Global one-step inversion of satellite occultation measurements: A practical method

Filip Vanhellemont, Didier Fussen, and Christine Bingen
Belgian Institute for Space Aeronomy, Brussels, Belgium

Received 21 September 2003; revised 14 March 2004; accepted 23 March 2004; published 12 May 2004.

[1] The retrieval of gas densities and aerosol extinction coefficients from satellite occultation measurements is usually handled in two steps: the separation of species (spectral inversion) and the separation of the spatial contributions of different atmospheric layers (spatial inversion). The reason for these sequential steps and their associated approximations is that a single-step inversion is computationally very demanding. In this paper we describe a global one-step method that can be applied to satellite data using reasonable computer resources. *INDEX TERMS:* 0340 Atmospheric Composition and Structure: Middle atmosphere—composition and chemistry; 0394 Atmospheric Composition and Structure: Instruments and techniques; 1640 Global Change: Remote sensing; 1694 Global Change: Instruments and techniques; 3360 Meteorology and Atmospheric Dynamics: Remote sensing; *KEYWORDS:* remote sensing, occultation experiments, ozone

Citation: Vanhellemont, F., D. Fussen, and C. Bingen (2004), Global one-step inversion of satellite occultation measurements: A practical method, *J. Geophys. Res.*, *109*, D09306, doi:10.1029/2003JD004168.

1. Introduction

[2] Occultation experiments form a distinct class within the different types of methods to perform atmospheric constituent soundings. In essence, the method consists of measuring the atmospheric transmission of light as a function of time and wavelength while the host satellite is orbiting a planet (Earth). During the setting or rising of the light source with respect to the horizon the optical path crosses different atmospheric layers. After the measurement the data can, in principle, be used to obtain information on atmospheric constituent abundances: The wavelength variation of the transmittance delivers information on the different constituents (because of the characteristic spectral extinction signature of every constituent), while the time variation delivers spatial (altitudinal) information. The combined spectral/temporal variation of the measured signal can be processed to retrieve altitude profiles of gas number densities and aerosol extinction coefficients.

[3] In principle, every light source can be used (a city, a laser beam, etc.). However, if extraterrestrial sources are used, the occultation measurement becomes self-calibrating by definition since the light source irradiance can be measured unattenuated when the optical path does not cross the atmosphere. The choice of light source depends on a number of criteria. The Sun is usually chosen to achieve a high signal-to-noise ratio; the Moon and planets, on the other hand, allow measurement of nighttime species with reasonable sensitivity. If a good global geolocation coverage is needed, stars and planets are a suitable choice.

0148-0227/04/2003JD004168

[4] At present, the occultation method has become a standard technique in atmospheric sounding. Among the extensive list of previously and presently operational experiments we can mention the Global Ozone Monitoring by Occultation of Stars (GOMOS) experiment [*Bertaux et al.*, 1991], the Stratospheric Aerosol and Gas Experiment (SAGE) series [*Chu et al.*, 1989], Atmospheric Trace Molecule Spectroscopy (ATMOS) [*Gunson et al.*, 1996], Polar Ozone and Aerosol Measurement (POAM) [*Lucke et al.*, 1999], the Midcourse Space Experiment (MSX) [*Yee et al.*, 2002], the Halogen Occultation Experiment (HALOE) [*Russell et al.*, 1993], the Occultation Radiometer (ORA) [*Fussen et al.*, 2001a], and the Atmospheric Chemistry Experiment/Measurement of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation (ACE/MAESTRO) [*McElroy et al.*, 2000].

[5] As already mentioned, altitude profiles of gas densities and aerosol extinction coefficients can be derived from the measured transmittance. Usually, by making some approximations in the physical model, the processing is performed in two steps: spectral inversion (separation of species contributions) and spatial inversion (separation of spatial contributions). The reason for these two distinct steps is that a one-step inversion is computationally expensive. However, it is possible to achieve a relatively efficient one-step inversion through the use of a specific method that will be described in this paper. The main advantages of a one-step inversion are that the typical two-step model approximations are avoided and that the spectral and spatial interdependency of all constituents is fully present in the model. Furthermore, the one-step formulation is very straightforward and elegant and can easily be implemented in an algorithm. For clarity: By "one-step" we mean the direct retrieval of the gas densities and aerosol extinctions

from the measurements without intervening steps; we are not referring to specific algorithm implementation details such as the number of iterations.

2. General Problem

2.1. Physical Model

[6] In the following derivation we will assume that our experiment consists of a UV-visible spectrometer with n_λ spectral channels. We take the light source to be point-like (such as a star or a small surface element of the Sun or Moon). At all times, the positions of the source and the orbiting satellite uniquely determine the path of the light ray emitted by the source and arriving at the satellite. The point on the ray that is closest to the surface of the Earth is defined as the tangent point, with an associated tangent altitude h_λ that depends on the wavelength λ because of chromatic refraction [*Edlen*, 1953]. Because of this unique determination of the light ray, the time evolution of the measured signal can equivalently be expressed as a function of the tangent altitude.

[7] In a realistic experiment some spectral smearing is always present because of a number of factors, such as the limited spectral resolution of the detector. Taking this in account and assuming that single-scattering conditions are fulfilled, the measured transmittance can be expressed as a spectral convolution of the Beer-Lambert law,

$$T(h_\lambda, \lambda) = \int_{\lambda'} F(\lambda, \lambda') \exp[-\tau(h_{\lambda'}, \lambda')] \mathrm{d}\lambda', \tag{1}$$

where $F(\lambda, \lambda')$ is an instrument function, properly normalized to unity over its spectral range, and $\tau(h_{\lambda'}, \lambda')$ is the slant path optical thickness. If $F(\lambda, \lambda')$ is narrow enough, this equation can be reduced within excellent approximation to the form

$$T(h_\lambda, \lambda) = \exp[-\tilde{\tau}(h_\lambda, \lambda)] \tag{2}$$

by convoluting the extinction cross sections in the model with the instrument function [*Kyrola et al.*, 1993] (in the case of broader instrument functions the problem becomes more complex). The slant path optical thickness is given by

$$\tilde{\tau}(h_\lambda, \lambda) = \int_{\text{source}}^{\text{sat}} \beta_{\text{tot}}(s_{h\lambda}, \lambda) \mathrm{d}s_{h\lambda}, \tag{3}$$

with $s_{h\lambda}$ a length coordinate along the optical path with tangent altitude h_λ and β_{tot} the total optical extinction coefficient.

[8] If we assume homogeneous spherical atmospheric layers, the problem is simplified to an integration along altitude. This integral can be discretized by dividing the entire altitude range in n_z discrete segments and by specifying an interpolation rule for the total extinction within one segment. In this way, we derive

$$\tilde{\tau}(h_\lambda, \lambda) = \sum_{i=1}^{n_z} \beta_{\text{tot}}(z_i, \lambda) g(h_\lambda, z_i), \tag{4}$$

with $g(h_\lambda, z_i)$ a factor with the dimension of length, representing the contribution or weight (hence the name "weighting function") of the altitude layer i to the entire extinction along the optical path.

[9] The total extinction equals the sum of the extinction by the different species,

$$\beta_{\text{tot}}(z_i, \lambda) = \beta_{\text{air}}(z_i, \lambda) + \cdots + \beta_{\text{aero}}(z_i, \lambda), \tag{5}$$

where the subscript aero refers to aerosols. For gases the extinction is simply the product of the extinction cross section σ times the number density N. In general, absorption cross sections depend on temperature and therefore on altitude. For example, the ozone extinction coefficient reads

$$\beta_{\text{O}_3}(z_i, \lambda) = \sigma_{\text{O}_3}(z_i, \lambda) N_{\text{O}_3}(z_i). \tag{6}$$

[10] Aerosols have to be treated differently than gases since their extinction spectra result from the convolution of extinction cross sections with a particle size distribution. A multitude of ways exists to express the extinction in a simple form, but it is usually sufficient to use a fairly smooth function, such as a low-order polynomial. The reason is that typical particle size distributions in the atmosphere are broad enough to significantly smooth out wavelength structures in the aerosol cross sections. Also, the number of aerosol parameters that is retrievable from UV-visible measurements is quite limited [see, e.g., *Fussen et al.*, 2001b; *Vanhellemont and Fussen*, 2002]. We therefore describe the spectrum by a quadratic polynomial, parametrized by the aerosol extinction at three different wavelengths:

$$\beta_{\text{aero}}(z_i, \lambda) = \sum_{j=1}^{3} q_j(\lambda) \beta_{\text{aero}}(z_i, \lambda_j) \tag{7}$$

with

$$q_j(\lambda) = \prod_{k \neq j} \frac{\lambda - \lambda_k}{\lambda_j - \lambda_k}. \tag{8}$$

[11] The full aerosol extinction spectrum at a given local altitude is thus expressed by three variables. The total number of constituent variables (gases and aerosols) at a given altitude is denoted as n_c.

2.2. Scaling

[12] The full transmittance model is completely specified by the expressions in section 2.1. The unknown parameters of the model (gas densities and aerosol extinction coefficients) generally span several orders of magnitude, not only in altitude but also from one constituent to another. To avoid numerical problems, some scaling is required. A common way to do this is by expressing the unknowns with respect to a climatological profile. For example, for air,

$$N_{\text{air}}(z_i) = \tilde{N}_{\text{air}}(z_i) x_{\text{air}}(z_i), \tag{9}$$

with $\tilde{N}_{\text{air}}$ taken from a climatological profile. In this way, equation (5) transforms to

$$\beta_{\text{tot}}(z_i, \lambda) = \tilde{\beta}_{\text{air}}(z_i, \lambda) x_{\text{air}}(z_i) + \cdots + \tilde{\beta}_{\text{aero}}(z_i, \lambda_3) x_{\text{aero},3}(z_i). \tag{10}$$

2.3. Matrix Representation

[13] If we use $\mathbf{T}_i$ as the notation for the $(n_\lambda \times 1)$ vector containing the transmission spectrum at the ith

tangent altitude and combine equations (4) and (10), we obtain

$$\mathbf{T}_i = \exp(-K_i\mathbf{x}). \tag{11}$$

The notation exp that we use here does not represent a vector exponential function but does represent an element-by-element exponentiation. The $(n_z n_c \times 1)$ vector $\mathbf{x}$ contains all the unknown parameters at all altitudes, while the $(n_\lambda \times n_z n_c)$ matrix K_i consists of elements that are products of weighting functions with climatological extinction values.

[14] All transmittance spectra can be stacked together in one total vector with dimensions $(n_\lambda n_h \times 1)$,

$$\mathbf{T}_{\text{tot}} = \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_{n_h} \end{bmatrix}, \tag{12}$$

and the associated $(n_\lambda n_h \times n_c n_z)$ total matrix reads

$$K_{\text{tot}} = \begin{bmatrix} K_1 \\ \vdots \\ K_{n_h} \end{bmatrix}. \tag{13}$$

[15] Although it is evident, it is important to notice that K_{tot} is not a block diagonal matrix, which means that the transmittance spectra at different tangent altitudes are physically related. The reason for this is, of course, that two different optical paths cross a certain number of common atmospheric layers.

[16] Combining equations (11), (12), and (13), the total forward model is

$$\mathbf{T}_{\text{tot}} = \exp(-K_{\text{tot}}\mathbf{x}). \tag{14}$$

2.4. Example

[17] The full dimension of the forward model can easily be illustrated. Let us say that our spectrometer has n_λ = 1000 spectral channels and that a full spectral acquisition is performed every 0.5 s. A typical value for an occultation duration is 30 s, so the total occultation measurement consists of 60 spectra at different tangent altitudes (n_h = 60). Suppose we want to retrieve n_c = 9 parameter profiles (for example, air, O_3, NO_2, NO_3, BrO, OClO, and the aerosol extinction at three wavelengths), evaluated on a local altitude grid from 1 to 100 km with Δz = 1 km (n_z = 100). The model matrix K_{tot} is then a (60,000 × 900) matrix. Using double precision, this amounts to a matrix of 432 Mb.

3. Two-Step Inversion

3.1. Model Approximations

[18] The previous example (section 2.4) clearly shows that calculations using the full forward model are very demanding with respect to present-day computational resources. Usually, in practice, some approximations are used that significantly reduce the dimensions of the system. Scaling issues are not relevant for the principles that we want to demonstrate, so we will not discuss them here.

[19] For example, we can start by neglecting chromatic refraction, assuming that the refractive index of air is wavelength-independent. The weighting functions can then be evaluated at a reference wavelength

$$g(h_\lambda, z_i) \approx g(\bar{h}, z_i). \tag{15}$$

This approximation is justified because the small differences in tangent altitude at different wavelengths leave the path lengths approximately the same. Also, the constituents vary too weakly along these altitude differences to cause a significant change in slant path optical depth.

[20] Furthermore, effective cross sections are introduced to take into account the temperature dependence at different altitudes. For example, for NO_2 the effective cross section $\bar{\sigma}_{NO_2}(\bar{h}, \lambda)$ is defined as

$$\begin{aligned}\bar{\sigma}_{NO_2}(\bar{h},\lambda) &= \frac{\sum_{i=1}^{n_z} g(\bar{h},z_i)\sigma_{NO_2}(z_i,\lambda)N_{NO_2}(z_i)}{\sum_{i=1}^{n_z} g(\bar{h},z_i)N_{NO_2}(z_i)} \\ &= \frac{\sum_{i=1}^{n_z} g(\bar{h},z_i)\sigma_{NO_2}(z_i,\lambda)N_{NO_2}(z_i)}{\rho_{NO_2}(\bar{h})},\end{aligned} \tag{16}$$

with $\rho_{NO_2}(\bar{h})$ the NO_2 density, integrated along the optical path with tangent altitude $\bar{h}$. The effective cross section and integrated density can initially be evaluated using climatological values for the density and have to be updated with the actual retrieval during the estimation phase in an iterative way until convergence is reached.

[21] Using these approximations, the model optical thickness transforms to

$$\tau(\bar{h},\lambda) = \bar{\sigma}_{\text{air}}(\bar{h},\lambda)\rho_{\text{air}}(\bar{h}) + \cdots + q_3(\lambda)\tau_{\text{aero}}(\bar{h},\lambda_3), \tag{17}$$

with $\tau_{\text{aero}}(\bar{h}, \lambda_3)$ the aerosol slant path optical thickness at wavelength λ_3. In matrix notation the transmittance spectrum at the ith tangent altitude reads

$$\mathbf{T}_i = \exp(-C_i\mathcal{N}_i), \tag{18}$$

with C_i the $(n_\lambda \times n_c)$ matrix containing the equivalent cross sections at the specified tangent altitude and $\mathcal{N}_i$ the $(n_c \times 1)$ vector with the integrated densities and aerosol optical thickness. Notice that all quantities in equation (18) are expressed at one tangent altitude. Therefore the vectors $\mathcal{N}_i$ can be retrieved for all tangent altitudes separately.

[22] The integrated density or aerosol optical thickness profile can now be expressed as a function of local density profile or aerosol extinction. For air,

$$\mathcal{N}_{\text{air}} = M\mathbf{N}_{\text{air}}, \tag{19}$$

with M the $(n_h \times n_z)$ matrix with (approximate) weighting functions. Notice that equation (19) expresses a pure geometrical relation and is specified for one constituent, independent from the other ones. All constituent profiles can thus be retrieved separately.

[23] The forward model is now fully specified. By applying a few approximations to the general forward model, we have established a decoupling of the spectral and geometrical dependency. The decoupling is, however, formal. Physically, integrated gas densities and aerosol optical thicknesses at two different tangent altitudes are interrelated, and two different constituent profiles are, from a retrieval point of view, correlated.

3.2. Example

[24] To illustrate the computational advantage of the derived expressions, let us take our previous example. The (1000 × 9) matrix C_i has to be evaluated at every tangent altitude, so we have 60 matrices. Following this, the (60 × 100) matrix G has to be constructed once. The data in the model occupies about 4.5 Mb.

3.3. Alternative Method

[25] The discussion in section 3.1 demonstrated that a retrieval can be performed in two steps: first, the spectral inversion, then the spatial inversion. The order of the two steps can, of course, be reversed. At every wavelength the transmission as a function of h can be inverted to a local total extinction profile for that wavelength. After this step, spectral inversion is achieved by inverting the total extinction spectrum at every altitude. To use our example again: Spatial inversion consists of constructing M once and inverting 1000 transmittance profiles separately (at every wavelength). Then the local densities and aerosol extinctions can be retrieved by inverting 100 total extinction spectra. Although possible, this way of working is not preferable because it is computationally more expensive.

4. One-Step Global Inversion

4.1. Estimation

[26] Let us now return to the full forward model, expressed by equation (14). Suppose the experiment has performed a full occultation measurement $\mathbf{T}_{\text{meas}}$ with unknown measurement error ϵ (e.g., detector noise). We assume that ϵ is drawn from a normal distribution with zero mean and covariance matrix S_ϵ. Thus we have

$$\mathbf{T}_{\text{meas}} = \mathbf{T}_{\text{tot}} + \epsilon = \exp(-K_{\text{tot}}\mathbf{x}) + \epsilon. \tag{20}$$

[27] In this expression we implicitly assume that no forward model error is present [see, e.g., *Rodgers*, 2000]. The vector of unknowns $\mathbf{x}$ can be estimated by the application of a nonlinear least squares method, for example, the (computationally expensive) Levenberg-Marquard algorithm [*Press et al.*, 1992]. Here we will derive an iterative expression specifically for our problem.

[28] Remote sensing problems typically suffer from ill-posedness, with the result that even a small amount of measurement error leads to a huge error amplification in the state vector $\mathbf{x}$. The remedy is given by imposing some constraint on the solution vector. Different approaches exist, such as optimal estimation methods that make use of statistical a priori knowledge of the atmosphere. Here we will use a soft altitude-smoothing constraint, leading to a constraining method that is referred to as Twomey-Phillips-Tikhonov regularization [*Press et al.*, 1992; *Rodgers*, 2000; *Hansen*, 1998]. The merit function to be minimized is given by

$$(\mathbf{T}_{\text{meas}} - \mathbf{T}_{\text{tot}})^{\text{T}} S_\epsilon^{-1} (\mathbf{T}_{\text{meas}} - \mathbf{T}_{\text{tot}}) + \mu^2 \mathbf{x}^{\text{T}} L^{\text{T}} L \mathbf{x}, \tag{21}$$

where the first term represents the usual least squares merit function and the second term represents a quadratic smoothness constraint. Usually, the matrix L is taken to be the identity matrix or the first- or second-derivative matrix [*Press et al.*, 1992]. The parameter μ controls the trade-off between goodness of fit and solution smoothness.

[29] The forward model is nonlinear. We can linearize it around the vector $\mathbf{x}^0$ by taking a first-order Taylor expansion. For the ith component of $\mathbf{T}_{\text{tot}}$,

$$\mathbf{T}_{\text{tot},i} = \mathbf{T}^0_{\text{tot},i} + \sum_{j=1}^{n_z n_c} \left(x_j - x_j^0\right) \left[\frac{\partial \mathbf{T}_{\text{tot},i}}{\partial x_j}\right]_0, \tag{22}$$

and for the total matrix,

$$\mathbf{T}_{\text{tot}} = \mathbf{T}^0_{\text{tot}} - D^0_{\text{tot}} K_{\text{tot}} \left(\mathbf{x} - \mathbf{x}^0\right), \tag{23}$$

where D^0_{tot} is a square matrix with the transmission vector $\mathbf{T}^0_{\text{tot}}$ on its diagonal,

$$D^0_{\text{tot}} = \text{diag}\left(\mathbf{T}^0_{\text{tot}}\right). \tag{24}$$

[30] We can now take partial derivatives of the merit function with respect to the unknowns and set them equal to zero, a condition that has to be fulfilled at the merit minimum. Using straightforward algebra and associating $\mathbf{T}^0_{\text{tot}}$ and $\mathbf{x}^0$ with the results obtained from a previous iteration, we finally come to the iterative formula

$$\mathbf{x}^{(i+1)} = \left(K^T_{\text{tot}} D^{(i)}_{\text{tot}} S_\epsilon^{-1} D^{(i)}_{\text{tot}} K_{\text{tot}} + \mu^2 L^T L\right)^{-1} \cdot K^T_{\text{tot}} D^{(i)}_{\text{tot}} S_\epsilon^{-1} \left(\mathbf{T}_{\text{meas}} - \mathbf{T}^{(i)}_{\text{tot}} - D^{(i)}_{\text{tot}} K_{\text{tot}} \mathbf{x}^{(i)}\right). \tag{25}$$

[31] At the final iteration we obtain the solution $\mathbf{x}^{(f)}$. If we define the gain matrix G as

$$G = \left(K^T_{\text{tot}} D^{(f)}_{\text{tot}} S_\epsilon^{-1} D^{(f)}_{\text{tot}} K_{\text{tot}} + \mu^2 L^T L\right)^{-1} K^T_{\text{tot}} D^{(f)}_{\text{tot}} S_\epsilon^{-1}, \tag{26}$$

then the random error covariance matrix [*Rodgers*, 2000] equals

$$S^{\text{rand}}_{\mathbf{x}} = G S_\epsilon G^{\text{T}}. \tag{27}$$

Furthermore, if the variability of the true atmospheric state is expressed by the covariance matrix S_e, then the smoothing error covariance matrix [*Rodgers*, 2000] equals

$$S^{\text{smooth}}_{\mathbf{x}} = \left(G D^{(f)}_{\text{tot}} K_{\text{tot}} - I\right) S_e \left(G D^{(f)}_{\text{tot}} K_{\text{tot}} - I\right)^{\text{T}}, \tag{28}$$

with I the ($n_z n_c$ × $n_z n_c$) identity matrix. The matrix $A = G D^{(f)}_{\text{tot}} K_{\text{tot}} - I$ is often referred to as the averaging kernel matrix. The total covariance matrix then equals

$$S^{\text{tot}}_{\mathbf{x}} = S^{\text{rand}}_{\mathbf{x}} + S^{\text{smooth}}_{\mathbf{x}}. \tag{29}$$

4.2. Singular Value Decomposition (SVD) Reduction

[32] Equations (25) and (29) represent the solution estimate for the total problem. However, as stated in sections 1 and 2.4, in practice, the dimensions of the problem are too

large to perform efficient calculations. Some reduction of the problem is necessary.

[33] Let us investigate the nature of a measured transmittance spectrum at a certain tangent altitude. It is clear that n_λ channels do not necessarily deliver n_λ-independent pieces of information. For a large number of channels most will be redundant with each other. No matter how large the number of channels, a full spectrum results from the optical extinction by a fairly limited number of constituents.

[34] Considerations like these could lead to the conclusion that a simple channel selection is the way to reduce the model dimensions. However, we prefer to use all channels for statistical reasons: Although two different channels possibly measure the same information, we want to use both to decrease the measurement uncertainty.

[35] It thus seems more appropriate to apply channel binning, defining spectral ranges in which we sum all measurements. However, how the bins should be defined is far from evident. Intuitively, we have some idea; for example, one bin would be centered on the maximum of the ozone Chappuis absorption band. As for other species, the situation is often unclear.

[36] Channel selection and binning are two special cases where linear combinations of channels are taken. What we need is a general linear operator that concentrates the spectral information at a certain tangent altitude into a small amount of numbers. The operator should be based on the actual physical model itself. It can be found by taking the singular value decomposition (SVD) [*Golub and Van Loan*, 1996; *Rodgers*, 2000; *Press et al.*, 1992] of the model matrix. At the ith tangent altitude we have

$$K_i = U_i \Sigma_i V_i^{\mathrm{T}}, \tag{30}$$

where Σ_i is a $(n_\lambda \times n_z n_c)$ diagonal matrix of singular values and U_i and V_i are unitary matrices with respective dimensions $(n_\lambda \times n_\lambda)$ and $(n_c n_z \times n_c n_z)$.

[37] The information content of a spectrum at a certain tangent altitude is usually very limited. Just a few, say, p_i base vectors of U_i and their associated singular values are usually sufficient to describe the signal with very high accuracy. We therefore replace K_i by its reduced form

$$\tilde{K}_i = \tilde{U}_i \tilde{\Sigma}_i V_i^{\mathrm{T}}, \tag{31}$$

with $\tilde{U}_i$ a $(n_\lambda \times p_i)$ unitary matrix and $\tilde{\Sigma}_i$ a $(p_i \times n_z n_c)$ diagonal matrix.

[38] Taking all tangent altitudes together, the reduced form of equation (13) is

$$\tilde{K}_{\mathrm{tot}} = \tilde{U}_{\mathrm{tot}} Q_{\mathrm{tot}} \tag{32}$$

with

$$\tilde{U}_{\mathrm{tot}} = \begin{bmatrix} \tilde{U}_1 & & \\ & \ddots & \\ & & \tilde{U}_{n_h} \end{bmatrix} \quad Q_{\mathrm{tot}} = \begin{bmatrix} \tilde{\Sigma}_1 V_1^{\mathrm{T}} \\ \vdots \\ \tilde{\Sigma}_{n_h} V_{n_h}^{\mathrm{T}} \end{bmatrix}. \tag{33}$$

If we take $n_p = \sum_{i=1}^{n_h} p_i$, then $\tilde{U}_{\mathrm{tot}}$ is a $(n_\lambda n_h \times n_p)$ block diagonal matrix, while Q_{tot} has dimensions $(n_p \times n_z n_c)$. Replacing K_{tot} in equation (25) by equation (32) and using the notation

$$P_{\mathrm{tot}}^{(i)} = D_{\mathrm{tot}}^{(i)} \tilde{U}_{\mathrm{tot}}, \tag{34}$$

we get

$$\mathbf{x}^{(i+1)} = \left(Q_{\mathrm{tot}}^{\mathrm{T}} P_{\mathrm{tot}}^{(i)\mathrm{T}} S_\epsilon^{-1} P_{\mathrm{tot}}^{(i)} Q_{\mathrm{tot}} + \mu^2 L^T L \right)^{-1} \cdot Q_{\mathrm{tot}}^{\mathrm{T}} P_{\mathrm{tot}}^{(i)\mathrm{T}} S_\epsilon^{-1} \left(\mathbf{T}_{\mathrm{meas}} - \mathbf{T}_{\mathrm{tot}}^{(i)} - P_{\mathrm{tot}}^{(i)} Q_{\mathrm{tot}} \mathbf{x}^{(i)} \right). \tag{35}$$

At the final iteration the gain matrix G (equation (26)) is now

$$G = \left(Q_{\mathrm{tot}}^{\mathrm{T}} P_{\mathrm{tot}}^{(f)\mathrm{T}} S_\epsilon^{-1} P_{\mathrm{tot}}^{(f)} Q_{\mathrm{tot}} + \mu^2 L^T L \right)^{-1} Q_{\mathrm{tot}}^{\mathrm{T}} P_{\mathrm{tot}}^{(f)\mathrm{T}} S_\epsilon^{-1}. \tag{36}$$

The random error covariance matrix has the same form as equation (27), while the smoothing error covariance transforms to

$$S_{\mathbf{x}}^{\mathrm{smooth}} = \left(G P_{\mathrm{tot}}^{(f)} Q_{\mathrm{tot}} - I \right) S_e \left(G P_{\mathrm{tot}}^{(f)} Q_{\mathrm{tot}} - I \right)^{\mathrm{T}}. \tag{37}$$

[39] One practical remark: There is no need to use the full matrix $P_{\mathrm{tot}}^{(i)}$ in practical calculations, since it is a block diagonal matrix. For example, the calculation of $P_{\mathrm{tot}}^{(i)} Q_{\mathrm{tot}}$ can be carried out for all tangent altitudes separately.

4.3. Example

[40] Once again, let us return to our standard example. Suppose that for every tangent altitude, 10 base vectors are enough to describe the signal in an accurate way ($p_i = 10$, $n_p = 600$), so the matrix Q_{tot} has dimensions (600×900). The matrix $\tilde{U}_{\mathrm{tot}}$ consists of 60 matrices with dimension (1000×10). Using double precision, the model has a size of about 9.1 Mb.

5. Simulation

5.1. Model Definition

[41] The accuracy and efficiency of the one-step algorithm can be illustrated by the retrieval of a simulated measurement. We assume (as in our standard case) 60 tangent altitudes ranging from 11 to 99.5 km, so the tangent distance between two subsequent measurements equals 1.5 km. The spectrometer consists of 451 channels with wavelengths ranging from 250 to 700 nm, resulting in a spectral resolution of 1 nm.

[42] Furthermore, we will retrieve the constituent profiles on a regular altitude grid from 10 to 100 km, with $\Delta z = 1$ km. In our simulation, only the species that cause the major part of the optical extinction will be taken into account: air, ozone, NO_2, and aerosols ($n_c = 6$). The three wavelengths where aerosol extinction profiles will be retrieved are chosen at 340, 435, and 600 nm (wavelengths that have also been used by the SAGE II and ORA experiments).

5.2. Transmittance Simulation

[43] As mentioned in section 2.2, the model is scaled with a climatological profile for numerical reasons, and in our case we used the reference profiles from *Anderson et al.*

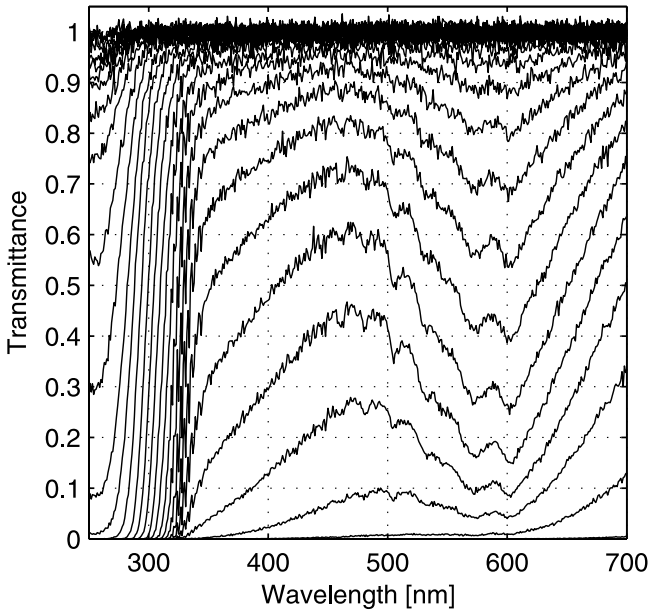

Figure 1. Simulated transmittance spectra at different tangent altitudes.

[1986]. We can use the same climatological profile to construct constituent profiles that represent the actual physical state of the atmosphere. Here we assume that the mean state is represented by the climatology, and the constituents can vary around this mean by 5%. The covariance matrix that represents this variability was introduced in equation (28) as the matrix S_e. The synthetic "true" profiles can now be generated. For example, for air,

$$N_{\text{air}}^{\text{true}}(z_i) = \tilde{N}_{\text{air}}(z_i)\left[1 + 0.05\sin\left(2\pi\frac{z_i}{20}\right)\right]. \tag{38}$$

Similar expressions were used for the other constituents, all with a sinusoidal perturbation with an amplitude of 5%.

[44] The transmittance spectrum at the ith tangent altitude can be calculated by constructing the model matrix K_i and using equation (11). Experimental noise then has to be added to the transmittance. We assume that the experimental error can be modeled as Poisson shot noise, for which the variance equals the mean of the signal $\sigma_S^2 = S$. Since the transmittance is calculated as $T = S/S_{\max}$ with $S_{\max}$ the unattenuated signal outside the atmosphere, the variance equals $\sigma_T^2 = T/S_{\max}$. In our simulation we take $S_{\max}$ = 10,000. The obtained variances form the diagonal elements of the measurement covariance matrix S_ϵ. The off-diagonal elements are zero since all measurements are assumed to be taken independently.

[45] The transmittance spectra at all tangent altitudes are shown in Figure 1. Some typical features are present. Strong ozone absorption by the Hartley bands at altitudes in the mesosphere can be observed in the UV. At stratospheric altitudes, ozone absorption by the Chappuis bands around 600 nm are clearly visible. The fine-scale absorption features around 460 nm are caused by NO_2 absorption bands.

5.3. Regularization

[46] In our retrievals the solution is stabilized by the application of an altitude-smoothing constraint for every constituent profile. We chose L to be the discrete second derivative operator. The amount of smoothing is controlled by the parameter μ, and some criterion has to be used to determine it. Quite a few criteria exist [*Hansen*, 1998], but to us the methods that use the actual statistics of the measurement are preferable since the theoretical knowledge of the variance of the measurement noise is used explicitly. Here we use the discrepancy principle [*Hansen*, 1998], which basically says that the sample variance of the retrieval residual should equal the theoretical estimate of the measurement noise variance, or, stated otherwise,

$$\chi^2 = (\mathbf{T}_{\text{meas}} - \mathbf{T}_{\text{tot}})^{\text{T}} S_\epsilon^{-1}(\mathbf{T}_{\text{meas}} - \mathbf{T}_{\text{tot}}) = m, \tag{39}$$

where m equals the number of measurements, in our case, $m = n_\lambda n_h$.

5.4. Calculation

[47] During the construction of the retrieval model, eight base vectors were chosen at every tangent altitude (so n_p = 480), a number that was sufficient to describe the transmittance spectra with very high accuracy (relative to the measurement error). Of course, an SVD calculation has to be performed at every tangent altitude, and this is computationally the most expensive step in the entire retrieval scheme. As initial guess for the state vector in the iteration, we chose the climatological profile. After five iterations the procedure was stopped since the χ^2 of the residual did not change significantly. The entire process, from model construction and SVD calculation to the iterative minimization, took less than 3 min.

5.5. Retrieval Results

[48] The physical meaning of the SVD reduction can be demonstrated by inspecting Figure 2. Shown are the eight column vectors of $P_{\text{tot}}^{(f)}$, associated with a tangent altitude of 32 km. From equation (35) we know that the scalar product of the column vectors of $P_{\text{tot}}^{(i)}$ with the (error-weighted) measured transmittance is calculated during the iteration. This is how the column vectors have to be interpreted: They are the coefficients (or weights) in the linear combination of channels. The first three columns are easily interpreted from Figure 2: They select spectral domains where O_3, air, and NO_2 are strong absorbers. Notice the cutoff in the UV below 300 nm: No channels are selected since the measured signal in the stratosphere and below is zero in this spectral window. The fourth column selects a spectral range that contains mainly information about the aerosol extinction (the relative contribution of aerosols to the total extinction is larger at long wavelengths). The interpretation of the next four columns is less clear, but their aim can be identified from the magnitude of the coefficients in the associated rows of Q_{tot}: They "select" aerosols (columns 5 and 6) and O_3 (columns 7 and 8). The order of the columns is also interesting: Since the first column is associated with the largest singular value, it is associated with the constituent of which the retrieval is the most accurate, in this case, O_3.

[49] Figure 3 shows the retrieved profiles for air, O_3, NO_2, and aerosol extinction at 600 nm in the lower and middle atmosphere on a linear scale, together with the true profiles. All constituent features are nicely retrieved. In

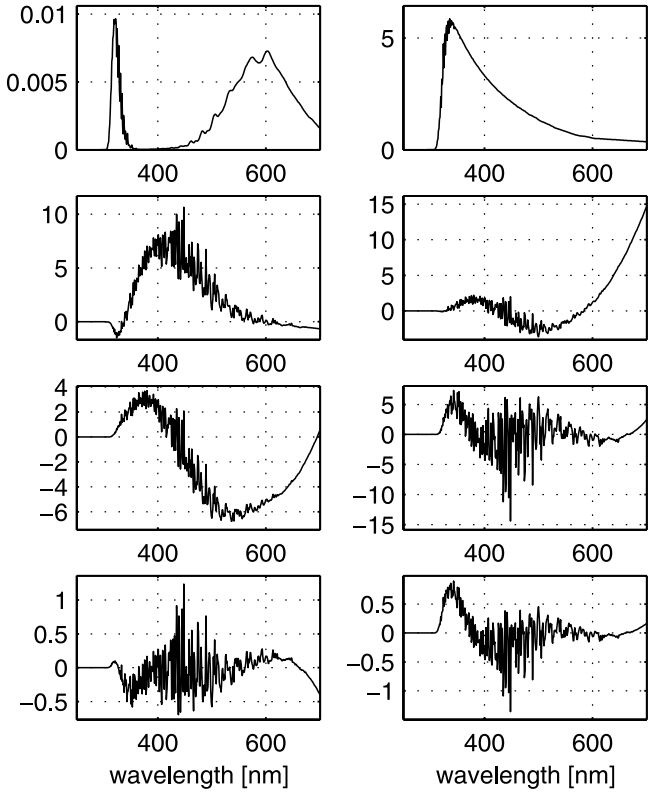

Figure 2. First eight columns (from left to right and top to bottom) of $P_{\mathrm{tot}}^{(f)}$ at $h = 32$ km, representing the weights for the channel "selection."

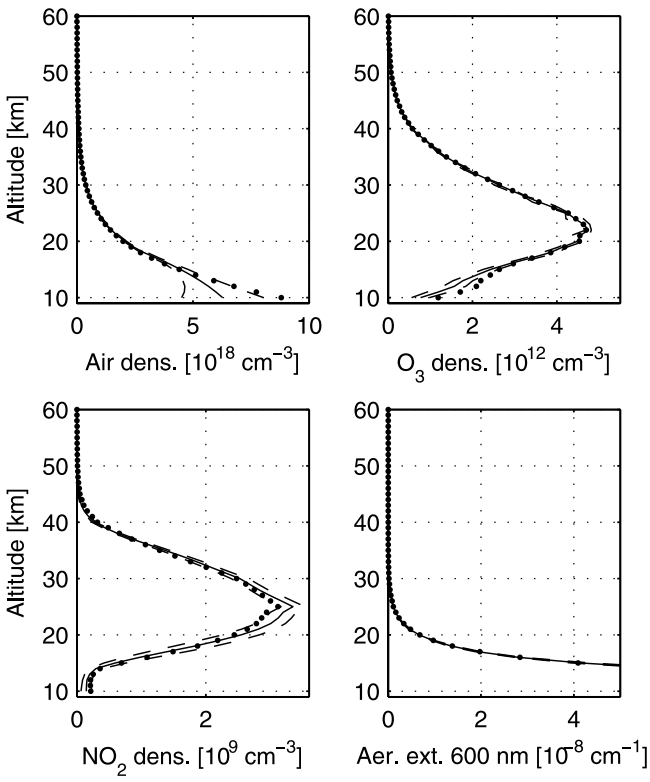

Figure 3. Retrieved profiles (solid lines) and associated total uncertainty (dashed lines) on a linear scale. The true profiles are represented by dots.

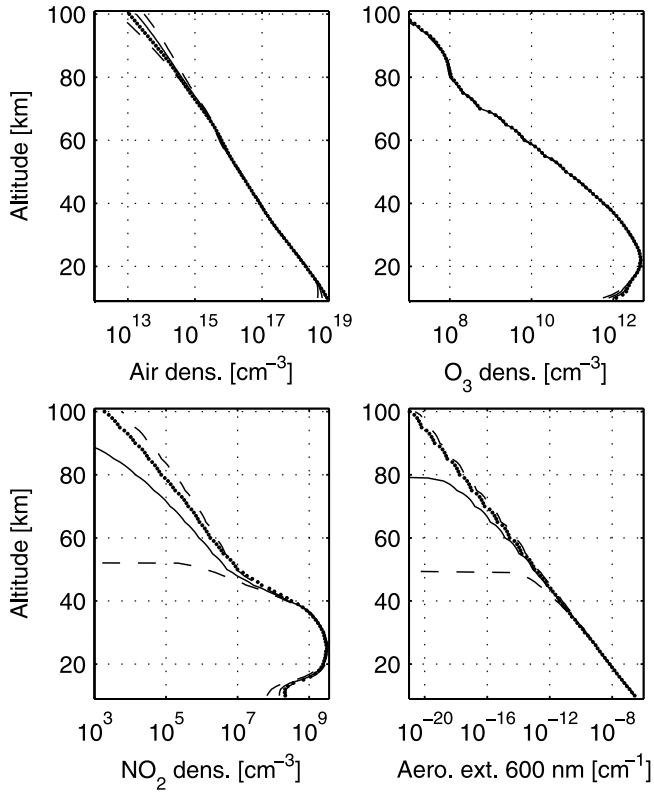

Figure 4. Same as Figure 3 but for the entire altitude range and on a logarithmic scale.

order to show the results at higher altitudes, we present the full profiles in Figure 4 on a logarithmic scale. Because of the strong absorption by the ozone Hartley bands in the mesosphere, an accurate ozone profile can be retrieved up to the top of the atmosphere. The air density profile is retrieved moderately well in the upper atmosphere. The results for NO_2 and aerosols in this altitude region have to be treated with caution: Although the retrieved profiles are smooth, the huge uncertainty shows that the retrievals are of extremely low accuracy.

[50] The retrieval error can be better investigated by calculating the percentage relative difference between retrieved and true profiles. For the air retrieval $\hat{N}_{\mathrm{air}}$ it is calculated as

$$100\left(\hat{N}_{\mathrm{air}} - N_{\mathrm{air}}^{\mathrm{true}}\right)/N_{\mathrm{air}}^{\mathrm{true}}, \tag{40}$$

with associated uncertainty $100\hat{\sigma}_{\mathrm{air}}/N_{\mathrm{air}}^{\mathrm{true}}$. Figure 5 shows the results for all profiles. Air is typically retrieved from the lower stratosphere to the middle mesosphere with an accuracy of about 3–5%. Ozone can be retrieved very well in the same region, having an uncertainty of only 2–3%. Since optical absorption by NO_2 is only significant in the stratosphere, reasonable results are achieved there within an uncertainty of 5–10%. And since we have chosen a true aerosol profile that is representative for low volcanic conditions, the retrieval is only reasonable in the lower stratosphere, having an uncertainty of about 5%.

[51] For comparison, we also calculated the two-step retrievals. Initially, the effective cross sections were evaluated using the climatological profiles. At every tangent

altitude, integrated densities were retrieved with the Levenberg-Marquard algorithm, and a linear inversion with vertical regularization was applied to obtain the local gas density and aerosol extinction profiles. To make the comparison valid, we used the same regularization parameter as in the global inversion. The obtained results were used to calculate the effective cross sections for the next iteration. The results that were obtained after three iterations are also shown in Figure 5. The global inversion retrieval offers no significant improvement for ozone in comparison with the two-step retrieval, except at high altitudes. For the other constituents, small to considerable improvements are observed almost everywhere.

[52] It is interesting to investigate the separate error contributions for the constituent profiles, which are shown in Figure 6. The most important conclusion that we can draw from Figure 6 is that accurate results can be expected when the random error has more or less the same magnitude as the smoothing error. This is the basic idea of regularization: We apply a constraint that introduces smoothing error but significantly reduces random error, such that the total error decreases.

6. Conclusion

[53] In this paper we have developed a new method for performing global occultation retrievals. The typical drawbacks of the classical two-step methods are avoided. No approximate weighting functions and cross sections are used in the formulation of the physical model, and instead, the model dimensions are significantly reduced by taking linear combinations of channels. The weights in these linear combinations are determined by the use of the singular value decomposition of the model matrix. Furthermore, the global approach of the one-step inversion scheme ensures that all spectral channels and tangent altitudes are exploited at the same time during the inversion scheme and that no artificial separation of spectral and spatial inversion occurs, and therefore all correlations between the different constituents at all altitudes are taken into account.

[54] A realistic simulation showed that the retrievals are of very good quality and of better accuracy than the two-step results. We therefore think that the method is a very strong contender among the different methods that are presently used to retrieve atmospheric constituent profiles from occultation measurements.

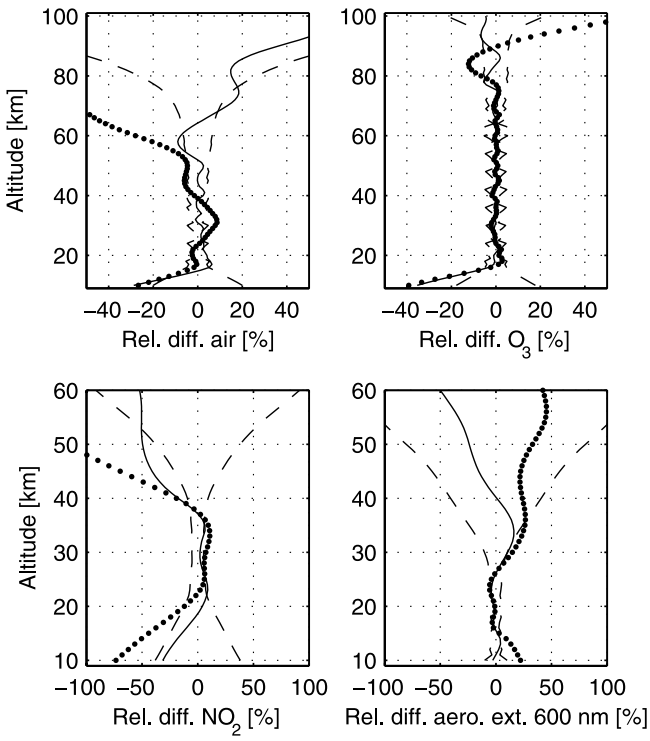

Figure 5. Relative differences between the one-step retrieval and the true profiles (solid lines), with associated total uncertainty (dashed lines). The results for the two-step retrieval (dots) are shown for comparison.

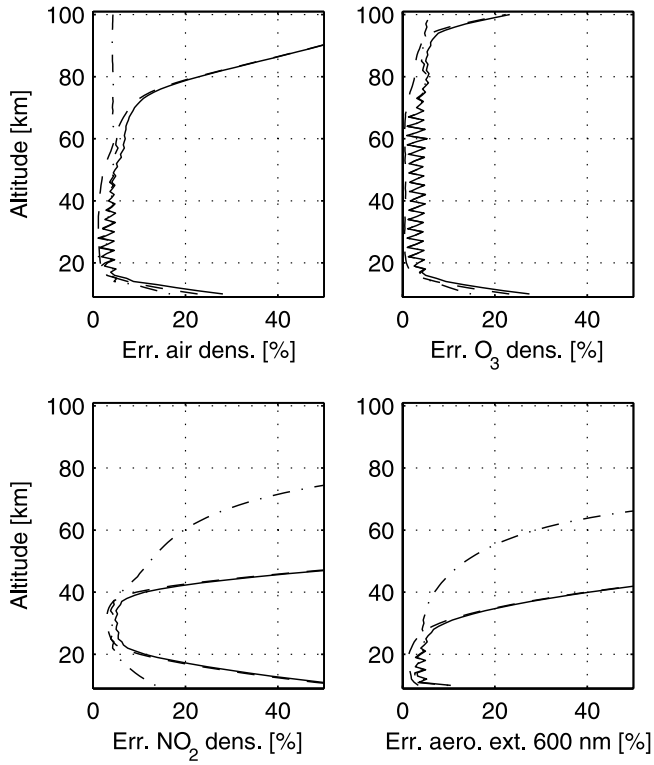

Figure 6. Separate error contributions for all profiles: random error (dashed lines), smoothing error (dash-dotted lines), and combined total error (solid lines).

[55] **Acknowledgments.** F.V. was financially supported by the Prodex contract SADE, granted by the OSTC of the Belgian government, while C.B. was financed by project MO/35/009 of the OSTC.

References

Anderson, G., S. Clough, F. Kneizys, J. Chetwynd, and E. Shettle (1986), AFGL atmospheric constituent profiles (0–120 km), *Tech. Rep. AFGL-TR-86-0110*, Air Force Geophys. Lab., Hanscom Air Force Base, Bedford, Mass.

Bertaux, J., G. Megie, T. Widemann, E. Chassefiere, R. Pellinen, E. Kyrola, S. Korpela, and P. Simon (1991), Monitoring of ozone trend by stellar occultations: The GOMOS instrument, *Adv. Space Res.*, *11*(3), 3237–3242.

Chu, W., M. McCormick, J. Lenoble, C. Brogniez, and P. Pruvost (1989), SAGE II inversion algorithm, *J. Geophys. Res.*, *94*, 8339–8351.

Edlen, B. (1953), The dispersion of standard air, *J. Opt. Soc. Am.*, *43*, 339–345.

Fussen, D., F. Vanhellemont, and C. Bingen (2001a), Remote sensing of the Earth's atmosphere by the spaceborne occultation radiometer, ORA: Final inversion algorithm, *Appl. Opt.*, *40*(6), 941–948.

Fussen, D., F. Vanhellemont, and C. Bingen (2001b), Evolution of stratospheric aerosols in the post-Pinatubo period measured by the Occultation Radiometer ORA, *Atmos. Environ.*, *35*, 5067–5078.

Golub, G., and C. Van Loan (1996), *Matrix Computations*, 3rd ed., Johns Hopkins Univ. Press, Baltimore, Md.

Gunson, M. R., et al. (1996), The Atmospheric Trace Molecule Spectroscopy (ATMOS) experiment: Deployment on the ATLAS space shuttle missions, *Geophys. Res. Lett.*, *23*, 2333–2336.

Hansen, P. C. (1998), *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, Pa.

Kyrola, E., E. Sihvola, Y. Kotivuori, M. Tikka, T. Tuomi, and H. Haario (1993), Inverse theory for occultation measurements: 1. Spectral inversion, *J. Geophys. Res.*, *98*, 7367–7381.

Lucke, R., et al. (1999), The Polar Ozone and Aerosol Measurement (POAM) III instrument and early validation results, *J. Geophys. Res.*, *104*, 18,785–18,799.

McElroy, T., J. Drummond, and C. Midwinter (2000), MAESTRO: Photodiode-array spectrometer enhancement for the ACE satellite, paper presented at 11th CASI conference on Astronautics, Ottawa, 7–9 November.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, New York.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, *Ser. Atmos. Oceanic Planet. Phys.*, vol. 2, 1st ed., World Sci., River Edge, N. J.

Russell, J. I., L. Gordley, J. Park, S. Drayson, W. Hesketh, R. Cicerone, A. Tuck, J. Frederick, J. Harries, and P. Crutzen (1993), The Halogen Occultation Experiment, *J. Geophys. Res.*, *98*, 10,777–10,797.

Vanhellemont, F., and D. Fussen (2002), Derivation of stratospheric aerosol properties from the ORA experiment, *Adv. Space Res.*, *29*(11), 1741–1746.

Yee, J.-H., et al. (2002), Atmospheric remote sensing using a combined extinctive and refractive stellar occultation technique: 1. Overview and proof-of-concept observations, *J. Geophys. Res.*, *107*(D14), 4213, doi: 10.1029/2001JD000794.

C. Bingen, D. Fussen, and F. Vanhellemont, Belgian Institute for Space Aeronomy, Ringlaan 3, B-1180, Brussels, Belgium. (christine.bingen@bira-iasb.oma.be; didier.fussen@oma.be; filip.vanhellemont@bira-iasb.oma.be)