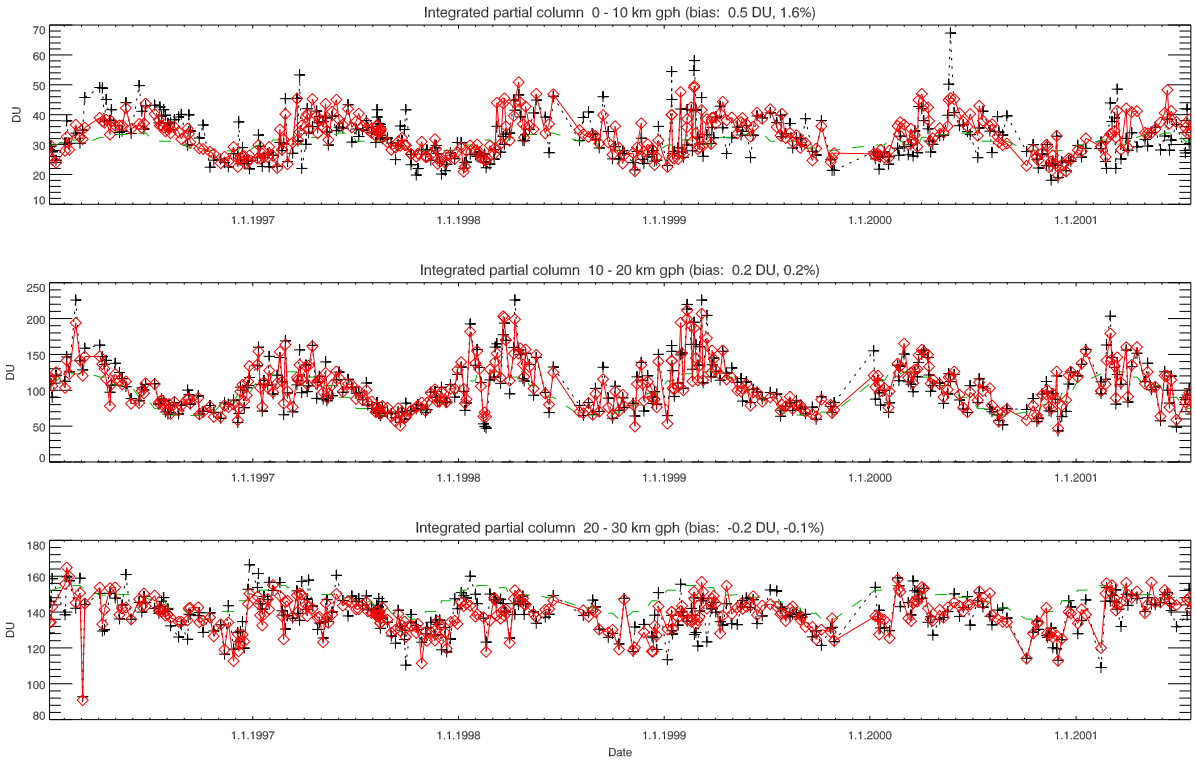

Figure 5: *Timeline of partial ozone columns from NNORSY (◇) and collocated Hohenpeissenberg (47.8°N, 11.0°E) ozone sondes (+). Multiple collocations within 160 km were averaged into single points. The F&K climatology is plotted for comparison as a dashed curve.*

and references therein), but most applications deal with fairly small data sets, for which an enhanced network can be constructed to approximate parameters of the output probability density (Bishop, 1995a). These methods commonly involve massive increases in the number of output neurons, which in our case would unbalance network size and training data set size.

Thus, two other alternatives have been investigated. One of them is a brute force approach, in which the retrieval is repeated M times while the input data is disturbed using realistic assumptions about measurement noise. The standard deviation $\tilde{\sigma}$ can then be calculated as

$$\tilde{\sigma}_k = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(d_{m,k}-\bar{d_k})^2}, \quad \text{where} \quad (12)$$

$$\mathbf{d}_m = \mathbf{o}_m - \mathbf{o}_t \quad (13)$$

is the difference between disturbed profile $\mathbf{o}_m$ and the undisturbed profile $\mathbf{o}_t$. Of course, this slows down the retrieval calculation by the factor M, but this is still acceptable – one retrieval with errors can be calculated in less than one second on a 400 MHz UltraSPARC2 processor. This method results in some kind of precision estimate. In the experiments conducted thus far, error covariances in the input data are not taken into account, and the assumptions about spectral measurement noise are greatly simplified, adding to the uncertainty of the method. As a worst case scenario, we used 10% noise for GOME Channel 1a, 5% for Ch. 1b, 3% for Ch. 2 and 2% for Ch. 3 and 4 [*R. van Oss, personal communication*], 1° uncertainty for the angular input data and 3 K for the UKMO temperature profiles.

The second method investigated has been inspired by Satchwell (1994, verbal presentation, cited in Bishop, 1995a). We trained a second neural network with exactly the same structure and input data as the retrieval network, but replaced the training collocations with the absolute difference between the retrievals and the collocations. The second network learns the dependence of error magnitude on the input data, giving an error estimate similar to the one provided in the previous two Sec-

tions, but this time for individual profiles. Due to the neural network property of reproducing the conditional average of the output, these error estimates when applied to the entire training data set sum up to give essentially the same picture as in Fig. 3. However, Dybowski (1997) notes that this method may result in a slight overestimation of high errors and underestimation of low errors, due to the optimization behaviour of the training algorithm. This remains to be checked for our data; visual comparison of the trained errors with the noise errors did not show any obvious anomalies, though.

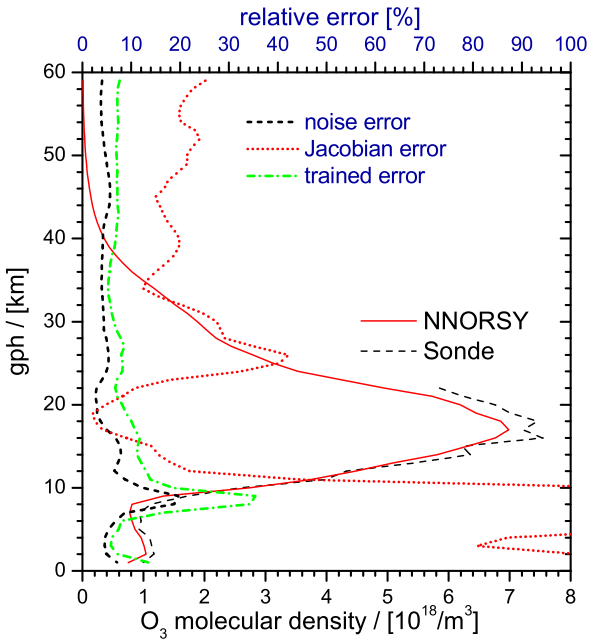

Figure 6: *Different NNORSY relative error estimates for a sample ozonesonde collocation from station Legionov, Poland (52.7°N; 19.3°O), 10th Feb. 1999. The noise error is derived from disturbing the input data, the trained error results from a second neural network. Classical error propagation using the neural Jacobian leads to questionable results. See text for further details on the methods employed.*

For the midlatitude high ozone situation depicted in Fig. 6, the brute force approach results in a relative standard deviation of around 5-10%, rather constant over height with the exception of a peak of 20% at 9 km. These statistics are based on $M = 5000$ noisy retrievals. The error-trained network yields a very similar error profile shape, but reaches higher values, as was expected since it encompasses more error sources. Similar behaviour of these two methods has been observed in different situations (Müller, 2002), with the highest relative errors occuring in situations and height ranges where low ozone concentrations coincide with high ozone variability and/or low training data density, i. e. inside the Antarctic ozone hole, around the midlatitude tropopause and in the tropical upper troposphere. As noted above, the Jacobian-calculated errors are generally larger by one or two orders of magnitude, and do not seem to represent the deviation of the retrievals from collocated profiles correctly. The best way to assess the quality of the error estimates gained thus far would probably be to investigate their impact when assimilating the NNORSY ozone profiles into a chemical transport model with independent validation.

Another interesting aspect of the noise error method is that it allows the output error correlation to be assessed simply by statistics on the noisy retrievals. In Fig. 7, we compare this correlation to the correlation of the Jacobian column vectors. Although not directly comparable in a mathematical sense, both should yield a measure of how the output profile levels are linked to each other. The resulting correlation matrices turn out to be fairly similar. Both show a largely symmetric and well-defined decrease of correlation around the diagonal, and distinct off-diagonal anticorrelations of the levels below 20–22 km to greater heights, which probably correspond to vertical ozone peak shifts: For instance, if the NNORSY retrieval wrongly shifts the ozone profile to a lower height, the steep flanks around the ozone peak will cause values below the peak to have a high tendency, while at the same time values above the peak will tend to be too low, hence the observed anticorrelation. It also follows that the similarities observed in the neural Jacobian column profiles do carry some information on the system's behaviour, which expresses in their height correlation, but cannot be exploited for error magnitude estimates, as shown above.

4.2. *Vertical Resolution*

The estimation of vertical resolution via averaging kernels (AKs) (Rodgers, 1990), as commonly used for classical retrieval schemes like OE, does not seem to be a ideal solution for the NNORSY system. The AK matrix $\mathbf{A}$ is defined here as

$$\mathbf{A} = \mathbf{D_y K}, \tag{14}$$

where $\mathbf{D_y}$ is the neural network Jacobian from Eq. 10, and $\mathbf{K}$ has to be calculated according to

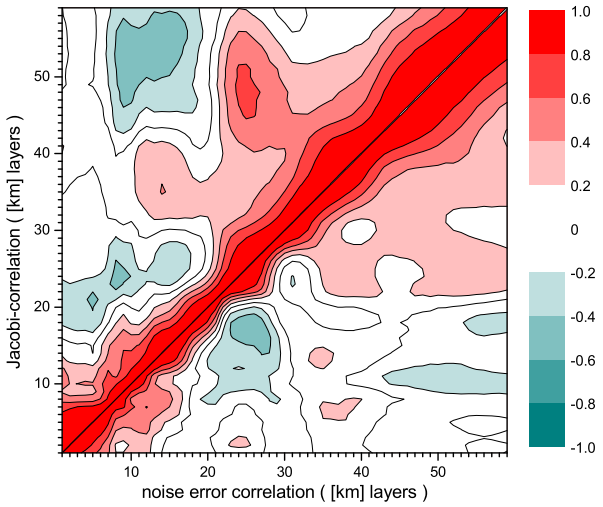

Figure 7: *Combined error correlation matrix, averaged over the test data set. The upper left half of this matrix is derived from Jacobian column correlation, the lower right from noise error correlation.*

Eq. 3 using a separate forward model. Each row of $\mathbf{A}$ then contains an AK for the respective height level.

While Jiménez (2000) found neural AKs for a simulated microwave limb-sounding retrieval to be comparable to OE AKs, if somewhat noisier, he also points out that the problem considered was fairly linear. For neural forward models, localized height information has been observed to spread out through the network during the different normalization procedures and nonlinear transfer functions it is submitted to (Aires *et al.*, 1999; Chevallier and Mahfouf, 2000). This leads to rather ill-conditioned Jacobians, as has been seen for NNORSY as well in the previous section. Using Jacobian-derived AKs for resolution estimation might also be inappropriate in our case for another reason, which is the wealth of statistical information present in the retrievals; in fact, some high-resolution features seen in the NNORSY profiles cannot possibly stem from spectral information alone. Therefore viewing the system through the classical AK lens might give a twisted impression of the observed geophysical vertical resolution altogether.

In view of these issues, we opted for a largely heuristical approach to vertical resolution estimation. From visual impression, it is obvious that the NNORSY profiles are oversampled, i. e. effective resolution is worse than the 1 km height sampling used for the output neurons. Ozonesondes and the occultation instruments used for collocation thus have a better resolution than NNORSY in most cases. A considerable part of the differences between retrievals and collocations in Fig. 3 can therefore be attributed to smoothing (cf. Rodgers, 2000). Convolution of the collocated profiles with a smoothing function should be able to reduce this error contribution. However, in the limiting case of an infinitely bad resolution, the smoothed profile would be constant over height, which obviously gave rise to large smoothing errors again, this time from the collocated profiles being too badly resolved. Somewhere between the extremes of full resolution and completely flattened profiles, there should be a minimum error, the corresponding FWHM of which could be viewed as the effective resolution of the retrieval at this height level.

We use Gaussians of variable full width half maximum (FWHM) δ as smoothing functions, which is equivalent to constructing AK matrices $\mathbf{A}_\delta$ from Gaussian-shaped AKs. Then, for δ varied in steps of 0.1 km, we calculate the RMS error of retrievals $\hat{\mathbf{R}}(\cdot)$ against smoothed test data set collocations,

$$\mathrm{RMSE}_\delta = \sqrt{\frac{1}{T}\sum_{p=1}^{T}(\hat{\mathbf{R}}(\mathbf{y}^p,\mathbf{c}^p,\mathbf{w}) - \mathbf{A}_\delta \mathbf{x}^p)^2}. \qquad (15)$$

In contrast to the standard deviation, the RMSE was chosen to take into account biases as well, which may occur whenever steep gradients in the profile are smoothed out. To reduce fringe effects from the convolution with incomplete profiles, which were observed to produce strong fluctuations, values at height levels with sonde and limb data missing were filled with corresponding retrieved values. Obviously, as soon as too many missing values have to be replaced by retrievals in a given height, the results will get distorted. It is clear that the smoothing error of a retrieval with respect to a smoothed version of itself can only rise, therefore the minimum error FWHM tends towards zero when the number of missing values increases. Note that this argument would not hold if we had used a climatology to replace missing values.

Plotting the RMSE against δ and height results in the contour seen in Fig. 8a. This plot was calculated from test collocations, but for training data (not shown) there is very little difference, which hints at a good generalization performance of the system — considerably larger test data set errors would have meant the network only “memorized” the training data.

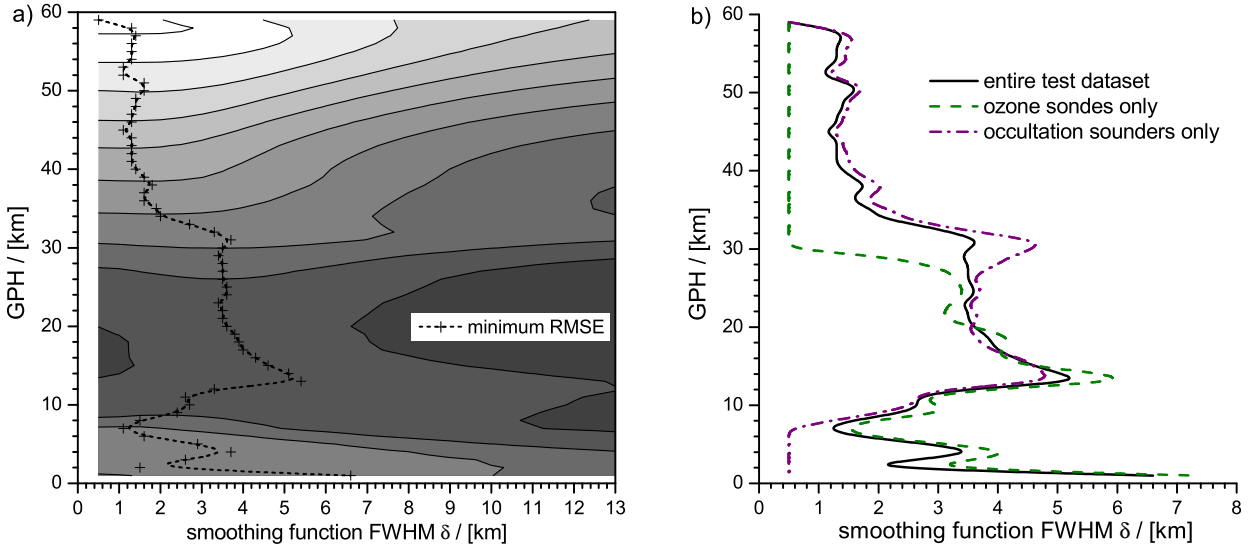

Figure 8: *(a) Test data set absolute RMSE variation (in arbitrary units) when convoluting target profiles with a Gaussian of given FWHM. (b) FWHM at which minimum RMSE occurs, calculated for different data (sub-)sets.*

A distinct saddle can be observed in the contour, the corresponding FWHM ranging between 3–5 km at 11 to 32 km gph, and between 1–3 km below and above, with the exception of the lower troposphere, where the minimum error occurs at 3–8 km FWHM. For once, it is clear that the method only works where the ozone profile exhibits some structure. Therefore interpretation of the minimum curve requires some care. To assist in the interpretation, Fig. 8b differentiates the results by collocation type. As can be seen, there is good agreement between the occultation and sonde curves between 8 and 22 km gph, with the curves diverging above and below. Due to the abovementioned treatment of missing data, we may safely assume the maximum FWHM from both curves to closer resemble the true vertical resolution.

Above 30 km, the collocated profiles are generally very smooth in the first place, therefore we cannot draw certain conclusions from the results here. On the other hand, vertical resolution is not as critical when there is no fine structure to detect. The other extreme can be observed below 12 km gph. Here, the convolution with a Gaussian of even relatively small FWHM leads to a shifting of ozone from the high ozone lower stratosphere into the troposphere, where it degrades retrieval accuracy. Thus, the minimum FWHM is unrealistically small. In fact, OE-derived AKs at these height levels tend to be quite asymmetric (Hoogen *et al.*, 1999), such that the approximation with a Gaussian may not reflect the physical conditions correctly. Using only sonde data from the tropics (not shown), where the tropopause is located at 15 km altitude and above, the RMSE minimum below 12 km altitude occurs at 5–8 km FWHM.

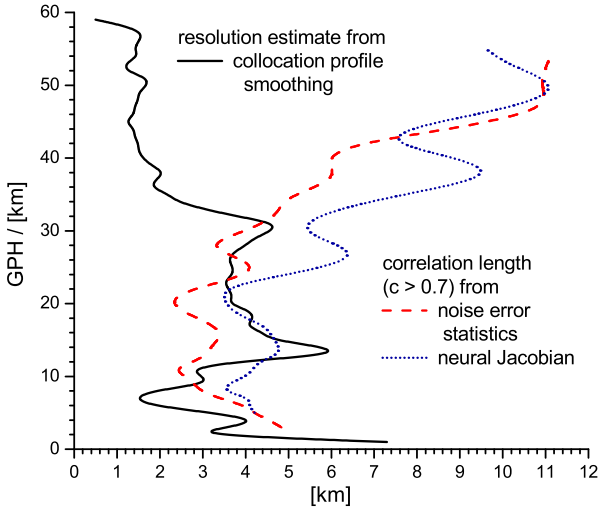

Figure 9: *Comparison of the different vertical resolution estimates employed.*

It might be worthwhile to compare the empirical resolution estimates obtained here with the height

level correlation lengths which can be derived from Fig. 7. Since the gradient between the contours for correlations of c =0.6 and 0.8 is particularly steep and the contours are quite well-defined, we choose the width of the diagonal ridge for $c \geq$0.7 as a measure of correlation length. Calculating this length separately for Jacobian and noise error correlation, and plotting it together with the maximum of the curves in Fig. 8b results in the graphs shown in Fig. 9. As can be observed, these resolution estimates agree reasonably well below 32 km, but diverge strongly above that level. The 10–11 km vertical resolution assessed through correlation lengths is probably quite realistic for larger heights, because it corresponds with the very smooth shape of the occultation profiles. Note however that this is a limitation of the training data, not necessarily of the retrieval system: From radiative transfer theory, there should be enough information in the GOME spectra to achieve better vertical resolution in this height range (Burrows *et al.*, 1999; Hoogen *et al.*, 1999). Inclusion of high-resolution lidar data into the training dataset might prove helpful to exploit this information in future versions of NNORSY. In the troposphere, looking at individual correlation matrices (not shown) reveals a lot of off-diagonal correlations and usually a fairly irregular diagonal ridge, therefore correlation lengths in this height range may not be dependable.

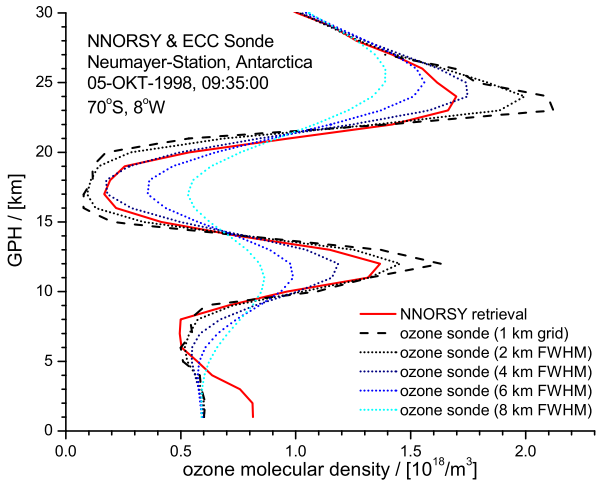

Figure 10: *NNORSY ozone profile retrieval with a collocated ozone sonde measurement, smoothed by Gaussians of varying FWHM.*

It may be surprising to see that these vertical resolution estimates are well below the theoretical limit derived from radiative transfer theory, which lies around 6–8 km (Hoogen *et al.*, 1999). We attribute this contradiction to the direct use of statistical and temperature information, and the absence of linear approximations in the system. To emphasize our point, Fig. 10 shows a NNORSY retrieval of a fairly extreme ozone hole case, where it can be seen that the collocated ozone sonde does indeed correspond best to the retrieval when smoothed with an FWHM of 2–4 km.

To summarize the findings presented, it seems that NNORSY ozone profiles have an effective vertical resolution (FWHM) of around 4 km in the 12–31 km GPH range, where the results are most reliable. Above 31 km, resolution degrades to about 10–11 km at 50 km GPH. This is a consequence of the smooth training profiles used, and does not necessarily reflect full exploitation of GOME spectral information. In the troposphere, resolution is difficult to assess with both methods employed, but is probably slightly less than in the stratosphere and distorted by long range correlations to greater heights.

5. CONCLUSIONS AND OUTLOOK

It has been demonstrated that the regularization imposed by using a neural network approach is sufficient to solve the inverse retrieval problem for GOME ozone profiles with stability and results comparable to classical methods. NNORSY is taking part in the international effort of comparing different retrieval systems in the frame of the ESA/ESRIN GOME Ozone Profiling Working Group, where we expect to gain more insight in the chances and problems of our method.

However, it is already clear that one big advantage of NNORSY is speed, making it well suited for real-time applications and global data reprocessing, even at full GOME horizontal resolution. An experimental near real-time system is already in operation at the German Remote Sensing Datacenter (DFD) of the DLR, the data can be accessed via the internet[†]. Error characterization still needs to be somewhat improved on a global and case-by-case basis for the operational product.

While the method is implicitly correcting for cloud albedos and degradation in GOME channels, it depends crucially on high quality training data. Therefore it cannot be used for species which are only sparsely measured. However, in case a forward model can accurately simulate measurements for

[†] http://auc.dfd.dlr.de/GOME_NRT/profile.html

a given sensor, using a partially or entirely simulated training data set is a viable option. In fact, this approach has recently led to the development of a new type of ground-based UV spectral radiometer[‡] based entirely on MPLs trained with simulated UV ground spectra (Schwander *et al.*, 2001). As a globally optimizing retrieval system, neural networks react to input data in a completely different way than local approaches like Optimal Estimation. If a bias-free training data set can be constructed, their retrievals results will also be globally bias-free, which might prove useful in applications like data assimilation.

For NNORSY-GOME ozone profiles, further improvements in the fields of spectral calibration and state space sampling distribution could lead to a yet better data quality in the future, especially concerning extreme cases of ozone profiles. This might include exploiting additional profile data sources – lidars, advanced sounders – and improving the outlier detection algorithms for both ozone training data and GOME spectra. Biases and errors occuring around the problematic tropopause region could probably be considerably reduced by defining a variable output height grid relative to the tropopause height, for this would introduce greater consistency e. g. between tropical and extratropical ozone profiles. There is also improvement potential in the refinement of the network architecture and training method. For instance, using an error function specifically adapted to the retrieval problem or training different (specialist) networks and combining their results might yield more accurate results and error estimations.

An adaptation of NNORSY towards state of the art sensors (SCIAMACHY, OMI, GOME-2) is planned, and can be carried out with considerably reduced effort, because instrument calibration specifics are learned automatically by the network, and much of the work in setting up software and training data has already been done. Since both the number of available training data and the computational effort for classical retrieval and/or assimilation of data from upcoming satellite instruments increase steadily, we expect to see growing use of neural network type methods in satellite meteorology, especially in the operational regime.

References

Aben, I., M. Eisinger, E. Hegels, R. Snel, and C. Tanzi, 2000: *GOME Data Quality Improvement (GDAQI), Final Report*, Tech. Rep. TN-GDAQI-003SR/2000, Space Research Organization Netherlands (SRON), Utrecht, Netherlands.

Aires, F., C. Prigent, W. B. Rossow, and M. Rothstein, 2001: A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations, *J. Geophys. Res.*, **106**(D14), 14887–14908.

Aires, F., M. Schmitt, A. Chedin, and N. Scott, 1999: The "weight smoothing" regularization of MLP for Jacobian stabilization, *IEEE-NN*, **10**(6), 1502.

Bargen, A. V., W. Thomas, and D. Loyola, 1999: *GOME Data Processor Update Report for GDP 0-to-1 Version 2.0 and GDP 1-to-2 Version 2.7*, Tech. Rep. ER-TN-DLR-GO-0043, Deutsches Zentrum für Luft- und Raumfahrt, Deutsches Fernerkundungs-Datenzentrum, Oberpfaffenhofen, Germany.

Baron, A., 1993: Universal approximation bounds for superposition of sigmoidal functions, *IEEE Trans. Information Theory*, **39**, 930–945.

Baron, A., 1994: Approximation and estimation bound for artificial neural networks, *Machine Learning*, **14**, 115–133.

Bhatt, P. P., E. E. Remsberg, L. L. Gordley, J. M. McInerney, V. G. Brackett, and J. M. Russell III, 1999: An evaluation of the quality of Halogen Occultation Experiment ozone profiles in the lower stratosphere, *J. Geophys. Res.*, **104**(D8), 9261–9275.

Bishop, C. M., 1995a: *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Bishop, C. M., 1995b: Training with noise is equivalent to Tikhonov regularization, *Neural Computation*, **7**(1), 108–116.

Bovesmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noel, V. Rozanov, K. V. Chance, and A. H. P. Goede, 1999: SCIAMACHY - Mission objectives and measurement modes, *J. Atmos. Sciences*, **56**, 125–150.

Burrows, J., E. Hölzle, A. Goede, H. Visser, and W. Fricke, 1995: SCIAMACHY - Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, *Acta Astronautica*, **35**, 445–451.

[‡] http://www.sprafimo.de

Burrows, J., M. Weber, M. Buchwitz, V. Rozanov, A. Ladstädter-Weissenmayer, A. Richter, R. de Beek, R. Hoogen, K. Bramstedt, K.-U. Eichmann, M. Eisinger, and D. Perner, 1999: The Global Ozone Monitoring Experiment (GOME): Mission concept and first scientific results, *J. Atmos. Sciences*, **56**, 151–175.

Chance, K. V., J. P. Burrows, D. Perner, and W. Schneider, 1997: 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. Rad. Transfer*, **57**(4), 467–476.

Chevallier, F., F. Chéruy, N. A. Scott, and A. Chédin, 1998: A neural network approach for a fast and accurate computation of a longwave radiative budget, *J. Appl. Meteor.*, **37**, 1385–1397.

Chevallier, F. and J.-F. Mahfouf, 2000: Evaluation of the Jacobians of infrared radiation models for variational data assimilation, *J. Appl. Meteor.*, submitted May 2000.

Cunnold, D. M., W. P. Chu, R. A. Barnes, M. P. McCormick, and R. E. Veiga, 1989: Validation of SAGE II ozone measurements, *J. Geophys. Res.*, **94**(D6), 8447–8460.

Cunnold, D. M., H. J. Wang, L. W. Thomason, J. M. Zawodny, J. A. Logan, and I. A. Megretskaia, 2000: SAGE (version 5.96) ozone trends in the lower stratosphere, *J. Geophys. Res.*, **105**(D4), 4445–4457.

de Beek, R., 1998: *Bestimmung von Ozonvertikalprofilen aus Messungen des Satelliteninstrumentes GOME im ultravioletten und sichtbaren Spektralbereich*, Ph.D. thesis, University of Bremen, Institute of Environmental Physics (IUP), Berichte aus der Umweltphysik D46 (in German).

Deniel, C., F. Dalaudier, E. Chassefiere, R. M. Bevilacqua, E. P. Shettle, J. D. Lumpe, K. W. Hoppel, J. S. Hornstein, D. W. Rusch, and C. E. Randall, 1997: A comparative study of POAM II and ECC sonde ozone measurements obtained over northern europe, *J. Geophys. Res.*, **102**(D19), 23629–23642.

Dybowski, R., 1997: *Assigning Confidence Intervals to Neural Network Predictions*, Tech. rep., Division of Infection, UMDS (St Thomas' Hospital), London.

Eichmann, K.-U., 2001: *Die Ozonverteilungen der Nordhemisphäre 1997–2000 Gemessen mit GOME: Einfluss Von Dynamik und Chemie in der Stratosphäre*, Ph.D. thesis, University of Bremen, Institute of Environmental Physics (IUP), (in German).

Fortuin, J. P. F. and H. Kelder, 1998: An ozone climatology based on ozonesonde and satellite measurements, *J. Geophys. Res.*, **103**(D24), 31079–31734.

Hasekamp, O., J. Landgraf, F. Helmich, R. van Oss, and R. van der A, 1999: Two inversion methods for ozone profile retrieval from GOME, in: *ESAMS'99 - European Symposium on Atmospheric Measurements from Space, Noordwijk, Netherlands, 18–22 January 1999*, pp. 549–553, ESA, WPP-161, Noordwijk.

Hoogen, R., V. V. Rozanov, and J. P. Burrows, 1999: Ozone profiles from GOME satellite data: Algorithm description and first validation, *J. Geophys. Res.*, **104**(D7), 8263–8280.

Hornik, K., M. Sinchcombe, and H. White, 1989: Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359–366.

Jiménez, C., 2000: *Inversion of Microwave Limb Sounding Observations of the Atmosphere by a Neural Network Technique*, Tech. Rep. 364L, Dept. Radio and Space Science, Chalmers University of Technology, Göteborg, Sweden.

Kaifel, A. K. and M. D. Müller, 2001: Results of TOVS ozone retrieval with neural networks, in: J. Le Marshall and J. D. Jasper (Eds.), *Techn. Proc. 11th Int. ATOVS Study Conference, Budapest, Hungary, 20–26 September 2000*, pp. 153–165, International ATOVS Working Group, Bureau of Meteorlogy Research Centre, Melbourne, Australia.

Krasnopolsky, V. M., 1997: *Neural Networks for Standard and Variational Satellite Retrievals*, Tech. rep., Contribution 148, NOAA Environmental Modeling Center, Ocean Modeling Branch, Washington (DC).

Lu, J., V. A. Mohnen, G. K. Yue, R. J. Atkinson, and W. A. Matthews, 1997: Intercomparison of stratospheric ozone profiles obtained by SAGE II, HALOE, and ozonesondes during 1994-1995, *J. Geophys. Res.*, **102**(D13), 16137–16144.

Lucke, R. L., D. R. Korwan, R. M. Bevilacqua, J. S. Hornstein, E. P. Shettle, D. T. Chen, M. Daehler, J. D. Lumpe, M. D. Fromm, D. Debrestian, B. Neff, M. Squire, G. König-Langlo, and J. Davies, 1999: The Polar Ozone and Aerosol Measurement (POAM) III instrument and early validation results, *J. Geophys. Res.*, **104**, 18.785–18.799.

Lumpe, J. D., R. M. Bevilacqua, K. W. Hoppel, and C. E. Randall, 2002: POAM III retrieval algorithm and error analysis, *J. Geophys. Res.*, in press.

Müller, M. D., 2002: *Bestimmung von atmosphärischem Gesamtozon und Ozonprofilen aus GOME Daten mit Hilfe des* Neural Network Ozone Retrieval System *(NNORSY)*, Ph.D. thesis, University of Bremen, Institute of Environmental Physics (IUP), (in German).

Müller, M. D. and A. K. Kaifel, 1999: Efficient processing of multi-year global TOVS data using ITPP, 3I and neural networks, in: J. LeMarshall and J. D. Jasper (Eds.), *Techn. Proc. 10th Int. ATOVS Study Conference, Boulder, Colorado, 27 Jan – 2 Feb 1999*, pp. 397–407, Int. ATOVS Study Working Group (ITWG), Bureau of Meteorology Research Centre, Melbourne, Australia.

Müller, M. D., A. K. Kaifel, M. Weber, and J. P. Burrows, 2002: Partial training of neural networks with incomplete target data, applied to atmospheric science, *Neural Networks*, submitted.

Müller, M. D., A. K. Kaifel, M. Weber, and S. Tellmann, 2001: Real-time total ozone and ozone profiles retrieved from GOME data using neural networks, in: *Proc. 2001 EUMETSAT Meteorological Satellite Data User's Conference, Antalya, 1-5 October 2001*, EUMETSAT, Darmstadt, Germany.

Papadopoulos, G., P. J. Edwards, and A. F. Murray, 2001: Confidence estimation methods for neural networks: A practical comparison, *IEEE-NN*, **12**(6), 1278–1287.

Plutovski, M. E. P., 1996: *Survey: Cross-Validation in Theory and in Practice*, Tech. rep., Department of Computational Science Research, David Sarnoff Research Center, Princeton, New Jersey, USA.

Reid, S. J., A. F.Tuck, and G. Kiladis, 2000: On the changing abundance of ozone minima at northern midlatitudes, *J. Geophys. Res.*, **105**, 12169–12180.

Riedmiller, M. and H. Braun, 1993: A direct adaptive method for faster backpropagation learning: The Rprop algorithm, in: *Proc. of the ICNN*, pp. 586–591, IEEE, San Francisco.

Rodgers, C. D., 1976: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, *Rev. Geophys. Space Phys.*, **14**(4), 609–624.

Rodgers, C. D., 1990: Characterisation and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, **95**(D5), 5587–5595.

Rodgers, C. D., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific Publishing, Singapore.

Rozanov, V., Y. Timofeyev, M. Biryulina, J. Burrows, R. Spurr, and D. Diebel, 1993: Accuracy of atmospheric constituent retrieval from multichannel remote sensing instruments, in: S. Keevallik and O. Kärner (Eds.), *IRS '92: Current Problems in Atmospheric Radiation*, pp. 394–397, Deepak Publishing, Ganotibm VAm USA.

Rumelhart, D. E., G. Hinton, and R. Williams, 1986: Learning internal representations by error backpropagation, in: D. Rumelhart and J. McClellant (Eds.), *Parallel Distributed Processing, Vol. 1*, vol. 1, pp. 318–362, MIT Press.

Russell, L. J., L. L. Gordley, J. H. Park, S. R. Drayson, W. D. Hesketh, R. J. Cicerone, A. F. Tuck, J. E. Frederick, J. E. Harries, and J. P. Crutzen, 1993: The halogen occultation experiment, *J. Geophys. Res.*, **98**, 10777–10797.

Schwander, H., A. Kaifel, A. Ruggaber, and P. Koepke, 2001: Spectral radiative-transfer modeling with minimized computation time by use of a neural-network technique, *Appl. Opt.*, **40**(3), 331–335.

Singer, S. F. and R. C. Wentworth, 1957: A method for the determination of the vertical ozone distribution from a satellite, *J. Geophys. Res.*, **62**, 299–308.

Slijkhuis, S. and D. Loyola, 1999: *GOME Data Processor Extraction Software User's Manual*, Tech. Rep. ER-SUM-DLR-GO-0045, Deutsches Zentrum für Luft- und Raumfahrt, Deutsches Fernerkundungs-Datenzentrum, Oberpfaffenhofen, Germany.

SPARC, 1998: *SPARC/IO$_3$C/GAW Assessment of Trends in the Vertical Distribution of Ozone*, WCRP-SPARC Report WMO TD No. 935, World Meteorological Organization.

Steele, H. M. and R. P. Turco, 1997: Separation of aerosol and gas components in the Halogen Occultation Experiment and the Stratospheric Aerosol and Gas Experiment (SAGE) II extinction measurements: Implications for SAGE II ozone concentrations and trends, *J. Geophys. Res.*, **102**(D16), 19665–19681.

Tamura, S. and M. Tateishi, 1997: Capabilities of a four-layered feedforward neural network: Four layers versus three, *IEEE Transactions on Neural Networks*, **8**(2), 251–255.

Thompson, A. M., J. C. Witte, R. D. McPeters, S. J. Oltmans, F. J. Schmidlin, J. A. Logan, M. Fujiwara, V. W. J. H. Kirchhoff, F. Posny, G. J. R. Coetzee, B. Hoegger, S. Kawakami, T. Ogawa, B. J. Johnson, H. Voemel, and G. Labow, 2001: The 1998-2000 SHADOZ (Southern Hemisphere Additional Ozonesondes) tropical ozone climatology: Comparisons with TOMS and ground-based measurements, *submitted to J. Geophys. Res.*

Wardle, D. I., E. W. Hare, D. V. Barton, and C. T. McElroy, 1998: The World Ozone and Ultraviolet Radiation Data Centre - Content and submission, in: R. Bojkov and G. Visconti (Eds.), *Proceedings of the XVIII Quadrennial Ozone Symposium, 1996*, L'Aquila, Italy.