

# Application of the Spectral Structure Parameterization technique: retrieval of total water vapor columns from GOME

R. Lang<sup>1,2,3</sup>, J. E. Williams<sup>1</sup>, W. J. van der Zande<sup>1</sup>, and A. N. Maurellis<sup>2</sup>

<sup>1</sup>FOM-Institute for Atomic and Molecular Physics, Amsterdam, The Netherlands

<sup>2</sup>SRON National Institute for Space Research, Utrecht, The Netherlands

<sup>3</sup>Department of Physics and Astronomy, Vrije Universiteit, Amsterdam, The Netherlands

Received: 26 June 2002 – Published in Atmos. Chem. Phys. Discuss.: 31 July 2002

Revised: 17 December 2002 – Accepted: 23 January 2003 – Published: 11 February 2003

**Abstract.** We use a recently proposed spectral sampling technique for measurements of atmospheric transmissions called the Spectral Structure Parameterization (SSP) in order to retrieve total water vapor columns (WVC) from reflectivity spectra measured by the Global Ozone Monitoring Experiment (GOME). SSP provides a good compromise between efficiency and speed when performing retrievals on highly structured spectra of narrow-band absorbers like water vapor. We show that SSP can be implemented in a radiative transfer scheme which treats both direct-path absorption and absorption by singly-scattered light directly. For the retrieval we exploit a ro-vibrational overtone band of water vapor located in the visible around 590 nm. We compare our results to independent values given by the data assimilation model of ECMWF. In addition, results are compared to those obtained from the more accurate, but more computationally expensive, Optical Absorption Coefficient Spectroscopy (OACS).

## 1 Introduction

The retrieval of concentrations of narrow-band line absorbers, such as water vapor (WV), from space-borne spectral instruments is often complicated by the relatively low resolution of the detector with respect to the width of an individual absorption line. For example, GOME spectral sampling in the visible region (channel 3) is 0.2 nm which covers up to 12 individual absorption lines of WV in the spectral region around 580 nm (ESA, 1995). To allow accurate modeling of such absorption spectra requires a large number of spectral realization points in order to resolve the narrowest lines, especially at high altitudes where pressure broadening is absent. In principle, band models, exponential sum fitting methods or other opacity sampling techniques, like the well known k-distribution method (Lacis and Oinas, 1991; Kato

et al., 1999), and the more recently proposed Optical Absorption Coefficient Spectroscopy technique (OACS) (Maurellis et al., 2000a; Lang et al., 2002) can be used to solve this problem. In practice, both of the latter two methods use opacity sampling probability functions to represent the effective absorption within a specific spectral sampling interval. These probabilities are then used either directly (OACS), or indirectly, via a transformation from wavelength into cross-section space (k-distribution), as weighting functions of a set of opacity basis functions, whose summation replaces a computationally expensive spectral sampling integral. Both of these methods have demonstrated their ability to solve the problem of spectral sampling in an accurate and efficient way for atmospheric water vapor absorption. In the case of OACS by utilizing GOME measurements in the absorption region around 590 nm (Lang et al., 2002) and, in the case of the k-distribution method, on synthetic spectra in the IR regions of the SCanning Imaging Absorption spectroMeter for Atmospheric CHartography (SCIAMACHY) instrument on ESA's ENVISAT (Buchwitz et al., 2000).

In contrast, in the Spectral Structure Parameterization (SSP) recently proposed by Maurellis et al. (2000b), and applied here for the first time to retrievals from reflectivity measurements, the number of opacity functions and their weights are reduced to only one basis function with one weighting parameter, called the spectral structure parameter  $w$ , which characterizes the spectral structure of the absorber within a specific wavelength range and at a specific altitude. SSP lends itself therefore more to the basic concept of band models, where the average absorption over a specific wavelength interval is represented by averaged line parameter values. Unlike band models, SSP is well suited for relatively small sampling regions containing only a small number of distinct absorption lines, such as those found in the regions covered by the detector pixels of both GOME and SCIAMACHY. In addition, the implementation of SSP in a radiative transfer scheme including direct path absorption, as well as the con-

Correspondence to: R. Lang (lang@mpch-mainz.mpg.de)

tribution of single-scattering to the measured reflectivities, is relatively simple and requires only a small amount of computational effort. Consequently, SSP retrieval of WVC is fast and may be performed on a global scale in a reasonable time, as demonstrated in this paper.

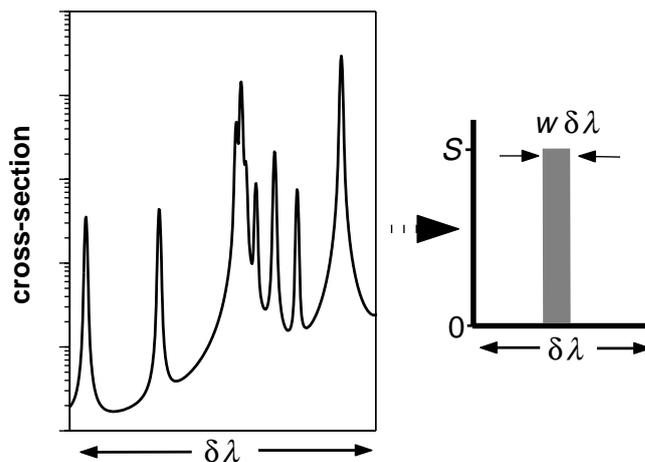
Retrieval of WVC from GOME has also been demonstrated by Noël et al. (1999) (at 700 nm) and Casadio et al. (2000) (at 740 nm). We focus on a ro-vibrational overtone band of WV in the visible between 580 and 605 nm covered by the GOME instrument. This band was first exploited for the retrieval of WVC by Maurellis et al. (2000a) and later by Lang et al. (2002), using the OACS sampling technique. SSP is a retrieval method, which is easy to implement, fast, self-contained and based on first principles. The only inputs needed for the retrieval of WVC by SSP are a pressure and temperature profile, together with some information on background absorbers, which are, for global retrievals, taken from climatology models or GOME level 2 data. SSP retrieves subcolumn profiles over 18 atmospheric levels from which a total column is calculated. The only constraints applied to the fits are three fixed upper profile constraints for tropical, mid-latitude and high-latitude cases.

In the following section we briefly summarize the basic concept of SSP for homogeneous atmospheres. Then we show how SSP can be applied to satellite-based, nadir viewing reflectivity measurements probing nonhomogeneous atmospheres, and including light paths where photons are reflected at the earth surface or scattered once (single scattering) in the atmosphere. Thereafter, we present results of SSP forward-modeled reflectivity spectra, which we compare to both line-by-line (lbl) forward-modeled spectra, as well as GOME measurements. SSP retrieval results from lbl forward-modeled spectra are used to assess the method related bias and to introduce a retrieval bias correction term especially for high WVC values. The impact of multiple scattering and aerosol loading on the retrieval values will also be discussed in some detail. We then expand our retrievals to a single GOME track and a global coverage (i.e. three days of continuous GOME measurements). We use a Singular Value Decomposition (SVD) scheme of the Jacobian matrix to assess the retrieval error of an individual fit. The WV total column results are compared and discussed with respect to co-located WVC given by the European Center of Medium Range Weather Forecast (ECMWF) data assimilation model, after which we present our conclusions.

## 2 Sampling of homogeneous absorption

For homogeneous absorption over a direct path in the absence of any scattering events, the spectrally sampled transmittance may be defined as

$$\langle T \rangle_{\delta\lambda} = \int_{\delta\lambda} \exp(-\sigma(\lambda)N) \frac{d\lambda}{\delta\lambda}, \quad (1)$$



**Fig. 1.** Schematic of the construction of an effective pixel coverage parameter  $w$  and effective mean value for the absorption cross-section  $S$  (right hand panel) from a typical absorption spectrum which has been sampled over a wavelength range  $\delta\lambda$  (left hand panel).

where  $\sigma(\lambda)$  is the absorption cross-section,  $N$  is the total column density and  $\delta\lambda$  is the sampling width. The transmittance  $\langle T \rangle_{\delta\lambda}$  may also be expressed as:

$$\langle T \rangle_{\delta\lambda} = 1 + w \exp(-SN) - w. \quad (2)$$

Here,  $S$  has units of cross-section and determines the height of a box with fractional width  $w$  ( $0 < w \leq 1$ ), which covers the area of absorption equivalent to the area covered by the individual lines of the absorber within the spectral region  $\delta\lambda$  (see Fig. 1) with

$$\langle \sigma(\lambda) \rangle_{\delta\lambda} = wS, \quad (3)$$

and the pixel-averaged optical thickness  $\langle \tau \rangle_{\delta\lambda} = wSN$ .  $w$  and  $S$  are now defined by the simultaneous solution of Eqs. (2) and (3) with the result that they are implicitly dependent on  $N$ . A Taylor expansion to second order of the exponents in Eqs. (1) and (2) reveals that

$$w \approx \frac{\langle \sigma \rangle_{\delta\lambda}^2}{\langle \sigma^2 \rangle_{\delta\lambda}} \quad \text{and} \quad S \approx \frac{\langle \sigma^2 \rangle_{\delta\lambda}}{\langle \sigma \rangle_{\delta\lambda}}. \quad (4)$$

The latter is a good approximation of  $w$  and  $S$  in cases of a wavelength-averaged optical thickness of lower than 0.1 (Maurellis et al., 2000b).

## 3 Application to nadir measurements in nonhomogeneous atmospheres

For the specific scenario of modeling the WV absorption from a nadir viewing geometry, we subdivide the atmosphere into 18 homogeneous levels  $\ell$ , containing WV subcolumn

densities  $N_\ell$ . For the retrieval, pressure and temperature profiles, as well as oxygen and nitrogen profiles for the calculation of Rayleigh scattering albedo, are taken from the Neutral Atmosphere Empirical Model MSISE90 (Hedin et al., 1991) for a given date, time, and geolocation at fixed altitude levels above the surface. We use 18 homogeneous atmospheric layers  $\ell$  to cover altitudes from the ground up to 9 km, which contain more than 99% of the atmospheric water vapor. The WV density drops generally by more than 2 orders of magnitude over the first 10 km (c.f. Lang et al., 2002). In general, the  $w$  width parameter changes with pressure and temperature over altitude due to the pressure and temperature dependent width of the lines. The same holds for the detector pixel averaged line strength  $S$ . For the treatment of nonhomogeneous paths traversed by photons undergoing higher orders of scattering we now assume that the absorption over the different light paths throughout the atmosphere is dominated by the level with the maximum absorption, as is the shape of the measured absorption line. The choice of a dominant absorbing layer is dependent on the shape of the profile of the absorber of interest. Note, that due to the strong exponential decrease of the WV profile with respect to altitude, we may assume that the maximum impact on the absorption over a specific light path will occur at the lowest point of the path. We write the unitless normalized reflectivity  $R$  as

$$R = \frac{\pi I}{\mu_o F_o}, \quad (5)$$

where  $I$  is the earth radiance,  $F_o$  the solar irradiance and  $\mu_o$  the cosine of the solar zenith angle (SZA). A solution of the transport equation of scalar radiative transfer in its plane parallel approximation (Lang et al., 2002) reveals that the measured reflectivity of a nadir viewing instrument may be separated into three distinctive parts, viz.,

$$R_j = \langle R_{\text{surf}} \rangle_j + \langle R_{\text{ss}} \rangle_j + \langle R_{\text{ms}} \rangle_j, \quad (6)$$

where  $j$  is a specific spectral interval with width  $\delta\lambda$ ,  $R_{\text{surf},j}$  is the reflectivity of the direct light path reflected at the earth surface,  $R_{\text{ss},j}$  is the reflectivity of singly scattered and  $R_{\text{ms},j}$  of multiply scattered light. Using the assumption above on the dominant  $w$ -parameter, we define  $w_{\text{max}} = w(N(\bar{\ell}))$ , where  $N(\bar{\ell})$  is the maximum subcolumn density for a specific path at level  $\bar{\ell}$ . Now, we may write the reflected light coming from the surface, utilizing Eq. (2), as

$$\langle R_{\text{surf}} \rangle_j = \Lambda \left\{ 1 + w(\bar{\ell}_{\text{surf}})_j \exp \left[ - \sum_{\ell} \tilde{\mu} S(\ell)_j N(\ell) \right] - w(\bar{\ell}_{\text{surf}})_j \right\}, \quad (7)$$

where  $\tilde{\mu}$  is the geometric path-length factor for a nadir viewing instrument, with  $\tilde{\mu} \equiv \left( \frac{\mu_o + \mu}{\mu_o \mu} \right)$ ,  $\mu$  is the cosine of the angle between the zenith and the instrument,  $\Lambda$  is the surface albedo of a Lambertian surface and  $w(\bar{\ell}_{\text{surf}})_j$  is the dominant  $w$ -parameter  $w_{\text{max}}$  for the specific case of the direct, surface

reflected path. Because, in the case of WV absorption, we expect the maximum optical thickness at the surface layer we may set  $\bar{\ell}_{\text{surf}} = 1$ . Using the same reasoning for the single-scattering contribution we can then write, by again utilizing Eq. (2),

$$\langle R_{\text{ss}} \rangle_j = \frac{p(\Theta)}{4\mu_o} \int_0^\infty \langle \beta(\ell(z))_{\text{sca}} \rangle_j \times \left\{ 1 + w(\bar{\ell}_{\text{ss}})_j \exp \left[ - \sum_{\ell=\ell(z)}^{\ell_{\text{top}}} \tilde{\mu} S(\ell)_j N(\ell) \right] - w(\bar{\ell}_{\text{ss}})_j \right\} dz \quad (8)$$

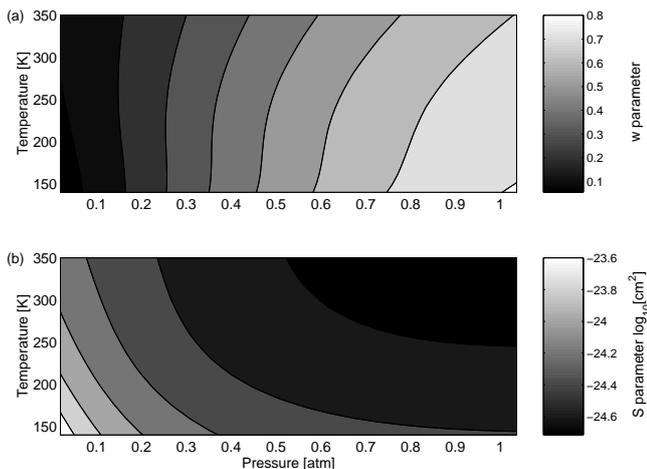
where  $\bar{\ell}_{\text{ss}} = \ell(z)$ , because  $\ell(z)$  is the lowest altitude point for each single-scattering path and, therefore, in the case of WV, the point with most impact on the absorption. In addition, we assume that Rayleigh scattering is the dominant form of single scattering and  $\langle \beta(\ell(z))_{\text{sca}} \rangle_j$  is the mean Rayleigh scattering coefficient within  $\delta\lambda_j$  at altitude  $z$ , with a corresponding phase function of  $p(\Theta) = \frac{3}{4}(1 + \cos^2(\Theta))$ , where  $\Theta$  denotes the SZA.

The multiple-scattering contribution  $\langle R_{\text{ms}} \rangle_j$  to the reflected light is represented by a first-order polynomial with free parameters  $C$  and  $D$ , which accounts for the broadband effect of multiple scattering. In the previous study by Lang et al. (2002) it was shown that the error on the retrieved WVC using this kind of treatment in OACS for the multiple-scattering contribution is lower than 20% in cloud-free situations for conditions representative of two worst-case scenarios: with high WVC ( $1.4 \times 10^{23}$  molec/cm<sup>2</sup>), low surface albedo (0.03 (maritime), 0.1 (rural)) and high aerosol loading (see also Sect. 7 for the impact of aerosol loading and multiple scattering on the SSP retrieval). Similarly, the unknown surface albedo is represented by a first order polynomial with free parameters  $A$  and  $B$ . The parameters  $A$  to  $D$  have to be adjusted to the real measurement when comparing SSP forward model results to the measurements, as well as during the retrieval and fitting of WVC values.

For details concerning the different contributions to the total reflectivity, as well as a detailed discussion about the impact of multiple scattering within this wavelength region, the reader is referred to the aforementioned study by Lang et al. (2002) and, for the specific case of SSP, to Sect. 7 of this work.

#### 4 Construction of look-up-tables for $w$ and $S$

In this study we utilize Eq. (4) in order to construct a look-up-table for each of the parameters  $w$  and  $S$ . For a range of 22 temperatures ( $T$ ) and 31 pressure values ( $p$ ), covering all  $p$  and  $T$  values thought relevant for our altitude region of interest for all possible geolocations, we then calculate realizations of the cross-section  $\sigma$  utilizing line-parameters from the HITRAN '96 database (Rothman et al., 1998) and Voigt absorption line shapes (Humlicek, 1982). For those realizations



**Fig. 2.** Contour plot of the  $w$ - (a) and the  $S$ -parameter (b) look-up-table for one spectral sampling bin  $j$  in dependence of temperature and pressure. A standard pressure and temperature grid is used for the construction of the look up tables, covering all possible pressure and temperature ranges within the altitude region of interest.

we use 50 000 numerical grid points over the whole spectral region between 585 and 600 nm, in order to resolve the narrowest lines with at least 5 points. We use a sampling width  $\delta\lambda$ , that is three times smaller than the sampling width of a GOME detector pixel and more than 4 times smaller than the full-width-half-maximum (FWHM) of the instrumental response function, which satisfies the Nyquist criterion (see also following section) and results in a total of 207 spectral bins. By doing so we create for each parameter  $w$  and  $S$ , a 22 by 31 by 207 look-up-table matrix which we can interpolate over  $p$  and  $T$  for a specific measurement scenario at each altitude level  $\ell$ . Fig. 2 shows  $w$  and  $S$  for one wavelength bin  $j$  as a function of the  $p$  and  $T$  reference profiles. From this we see that  $w$  and  $S$  vary smoothly with respect to  $p$  and  $T$ , meaning that interpolation on intermediate  $p$ ,  $T$ -tuples leads to reasonable results. As the pressure decreases the  $w$  parameter also decreases (upper panel Fig. 2), i.e. the spectrum becomes more structured due to the pressure-related narrowing of the lines. The line width is only weakly dependent on the temperature (Rothman et al., 1998) which can also be seen from the  $w$  contour plot. The  $S$  parameter (lower panel Fig. 2) changes significantly only for the lowest pressures. For low pressures, when the  $w$  parameter decreases, the  $S$  parameter increases, because the area of absorption or the averaged line intensity within a certain wavelength bin has to be conserved. Due to the temperature dependence of the line-intensity value of individual absorption lines (Rothman et al., 1998), the variation in temperature for the  $S$  parameter is stronger than that for the  $w$  parameter.

In general, the  $w$ ,  $S$ -tables are smaller than the usual tables used for the  $k$ -distribution method, and significantly smaller than the look-up-tables used for the OACS method, because

the extra dimension for the summation over cross-section probability, or cross-section bin space, necessary in the latter two methods, is not required for SSP. Therefore, the interpolation on the look-up-tables can be performed very rapidly even for small  $\delta\lambda$ , and thus, for high wavelength resolution.

## 5 Forward modeling and retrieval

The spectral region between 585 and 600 nm is covered by 69 detector pixels of the GOME instrument. Within this region absorptions of the additional absorbers ( $\text{O}_2$ )<sub>2</sub>,  $\text{O}_3$ ,  $\text{NO}_2$  and Sodium have to be taken into account, together with the loss of light scattered out of the light path by means of Rayleigh scattering. Apart from sodium and  $\text{NO}_2$ , which contribute very little to the total absorption in this spectral region, all of the remaining absorbers contribute to a smooth background absorption. The contribution of the Fraunhofer line absorption to the so-called  $F_0$  effect and the contribution of the Ring effect, from rotational Raman-scattering, to the total reflectivity is also assumed to be small within this spectral region (for a detailed discussion the reader is referred to Lang et al., 2002).

GOME measures the earth radiance  $I$  in nadir viewing geometry for a footprint of 40 by 320 km. The solar irradiance  $F_0$  is measured once for every day (for a detailed description of the instrument see ESA, 1995; Burrows et al., 1999). The instrumental function  $\mathcal{H}(\lambda, \lambda'; \Sigma)$  of channel 3 is represented here by a Gaussian function with a FWHM  $\Sigma$ . The reflectivity measured by the  $k$ th detector pixel is then defined as

$$R_k^{\text{GOME}} = \int_{\Delta\lambda_k} \int_{-\infty}^{+\infty} \frac{\pi I(\lambda)}{\mu_o F_0} \mathcal{H}(\lambda, \lambda'; \Sigma) d\lambda' \frac{d\lambda}{\Delta\lambda_k}, \quad (9)$$

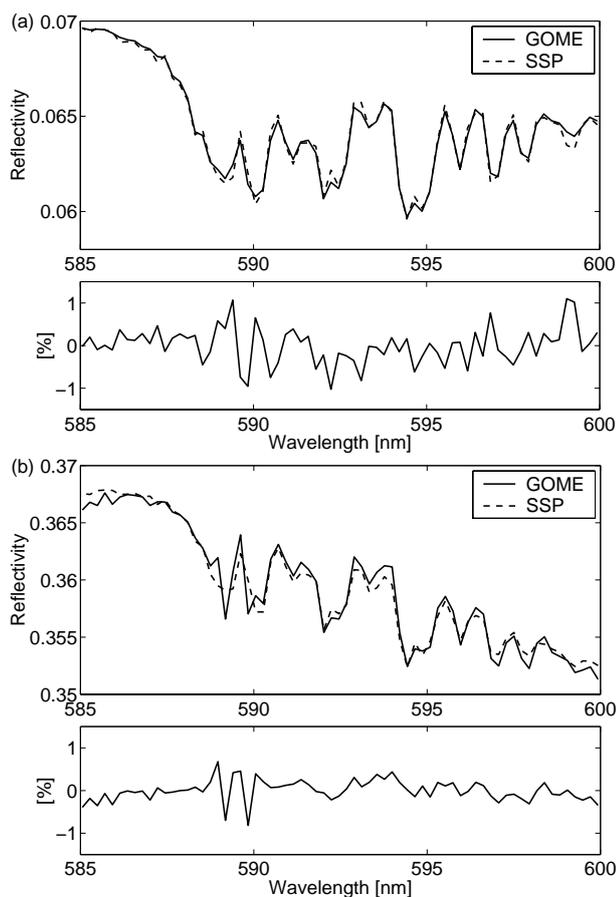
where  $\Delta\lambda_k$  is the wavelength coverage of a GOME detector pixel of about 0.21 nm in case of channel 3. Here, it is assumed that the solar irradiance  $F_0$  is constant within  $\Delta\lambda_k$ . The modeled reflectivity  $R_k$  can be approximated by

$$R_k^{\text{Model}} = \sum_j^K \sum_i^M R_j \mathcal{H}(\lambda_j, \lambda_i; \Sigma) \delta\lambda_i \frac{\delta\lambda_j}{\Delta\lambda_k}, \quad (10)$$

where  $M$  is the number of sampling bins  $i$  over the total spectral region of interest,  $K$  is the number of sampling bins  $j$  within the spectral region covered by one GOME detector pixel and  $|\delta\lambda_{i,j}| = K^{-1} \Delta\lambda_k$ .  $R_j$  is the sampled reflectivity from Eq. (6) with

$$R_j = \langle R_{\text{surf}}(N_\ell, A, B) \rangle_j + \langle R_{\text{ss}}(N_\ell) \rangle_j + (C\lambda_j + D). \quad (11)$$

Here, it must be noted that Eq. (10) is only a good representation of Eq. (9) when  $\delta\lambda \ll \Sigma$ , satisfying the Nyquist criterion. In our case  $\delta\lambda$  (around 0.07 nm) is a third of the spectral width covered by a single GOME detector pixel ( $K = 3$ ), whilst  $\Sigma$  is about 0.29 nm.



**Fig. 3.** The upper panels (a) show a GOME measurement (solid line) at 34° N latitude and 110° W longitude over the pacific with a SZA of 23° together with the result of a SSP forward model (dashed line) and their residual (GOME-SSP)/GOME. The albedo was adjusted using a first order polynomial, resulting in a value of 0.04 at 590 nm. The WV density profile was taken from ECMWF with a high WVC of  $1.34 \times 10^{23}$  [molec/cm<sup>2</sup>]. The lower panels (b) show the same but for a rural GOME measurement at 62° N latitude and 100° W longitude with a high SZA of 73° and a low WVC of  $8.9 \times 10^{21}$ . For this pixel the albedo was adjusted resulting in a value of 0.32 at 590 nm.

In Fig. 3 we present a direct comparison between sample GOME measurements and forward calculations using Eq. (10) for two representative measurement geometries and geolocations with both high and low WVC. Realistic WV profiles used for the forward calculations are taken from the ECMWF database, with the resulting residuals being smaller than 1%, except for the region near the solar sodium Fraunhofer lines which are not explicitly modeled here.

For the SSP retrievals we fit Eq. (10) to GOME reflectivity spectra with the aid of a robust, non-linear, large-scale trust-region method (Byrd et al., 1988), which solves the op-

timization problem

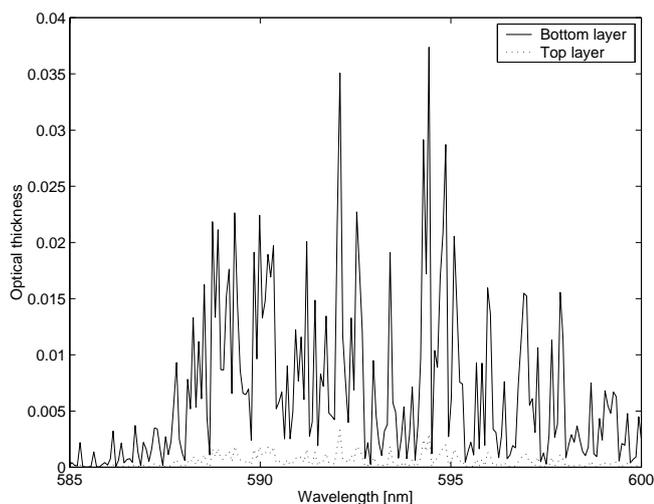
$$\min_{N_\ell, (A, \dots, D)} \sum_k \left[ R_k^{\text{GOME}} - R_k^{\text{Model}}(N_\ell, (A, \dots, D)) \right]^2. \quad (12)$$

Each SSP fit is initialized with a flat WV subcolumn profile  $N_{\ell,0} = 10^{16}$ , where the  $N_\ell$  are given in units of sub-column density. Each fit is constrained by an upper limit  $N < N_{\ell,\text{max}}$  in the form of a step function.  $N_{\ell,\text{max}}$  consists of three steps over the total number of levels with decreasing sub-column values from low to high altitude. The lower profile constraint is set to  $N_{\ell,\text{lower}} = 0$ . The constraints prevent the fit from giving too much weight to the higher altitude levels, which otherwise would increase the relative contribution of the singly scattered photons to unrealistic values by setting the surface albedo to zero and, in doing so, decrease the total mean free path length (for a detailed discussion see Sect. 8). We scale the upper profile constraint differently without changing their altitude-dependent shape for geolocations between 30° N and 30° S latitude (expected total WVC  $< 4 \times 10^{23}$  molec/cm<sup>2</sup>), between 30° and 60° latitude (expected total WVC  $< 2 \times 10^{23}$  molec/cm<sup>2</sup>) and latitudes above or below 60° (expected total WVC  $< 2 \times 10^{22}$  molec/cm<sup>2</sup>).

## 6 Method accuracy and bias adjustment

Two important assumptions were made in the derivation of Eqs. (7) and (8): (i) that the approximation of the  $w$  and  $S$  parameter (Eq. 4) is an equality, which is only true for optical thicknesses lower than 0.1, where the non-linear nature of the absorption of individual lines per layer becomes weakened; and (ii) the assumption on the dominant spectral structure parameter ( $w_{\text{max}}$ ) for the individual light paths of singly-scattered photons (Sect. 3). We call the bias introduced due to these assumptions the method-related bias of SSP.

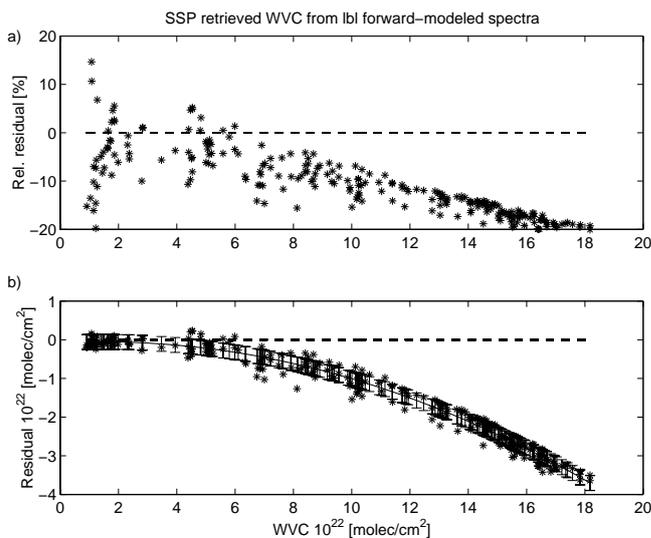
In principle, by reducing the thickness of the individual layers (i.e. utilizing more altitude layers) the optical thickness per layer may be always reduced below the 0.1 limit. However, increasing the number of altitude layers increases the number of fit parameters and makes the optimization procedure computationally more expensive without gaining much additional information about the profile. Fig. 4 shows that, by using 18 homogeneous altitude layers, the spectrally averaged optical thickness per layer is significantly below 0.1 even for measurement cases with high total WV content (for the measurement modelled in Fig. 4 the total WVC is  $1.83 \times 10^{23}$  molec/cm<sup>2</sup>). Therefore, for our case, the derivation of  $w$  and  $S$  is based on good assumptions, which do not contribute to the method-related bias. However, our assumption of the dominant spectral structure parameter  $w_{\text{max}}$ , which is equal to the  $w$  value for the layer with the highest optical thickness in the path, creates a bias. The fixed  $w_{\text{max}}$  is higher (except for the lowest layer) than the real  $w$  parameter per layer would be. Subsequently (from Eq. 2) an underestimation follows for high WVC.



**Fig. 4.** Spectral averaged optical thickness for the bottom and top atmospheric layer used in the retrieval with subcolumn densities of  $1.8 \times 10^{22}$  and  $1.3 \times 10^{21}$  molec/cm<sup>2</sup>, respectively. In this case all of the 18 subcolumn densities would sum up to a high total WVC of  $1.83 \times 10^{23}$  molec/cm<sup>2</sup>.

In order to quantify this bias and to correct for it, we perform SSP retrievals of 241 lbl forward-modeled reflectivity spectra on the basis of Eq. (9) from a range of WV profiles, with the total WVC ranging between  $8 \times 10^{21}$  and  $1.8 \times 10^{23}$  molec/cm<sup>2</sup>. These profiles were taken from the ECMWF database and chosen to be representative of a wide range of geolocations. The contributions from the additional background absorbers at these different geolocations were also accounted for. Surface albedos and multiple-scattering contributions (parameters A-D) used in the forward-lbl modeling results are evaluated from GOME measurements taken at corresponding geolocations. In order to do so the lbl-model results are fitted to the GOME measurements by keeping the ECMWF water vapor profile fixed. This procedure provides realistic parameters A to D to be used for the lbl-modelled spectra.

The results are summarized in Fig. 5, which shows the relative and absolute differences between WVCs retrieved using SSP and that used for the forward-lbl modeling. For low and mid-range columns ( $\leq 8 \times 10^{22}$  molec/cm<sup>2</sup>) the maximum differences between the retrieved and the true values is about 10%. For instances of very low WVC ( $< 1 \times 10^{22}$  molec/cm<sup>2</sup>) that occur predominantly at high latitudes, the differences are larger not only because of the low WVC values, but also due to the high SZA, which are usually associated with such GOME measurements. Note that together with the forward-lbl calculations for the lowest WVC (i.e.  $8 \times 10^{21}$  molec/cm<sup>2</sup>), a SZA of  $61^\circ$  is associated, which is representative of solar zenith angles in GOME measurements taken at latitudes  $> 70^\circ$ . The accuracy for low and mid columns up to  $8 \times 10^{22}$  molec/cm<sup>2</sup>, calculated from the mean



**Fig. 5.** Relative (a) and absolute (b) differences between the retrieved WVC by SSP and the WVC used in the lbl forward model (stars). The solid line in the lower panel (b) shows a second order polynomial fit through the absolute difference values together with the errorbars derived from the distribution matrix (for details see text and Sect. 8.1). The polynomial values are used to adjust for the systematic method related bias of the SSP retrieval.

relative differences, is  $-5.1\%$ .

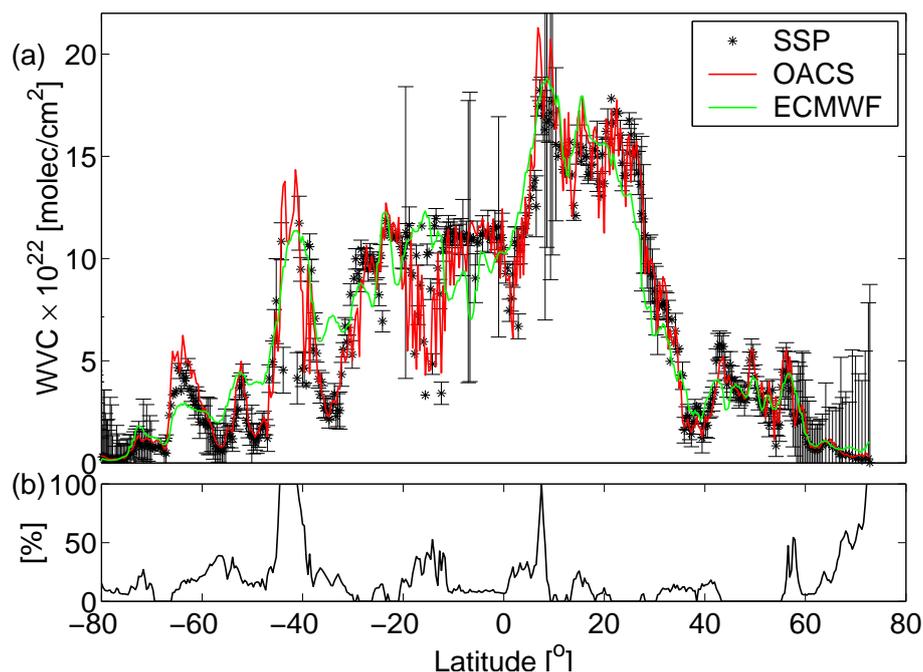
For high WVC ( $> 8 \times 10^{22}$  molec/cm<sup>2</sup>) the systematic retrieval bias increases up to a maximum of  $-20\%$ , which is referred to as the underestimation of the WVC by the dominant layer assumption. An empirical correction term was determined by fitting a second-order polynomial (solid line in Fig. 5b) to the absolute differences, which results in

$$\begin{aligned} \text{WVC}_{\text{corr}} = & -[0.018 \pm 4 \times 10^{-7}] * (\text{WVC} - [0.096 \pm 0.007])^2 \\ & + [0.164 \pm 1.9 \times 10^{-4}] * (\text{WVC} - [0.096 \pm 0.007]) \\ & - [0.283 \pm 0.190], \end{aligned} \quad (13)$$

where the WVC is in units of  $10^{22}$  molec/cm<sup>2</sup>. The error bars are calculated from evaluating the distribution matrix of the second-order polynomial fit together with the  $2\sigma$  distribution around the mean values. For details on such a fitting-error analysis see Sect. 8.

## 7 Aerosol loading and multiple scattering

Due to the large footprint of GOME (40 by 320 km), most ground pixels are contaminated by cloud cover to some degree. For instances where the cloud cover  $< 10\%$  significant errors in the retrieved WVC may still result due to the presence of background aerosol. Such aerosol may affect the net absorption, as well as the path length of the scattered light and, therefore, make the assumptions made with respect to scattering and surface albedo invalid.



**Fig. 6.** SSP WVC fit results over latitude ((a), stars) for a GOME pass on 23 October 1998, covering longitudes from  $90^{\circ}$  W at  $73^{\circ}$  N latitude to  $150^{\circ}$  W at  $80^{\circ}$  S latitude. 1.5% of the results are neglected because the errors were bigger than  $1 \times 10^{23}$  molec/cm<sup>2</sup>. In 19% of the cases the calculated error was rejected and is not shown because an optimal regularization value using the L-curve regularization scheme could not be found (for details see text). SSP results are corrected for model bias (cf. Fig. 5). For validation, the WVC values from OACS using GOME measurements (red curve) and from ECMWF given at 18:00 UTC are shown (green curve). The cloud coverage in percentage per GOME observation is taken from GOME GDP level-2 data and is indicated by the solid curve in the lower panel (b).

The effect of multiple scattering and aerosols on the retrieved WVC has been estimated by solving the full scalar radiative transfer equation, including the multiple-scattering source term, by employing a doubling-adding model (DAM) (de Haan et al., 1987), which has been reduced to the scalar representation of the radiation field. The reflectance  $R(\lambda)$  is calculated in a line-by-line mode for a spectral resolution of  $0.01 \text{ cm}^{-1}$ .

We investigate four atmospheric scenarios: a maritime case and a rural boundary layer case both with and without aerosol loading (clear-sky scenario). The clear-sky scenario quantifies the contribution of multiple Rayleigh scattering and its impact on the retrieved WVC. All scenarios include the effects of Rayleigh scattering, ozone absorption and Lambertian surface reflection. In the case of the maritime scene we use a surface albedo of 0.03, whereas for the rural scene a surface albedo of 0.1 is used. For maritime aerosol loading a constant particle density of 4000 particles per cm<sup>3</sup> is assumed. For the middle and upper troposphere, we assume a tropospheric background aerosol, for which the particle density decreases with the third power in pressure; aerosol optical properties are taken directly from Shettle and Fenn (1979). For the rural case we have chosen a constant, but much higher, particle density of 15 000 particles per cm<sup>3</sup>.

Table 1 lists the relative contribution of ground-reflected,

singly and multiply scattered photons as a percentage of the total reflectivity using the DAM-model for the four different aerosol loading cases described above, and with a medium WVC of  $7.72 \times 10^{22}$  molec/cm<sup>2</sup>, for which the systematic method related bias is negligible. Without aerosols, the maximum contribution of multiply scattered photons is about 8% at 590 nm. Single scattering is the dominant source of reflectivity for the maritime aerosol scenario (47%) because of low surface albedo, with the ground-reflected component being comparable to, or lower than, the contribution due to multiple scattering. This may be contrasted with the rural aerosol scenario, in which multiple scattering is the dominant source of scattered light (51%) due to the high aerosol optical density. Aerosols reduce the ground-reflected component to 13%.

The first-order polynomial used in this retrieval to implement multiple scattering (Eq. 6) does not account for a differential contribution reflecting the absorption by water vapor, but only for the broadband effect. The retrieval of parameters C and D may, therefore, influence the retrieval of the surface albedo (parameters A and B) and the WVC.

Table 1 also includes the SSP retrieved WVC and surface-albedo values from the DAM forward-modeled reflectivity spectra for the four scenarios. The results show that SSP retrieves accurate values for both of the clear-sky situations. It

**Table 1.** Relative contribution of ground-reflected, singly-scattered and multiply-scattered photons to the total reflectivity calculated by the DAM model assuming a SZA of 40° and a WVC of  $7.72 \times 10^{22}$  [molec/cm<sup>2</sup>] for each case

	Model Input		Model Output			SSP Retrieval Results			
	AOD <sup>a</sup>	Λ	GR <sup>b</sup>	SS <sup>c</sup>	MS <sup>d</sup>	WVC <sup>e</sup>	Δ WVC [%]	Λ	ΔΛ [%]
Maritime Clear Sky	0	0.03	42.2	49.6	8.2	7.68	-0.33	0.036	+20.0
Maritime Aerosol	0.30	0.03	17.9	46.9	35.2	6.76	-12.3	0.052	+73.3
Rural Clear Sky	0	0.10	67.8	23.9	8.3	7.69	-0.26	0.110	+10.0
Rural Aerosol	0.63	0.10	13.0	25.1	61.9	9.14	+18.6	0.121	+21.0

<sup>a</sup> Aerosol Optical Depth<sup>b</sup> Percentage of ground reflected light at 592 nm<sup>c</sup> Percentage of single scattering at 592 nm<sup>d</sup> Percentage of multiple scattering at 592 nm<sup>e</sup>  $10^{22}$  molec/cm<sup>2</sup>**Table 2.** Same as Table 1 but for a high WVC of  $1.43 \times 10^{23}$  molec/cm<sup>2</sup>. The relative difference Δ WVC is calculated between the model value and the bias-corrected retrieved WVC

	Model Input		Model Output			SSP Retrieval Results				
	AOD <sup>a</sup>	Λ	GR <sup>b</sup>	SS <sup>c</sup>	MS <sup>d</sup>	WVC	corr. WVC <sup>e</sup>	Δ WVC [%]	Λ	ΔΛ [%]
Maritime Clear Sky	0	0.03	43.8	48.2	8.0	12.5	13.5	-5.2	0.039	+30.0
Maritime Aerosol	0.21	0.03	20.0	53.2	26.8	14.0	15.5	+8.5	0.038	+26.6
Rural Clear Sky	0	0.10	69.1	22.8	8.1	13.3	14.5	+1.6	0.102	+2.0
Rural Aerosol	0.51	0.10	17.9	30.9	51.2	14.6	16.3	+14.1	0.112	+12.0

<sup>a</sup> Aerosol Optical Depth<sup>b</sup> Percentage of ground reflected light at 592 nm<sup>c</sup> Percentage of single scattering at 592 nm<sup>d</sup> Percentage of multiple scattering at 592 nm<sup>e</sup>  $10^{22}$  molec/cm<sup>2</sup>

underestimates the WVC by 12%, in the case of maritime aerosol loading, and overestimates it by 18% in case of rural aerosol loading. The error in the retrieved albedo is generally of the order of 20%, except for the maritime case, where the aerosol layer above the surface alters the retrieved surface-albedo value significantly, resulting in a decrease in the retrieved WVC (see also Sect. 8).

For high WVC, the retrieval using SSP is affected by the method-related bias (see previous section). After application of the empirical correction (Table 2), the error due to multiple scattering for clear sky cases is less than 6% (maritime) and 2% (rural). The impact due to aerosol loading on the retrieved column is less than 9% (maritime) and 14% (rural), after correction for the method-related bias.

## 8 Results for individual GOME passes

Figure 6a shows bias-corrected (Eq. 13) WVC results (stars) retrieved using SSP for the GOME pass on 23 October 1998, which started from 80° S, 90° W and finished at 80° N, 150° W, passing over the North American continent. In total, 476 nadir-view ground-pixel measurements are used (center pixels only). For comparison, we show WVC values retrieved using the OACS method (using the same WV absorption band) for identical GOME ground pixels (solid red line), together with co-located ECMWF values (solid green line).

The cloud-cover fraction for each individual ground pixel is shown in Fig. 6b and was taken as reported by the GOME Data Processor (GDP) level-2 data (Balzer and Loyola, 1996; DLR, 1999). This fraction is derived by the initial cloud-fitting algorithm (ICFA) (Kutz et al., 1994) with the cloud-top pressure being taken from the International Satellite Cloud Climatology project (ISCCP) (Rossow et al., 1991). In

general, the SSP values compare well with the corresponding ECMWF values, except in situations where the cloud cover > 10% of the GOME ground pixel. This may be partly due to an enhanced photon path length which occurs in the cloud layer as a consequence of the increased amount of multiple scattering events. The extent of multiple scattering depends on the optical depth of the cloud, which is not provided in the ICFA cloud product, meaning that a rigorous description of the scattering events is difficult to implement. For this reason, and in the interests of computational speed, the SSP method uses a first-order polynomial to account for the contribution to the total reflectivity by multiple scattering (c.f. Eq. 11). In addition, depending on the optical depth, the presence of cloud also reduces the transmission of light through to the lower layers of the atmosphere, where most of the water vapor is located. These two effects may actually cancel each other in some instances, as for the first, the SSP model underestimates the absorption resulting in an overestimation of the WVC by the fitting procedure, whilst for the second, SSP overestimates the absorption resulting in an underestimation of the WVC. However, due to the complexity of cloud modeling, the fact that important cloud parameters are unknown for most of the retrieval situations affected by clouds, a quantitative analysis of this effect can not be given or would be way beyond the scope of the paper. We therefore restrict further analysis of the results predominantly to measurements of cloud-free situations (cloud fraction < 10%) as is reported by the GDP level-2 cloud-fraction value.

### 8.1 Retrieval error analysis

For each SSP-retrieved WVC we evaluate a total fitting error, including the instrument shot-noise contribution, as well as the systematic model and input-error contributions in the following way. After the  $i$ th iteration step the optimization method (Sect. 5 and Eq. 12) finds a solution very close to the measurement. In this regime the response of the retrieval problem is assumed to be close to linearity and we may define the linearized measurement  $y$  as

$$\mathbf{y} = \mathbf{R}^{\text{GOME}} - \mathbf{R}^{\text{SSP}}(N_i) + \mathbf{K}N_i, \quad (14)$$

where  $\mathbf{K}$  is the Jacobian matrix after the  $i$ th iteration step. The total retrieval error,  $\delta = N_i - N$ , may be divided into different contributions (Eriksson, 2000; Rodgers, 2000)

$$\begin{aligned} \delta &= (\mathbf{A} - \mathbf{I})(N - N_i) \\ &+ \mathbf{D}_y \mathbf{K}_b (\mathbf{b} - \mathbf{b}_i) \\ &+ \mathbf{D}_y \Delta \mathbf{r}(N, \mathbf{b}) \\ &+ \mathbf{D}_y \epsilon, \end{aligned} \quad (15)$$

where the rows of  $\mathbf{A}$  contain the averaging kernels,  $\mathbf{I}$  is the identity matrix,  $\mathbf{b}$  contains the model parameters,  $\mathbf{r}(N, \mathbf{b})$  is the difference between the used model and a full lbl solution of the equation of radiative transfer,  $\epsilon$  is the measure-

ment noise and  $\mathbf{D}_y$  is the contribution-function (or distribution) matrix with

$$\mathbf{D}_y = \frac{\partial N_i}{\partial \mathbf{y}}. \quad (16)$$

The different error contributions of Eq. (15) are called smoothing error, model-parameter error, forward-model error and retrieval noise. For the purpose of evaluating the smoothing-error contribution the true profile  $N$  is needed. This contribution is therefore difficult to evaluate for all the retrievals. Here, we therefore neglect this contribution in the calculations of the individual total retrieval errors, and refer instead to the validation of the results by other, independent measurements. The mean error in the parameters ( $\mathbf{b} - \mathbf{b}_i$ ) used in the model, i.e. errors in the temperature and pressure profiles, and also the density values of additional absorbers like  $\text{O}_3$  and  $(\text{O}_2)_2$ , is estimated to be a total of 5% (dominated by the expected error in the ozone density of the same order Burrows et al., 1999). The error in the forward model  $\Delta \mathbf{r}(N, \mathbf{b})$  is calculated from differences between SSP and DAM (see Fig. 3) for WVC of  $7.72 \times 10^{22}$  molec/cm<sup>2</sup> and  $1.43 \times 10^{23}$  molec/cm<sup>2</sup> and for a surface albedo of 0.03 and 0.1 (see Tables 1 and 2). Different  $\Delta \mathbf{r}(N, \mathbf{b})$  can be used in the retrieval calculations assuming a clear-sky situation or an aerosol-loaded situation as shown in Tables 1 and 2. The error of the bias adjustment (Eq. 13) is added to this results. The retrieval noise stems from the instruments shot-noise contribution, which is provided for each individual measurement by the GOME GDP data product. This contribution  $\epsilon$  is usually significantly smaller than 1% of the transmittance value (DLR, 1999).

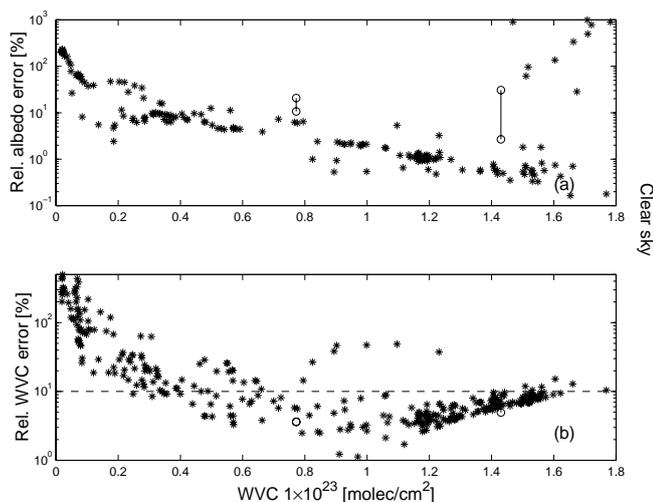
The distribution matrix of the individual retrieval is evaluated in the following way. First a singular value decomposition (SVD) of the Jacobian  $\mathbf{K}$ ,

$$\mathbf{K} = \mathbf{U} \Sigma \mathbf{V}, \quad (17)$$

is performed. Here the orthonormal columns of  $\mathbf{U}$  and  $\mathbf{V}$  form a basis of the measurement and solution space. The diagonal matrix  $\Sigma$  contains the singular values  $\sigma_f$  ( $f = 1, \dots, k$ ), which weight the individual basis vector contributions with  $k$  being the total number of independent variables. It is now possible to construct a parametric plot of  $\|N_i\|$  over  $\|\mathbf{K}N_i - \mathbf{y}\|$  parameterized by a parameter  $\gamma$  (L-curve), which exhibits a corner where both  $\|N_i\|$  and  $\|\mathbf{K}N_i - \mathbf{y}\|$  are minimal for a specific  $\gamma_c$  (Hansen & O'Leary, 1994). This corner value  $\gamma_c$  can be evaluated analytically. From this optimal corner value a filter matrix  $\Phi$  can be constructed by

$$\Phi = \text{diag} \left( \frac{\sigma_f^2}{\sigma_f^2 + \gamma_c^2} \right), \quad (18)$$

which, when applied to the solution matrix  $\mathbf{V}$ , filters out the basis vectors for which  $\sigma_f < \gamma_c$  and leaves the rest intact. In other words, it filters out those contributions of the solution space, which effect the norm of the retrieval problem



**Fig. 7.** Relative retrieval errors for the retrieved surface albedo (a) and the retrieved WVC (b) from the GOME measurements used in Fig. 10 and 11. The errors are calculated using the Jacobian of the last iteration step of the optimization method (Eq. 15). In order to evaluate the distribution matrix of the problem a SVD of the Jacobian together with a parametric L-curve scheme are used (for details see text). Errors include contributions of the shot noise of the instrument, the error introduced by the bias correction of the SSP values (see Fig. 5) and the impact of multiple scattering in clear sky situations as evaluated from SSP fits to the DAM model results (open circles; cf. Table 1 and 2).

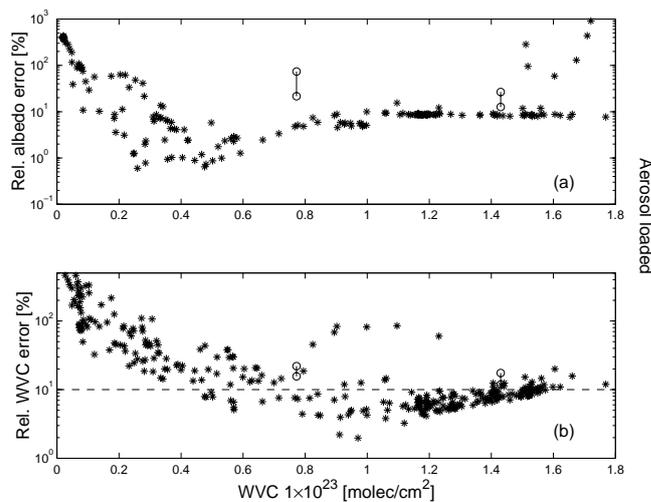
$\|\mathbf{KN}_i - \mathbf{y}\|$  less than the error between the model and the measurement. The averaging kernel matrix can then be written as

$$\mathbf{A} = \mathbf{V}\Phi\mathbf{V}. \quad (19)$$

From the averaging kernel matrix, the distribution matrix can be calculated by  $\mathbf{D}_y = \mathbf{AK}^{-1}$ . The individual retrieval errors on the WVC can now be calculated using Eq. (15). Also the error on the retrieved surface albedo can be calculated following the same procedure.

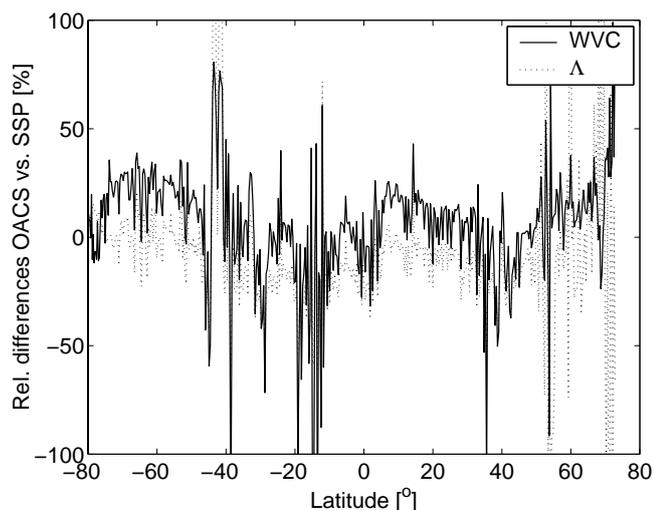
Figure 6a shows the error on the retrieved SSP WVC for most of the retrieved values of the GOME pass. Different values of  $\Delta r(\mathbf{N}, \mathbf{b})$  are used for each error depending on the value of the retrieved WVC and surface albedo. Furthermore, and in all cases shown in Fig. 6a, an aerosol-loaded atmosphere is assumed either for an ocean or for a rural scenario, depending on if the retrieved surface albedo is below or above 0.1. In 1.5% of the cases the retrieved values are rejected because the errors were bigger than  $1 \times 10^{23}$  molec/cm<sup>2</sup>. In 19% of the cases the calculated error was rejected and is not shown because an optimal corner value using the parametric L-curve could not be found.

Figures 7 and 8 show the calculated total retrieval errors using Eq. 15 for both the retrieved surface albedo and the retrieved WVC and for the single GOME pass shown in Fig. 6, plus an additional GOME pass on 23 October 1998 from



**Fig. 8.** Same as Fig. 7 but for scenarios with various aerosol loadings, for which the contribution of this systematic error is evaluated from SSP fits to the DAM model results (open circles; cf. Table 1 and 2).

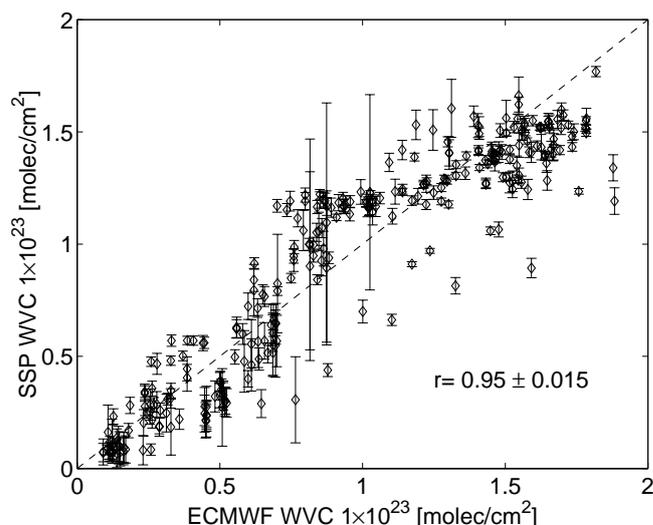
17:52 to 18:32 UTC between 90° W and 150° W longitude. The errors on the retrieved WVC decrease from more than 100% (for WVC below  $1 \times 10^{22}$  molec/cm<sup>2</sup>) to errors below 10% (for medium WVC around  $1 \times 10^{22}$  molec/cm<sup>2</sup>), with only a very few exceptions. Similarly, the error on the retrieved surface albedo decreases below 10% for medium WVC (Fig. 7a and b). For higher WVC columns errors increase slightly for the retrieved WVC, which we relate to the contribution of the error on the bias-correction (Fig. 5). Most of the errors on the retrieved surface albedo get even smaller for high WVC (Fig. 7a), with some exceptions leading to an increasing spread in the values for high WVC. This effect is also seen in the retrieved differences between the surface albedo used for the DAM spectra and the SSP retrieved values from these spectra (Table 1 and 2, and open circles in Fig. 7a). For the retrieved WVC these expected errors (open circles in Fig. 7b) are close to the individual retrieval errors. The impact of an aerosol-loaded atmospheric scenario is (in the case of the individual error on the retrieved WVC) small compared to the error-value dependence on the retrieved mean value. Aerosol loading (Fig. 8) leads to an overall shift of the WVC retrieval error to higher values of, on average, 35% of the clear-sky error value (Fig. 8b). In contrast, the error on the retrieved surface albedo stays at around 10% for an aerosol-loaded atmosphere going towards higher WVC (Fig. 8b). We relate this to an increased error in the modeling of the multiple-scattering contribution in the case of aerosol-loaded scenarios, which also effects the albedo retrieval. Fig. 7 and 8 show a general correlation between the accuracy of the retrieved WVC and the accuracy of the retrieved surface albedo. For increasing WVC there is an overall tendency for decreasing errors on both retrieved



**Fig. 9.** Relative differences between SSP and OACS for retrieved values of WVC ( $[\text{OACS}-\text{SSP}]/\text{OACS}$ ; solid line) and surface albedo  $\Lambda$  ( $[\text{SSP}-\text{OACS}]/\text{OACS}$ ; dotted line). An underestimation of the surface albedo leads to an overestimation in the retrieved WVC. The relative differences result form the same GOME track as shown in Fig. 6 including all ground pixel.

values. This correlation breaks down only in cases of a significant impact of multiple scattering or by an increased contribution of the bias-correction error.

The correlation of the retrieved surface albedo with the retrieved WVC may be explained as follows: A critical parameter which governs the accurate retrieval of WVC is the shape of the resulting subcolumn profile retrieved by SSP. Due to the effect of the altitude dependence of both  $p$  and  $T$  on the spectral structure of WV, the effective absorption is governed by the specific atmospheric paths through which the photons travel. The retrieval of an unrealistic shape of the retrieved subcolumn profile may, therefore, unrealistically alter the weights of various atmospheric paths, i.e. the relative contribution between ground-reflected, singly- and multiply-scattered light. This yields an error in the retrieved surface albedo, which weights the direct, ground-reflected, light-path contribution (Eq. 6) or an error in the retrieved broadband multiple-scattering contribution. This error propagates into the final retrieved column (Fig. 7 and 8). From Fig. 9 we can see that the differences between the WVC values retrieved by OACS and SSP, and the differences between the surface albedo values retrieved by both methods, are very well anti-correlated meaning that, in comparison to OACS, a larger retrieved surface albedo corresponds to a lower retrieved WVC (note the difference in sign for the calculation of the difference values; caption of Fig. 9). Introducing upper fit constraints assists both methods in finding the accurate profile shapes, but the OACS method is, due to its more accurate treatment of the representation of the spectral structure at each altitude level, superior to the SSP method, resulting

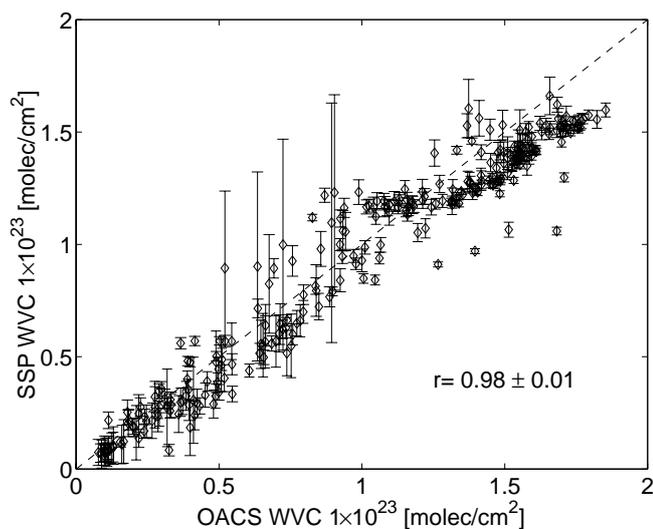


**Fig. 10.** Scatter plot between ECMWF total WVC and SSP WVC retrieval results for GOME measurements with a GDP level-2 data reported cloud fraction lower than 10%. The GOME measurements are performed between  $70^\circ$  S and  $70^\circ$  N in latitude: on 30 September 1999 from 0:09 to 1:05 UTC between  $120^\circ$  E and  $180^\circ$  E longitude and on 23 October 1998 from 17:52 to 18:32 UTC between  $90^\circ$  W and  $150^\circ$  W longitude. Co-located ECMWF data is given for 00:00 UTC and 18:00 UTC, respectively. Errorbars are calculated from retrieval Jacobians as for Fig. 6 using a SVD scheme (for details see text). Pearson's  $r$  correlation number is calculated to be  $0.95 \pm 0.015$ .

in higher sensitivity and better agreement with the independent data-set from ECMWF. However, a big benefit of SSP is that it is about 4 to 5 times faster than the OACS method even though it uses a higher sampling resolution than the one chosen for OACS in Lang et al. (2002).

## 8.2 Comparison with ECMWF and OACS

Figure 10 shows a scatter plot between SSP retrieved WVC from GOME and co-located WVC data reported by ECMWF for both GOME passes discussed in this section. The GOME measurements coincide with ECMWF data within a 1 hour time window. We compare 420 GOME measurements in total with a GDP level-2 data reported cloud fraction of less than 10%. Pearson's  $r$  correlation number is evaluated to  $0.95 \pm 0.015$ . The error on Pearson's  $r$  is calculated using a transformation to Fisher's  $Z$ -distribution, for which the confidence interval can be calculated on a 99% confidence level as  $Z$  can be assumed to be normally distributed. From the back-transformation the error on  $r$  is obtained. Fig. 11 shows the same SSP results but now in comparison to OACS retrieval results using the same GOME measurements. As expected, using exactly the same measurements and a similar radiative-transfer scheme (see Sect. 3), the correlation is better with a Pearson's  $r$  number of  $0.98 \pm 0.01$ .



**Fig. 11.** Same SSP WVC retrieval results as shown in Fig. 10 but in correlation with OACS WVC results. Here, Pearson's  $r$  correlation number is calculated to be  $0.98 \pm 0.01$ .

## 9 Global retrieval results

Finally we present the results of a global WVC retrieval utilizing 47 000 ground pixels (using east, center and west pixels) from GOME between  $70^\circ$  S and  $70^\circ$  N latitude on 22–24 October 1998 using the SSP method. The globally bias-corrected results (Eq. 13) are shown in the global plot in Fig. 12a. For comparison we show the global WVC, as given by ECMWF, on 23 October 1998, 18:00 UTC (Fig. 12b). Fig. 12c shows the global WVC as retrieved by SSP for cloud-free pixels, i.e. where the GDP reported cloud-fraction is lower than 10%.

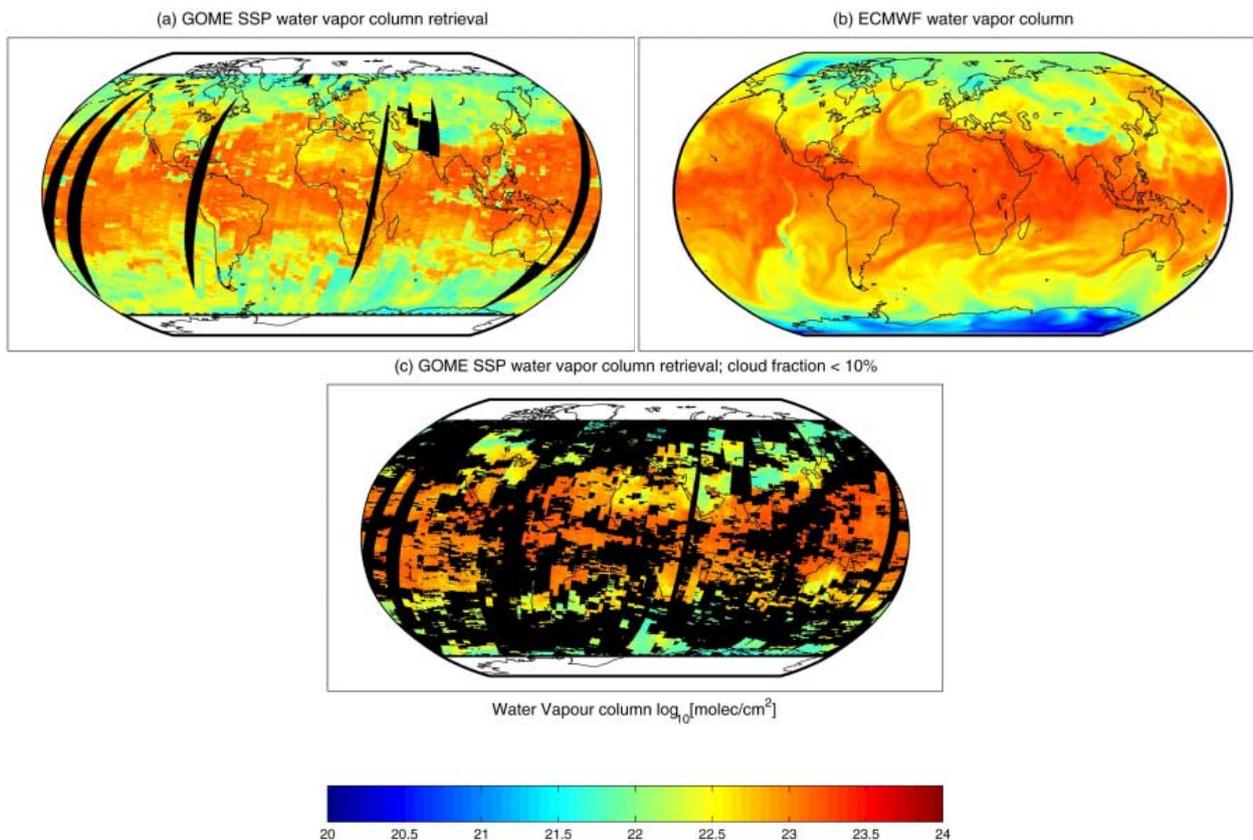
The general latitudinal dependence of WV concentration, with high WVC in the equatorial regions and low WVC towards the poles, is well represented by the SSP retrieval results. Individual features common to both the ECMWF model results and the SSP results may also be identified as, for example, the low WVC above north-west China and the Gobi desert as well as in the United States at this time above western North Carolina. A similar transition from low to high WVC can be observed from the southern part of the United States to Mexico and the Caribbean Sea. However, GOME-SSP WVC results show significantly lower values compared to ECMWF above the Sahara, Saudi Arabia and the Andes around  $40^\circ$  S latitude. SSP retrieves high WVC above the Philippine Sea north of New Guinea where the ECMWF model reports only medium WV concentrations.

## 10 Discussion

Differences in the surface albedo retrieved by the OACS sampling technique and SSP, with both exploiting the same WV

absorption band, lead to small overall differences between the WVC retrieved by both methods. In addition, the fitting error analysis also shows that both variables are somewhat correlated. The difficulties in the retrieval of the correct surface albedo is a general problem for nadir remote-sensing techniques, especially for cases of high aerosol optical density. This is due to the uncertainty in the contribution to the backscattered light introduced by aerosol layers, with the aerosol profile (i.e. the distribution with respect to height) being a critical parameter. Moreover, surface albedo is also dependent on wavelength, geolocations and season. In our wavelength region of interest, the first-order polynomial used here is a good approach for the wavelength dependence of the albedo for most cases (c.f. Koelemeijer et al. (1997)). We find that the accuracy of the retrieved surface-albedo values and their gradient is linked to both an accurately retrieved WV profile shape and the contribution of multiply scattered photons, which for SSP is approximated by a first-order polynomial. For this reason, we introduced three different step functions, depending on geolocation, as upper constraints for the retrieved profile, in order to prevent the optimization method from finding a local fit minimum by setting the ground albedo to zero and unrealistically increasing the scattering components.

An additional concern for any WV-retrieval approach is the accuracy and completeness of the cross-section database used for the construction of the  $w$ - $S$  look-up table. From recent measurement studies of water vapor absorption bands other than the one used in this study, Learner et al. (2000) recently showed discrepancies ranging from about 100% for small lines to about 20% for strong lines for the main water vapor absorption bands in the region between 685 and 1110 nm. They also found systematic differences in various bands ranging from 6% to 33%. Even though measurements by Learner et al. (2000) in the wavelength range between 585 and 600 nm have not been studied, large differences in some of the water absorption bands between 685 and 1110 nm suggest the presence of potentially large uncertainties in the reference cross-sections of the HITRAN'96 database used in this study. In addition, the presence of many weak absorption lines, not accounted for in the HITRAN'96 database, may contribute to an additional background absorption, which may affect the retrieval of both surface albedo and broadband multiple-scattering contribution and, therefore, also the retrieval of WVC. The water-vapor line-parameter data for our band region given in the new HITRAN database (HITRAN-2000) is very similar to that in HITRAN'96. No additional lines are added but line intensities differ slightly due to corrections described by Giver et al. (2000). The most recent update (HITRAN-2001) does not differ from HITRAN-2000 concerning the water-vapor absorption within our band. A recent study by Veihelmann et al. (2002) has shown that the differences between HITRAN'96 and HITRAN-2000 are negligible when compared to high-resolution fourier-transform sun-spectrometer measurements



**Fig. 12.** Global total water vapor column plots for all retrieved GOME ground pixel using SSP (a). The measurements were performed from 22–24 October 1998. The right panel (b) shows global WVC as given by ECMWF at 23 October, 1998, 18:00 UTC. The lower panel (c) shows SSP columns for GOME observations where the cloud fraction reported by GOME level-2 data is lower than 10%. Data gaps for GOME retrievals shown in (a) are due to calibration periods or data processing failures. At 24 October from 14:00 UTC GOME was in Narrow Swath Mode until 11:00 UTC the next day. For these tracks the data was extrapolated to the standard swath width. Data gaps in (c) are predominantly due to the removal of cloudy ground pixel.

within our absorption band.

The contribution of aerosols and clouds to the error in the retrieved column, as given in Sect. 7, can only be estimated. In the case of aerosols, this is due to their high variability in optical density, concentration profile and scattering probabilities, all of which depend on geolocation and season. In the case of clouds, the uncertainties arise from the high variability with respect to cloud-top height, form and density of clouds. Furthermore, the scatter between SSP-retrieved WVC values and those given by ECMWF, as visible in the scatter plot (Fig. 10), may also be affected by systematic errors in the cloud-fraction value. We call a ground pixel cloud free when the reported cloud-fraction value is  $< 10\%$ . However, differences in the reported cloud-fraction by GDP level-2 data with respect to the co-located Along Track Scanning Radiometer-2 (ATSR-2) can be as much as 18% (Koelemeijer et al., 1999).

The individual fitting errors as analyzed in Sect. 8.1 are varying between more than 100% for low and 10% for

medium and high WVC. They are similar to, or smaller than, the errors expected from SSP WVC retrieval from DAM forward model (Tables 1 and 2 and Fig. 7 and 8). The contribution of the shot noise of the instrument to the total fitting error (Eq. 15) is very small when compared to the systematic error contribution. Using an SVD and L-curve scheme to evaluate the distribution matrix is, therefore, necessary due to error contributions which cannot be expected to be normally distributed (Eriksson, 2000).

## 11 Conclusion

In this study the Spectral Structure Parameterization Technique (SSP) has been implemented and tested for retrieval of total WVC from GOME measurements utilizing a radiative-transfer scheme, which includes molecular clear-sky direct-path absorption as well as molecular clear-sky single scattering, together with a broadband approximation of higher orders of scattering. The radiative-transfer scheme was pre-

viously used for retrieval of total WVC using the Optical Absorption Coefficient Spectroscopy. SSP is less accurate and less sensitive than the OACS retrieval method but significantly faster and easier to implement. Retrieval of WVC using SSP for one ground pixel takes less than 30 seconds on a Pentium III, 800 MHz, using an uncompiled MATLAB retrieval code, and requires less than 128 MB internal memory.

Whereas OACS utilizes a spectral-sampling technique based on cross-section probability-density distribution functions, SSP utilizes average line parameters within the sampling region, and is, therefore, conceptually closer to band models. From retrieval studies of synthetic spectra we conclude that the sensitivity of SSP is better than 10% and the accuracy is about 5% for WVC lower than  $8 \times 10^{22}$  molec/cm<sup>2</sup>. For WVC above this limit, a systematic model bias of up to 20% due to the assumption of a dominant absorbing layer applied to the modeling of nonhomogeneous paths (Sect. 3 and Eq. 8). The model bias has been corrected for in the current study by applying an empirical second-order polynomial (Sect. 6).

We studied the impact of aerosols and the differential contribution of multiple scattering for four different scenarios utilizing synthetic spectra from a doubling-adding radiative-transfer method. For medium WVC, for which the bias-correction is small, the multiple-scattering impact on the SSP retrieved WVC is below 1%. Neglecting aerosol results in a retrieval error of less than 13% (maritime) and 19% (rural). For bias-corrected high WVC (Table 2) under clear-sky conditions, the error due to multiple scattering is less than 6% (maritime) and 2% (rural). Here, the impact due to aerosols on the retrieved column is less than -9% (maritime) and +14% (rural), after correction for the method-related bias.

The scatter plot comparison between SSP and ECMWF (Fig. 10), as well as the single GOME pass comparison (Fig. 6), are correlated in time and geolocation. The scatter plot shows good correlation between the two data-sets of correlated cloud-free ocean pixel with a Pearson's  $r$  number of 0.95. However, in some cases, the scatter can be as high as 50%, which points to local differences between the two data products. ECMWF WV profiles are provided without any error estimate. ECMWF is known to underestimate WVC values given by another independent remote sensing instrument, the Special Sounder Microwave Imager (SSM/I), by 30–50% in the tropics and to overestimate WVC in the sub-tropics by as much as 60% depending on season and geolocation (Vesperini et al., 1998). ECMWF values over the ocean are predominantly based on data from the TIROS N operational vertical sounder (TOVS). The standard deviation between TOVS WVC and radiosonde data in the 1000–850 hPa layer is known to be on the order of 20% and less than 40% in the layer between 500 and 300 hPa (Chaboureaud et al., 1998). We therefore conclude that the maximum scatter of 50% we find by comparing ECMWF with SSP WVC is within the range of differences of ECMWF data to other datasets. Scatter plot comparisons between SSP and OACS

(Fig. 11) show even higher correlation (Pearson's  $r=0.98$ ). This is due to the fact that both retrieval methods use exactly the same measurements and similar radiative-transfer schemes.

The error on the individual fits is estimated using an SVD of the Jacobian matrix and by applying an L-curve scheme to evaluate the averaging kernel matrix and the distribution matrix. The impact of most systematic errors and the shot-noise contribution are considered. The retrieval errors decrease from more than 100% for very low WVC to less than 10% for mid and high columns for an aerosol-loaded atmosphere. The error of both the retrieved surface-albedo value and the retrieved WVC show some correlation. We refer this correlation to the assumptions made for the modeling of multiple scattering and the uncertainties in the retrieved WV-profile shape. The estimated retrieval errors are similar or smaller than the errors expected from SSP retrievals from DAM modeled spectra. Only retrievals with a GDP reported cloud fraction of less than 10% are considered during the error analysis.

A global comparison between SSP-retrieved WVC and data from ECMWF shows good general agreement in the latitudinal dependence of the WV concentration. However, significant regional differences, for example, over the Sahara and the Philippine Sea are found. The larger spatial coverage of the GOME measurements over sea and land with respect to ECMWF, where the density of ground measurements is low, is clearly an advantage of this kind of SSP retrieval from GOME. However, ECMWF data is provided for a specific time and day, whereas global GOME retrieval results, as shown in Fig. 12, are collected over three days. Wide parts of the differences in the global data set may, therefore, also be related to changes in the global WV distribution over the analysis period.

## 12 Outlook

In general, SSP is well suited for fast clear-sky retrieval of narrow-band absorbers from nadir viewing satellite instruments like GOME. The accuracy in the retrieval of the surface albedo and the impact of aerosols and multiple scattering limits the accuracy of the retrieved WVC to some extent. The latter may be improved by introducing surface albedo from global databases once available and tested, although the typical footprint of a satellite is so large that a scaled value to account for different surface types within a typical ground pixel is possibly the way forward. The current implementation of SSP in a direct path and single-scattering radiative-transfer scheme provides the possibility for implementation of specific aerosol optical properties. The treatment of clouds for such nadir-viewing measurements in the visible is, as yet, an unsolved problem. SSP provides WVC (and potentially profile) retrieval from instruments like GOME and SCIAMACHY, which were not originally intended for WV

retrieval. SSP retrieved WV values may be used to refine other retrievals from such satellite data because WV “contaminates” the entire spectrum. In addition, SSP is also suitable for the retrieval of spectrally-overlapping narrowband absorptions from different species. This situation occurs in the IR regions around  $2\ \mu\text{m}$  covered by the SCIAMACHY instrument. Multi-species retrieval may be performed by introducing an additional, averaged line parameter, which accounts for the degree of overlap between the different absorbers within a specific spectral sampling width.

*Acknowledgements.* We would like to thank Ilse Aben (SRON) for useful discussions and the reading of the manuscript. We also would like to thank Jochen Landgraf (SRON) for providing DAM model spectra. We like to thank P. F. J. van Velthoven (KNMI) for assistance with ECMWF data. ESA is acknowledged for providing GOME data (ESA 1995–1999) processed by DFD/DLR. This work is part of the research program of the “Stichting voor Fundamenteel Onderzoek der Materie (FOM)”, which is financially supported by the “Nederlandse organisatie voor Wetenschappelijke Onderzoek (NWO)” and is supported by SRON through project grants EO-023 and EO-046.

## References

- Balzer, W. and Loyola, D.: Product Specification Document of the GOME Data Processor, Technical Document ER-PS-DLR-60-0016, Deutsche Forschungsanstalt für Luft- und Raumfahrt, pp. 21, 1996.
- Buchwitz, M., Rozanov, V. V., and Burrows, J. P.: A near-infrared optimized DOAS method for the fast global retrieval of atmospheric  $\text{CH}_4$ ,  $\text{CO}$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{O}$ , and  $\text{N}_2\text{O}$  total column amounts from SCIAMACHY Envisat-1 nadir radiances, *J. Geophys. Res.*, 105, 15 231–15 245, 2000.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißmayer, A., Richter, A., deBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., and Eisinger, M.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results, *J. Atmos. Sci.*, 56, 151–175, 1999.
- Byrd, R. H., Schnabel, R. B., and Shultz, G. A.: Approximate Solution of the Trust Region Problem by Minimization over Two-Dimensional Subspaces, *Mathematical Programming*, 40, 247–263, 1988.
- Casadio S., Zehner, C., Piscane, G., and Putz, E.: Empirical Retrieval of Atmospheric Air Mass Factor (ERA) for the Measurement of Water Vapor Vertical Content using GOME Data, *Geophys. Res. Lett.*, 27, 1483–1486, 2000.
- Chaboureaud, J.-P., Chédin, A., and Scott, N. A.: Remote Sensing of the vertical distribution of atmospheric water vapor from the TOVS observations: Method and validation, *J. Geophys. Res.*, 103, 8743–8752, 1998.
- DLR: GOME data Processor Extraction Software User’s Manual, Doc.No.: ER-SUM-DLR-GO-0045, DLR/DFD, Oberpfaffenhofen, Germany, 1999.
- Eriksson, P., Analysis and comparison of two linear regularization methods for passive atmospheric observations, *J. Geophys. Res.*, 105, D14, 18 157–18 167, 2000.
- ESA: The Global Ozone Monitoring Experiment Users Manual, (Ed) Bednarz, F., ESA Publication SP-1182, ESA Publication Division, ESTEC, Noordwijk, The Netherlands, 1995.
- Giver, L.P., Chackerian, C. and Varanasi, P.: Visible and near-infrared ( $\text{H}_2^{16}\text{O}$ ) line intensity corrections for HITRAN-96, *J. Quant. Spect. & Rad. Transfer*, 66, 101–105, 2000.
- de Haan, J., Bosma, P., and Hovenier, J.: The adding method for multiple scattering calculations of polarized light, *Astron. Astrophys.*, 181, 371–391, 1987.
- Hansen, P. C., and O’Leary, D. P.: The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, 14(6), 1487–1503, 1993.
- Hedin, A. E.: Extension of the MSIS Thermosphere Model into the Middle and Lower Atmosphere, *J. Geophys. Res.*, 96, 1159–1172, 1991.
- Kato, S., Ackerman, T. P., Mather, J. H., and Clothiaux, E. E.: The  $k$ -distribution method and correlated- $k$  approximation for a shortwave radiative transfer model, *J. Quant. Spect. & Rad. Transfer*, 62, 109–121, 1999.
- Koelemeijer, R. B. A., Stammes, P., and Stam, D.: Spectral Surface Albedo Derived From GOME Data, Proc. 3rd ERS Symp. on Space at the service of our Environment, Florence, Italy, 17–21 March 1997, ESA SP-414, 3 Vols, 1997.
- Koelemeijer, R. B. A., and Stammes, P.: Validation of GOME cloud cover fraction relevant for accurate ozone column retrieval, *J. Geophys. Res.*, 104, 18,801–18,814, 1999.
- Kutz, A., and Chance, K.V.: Analysis of Cloud Top height and cloud Coverage from Satellites Using the  $\text{O}_2$  A and B Bands, *J. Geophys. Res.*, 99, D7, 14 481–14 491, 1994.
- Lacis, A. A. and Oinas, V.: A description of the correlated  $k$ -distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres, *J. Geophys. Res.*, 96, D5, 9027–9063, 1991.
- Lang, R., Maurellis, A. N., van der Zande, W. J., Aben, I., Landgraf, J., and Ubachs, W.: Forward Modeling and Retrieval of Water Vapor from GOME: Treatment of Narrow Band Absorption Spectra, *J. Geophys. Res.*, 107, D16, 10.1029/2001JD001453, 2002.
- Learner, R., Schermaul, R., Tennyson, J., Zobov, N., Ballard, J., Newnham, D., and Wickett, M.: Measurement of  $\text{H}_2\text{O}$  Absorption Cross-Sections for the Exploitation of GOME data, ESTEC Contract No 13312/9/NL/SF, Final Presentation, 2000.
- Maurellis, A. N., Lang, R., van der Zande, W. J., Ubachs, U., and Aben, I.: Precipitable Water Column Retrieval from GOME Data, *Geophys. Res. Lett.*, 27, 903–906, 2000a.
- Maurellis, A. N., Lang, R., and van der Zande, W. J.: A New DOAS Parameterization for Retrieval of Trace Gases with Highly-Structured Absorption Spectra, *Geophys. Res. Lett.*, 27, 4069–4072, 2000b.
- Humlicek, J.: Optimized computation of the Voigt and complex probability function, *J. Quant. Spect. Rad. Transfer*, 27, 437, 1982.
- Noël, S., Buchwitz, M., Bovensmann, H., Hoogen, R., and Burrows, J. P.: Atmospheric Water Vapor Amounts Retrieved from GOME Satellite Data, *Geophys. Res. Lett.*, 26, 1841–1844, 1999.
- Rossow, W. B., and Schiffer, R. A.: ISCCP Cloud Data Products, *Bull. Am. Met. Soc.*, 72, 2–20, 1991.
- Rothman, L. S., Rinsland, C. P., Goldman, A., Massie, S. T., Ed-

- wards, D. P., Flaud, J.-M., Perrin, A., Camy-Peyret, C., Dana, V., Mandin, J.-Y., Schroeder, J., McCann, A., Gamache, R. R., Wattson, R. B., Yoshino, K., Chance, K. V., Jucks, K. W., Brown, L. R., Nemtchino, V., and Varanasi, P.: The HITRAN Molecular Spectroscopic Database and HAWKS (HITRAN Atmospheric Workstation): 1996 Edition, *J. Quant. Spect. Rad. Transfer*, 60, 665–710, 1998.
- Rodgers, C., *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sc., River Edge, N. J., 2000.
- Shettle, E. P. and Fenn, R. W.: Models for aerosols of the lower atmosphere and the effects of the humidity variations on their optical properties, Air Force Geophys. Lab. (OP), *Envir. Res. Pap.* 676, AFGL-TR-79-0214, Hanscom, Massachusetts (technical report), 1979.
- Veihelmann, B., Lang, R., Smith, K. M., Newnham, D. A., and van der Zande, W. J.: Evaluation of spectroscopic databases of water vapor between 585 and 600 nm, *Geophys. Res. Lett.*, 29, 15, 10.1029/2002GL015330, 2002.
- Vesperini, M.: Humidity in the ECMWF model: Monitoring of operational analyses and forecasts using SSM/I observations, *Q. J. R. Meteorol. Soc.*, 124, 1313–1327, 1998.