



Regularisation model study for the least-squares retrieval of aerosol extinction time series from UV/VIS MAX-DOAS observations for a ground layer profile parameterisation

A. Hartl¹ and M. O. Wenig²

¹School of Energy and Environment, City University of Hong Kong, Hong Kong, Hong Kong

²Meteorological Institute, Ludwig-Maximilians-Universität München, Munich, Germany

Correspondence to: A. Hartl (ahartl@cityu.edu.hk)

Received: 6 February 2013 – Published in Atmos. Meas. Tech. Discuss.: 14 March 2013

Revised: 28 May 2013 – Accepted: 8 July 2013 – Published: 12 August 2013

Abstract. The retrieval of tropospheric aerosol extinctions from MAX-DOAS observations of O₄ using a small number of three or four extinction profile parameters suitable for boundary layer reconstruction is investigated with respect to the following questions. First, to what extent does this nominally over-constrained pure least-squares problem for the inversion of the radiative transfer equation require regularisation and how should parameters of the regularisation be chosen? Second, how can a lack of information in the under-constrained case be best compensated by using the information contained in a sequence of observations and by explicitly including intensities into the fit?

The forward model parameterises the optical properties of the boundary layer aerosol by its extinction profile, single-scattering albedo and a Henyey–Greenstein phase function. Forward calculations are carried out online, i.e. without lookup tables. The retrieval uses a Tikhonov regularisation combined with an approximate L-curve criterion and empirical a priori information from the retrieval sequence based on previous valid solutions. The consistency of the approach is demonstrated in selected model case studies assuming a polluted urban scenario and westward viewing direction of the instrument. It is shown that a dynamic choice of the regularisation parameter is crucial for high aerosol load and large diurnal variations. The quality of the retrieval can be improved significantly if the retrieval sequence and thus the a priori is chosen according to the information content of the measurement series. Additional intensities improve the solution for all solar angles if suitably weighted. This flexible retrieval algorithm allows for reconstruction of aerosol profiles in the

boundary layer for a wide range of viewing directions and extinctions. It can thus be applied to observational geometries scanning the sky in two angular dimensions and to retrieve further aerosol optical parameters in the boundary layer.

1 Introduction

The experimental method to obtain tropospheric aerosol extinction and trace gas concentration profiles by applying the differential optical absorption spectroscopy (DOAS, Platt and Stutz, 2008) to ground-based observation of sun light intensities under different viewing directions has been an area of considerable activity in the past years – in field measurements in urban and remote areas (e.g. Wittrock et al., 2004; Irie et al., 2009; Lee et al., 2009; Li et al., 2010; Clémer et al., 2010; Frieß et al., 2011; Halla et al., 2011; Shaiganfar et al., 2011), within intensive measurement campaigns (e.g. Heckel et al., 2005; Sinreich et al., 2007; Brinksma et al., 2008; Roscoe et al., 2010; Irie et al., 2011; Vlemmix et al., 2011; Wagner et al., 2011; Zieger et al., 2011), as well as in comparative radiative model studies (Hendrick et al., 2006; Wagner et al., 2007). While the basic idea of this multi-axis DOAS, or MAX-DOAS, technique – namely to infer altitude information on an atmospheric absorber from its absorption signal along several light paths – is fairly straightforward, and the experimental setup relatively inexpensive, the actual conversion requires inversion of the underlying radiative transfer equation. It strongly depends on tropospheric aerosol and, in general, the equation cannot be

linearised. Evolving in a series of studies (Wagner et al., 2002; Hönninger et al., 2004; Sinreich et al., 2005; Wagner et al., 2004; Frieß et al., 2006), it has therefore been suggested to fit forward-modelled absorption signals of an absorber with a known concentration profile, the oxygen dimer complex O_4 , to the measured data in order to retrieve aerosol extinction profile parameters.

This method has been used in several of the field measurements referred to above with different experimental setups and different retrieval procedures. Instruments, amongst other factors, may differ in the number of O_4 wavelength absorptions bands their spectrometers cover and in the number and orientation of viewing angles under which their telescopes scan the sky. Apart from details of the DOAS fit itself – here the O_4 absorption cross section is a source of uncertainty (Wagner et al., 2009; Clémer et al., 2010) – retrieval methods mainly differ in the way they parameterise the aerosol extinction and other aerosol parameters, in which radiative transfer model they use and in their fit algorithms. Currently there are mainly two kinds of approaches. One combines a linear parameterisation of the extinction profile by discrete layers and a statistical parameter estimation which in the area of satellite profile retrieval is commonly referred to as the optimal estimation method (Rodgers, 2000) to result in a non-linear least-squares problem. This approach bears the advantage of providing an estimate for the uncertainty of the retrieval, as well as diagnostic tools for the information content of the measurement. Frieß et al. (2006) employed this approach in their comprehensive model studies to show, among other things, that the retrieval improves with increasing number of O_4 wavelength bands and if (relative) intensities are explicitly added to the fit quantities. Clémer et al. (2010) retrieved monthly aerosol extinction profiles in Beijing at four O_4 wavelengths separately using this method. Both studies come to the conclusion that the number of extinction profile parameters constrained by a MAX-DOAS measurement is rather limited, and although the non-linearity of the problem makes it hard to give a generally valid number, it is even for ideal conditions somewhere below four. The second kind of approach may be characterised by making use of empirically motivated profile shapes such as linear or exponential functions and combinations of them, thus leading to a smaller set of parameters such as ground extinction, layer height, aerosol optical depth and so forth, or, alternatively, scaling factors. Li et al. (2010) used this kind of parameterisation assuming a well-mixed ground layer for their measurements in Guangzhou, southern China. Wagner et al. (2011) proceeded similarly but also accommodated a linear decrease from ground for their retrieval of aerosol and trace gas profiles in Milan. In both studies a pure (i.e. unregularised) non-linear least-squares fit is used to determine the four to six (for the former) or three (for the latter) profile parameters. Irie et al. (2008), on the other hand, introduced a set of three scaling factors for the total aerosol optical depth to parameterise partial optical depths, and combined

this with the optimal estimation method to retrieve four profile parameters in total. The forward radiative transfer model in the least-squares fit uses pre-calculated “offline” look-up tables. All studies using profile shape parameterisation with a small set of parameters employ look-up tables, whereas those mentioned using the discrete layer parameterisation carry out their calculation “online”.

The inversion of the atmospheric radiative transfer equation as an integral equation potentially constitutes an ill-posed inverse problem (e.g. Groetsch, 1993; Twomey, 1997; Rodgers, 2000), meaning, for example, that even if all fit parameters are well constrained by the observation, they might be extremely sensitive to measurement errors. This study first addresses the question as to what extent a nominally “over-constrained” least-squares problem for the inversion of ground-based MAX-DOAS measurements of O_4 to aerosol extinction coefficients requires modification to stabilise its solutions (regularisation). We use a simple parameterisation that represents the lower part of the profile linearly and the upper part exponentially to arrive at this formulation. The regularisation is studied using a basic Tikhonov scheme by comparing ad hoc choices of the regularisation parameter to an approximated L-curve criterion. We secondly examine the “under-constrained” problem and how to provide further (“a priori”) information from the context of a measurement series and by adding intensities to the fit (both similar to Frieß et al. (2006) but with slightly different conclusions for our case). By way of addressing these questions in selected, detailed case studies building upon each other in successive sections of this paper, we finally arrive at a robust retrieval algorithm that can be used for any observation geometry and level of aerosol optical depth.

The measurement scenario is the same for all case studies. We assume a situation in urban air pollution monitoring in the UV/VIS spectral range and choose the day and location of the 1-day time series arbitrarily as 5 June 2010 in Shanghai. Aerosol optical properties other than the extinction coefficient are fixed and chosen to be representative for air pollution in a Chinese megacity (see Sect. 3). All radiative transfer calculations are carried out online using the model SCIATRAN (Rozanov et al., 2005, <http://www.iup.uni-bremen.de/sciatran>). The retrieval combines data of the two O_4 absorption bands around 360 and 477 nm.

2 Method

2.1 Principle

The DOAS technique makes use of the fact that atmospheric scattering and absorption processes have different and distinct wavelength dependencies which can be analysed in a spectrum by separating broad and narrow wavelength bands to retrieve trace gases through their unique narrow-band absorption structures. Radiative calibration, in other words

absolute intensities, is dispensed with by dividing intensities I of a spectrum by those of a reference. Defining the optical depth of a certain absorber at wavelength λ through

$$\tau(\lambda) = -\ln\left(\frac{I(\lambda)}{I'(\lambda)}\right), \quad (1)$$

where I' are the intensities without this absorber, the DOAS analysis ultimately yields differential optical depths of an absorber as

$$\Delta\tau(\lambda) = \tau(\lambda) - \tau_{\text{ref}}(\lambda). \quad (2)$$

We refer the reader to Platt and Stutz (2008) for details of this procedure. Slant column densities (SCDs) S of an absorber are defined by dividing optical densities by the absorption cross section σ_{abs} , i.e. $S = \tau/\sigma_{\text{abs}}$. Differential slant column densities (DSCDs) ΔS are understood here in terms of the above differential optical densities as $\Delta S = \Delta\tau/\sigma_{\text{abs}}$.

We consider an experimental setup where the instrument records intensities $I(\alpha)$ for a certain wavelength range in a cycle of elevation angles α between 0° (horizon) and 90° (zenith) in a fixed azimuthal plane $\phi = \phi_{\text{obs}}$, and take the aforementioned reference spectrum to be recorded for each cycle in zenith direction $\alpha_{\text{ref}} = 90^\circ$. Through this choice, the stratospheric contributions in Eq. (2) approximately cancel, and one is left with the desired tropospheric part to the differential absorption optical depth. DOAS evaluation of the measured intensities in this way with respect to the absorber O_4 at one or several wavelength bands makes it possible to infer tropospheric aerosol extinctions by inverse modelling because the O_4 profile is known. Details of this reasoning can be found in Wagner et al. (2002), Hönninger et al. (2004), Wagner et al. (2004), Sinreich et al. (2005) and Frieß et al. (2006).

The observational data $\mathbf{d} = \mathbf{d}(\theta, \phi)$ for a measurement at solar zenith and azimuth angles θ and ϕ is thus given by an m -dimensional vector

$$\mathbf{d} = (\Delta S_1, \dots, \Delta S_m)^T, \quad m = m_\alpha m_{\text{O}_4}, \quad (3)$$

where m_α is the number of elevation angles (exclusive of the reference direction) and m_{O_4} is the number of O_4 wavelengths bands used simultaneously in the profile retrieval. Frieß et al. (2006) suggested to use both O_4 optical depths (or SCDs) and relative intensities (RIs) $\tilde{I} = I/I_{\text{ref}}$ as fit quantities. In this case \mathbf{d} becomes

$$\mathbf{d} = (\Delta S_1, \dots, \Delta S_{m/2}, \tilde{I}_1, \dots, \tilde{I}_{m/2})^T, \quad m = 2m_\alpha m_{\text{O}_4}. \quad (4)$$

We assume that the aerosol extinction profile $k(\lambda, z)$ at a certain wavelength λ (this wavelength λ_{fm} as parameter of the forward model (fm) will be specified in Sect. 2.3) is given by n parameters x_i

$$\mathbf{x} = (x_1, \dots, x_n)^T \mapsto k(\lambda, z, \mathbf{x}) \quad (5)$$

and that for this wavelength the forward model $\mathbf{F}(\mathbf{x})$ produces the quantities corresponding to the measurement data \mathbf{d} ; that is

$$\mathbf{d} + \boldsymbol{\epsilon} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \boldsymbol{\delta}, \quad (6)$$

with $\boldsymbol{\epsilon}$, $\boldsymbol{\delta}$ being the data and model errors, respectively, and \mathbf{b} including all forward-model parameters but \mathbf{x} . Then \mathbf{x} can be retrieved by fitting the forward-modelled quantities $\mathbf{F}(\mathbf{x})$ to the observational data using some cost function of \mathbf{d} and \mathbf{F} .

2.2 Least-squares fit – regularisation, a priori and information content

For parameterisations with (a small number of) parameters such that $n \ll m$ the most straightforward approach to the inverse problem of retrieving \mathbf{x} in Eq. (6) is a non-linear least-squares principle, i.e. a quadratic cost function $\chi^2(\mathbf{x})$ of the form

$$\min_{\mathbf{x} \in R^n} (\mathbf{d} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_\epsilon^{-1} (\mathbf{d} - \mathbf{F}(\mathbf{x})), \quad (7)$$

where \mathbf{S}_ϵ is the covariance matrix of the measurement data and \mathbf{b} is dropped for the moment. This approach requires that the underlying inverse problem is well posed in the sense that the desired parameters are actually constrained by the measurement and that they are reasonably stable for a given level of noise. To examine whether the problem is ill posed in this sense, we first linearise it, assuming that \mathbf{x} is a point in parameter space in the vicinity of the true optimum $\hat{\mathbf{x}}$ so that

$$\mathbf{d} + \boldsymbol{\epsilon} = \mathbf{F}(\mathbf{x}) + \mathbf{F}_x(\mathbf{x})(\hat{\mathbf{x}} - \mathbf{x}) + \dots \quad (8)$$

(derivatives are written throughout as $\mathbf{F}_x = \partial \mathbf{F} / \partial \mathbf{x}$ etc.). We now apply the singular value decomposition (SVD) to the $m \times n$ Jacobian matrix \mathbf{F}_x . For an arbitrary $m \times n$ matrix \mathbf{M} of rank r it can be written as

$$\mathbf{M} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (9)$$

where the $m \times n$ matrix $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \quad (10)$$

and $\Sigma_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ contains the r non-negative singular values by convention ordered as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$. \mathbf{U} and \mathbf{V} are orthonormal matrices of dimension $m \times m$ and $n \times n$, respectively, and \mathbf{u}_i , \mathbf{v}_i their column vectors. Inserting the singular vector expansion for \mathbf{F}_x , Eq. (9), into Eq. (8) allows for similar conclusions to the linear case to be drawn (e.g. Twomey, 1997; Hansen, 1998). Only if the contribution $\sigma_i \mathbf{v}_i^T \hat{\mathbf{x}}$ of a singular value is significantly larger than the error components ϵ_i (and δ_i) does it add information to the

retrieval process, otherwise the inversion of the normal equation of the least-squares problem may magnify errors up to the order of the condition number σ_r/σ_1 . This is irrespective of the rank of the Jacobian matrix.

We will show later that for our aerosol parameterisation the pure least-squares solution can indeed suffer from instability in some of its parameters, and consider here the following modified least-squares principle to regularise the solution of Eq. (7)

$$\min_{\mathbf{x} \in \mathbb{R}^n} \left\{ (\mathbf{d} - \mathbf{F}(\mathbf{x}))^T \mathbf{S}_\epsilon^{-1} (\mathbf{d} - \mathbf{F}(\mathbf{x})) + \gamma^2 (\mathbf{x} - \mathbf{x}_a)^T (\mathbf{x} - \mathbf{x}_a) \right\}. \quad (11)$$

This formulation of the so-called Tikhonov regularisation implies that the regularisation parameter γ is kept constant when solving the minimisation problem. The vector \mathbf{x}_a is usually referred to as a priori of \mathbf{x} . In the context of statistical parameter estimation for Gaussian probability distributions (e.g. Tarantola, 1987, 2005; Rodgers, 2000) Eq. (11), corresponds to the assumption of uncorrelated a priori distributions with mean \mathbf{x}_a and the same variance $\sigma_a = \gamma^{-1}$. Without going into any detail, and assuming uncorrelated errors with variances of the same size σ_ϵ for the weighted least-squares problem, we note for the discussion of the local linearisation following Eq. (7) that, as a consequence of this regularisation, a singular value of the inverse for the weighted linear least-squares problem is changed from $(\sigma_i/\sigma_\epsilon)^{-1}$ to $f_i (\sigma_i/\sigma_\epsilon)^{-1}$, where the filter factor $f_i = (\sigma_i/\sigma_\epsilon)^2 / ((\sigma_i/\sigma_\epsilon)^2 + \gamma^2)$ “filters out” contributions with $\sigma_i/\sigma_\epsilon < \gamma$. Contributions with $\sigma_i/\sigma_\epsilon \gg \gamma$ are more or less unaffected by the regularisation. We note that substitution of σ_a^{-1} for γ yields the definition of independent measurements given in Rodgers (2000, chap. 2.4).

The right choice of the regularisation parameter γ (or σ_a) is evidently important, but even in the linear case an intricate matter. Simple regularisation schemes can be based on an estimate ϵ of the data error, e.g. a choice of γ such that $\|\mathbf{d} - \mathbf{F}(\mathbf{x})\| > \|\epsilon\|$ to avoid “fitting to noise” (discrepancy principle). Other methods do not use an estimate of the noise. The empirically developed L-curve criterion, for example, relies on the observation that the optimal regularisation parameter appears close to a γ that balances the data residual and the norm $\|\mathbf{x} - \mathbf{x}_a\|$ in Eq. (11). For details on these and more sophisticated regularisation schemes in the linear case, we refer to Hansen (1998) and the discussion in Sect. 4. The statistical inversion of Eq. (11) using fixed a priori uncertainties σ_a of the fit parameters as regularisation parameters implicitly assumes that the resulting problem is well posed. Iterative solution of the non-linear (or linear) least-squares problem can, of course, also involve regularisation parameters (or even schemes) which depend on the state of the iteration (see e.g. Doicu et al., 2010, for a comprehensive overview of atmospheric applications), but we would like to postpone the related discussion to Sect. 4 as part of our conclusions. In the present study we firstly want to investigate the relevance of regularisation for an “over-determined”

formulation of the inverse problem at hand and discuss the efficiency of a fixed, ad hoc choice of the regularisation parameter by comparing it to a very simple scheme related to the L-curve criterion for the linear case. For the unweighted least-squares problem, this scheme uses the fact that if e and \mathbf{x}_{est} are good estimates for $\|\epsilon\|$ and \mathbf{x} at the true values, then the choice $\gamma = e/\|\mathbf{x}_{\text{est}}\|$ is similar to that of the L-curve method (Hansen, 1998, Sect. 7.5.3). Assuming that the expected value of the data residual is a reasonable estimate for its real value, this choice here becomes for the weighted least-squares problem

$$\gamma = \frac{\sqrt{m}}{\|\mathbf{x}_{\text{est}}\|}, \quad (12)$$

where the estimate \mathbf{x}_{est} of \mathbf{x} will be specified in Sect. 2.3. In the statistical framework this may be expressed as $\sigma_a = \sqrt{n}/\sqrt{m} \cdot x_{\text{RMS}} \sim x_{\text{RMS}}$, with x_{RMS} being the RMS of \mathbf{x}_{est} .

The role of the a priori \mathbf{x}_a in the regularised framework is commonly discussed for a linearised approximation of \mathbf{F} and the retrieval function \mathbf{R} , so that the retrieved $\hat{\mathbf{x}} = \mathbf{R}(\mathbf{d})$ and the true \mathbf{x} are related through

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x} + (\mathbf{I}_n - \mathbf{A})\mathbf{x}_a + \text{errors}. \quad (13)$$

The explicit form of the averaging kernel matrix (or resolution matrix) $\mathbf{A} = \mathbf{R}_d \mathbf{F}_x$ can be found e.g. in Rodgers (2000, chap. 3) (following from setting the a priori variance-covariance matrix \mathbf{S}_a to $\gamma^{-2} \mathbf{I}_n$ or $\sigma_a^2 \mathbf{I}_n$, respectively). The “errors” include all contributions from measurement and forward-model errors; see Sect. 2.4 for details. The first expression on the right-hand side of Eq. (13) represents the contribution of the true state to the retrieval; the complementary second one describes how the “missing” components are provided by the a priori. For unregularised least-squares solutions $\gamma \rightarrow 0$ ($\sigma_a \rightarrow \infty$) of full-rank $r = n < m$ the a priori has no influence. In the rank-deficient case the solution becomes ambiguous; algorithms producing the generalised inverse yield the minimum-norm solution. In the statistical context the trace of \mathbf{A} is often used to quantify how many independent degrees of freedom d_s out of the maximal m the measurement has (see e.g. chap. 3 or Rodgers, 2000, or the discussion after Eq. 11). For uncorrelated errors ϵ with same variance σ_ϵ it here takes the form

$$d_s = \sum_{i=1}^m \frac{(\sigma_i/\sigma_\epsilon)^2}{(\sigma_i/\sigma_\epsilon)^2 + \sigma_a^{-2}}, \quad (14)$$

where σ_i are the singular values of $\mathbf{F}_x(\hat{\mathbf{x}})$ (assumed to have full rank), so that m in the sum can here be replaced by n .

2.3 Aerosol parameterisation, forward model and retrieval of a time series

The aerosol extinction profile $k(z)$ is divided with respect to altitude into a lower part $k(z, \mathbf{x}, \mathbf{b})$ ($z \leq z_{\text{TOR}}$) given mainly

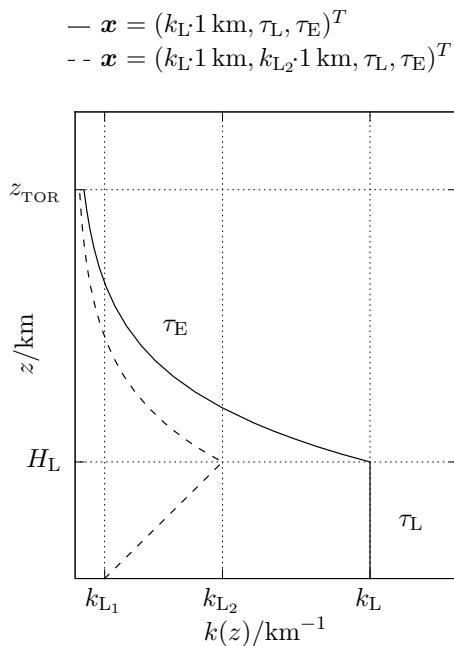


Fig. 1. Parameterisation of the aerosol extinction profile below retrieval height z_{TOR} : linear from ground to height H_L , exponential from H_L to z_{TOR} and continuous at $z = H_L$. τ_L and τ_E are the partial optical depths between ground and H_L and H_L and z_{TOR} , respectively. The three-parameter case with $k_{L_1} = k_{L_2}$ represents a well-mixed layer.

by the parameters \mathbf{x} to be retrieved and an upper part $k(z, \mathbf{b})$ ($z > z_{\text{TOR}}$) given by some of the forward-model parameters \mathbf{b} not subject to the retrieval (with TOR standing for “top of retrieval”). The profile below z_{TOR} consists of a linear part starting at the ground and an exponential part above with continuous transition at height H_L (see Fig. 1). The linear part is parameterised by extinction coefficients k_{L_1} at ground and k_{L_2} at height H_L and the partial aerosol optical depth (AOD) τ_L between. The exponential part is given by its optical depth τ_E . For this four-parameter representation we set

$$\mathbf{x} = (x_1, x_2, x_3, x_4)^T = (k_{L_1} \cdot 1 \text{ km}, k_{L_2} \cdot 1 \text{ km}, \tau_L, \tau_E)^T, \quad (15)$$

so that all x_i have the same unit (none) and are of similar size. For a well-mixed layer ($k_{L_1} \equiv k_{L_2}$) the three parameters are

$$\mathbf{x} = (x_1, x_2, x_3)^T = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T. \quad (16)$$

These aerosol profiles for \mathbf{x} are mapped to the forward quantities $\mathbf{F}(\mathbf{x})$ using the numerical radiative transport model SCIATRAN (version 2.2.2) (Rozanov et al., 2000, 2001, 2005, <http://www.iup.uni-bremen.de/sciattran>). In this study we are not particularly concerned with the details of the numerical model other than the fact that it has to be suitable for this application and the way it parameterises aerosol optical properties. SCIATRAN describes the atmosphere in terms

of altitude layers for profiles and their derivatives (weighting functions) accounting for multiple-scattering processes in a fully spherical geometry including refraction. It offers two methods to solve the radiative transfer equation: a combined differential-integral (CDI) method and a discrete ordinate method (DOM), where the choice of solver also depends on the output quantities desired. The former allows for iterative improvement of the multiple-scattering contribution with respect to sphericity, but not computation of aerosol weighting functions, whereas the latter produces these quantities, but does not allow for refinement of the multiple-scattering calculation. The model has been used in other MAX-DOAS studies (e.g. Wittrock et al., 2004, and also in Frieß et al., 2006, 2011) and has been validated for this application in comparison with other radiative transfer models for this application in Hendrick et al. (2006) and Wagner et al. (2007).

Aerosol optical properties can be described in several ways in SCIATRAN; we use the mode that parameterises the aerosol by a Henyey–Greenstein phase function P_{HG} with asymmetry parameter $g(z)$ and its absorption and scattering coefficient profiles $k_{a,s}(z)$ in discrete layers (the dependency on λ for all quantities is suppressed for the moment). Within this study, we assume that g and the aerosol single-scattering albedo ω_0 are constant below z_{TOR} so that the set of optical aerosol parameters $k_a(z)$, $k_s(z)$ and $P_{\text{HG}}(z)$ becomes $k(z, \mathbf{x})$, ω_0 , g for $z \leq z_{\text{TOR}}$ and $k_a(z)$, $k_s(z)$ and $g(z)$ for $z > z_{\text{TOR}}$, respectively. The profile parameterisations involve a variable height H_L and have to be accommodated to the fixed model grid given by altitude levels z_j . We do this by linear interpolation between the forward quantities $\mathbf{F}(z_j)$ and $\mathbf{F}(z_{j+1})$, where $z_j \leq H_L < z_{j+1}$.

The retrieval algorithm based on the regularised least-squares fit in Sect. 2.2 and the forward model described above can be carried out with or without explicit use of intensities, Eq. (4), and for wavelengths λ_i , $i = 1, \dots, m_{\text{O}_4}$ representing the O_4 bands either individually or simultaneously. In the latter case, profile parameters are calculated at one reference wavelength λ_{fm} which, together with Ångström exponents a_i for the conversion to extinction coefficients at λ_i , has to be specified as part of the forward-model parameters \mathbf{b} .

The optimisation problem presented in Eq. (11) is solved using an implementation of the Levenberg–Marquardt (LM) algorithm, which is well suited for this kind of non-linear least-squares problem (see, for example, Nocedal and Wright, 2006 and references therein), and which is provided by the MINPACK library (Moré et al., 1980; Cowell, 1984, <http://www.netlib.org/minpack>). This library contains a trust-region implementation of the Levenberg–Marquardt algorithm, and was chosen here because it includes routines both for the case where derivatives are available and for the case where these are not directly available. In the latter case the Jacobian matrix \mathbf{J} used to calculate the next step correcting \mathbf{x} in each iteration is estimated using forward differences. In the statistical framework the output at the optimum can be used to calculate the covariance matrix $\hat{\mathbf{S}}$ of the result. All

radiative transfer calculations during the iterative fit are carried out online, i.e. without using look-up tables.

The retrieval described so far refers to individual data \mathbf{d} and – unless other sources are available – both the a priori \mathbf{x}_a and the regularisation parameter in Eq. (12) have to be chosen more or less ad hoc. But since a MAX-DOAS measurement usually consists of a series of diurnal data, such an ad hoc choice is in fact only necessary for the very first retrieval of a sequence within a day. We label retrievals and corresponding parameters in this sequence by the index $k = 1, 2, \dots$. The retrieval algorithm for a time series of data $\mathbf{d}(T_l)$ for 1 day implemented here proceeds as follows: it starts with some non-committal choice of the initial parameters $\mathbf{x}_{a,k=1}$ and $\gamma_{k=1}$ (or $\sigma_{a,k=1}$), preferably at a point in the series considered to represent the best measurement conditions/highest information content. The result $\hat{\mathbf{x}}_k$ of the LM iteration (starting throughout with $\mathbf{x}_{0,k} = \mathbf{x}_{a,k}$) is regarded here as valid if the weighted data residual $\|\mathbf{r}_d\|$ lies within a certain range of its expected value

$$\mathbf{r}_{d_k}^T \mathbf{r}_{d_k} = (\mathbf{d}_k - \mathbf{F}(\hat{\mathbf{x}}_k))^T \mathbf{S}_{\epsilon,k}^{-1} (\mathbf{d}_k - \mathbf{F}(\hat{\mathbf{x}}_k)) \leq \Delta^2 \cdot m, \quad (17)$$

where $\Delta = 3$ in this model study. This choice represents a 3σ threshold, but the exact value is not very important in this model study, where the only errors are random measurement errors. If the retrieval fails, a second fit uses the initial a priori parameters. A valid retrieval $\hat{\mathbf{x}}_k$ is used as a priori for the next data point \mathbf{d}_{k+1} such that

$$\mathbf{x}_{a,k+1} = \hat{\mathbf{x}}_k \quad (18)$$

$$\gamma_{k+1} = \frac{\sqrt{m}}{\|\hat{\mathbf{x}}_k\|}, \quad (19)$$

where the latter equation again defines $\sigma_{a,k+1}$. This procedure is somewhat similar to the retrieval scheme using a Kalman filter in Frieß et al. (2006) but without the explicit requirements of the Bayesian interpretation and a model for the evolution of \mathbf{x} with k or time T_l .

Finally, we introduce a modification of the data weights σ_ϵ for the case that relative intensities \tilde{I} are used as additional fit quantities. Choosing fix errors for DSCDs and RIs here similar to Frieß et al. (2006) (the latter model study is for a northward viewing geometry), we observed that, in the presence of noise, for small relative azimuth angles the object function is dominated by contributions from \tilde{I} in such a way that the profile retrieval suffers in terms of height information. We found that including intensities can still improve the retrieval if their contribution is weighted (penalised) relative to the one from the O_4 slant columns. In the following, they are scaled with weights w such that the relative error for the \tilde{I}_i is not smaller than the one for the corresponding ΔS_i . That is, in the notation of Eq. (4)

$$\sigma_{\tilde{I}_i}^w = w^{-1} \sigma_{\tilde{I}_i} \quad \text{with} \quad w = \left(\frac{\sigma_{\Delta S_i}}{\Delta S_i} \right)^{-1} \frac{\sigma_{\tilde{I}_i}}{\tilde{I}_i}. \quad (20)$$

2.4 Error of the retrieval

We first address the error of the retrieval in parameter space, where we have to take into account that the simple profile parameterisation in the form of $k(z, \mathbf{x})$, Eqs. (5), (15) and (16), might not provide an adequate representation of the true extinction profile $k(z)$ (e.g. for an elevated layer or two distinct layers etc.). We here define the parameters \mathbf{x} of the best possible representation as the ones that minimise $k(\lambda_{\text{fm}}, z) - k(\lambda_{\text{fm}}, z, \mathbf{x})$ in a least-squares sense, which thus defines $h(\lambda_{\text{fm}}, z)$ in

$$k(\lambda_{\text{fm}}, z) = k(\lambda_{\text{fm}}, z, \mathbf{x}) + h(\lambda_{\text{fm}}, z). \quad (21)$$

Assuming that the altitude grid of the forward model given by z_j has a sufficient spatial resolution, $h(\lambda_{\text{fm}}, z)$ may be expressed by discrete parameters h_j . We now follow the error analysis in Rodgers (2000, chap. 3), treating \mathbf{h} in the same way as the other forward-model parameters \mathbf{b} , so that the error in the retrieval $\hat{\mathbf{x}}$ can be written as

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{I}_n - \mathbf{A})(\mathbf{x}_a - \mathbf{x}) + \mathbf{R}_d \epsilon + \mathbf{R}_d \mathbf{F}_h \mathbf{h} + \mathbf{R}_d \Delta \mathbf{F}(\mathbf{x}, \mathbf{b}) + \mathbf{R}_d \mathbf{F}_b (\mathbf{b} - \hat{\mathbf{b}}) \quad (22)$$

with derivatives of \mathbf{F} evaluated at $\mathbf{x} = \mathbf{x}_a$, $\mathbf{b} = \hat{\mathbf{b}}$, $\mathbf{h} = \mathbf{0}$ and $\mathbf{R}_d = \mathbf{R}_d(\mathbf{d})$. The first expression on the right-hand side, usually referred to as smoothing error (also regularisation error), was discussed in Sect. 2.2. The second one (perturbation error) describes the propagation of the data error, and the last two expressions are errors due to shortcomings in the formulation of the forward model itself and the choice of its parameters deviating from the optimal ones $\hat{\mathbf{b}}$. The third contribution represents the parameterisation error of k . Equation (22) results from various linearisation of both the forward and inverse model $\hat{\mathbf{x}} = \mathbf{R}(\mathbf{d})$ and therefore might not be accurate, for example, if \mathbf{x}_a is very different from $\hat{\mathbf{x}}$ or if the representation error \mathbf{h} is large. Expectation values of individual contributions to the covariance matrix for Eq. (22) in the statistical framework are discussed in Rodgers (2000, chap. 3). If \mathbf{x}_a and \mathbf{S}_a are good estimates for the true mean and its variances and covariances, then the covariance matrix of $\hat{\mathbf{x}}$ is

$$\hat{\mathbf{S}} = \left(\mathbf{R}_d^T \mathbf{S}_\epsilon^{-1} \mathbf{R}_d + \mathbf{S}_a^{-1} \right)^{-1} \quad (23)$$

if all forward-model errors are negligible.

The error for a function $f(\mathbf{x})$ at $\hat{\mathbf{x}}$ is in first order of $\hat{\mathbf{x}} - \mathbf{x}$ given by $f_x(\hat{\mathbf{x}})(\hat{\mathbf{x}} - \mathbf{x})$. In particular, the variance of the profile $k(z, \hat{\mathbf{x}})$ becomes

$$\sigma_k^2(z, \hat{\mathbf{x}}) = [k_x(z, \hat{\mathbf{x}})]^T \hat{\mathbf{S}} k_x(z, \hat{\mathbf{x}}) \quad (24)$$

under the same circumstances as for Eq. (23). The variance of $H_L(\hat{\mathbf{x}})$ follows correspondingly. The tropospheric (meaning here below z_{TOR}) AOD $\tau = \tau_L + \tau_E$ can be expressed as the special case of a linear function $\mathbf{t}^T \mathbf{x}$, so that

$$\sigma_\tau^2(\hat{\mathbf{x}}) = \mathbf{t}^T \hat{\mathbf{S}} \mathbf{t}, \quad (25)$$

with $\mathbf{t} = (0, 0, 1, 1)^T$ for the four-parameter representation, Eq. (15), and $\mathbf{t} = (0, 0, 1)^T$ for the three parameters in Eq. (16).

We would like to conclude the error discussion with a general remark on the error in data space. Using a linearisation around the optimum with assumptions similar to those leading to Eq. (22), one can derive the following expression for the error of the forward-modelled quantities $\hat{\mathbf{d}} = \mathbf{F}(\hat{\mathbf{x}})$

$$\hat{\mathbf{d}} - \mathbf{d} = (\mathbf{I}_m - \mathbf{A}_d) (\mathbf{F}(\mathbf{x}_a) - \mathbf{d}) + \mathbf{A}_d \epsilon'. \quad (26)$$

Here \mathbf{d} is the ideal, error-free data, $\mathbf{A}_d = \mathbf{F}_x \mathbf{R}_d$ the data-averaging kernel matrix or influence matrix, and ϵ' represents all measurement and forward-model errors. Using the picture of the linear case for the sake of clarity, one essential difference between the errors in Eqs. (22) and (26) lies in the null spaces of the maps \mathbf{F}_d and \mathbf{R}_d . For the least-squares retrieval, $\mathbf{I}_n - \mathbf{A}$ and $\mathbf{I}_m - \mathbf{A}_d$ project onto the data and parameter null spaces, respectively. In the over-determined, full-rank case the former is empty, whereas the latter is not, so that the potential of any discrepancy between the model and data leading to large data residuals is higher than in the under-determined case. In this sense the data residual, Eq. (26), can be used as a criterion for the consistency of model and error assumptions of the retrieval.

3 Results

3.1 Model and retrieval settings

In this model case study we consider mainly the retrieval of aerosol extinctions for the situation of urban air pollution monitoring, i.e. medium to high tropospheric aerosol optical depths. Unless stated otherwise in the following, the most important settings for the model experiment, the forward model and the retrieval are contained in Table 1. Representing restrictions of the view for a location in an area of high buildings, we choose the lowest elevation angle as 2° and the orientation of the telescope as west (north). Solar zenith and relative solar azimuth angles θ and ϕ_{rel} shown in Fig. 2 are arbitrarily chosen corresponding to a midsummer day (5 June 2010) in Shanghai ($31^\circ 12' \text{ N}$, $121^\circ 30' \text{ E}$). Apart from the extinction profile below the retrieval height z_{TOR} , all other aerosol optical properties are fixed throughout as specified in Table 1. The upper-tropospheric/stratospheric aerosol follows a standard background aerosol taken from the LOWTRAN database (Kneizys et al., 1988) with extinction coefficient $\sim 0.0115 \text{ km}^{-1}$ at 5 km altitude and a single-scattering albedo ω_0 of 0.99 (at 477 nm). The single-scattering albedo ω_0 and the Henyey–Greenstein parameter g_{HG} below z_{TOR} are set to values that are typical for air pollution containing absorbing aerosol (e.g. Lee et al., 2007; Garland et al., 2008, 2009). The Ångström parameter used here in the simultaneous fit of the first two O_4 bands to convert extinctions from the reference wavelength $\lambda_{\text{fm}} = 477$ to 360 nm is set to 1.25.

Table 1. Settings for model experiments and retrieval unless stated otherwise in the text. The model errors σ_{ϵ_0} for the differential optical depths and the relative intensities are adopted from Frieß et al. (2006) (for the wavelength 550 nm).

Model experiment	
$\alpha(m_\alpha)$	2, 4, 6, 8, 10, 12, 15, 90° (7)
ϕ_{obs}	west (north)
θ, ϕ	Shanghai, 5 June 2010 (see Fig. 2)
$\lambda_{\text{O}_4 i}(m_{\text{O}_4})$	360, 477 nm (2)
$\sigma_{\epsilon_0}(\Delta\tau)$	5×10^{-4}
$\sigma_{\epsilon_0}(\bar{I})$	5×10^{-4}
Forward model	
λ_{fm}	477 nm
z_{TOR}	5 km
$z_i (< z_{\text{TOR}})$	$\Delta z = 0.1 \text{ km}$
O_4 cross section	Greenblatt et al. (1990)
Model profiles	
aerosol $z < z_{\text{TOR}}$	$\omega_0 = 0.95, g_{\text{HG}} = 0.68, \text{Å} = 1.25$
aerosol $z > z_{\text{TOR}}$	LOWTRAN background
$p, T, \text{trace gases}$	standard atmosphere
Retrieval	
$m = m_\alpha m_{\text{O}_4} (\times 2)$	14 (28)
n	3 or 4
$\mathbf{x}_{a,0}$ (initial)	$0.1 \text{ km}^{-1} \cdot \exp(-z/1 \text{ km})$

The number n of fit parameters is three or four. The initial a priori $\mathbf{x}_{a,0}$ is an exponential with ground extinction coefficient $k(0) = 0.1 \text{ km}^{-1}$ and scale height $\zeta = 1 \text{ km}$.

We use the O_4 absorption cross section according to Greenblatt et al. (1990), but do not take into account the discrepancy between measured and modelled O_4 slant column densities reported in Clémer et al. (2010) (the investigation of this issue using results of this study will be part of our future work). Measurement errors are assumed to take the form of uncorrelated, Gaussian-distributed noise with standard deviation σ_{ϵ_0} as in Frieß et al. (2006).

3.2 Example for SVD analysis of the forward model

To examine the conditions under which a measurement of the type just described allows for a least-squares fit of the three (four) parameters without regularisation, we first consider the singular value decomposition of the linearised model, Eq. (8), for two profile shapes and levels for the optical depths of about 1.2 and 0.12, respectively. The first profile type is a well-mixed ground layer of height $H_L = 1.5 \text{ km}$ (O_4 DSCDs for these profiles are shown in Fig. 2); the second one is an exponential one.

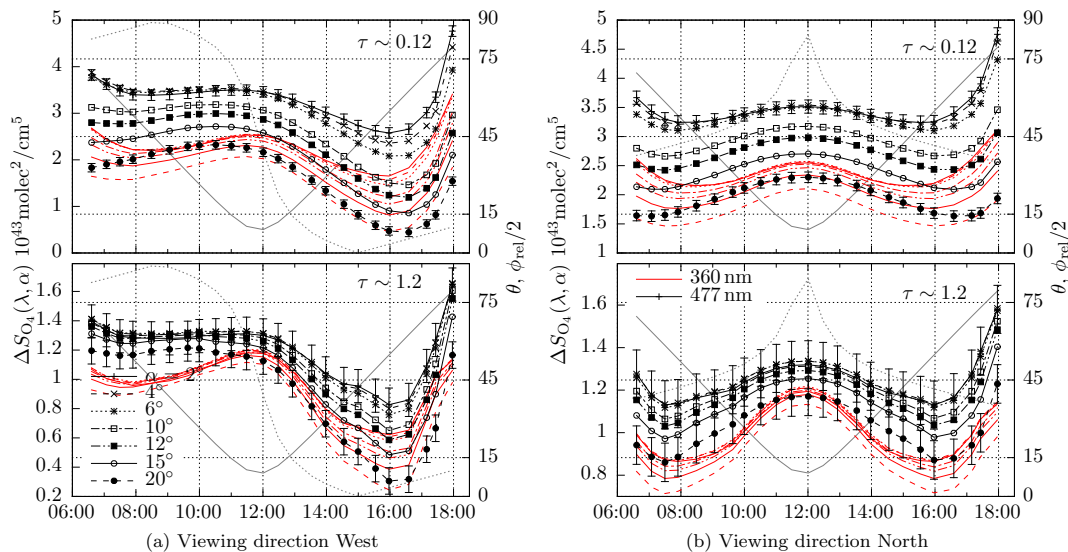


Fig. 2. Differential slant column densities of O_4 for the model experiment specified in Table 1 and a well-mixed ground layer of height $H_L = 1.5$ km for viewing direction west (a) and north (b). Top row with low aerosol load, and bottom row with high aerosol load. Layer parameters (see Fig. 1) are $k_L = 0.05 \text{ km}^{-1}$, $\tau_L = 0.075$, $\tau_E = 0.048$ (top panels) and $k_L = 0.5 \text{ km}^{-1}$, $\tau_L = 0.75$, $\tau_E = 0.48$ (bottom panels) at $\lambda_{\text{fm}} = 477 \text{ nm}$. The exponential part of the profile has a scale height $\zeta = 1 \text{ km}$. Solar zenith and relative solar azimuth angles θ and ϕ_{rel} referring to Shanghai ($31^\circ 12' \text{ N}$, $121^\circ 30' \text{ E}$) on 5 June 2010 are here and in following figures indicated by the solid and dashed line, respectively.

3.2.1 Singular values

Figure 3 shows the three non-zero singular values for the three-parameter representation, Eq. (16), for viewing direction west (Fig. 3a) and north (Fig. 3b). For both levels of aerosol optical depths the range of singular values comprises several orders of magnitude, as is characteristic of ill-posed problems and especially for the higher aerosol extinction; the smaller modes show a distinct variation with solar angles, indicating changes in the information content related to different profile parameters. Only singular values whose contribution according to Eq. (8) is above the measurement error represent a degree of freedom of \mathbf{x} that can be distinguished from noise. For the profiles with higher aerosol extinction (bottom panel in Fig. 3), where Fourier coefficients $\mathbf{v}_i^T \mathbf{x}$ are of the order 0.1 to 1, a measurement error of about $0.1 \times 10^{43} \text{ molec}^2 \text{ cm}^{-5}$ would mean that the contribution of the smallest singular value is comparable to the size of the noise. Singular values for the lower optical depths (top panel in Fig. 3) are about 10 times larger, but the Fourier coefficients are reduced by a similar factor. To avoid that the smallest mode affects the retrieval through error magnification, the least-squares formulation thus needs regularisation – at least for the errors and profiles assumed here. This will be further substantiated by model retrievals below, where it will be furthermore shown that the seemingly small variations of the second singular value for the ground layer profile have in fact significant influence on the retrieval.

3.2.2 Singular vectors in state space

To interpret the three modes distinct in all singular value decompositions of Fig. 3, the corresponding singular vectors for the layer profile in the top panel of this figure are shown in Fig. 4. The vector for the dominant first mode is almost completely in direction of the layer extinction coefficient k_L , whereas it does not contribute to the lower modes. This means that this parameter is both robust with respect to noise and the choice of the regularisation parameters. The second mode is modulated similarly to the differential slant columns. For westward viewing direction, it represents more or less the sum of τ_L and τ_E , i.e. the optical depth, except for times when the solar zenith (and the relative azimuth) angle is small. This means that the total optical depth $\tau_L + \tau_E$ is a relatively well determined parameter, unless the noise is so large or the signal so low that the second mode becomes affected (as shown for this example in Sect. 3.4). In the same way, it becomes clear that for the northward viewing direction the optical depth for this layer profile is less well determined for low solar elevation, illustrating the advantage of observation in different azimuthal directions.

3.3 Pure least-squares retrieval of a polluted ground layer

We now consider the unregularised least-squares problem Eq. (7) for the polluted ground layer with optical density $\tau \sim 1.2$.

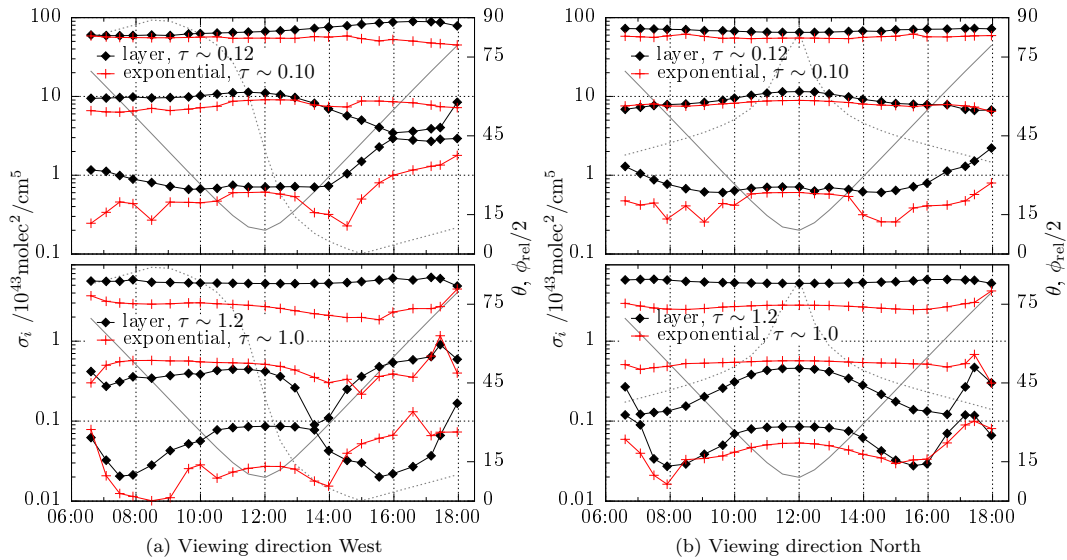


Fig. 3. The three non-zero singular values σ_i of the Jacobian matrix for the forward model (without \tilde{I}) using the profile parameters k_L , τ_L and τ_E for the same ground layer profiles as in Fig. 2. For comparison, also shown are the first three non-zero singular values σ_i for exponential profiles with scale height $\zeta = 1$ km and ground extinction coefficients k_L of 0.1 and 1 km^{-1} , respectively. (a) Viewing direction west, and (b) north.

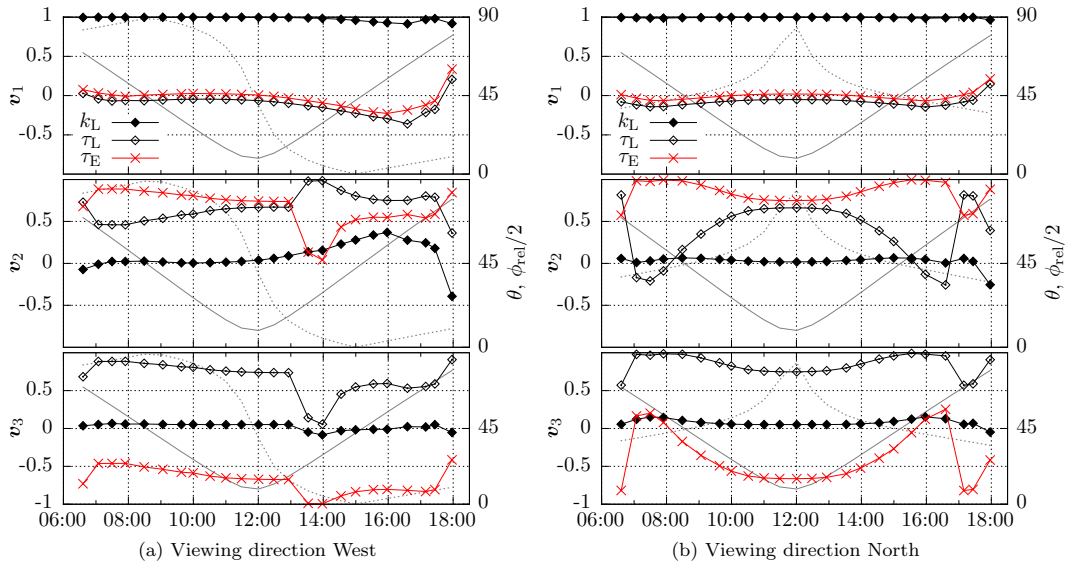


Fig. 4. Singular vectors related to the singular values in Fig. 3 for the ground layer profile with $\tau \sim 1.2$ (Fig. 2, bottom row), again for viewing directions west (a) and north (b).

3.3.1 χ^2 landscape

In the linear case the object function $\chi^2(\mathbf{x})$ in Eq. (7) describes ellipsoids in parameter space for a given value $c^2 = \chi^2$. Figure 5 illustrates how the non-linear case behaves, again for the ground layer profile of the previous section with $\tau \sim 1.2$, $k_L = 0.5 \text{ km}^{-1}$, $H_L = 1.5 \text{ km}$ (Fig. 2), this time represented by the four parameters $k_{L1} = k_{L2} = 0.5 \text{ km}^{-1}$, $\tau_L = 0.75$ and $\tau_E = 0.48$. Figure 5a shows the dependency of χ^2 on

the layer extinctions, and Fig. 5b its variation with optical depths. Both graphs illustrate that only in a relatively close neighbourhood of $\hat{\mathbf{x}}$ the linearisation is a sufficient approximation. In agreement with the discussion at the end of Sect. 3.2, the minimum is quite sharp in the sense that the residual function χ^2 grows rapidly with distance from $\hat{\mathbf{x}}$, and thus a sufficiently fine resolution in x_j is needed in the forward-model calculation in order to achieve an accurate value of the minimum. The function χ^2 shows the strongest

variation in the ground extinction k_L ; the penalising effect of the ground layer parameters is considerably larger than the exponential one.

3.3.2 Monte Carlo simulations

The previous discussion did not involve any errors. To show how even moderate measurement errors affect the solution of the unregularised least-squares problem, Eq. (7), we consider the same profile – i.e. a relatively high aerosol optical depth – for the westward viewing direction (the corresponding example for low optical depth will appear in the next Sect. 3.4). Without measurement errors the inverse problem for both parameterisations (Eqs. 16 and 15) can be solved almost exactly for all solar positions, as shown for the three-parameter representation with its parameters k_L , τ_L and τ_E in Fig. 6. A Monte Carlo simulation of the retrieval with uncorrelated Gaussian-distributed errors of variance $\frac{1}{2}\sigma_{\epsilon_0}$ (about 5 %) shows, as anticipated, that the layer extinction coefficient k_L corresponding to the first mode of the singular value decomposition is relatively stable against noise: $\sim 5\%$ error on the data has an effect of less than 10 % on the retrieval, except for solar positions with small θ and ϕ_{rel} (between about 13:00 and 15:00 LT) pointed out in the previous discussion of the singular vectors. The largest effect on any of the retrieval parameters occurs for the optical depth of the exponential τ_E . Here the magnification in the unregularised inversion of an error of only 5 % can lead to errors in the retrieval of more than 100 %, which merely reflects the fact that it is the least well determined parameter for the MAX-DOAS retrieval. The error of the fully retrieved AOD $\tau = \tau_L + \tau_E$ is considerably smaller than the individual errors, reflecting a strong correlation of τ_E and τ_L (see Eq. 25). As anticipated in the discussion of the singular value decomposition, the error of τ is largest (up to 50 %) when the second singular value is smallest (Fig. 3a, bottom panel). The 1- σ variation obtained from the Monte Carlo calculation generally agrees well with the variance of \hat{S} calculated for error-free data at the true minimum. The reason for their discrepancy around 14:00 LT is that the retrieval can produce exponential profiles as acceptable solutions for which the estimate \hat{S} obtained from linearisation does not hold. Acceptable here refers to the data residuum in Eq. (17), shown in Fig. 7 (red). Taking into account the limited number of random samples ($N = 100$), its mean value agrees well with the expected value of \sqrt{m} for all solar positions. To illustrate that this is not necessarily the case if explicit or implicit assumptions on the data or model error are incorrect, we also show χ for the error-free retrieval under the assumption that $\sigma_{\epsilon} = 10^{-3}\sigma_{\epsilon_0}$. The residual could in principle be zero. But for small ϕ_{rel} it lies outside the 2- σ range, indicating that numerical errors in the forward model are in fact larger than the assumed data error.

3.4 Regularised least-squares retrieval of ground layer profiles

In this section we demonstrate that the regularisation criterion according to Eqs. (12) and (19) is indeed a useful choice of the regularisation parameter and that the update of the a priori according to Eq. (18) works consistently in the presence of noise. We consider again the retrieval of $\mathbf{x} = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ for the two ground layer profiles with AOD $\tau \sim 0.12$ and $\tau \sim 1.2$ and westward viewing direction. The retrieval procedure is as described at the end of Sect. 2.3 except that here we consider different ad hoc choices of a constant regularisation parameter.

3.4.1 Information content and averaging kernels

To get a rough idea of the potential influence of the regularisation parameter on the retrieval for arbitrary (but appropriate) a priori \mathbf{x}_a in the present scenario, we first look at the information content in terms of the degrees of freedom d_s and at the averaging kernels \mathbf{A} . Figure 8 shows d_s for both profiles and two levels of noise calculated for the error-free retrieval at $\hat{\mathbf{x}}$. The range of the regularisation parameters chosen corresponds to the size of the true parameters x_j ($\sigma_a = 0.05$ for $\tau \sim 0.1$ and $\sigma_a = 0.5$ for $\tau \sim 1$). The general tendencies of d_s in σ_a and σ_{ϵ} can be understood from the special case in Eq. (14). The values of d_s become larger if either the signal-to-noise ratio increases or the regularisation becomes weaker (σ_a larger). For given σ_{ϵ} , the signal-to-noise ratio $\sigma_i/\sigma_{\epsilon}$ decreases, and consequently d_s is considerably smaller for the profile with larger AOD $\tau \sim 1$. Depending on the value of σ_a , a measurement at AOD around $\tau \sim 0.1$ can be expected to hold between 2 and almost 3 degrees of freedom (DOFs) for a noise level given by $\sigma_{\epsilon} \sim 10\%$, whereas for high AODs around $\tau \sim 1$ this number is 1 to 2.5. The temporal variation of d_s with solar angles is similar to that of the condition number (ratio of largest to smallest singular value) except when the second singular value varies significantly. Thus the minimum of this singular value for $\tau \sim 1$ at small θ and ϕ_{rel} (around 14:00 LT in Fig. 3a, bottom panel) occurs again as a distinct reduction in information content. As argued before, the second singular value is related to the AOD $\tau = \tau_L + \tau_E$, and the corresponding lack of information on τ clearly appears in the averaging kernels in Fig. 9.

This figure shows the rows of \mathbf{A} for $\sigma_a = 0.1$ in the case of the ground layer profile with $\tau \sim 0.12$ and the range $\sigma_a = 0.1 \dots 0.5$ for $\tau \sim 1.2$ (this choice of σ_a will become clear shortly; for $\tau \sim 0.1$ only one value of the regularisation parameter σ_a is shown because \mathbf{A} is less sensitive to σ_a than in the case of higher AOD $\tau \sim 1$). As anticipated in the discussion of the singular values in Sect. 3.2 for the profile with $\tau \sim 1$, the layer extinction coefficient k_L is a well-resolved parameter almost unaffected by regularisation. For low AOD $\tau \sim 0.1$ the averaging kernels for the optical depths τ_L and τ_E are “most diagonal” for large θ . Vice versa, the profile

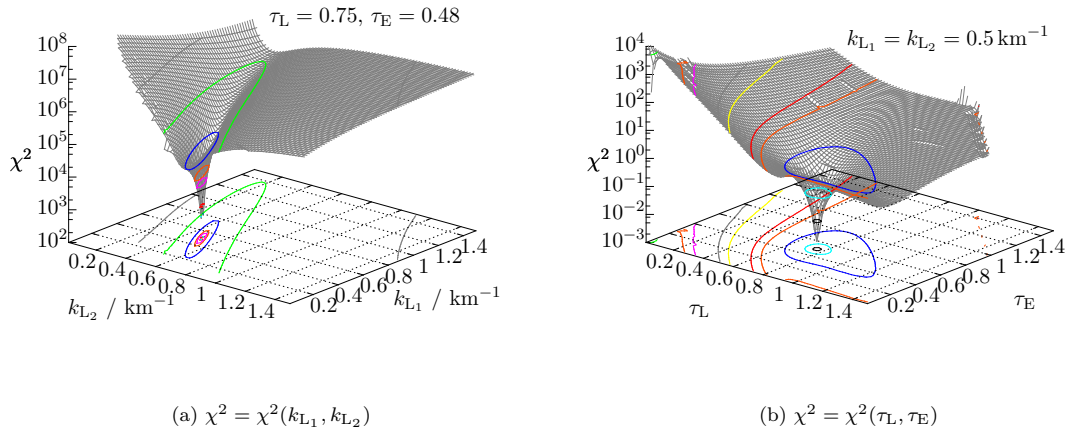


Fig. 5. Error-free least-squares object function (residual) $\chi^2(\mathbf{x})$, Eq. (7), for the well-mixed ground layer with $\tau \sim 1.2$ (Fig. 2) represented by the four parameters k_{L1} , k_{L2} , τ_L and τ_E , Eq. (15). **(a)** For constant τ_L , τ_E , and **(b)** for constant k_{L1} , k_{L2} . The viewing direction is west and $\theta \sim 10^\circ$, $\phi_{rel} \sim 170^\circ$.

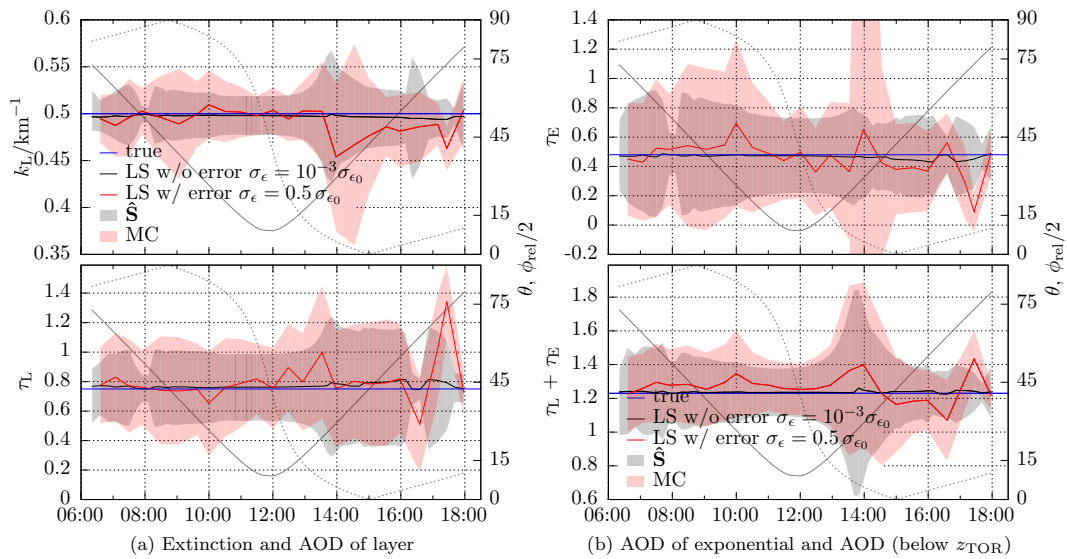


Fig. 6. Least-squares retrieval (without \tilde{I}) of $\mathbf{x} = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ for the well-mixed ground layer profile with true parameters $k_L = 0.5 \text{ km}^{-1}$, $\tau_L = 0.75$ and $\tau_E = 0.48$ (see Fig. 2, bottom row) and viewing direction west. The retrieval without errors added to the DSCDs assumes a precision of less than $10^{-2} \%$. The retrieval with random noise of about 5% added to the SCDs is the mean value of 100 Monte Carlo (MC) runs, its standard deviation given by the red shaded area. The grey shaded area indicates the corresponding 1- σ variation calculated from $\hat{\mathbf{S}}$, Eq. (23) (around the error-free retrieval).

information is lowest for small θ . Apart from the layer extinction k_L , the averaging kernels for higher AOD strongly depend on the regularisation level. The same regularisation parameter σ_a as before leaves almost only one DOF, namely k_L .

3.4.2 Monte Carlo simulations

We now turn to the retrieval results of simulated measurements with errors for the above fixed values of the regularisation parameter. Averages and standard deviations are again obtained from Monte Carlo simulations for the measurement

errors given in Table 1. Starting with the ground layer profile with low AOD $\tau \sim 0.1$, $(k_L, \tau_L, \tau_E) = (0.05 \text{ km}^{-1}, 0.075, 0.048)$ and singular values as in Fig. 3a (top panel), it can be expected from the discussion following Eq. (11) that for errors of the DSCD around $0.1 \times 10^{43} \text{ molec}^2 \text{ cm}^{-5}$ the regularisation parameter γ has to be larger than around 1 to be effective for the smallest singular value. Figure 10 contains the result of modelled retrievals for different levels of regularisation. The pure least-squares fit suffers from large errors in the retrieved optical depths. Regularisation parameters σ_a larger than about 0.1 (γ less than 10) have little impact on the retrieval. Values smaller than around 0.01 affect

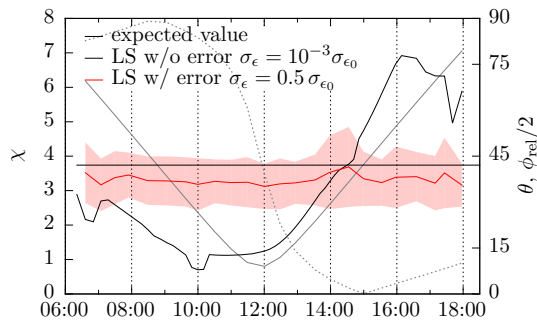


Fig. 7. Residual for the retrieval shown in Fig. 6.

the lowest two singular values, forcing corresponding parameters x_j strongly towards the a priori (“over-regularisation”). A choice of γ around 20 appears to be a good compromise between minimising the perturbation error and the bias introduced by regularisation (in this case these are the only contributions to the retrieval error in Eq. 22). In the statistical framework this choice of $\sigma_a = \gamma^{-1} = 0.05$ would correspond to an a priori variance being of the order of the true, unknown parameters.

Turning to the profile with high AOD $\tau \sim 1$ and parameters $(k_L, \tau_L, \tau_E) = (0.5 \text{ km}^{-1}, 0.75, 0.48)$ and assuming the same (or similar) measurement errors as before, the previous argument on the basis of the smallest singular value would again lead to choosing σ_a smaller than about 1. Results for the retrieval of Monte Carlo simulations in Fig. 11 suggest an optimal value of $\sigma_a \sim 0.2$; larger values increase the bias without real benefit for the variance of the perturbation error. In terms of a priori variances, this choice would again be similar to the size of the x_j .

The “optimal” regularisation parameters $\sigma_a \sim 0.05$ ($\gamma \sim 20$) for $\tau \sim 0.1$ and $\sigma_a \sim 0.2$ ($\gamma \sim 5$) for $\tau \sim 1$ found in the modelled retrievals agree reasonably with the values of $\gamma \sim 33$ (for $\tau \sim 0.1$) and $\gamma \sim 3.3$ (for $\tau \sim 1$) according to Eqs. (12) and (19) if one inserts the true parameters x . Since this criterion for the choice of the regularisation parameter is not very sophisticated, we are not too concerned about precise values of the “best” parameter. However, we want to point out that for the range of aerosol extinctions considered here, the choice of a constant regularisation parameter is not a good compromise (e.g. setting $\gamma = 10$ in both examples would hardly have any regularising effect for the profile with $\tau \sim 0.1$ and over-regularise the solution for $\tau \sim 1$). The same conclusion holds, of course, for the a priori uncertainties σ_a , so that the approach to choose a non-committal σ_a reflecting the natural uncertainty of the aerosol extinctions is similarly inadequate.

Comparing the “optimal” regularisation parameters with the corresponding d_s in Fig. 8 shows that, depending on the solar angles, for low $\tau \sim 0.1$ between 70 and 80 % out of the maximal 3 degrees of freedom are given by the measurement itself; the remaining 20–30 % are affected by noise, for

high $\tau \sim 1$ between 30 and 60 %. These numbers are mostly larger than the relative errors, and here do not seem to allow for any conclusion back on the choice of the regularisation parameter.

3.5 Use of relative intensities as fit quantities

The advantage of including (relative) intensities into the optimal estimate of aerosol extinction profiles has been demonstrated for the parameterisation by discrete layers and northward orientation of the telescope by Frieß et al. (2006). It can be expected that the retrieval of three (four) profile shape parameters benefits in a similar way.

3.5.1 Information content and averaging kernels

For ideal, error-free measurement data for the polluted layer with $\tau \sim 1.2$, this is indeed confirmed by the degrees of freedom for the measurement d_s shown in Fig. 12. Compared to the case without intensities (open diamonds in this figure), d_s enhances by 0.5 to 1.5 degrees of freedom when (unscaled) intensities are added to the fit (filled diamonds and squares in this figure). The highest increase of information occurs for large relative solar azimuth angles ϕ_{rel} in the morning; the lowest for small solar zenith angles θ around noon. The corresponding averaging kernels in Fig. 13 illustrate which profile parameters contribute to this increase. Comparing again the retrieval without and with (unscaled) \tilde{I} , Fig. 13b (top panel) indicates that for large and medium ϕ_{rel} , especially the upper, exponential part of the profile given by τ_E can be retrieved more accurately, whereas for small ϕ_{rel} the lower ground layer part, τ_L , profits most from the additional information provided by \tilde{I} (Fig. 13a, bottom panel). The change of the averaging kernels is less pronounced for small θ ; here mainly τ_E is better resolved. The total tropospheric AOD τ can now be retrieved reliably for all solar positions.

3.5.2 Monte Carlo simulations with relative intensities

Taking into account measurement errors for the simulated data of slant column densities and relative intensities according to Table 1 changes the situation as summarised in Fig. 14. The graphs are the result of Monte Carlo simulations for the ground layer profile with $\tau \sim 1.2$ similar to the previous ones but now including relative intensities in the retrieval. The retrieval of the aerosol optical depth $\tau = \tau_L + \tau_E$ is indeed greatly improved both with respect to its mean value and its variance for all solar positions when compared to the case without relative intensities in Fig. 11b (bottom panel). However, the retrieved profile parameters k_L , τ_L and τ_E do not show the behaviour expected from the above analysis of the information content for the error-free data. For very large relative azimuth angles ($\phi_{\text{rel}} > \sim 150^\circ$ in this example) all parameters are more or less insensitive to the choice of the regularisation parameter (within the range of $\sigma_a = 0.1 \dots 0.5$), showing similar variance and bias, which for τ_E is actually

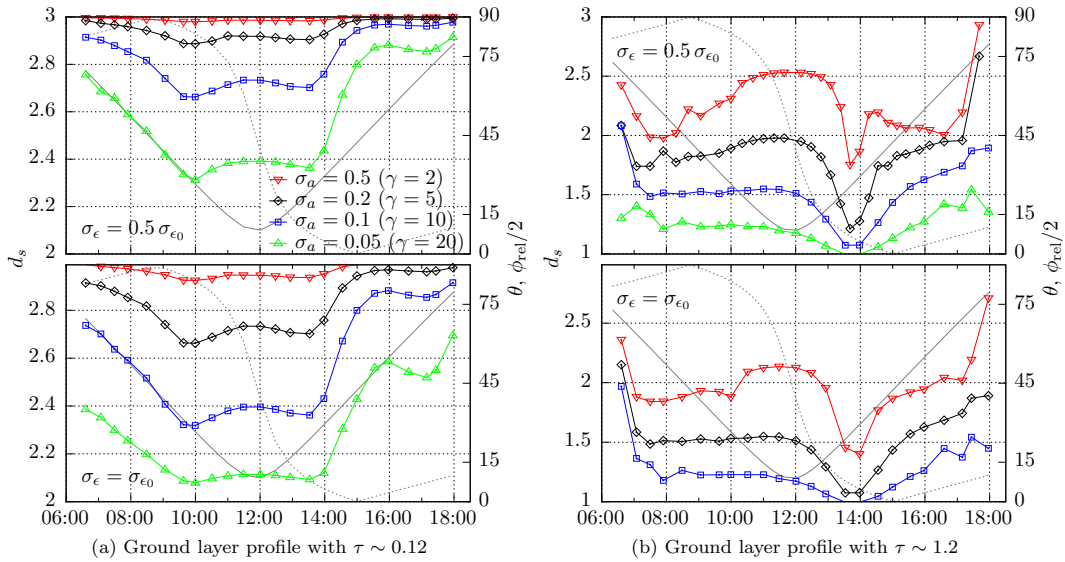


Fig. 8. Degrees of freedom of the signal $d_s = \text{tr}(\mathbf{A})$ for the regularised least-squares retrieval, Eq. (11), of $\mathbf{x} = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ for the ground layer profiles with (a) $\tau \sim 0.12$ and (b) $\tau \sim 1.2$ (see Fig. 2) for different levels of regularisation and levels of measurement noise (viewing direction west).

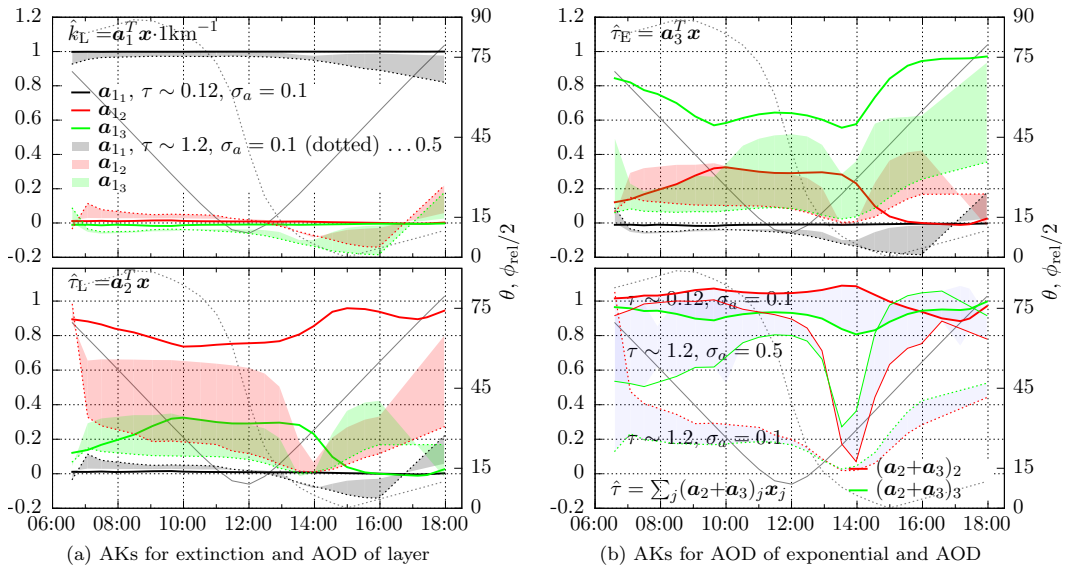


Fig. 9. (a) and (b), top panel: row vectors \mathbf{a}_i of the averaging kernel (AK) matrix for the regularised least-squares retrieval of $\mathbf{x} = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ as in Fig. 8. For the ground layer profile with $\tau \sim 0.12$ (bold lines) only the regularisation parameter $\sigma_a = 0.1$ is shown. For $\tau \sim 1.2$ the range of parameters $\sigma_a = 0.1 \dots 0.5$ is indicated by the shaded areas. (b) (bottom panel) illustrates how well the AOD $\tau = \tau_L + \tau_E$ can in principle be reconstructed. Ideally, the contributions from τ_L (red) and τ_E (green) should be equally 1.

larger than without intensities. With ϕ_{rel} getting smaller the parameters become increasingly sensitive to regularisation if the weights of \tilde{I} are not modified, i.e. $\sigma_{\tilde{I}}$ not scaled. Irrespective of the regularisation parameter or weighting scheme for \tilde{I} , the AOD τ_L and the height H_L of the layer are underestimated for $\phi_{\text{rel}} > \sim 90^\circ$, whereas they become overestimated for $\phi_{\text{rel}} < \sim 90^\circ$ if \tilde{I} is not specially weighted. Scaling $\sigma_{\tilde{I}}$ as proposed in Eq. (20) has only a slight effect on the profile

parameters for large ϕ_{rel} , but improves their retrieval significantly for small ϕ_{rel} , and in this particular example of high AOD is even more accurate than for large relative azimuth angles.

The dependency of the error-afflicted retrieval on the solar azimuth angle ϕ_{rel} in this example is, of course, related to a strong increase in the signal-to-noise ratio of the relative intensities for smaller ϕ_{rel} when looking into the direction of

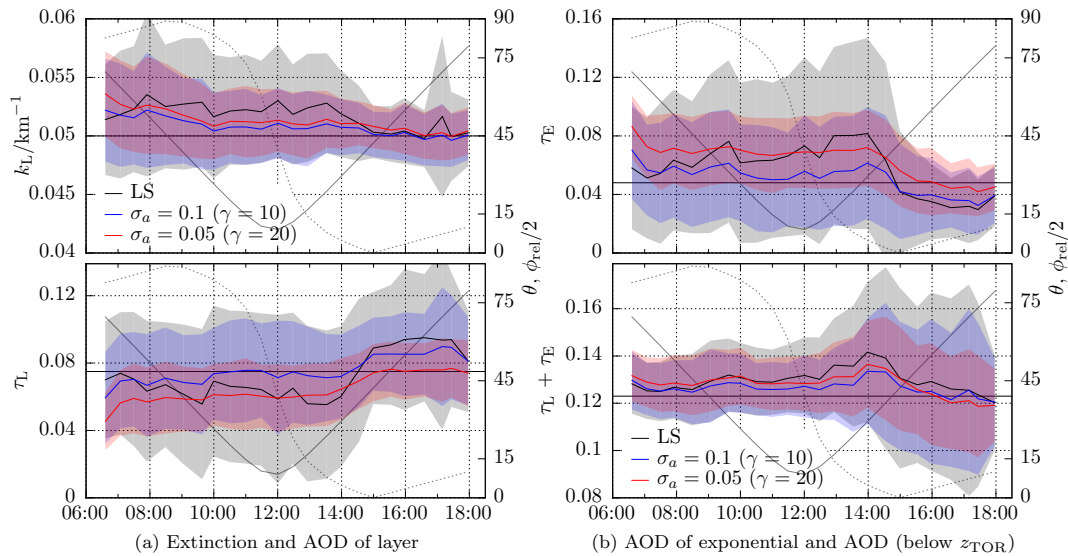


Fig. 10. Retrieval of $x = (k_L \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ for the ground layer profile with $\tau \sim 0.12$ (Fig. 2, top row, viewing direction west) for the unregularised (LS) and regularised problem and a random measurement error with $\sigma_\epsilon = \sigma_{\epsilon_0}$ (Table 1). Lines indicate averages of 100 samples, and shaded areas the standard deviations. Compare to Fig. 3 (a, top panel).

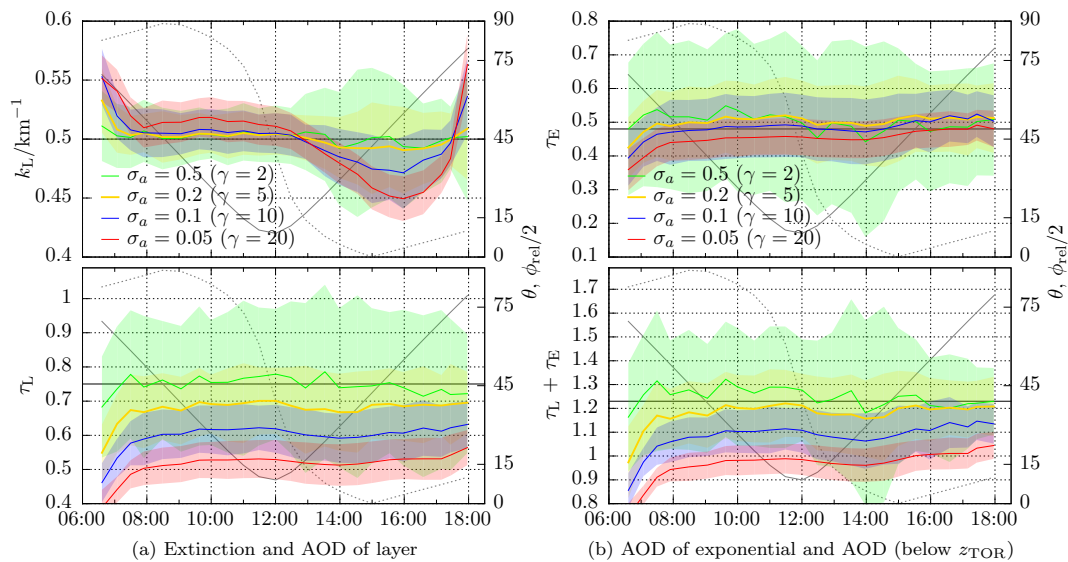


Fig. 11. As Fig. 10 for the ground layer profile with $\tau \sim 1.2$. See also Fig. 3 (a, bottom panel) and Fig. 6.

the sun. The fact that the contribution from the relative intensities to the cost function in Eq. (11) may outweigh the one from the differential slant columns, which carry most of the profile information, does not pose a problem in itself if measurement errors can be neglected. If measurement errors cannot be neglected, one first has to take into account that the two data sets of differential column densities and relative intensities are assumed to be completely uncorrelated within themselves and with respect to each other. In order to see to what extent the neglect of correlations between DSCDs and RIs in their relative weighting in the object function

becomes relevant, we look at the statistics of the data residual in Fig. 15. If all model assumptions were correct, the average data residual for the Monte Carlo runs would be around its expected value $\sqrt{m} \sim 5.3$, similar to Fig. 7. With the exception of small solar zenith angles θ , the actual mean residuals are significantly larger, and for small relative solar azimuth angles ϕ_{rel} frequently exceeds the threshold for valid retrievals in Eq. (17). The sometimes high number of invalid results distorts the statistics and causes the jumps in some of the curves in Fig. 15.

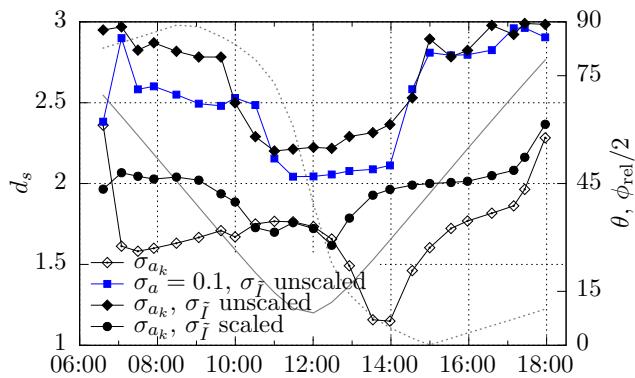


Fig. 12. As Fig. 8, but here for the ground layer profile with $\tau \sim 1.2$ and choice of the regularisation parameter as in Eq. (19). Shown are cases without \tilde{I} (open symbols), with \tilde{I} and fixed $\sigma_{\tilde{I}}$ (squares and diamonds) and scaled $\sigma_{\tilde{I}}$ according to Eq. (20) (filled circles).

3.6 Retrieval of profile time series

The retrieval algorithm of the previous section for static, well-mixed ground layer profiles is now applied in two examples of profile time series with Gaussian functions serving as model profiles. In the first example the series of profiles shown in Fig. 16a (top panel) starts with low AOD $\tau \sim 0.063$ at large relative solar azimuth ϕ_{rel} . The maximum extinction $k_0 = 0.05 \text{ km}^{-1}$, its height $z_0 = 1.5 \text{ km}$ and the half width $\sigma_0 = 0.5 \text{ km}$ of the Gaussian profile evolve linearly in time such that at the other end of the time series for small ϕ_{rel} , these parameters are $k_0 = 1.5 \text{ km}^{-1}$, $z_0 = 0 \text{ km}$ and $\sigma_0 = 1 \text{ km}$. The AOD increases by a factor of about 30 to $\tau \sim 1.86$. The second example of profile time series consists of the same profiles in reverse order in time (see Fig. 18a (top panel)) i.e. from high AOD for long light paths to low AOD for short light paths. The profiles in the forward model are represented now by the four parameters in $\mathbf{x} = (k_{L_1} \cdot 1 \text{ km}, k_{L_2} \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ (see Fig. 1) to be retrieved, and Monte Carlo simulations of measurements are carried out in the same manner as before.

3.6.1 Example for d_s decreasing with time

Statistics for the retrieval of the first profile series are illustrated in Fig. 16 for four different settings of the retrieval process. The first setting uses a fixed regularisation parameter $\sigma_a = 0.1$ – this value was found to be a reasonable choice for the range of profile parameters in Sect. 3.4 – and no relative intensities in the data vector. The second retrieval uses a variable regularisation parameter according to Eqs. (12) and (19), and the third and fourth additionally include relative intensities with unscaled and scaled weights. Values for σ_{a_k} range from less than about 0.02 for the lowest AOD to about 0.4–0.6 for highest AODs. Depending on the choice of the regularisation scheme, one obtains different numbers for the degrees of freedom of the signal d_s , but in either case

the fact that extinctions increase and the lengths of the light paths generally decrease within the time series results in decreasing signal-to-noise ratios (assuming constant measurement errors). For the first retrieval scenario with constant $\sigma_a = 0.1$ (green in this figure), d_s goes down continuously from around 3.8 for large ϕ_{rel} in the morning to only 0.8 for small ϕ_{rel} in the evening. For the second retrieval with variable σ_{a_k} (orange) the corresponding numbers are about 3.8 and 2.3.

In this case, it therefore makes sense to start the retrieval algorithm at the first data point with largest θ and ϕ_{rel} , and since the different choices of regularisation and weighting perform in general similarly to the previous constant profile examples, we here concentrate on some aspects we consider most noteworthy. The first thing to notice is that the reconstruction of the first two or three profiles consisting of elevated layers with low extinctions does not work well for the regularisation with σ_{a_k} , especially when including \tilde{I} . The small values of σ_{a_k} at this point possibly over-regularise the problem and result in exponential (for scaled $\sigma_{\tilde{I}}$, red) or no (for unscaled $\sigma_{\tilde{I}}$, blue) solutions, which might suggest that for very low AODs the regularisation criterion in Eqs. (12) and (19) is not appropriate. Another explanation is that the true profile $k(z)$ cannot be well represented by \mathbf{x} . The subsequent retrievals, however, confirm the advantage of the variable regularisation over the fixed $\sigma_a = 0.1$ as illustrated by the examples in Fig. 17. The latter now over-regularises and produces profiles that, in terms of standard deviations, are partly inconsistent with the true profile parameters. Without intensities the AOD is underestimated for small ϕ_{rel} , as before, due to the lack of information, i.e. low signal-to-noise ratio. A tendency to overestimate the extinction coefficient k_{L_2} evident in Figs. 18a (top panel) and 17 occurred frequently, also for retrievals of well-mixed ground layers parameterised by $\mathbf{x} = (k_{L_1} \cdot 1 \text{ km}, k_{L_2} \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ – usually accompanied by an underestimation of the ground value. The data residuum $\|\mathbf{r}_d\|$ behaves similarly to the retrievals shown before and, except for some violations of the validity criterion (Eq. 17) for the setting with σ_{a_k} and unscaled $\sigma_{\tilde{I}}$, in practice does not allow for discrimination between good or bad retrievals.

3.6.2 Example for d_s increasing with time

While for all profile series so far the retrieval sequence of one day started with data in the morning, the second example in Fig. 18a (top panel) will now underline how important the choice of the sequence of retrievals is in order to make optimal use of varying information content of the measurement and thus quality of the empirical a priori. Analysing the information content expressed in d_s for the variable regularisation with σ_{a_k} (without \tilde{I}) both in forward and backward direction, it turns out that d_s has a maximum value of almost 4 when starting with the last data point in time. It decreases to a value of about 2 at the first data point when going through

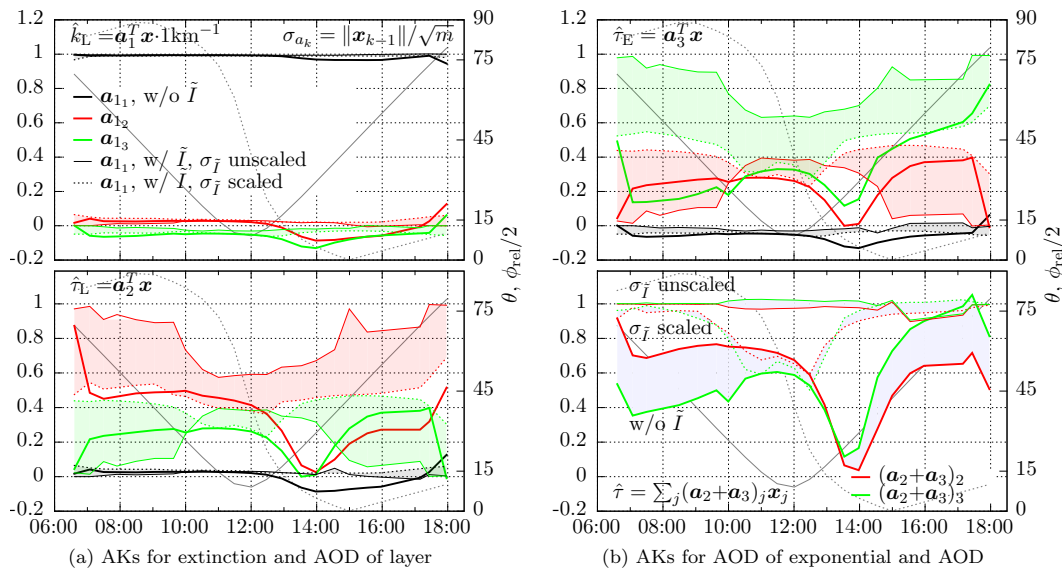


Fig. 13. As Fig. 9, but here for the ground layer profile with $\tau \sim 1.2$ and choice of the regularisation parameter as in eq. (19). Shown are the cases without \tilde{I} (bold lines), with \tilde{I} and fixed $\sigma_{\tilde{I}}$ (thin lines) and scaled $\sigma_{\tilde{I}}$ according to Eq. (20) (dashed lines).

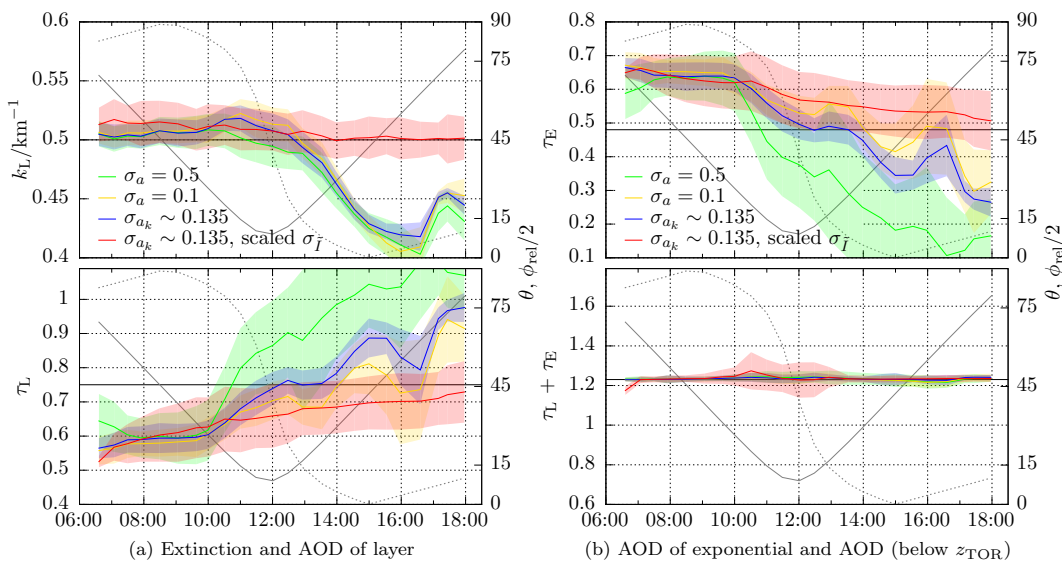


Fig. 14. As Fig. 11, but here with relative intensities \tilde{I} , for different choices of the regularisation parameter (fixed and according to Eq. 19) and $\sigma_{\tilde{I}}$ scaled according to Eq. (20) in one case (red line).

the sequence in reverse chronological order. The decrease is not strictly monotonic, but the effect on the retrieval quality is clearly obvious from Fig. 18. Monte Carlo results for simulated retrievals were carried out for the regularisation parameter σ_{a_k} with and without relative intensities in forward direction starting with the first data point (\rightarrow) and backward direction starting with the last data point (\leftarrow).

In the discussion of the results in this figure we restrict ourselves again to the most important observations. Starting the retrieval in forward direction, the lower information content at the beginning of the profile time series shows in

the “warm-up” character of the first retrieval needed to adjust both x_a and σ_a . The ground extinction coefficient k_{L1} (corresponding to the largest singular value for the polluted ground layer profile) can be estimated relatively well. This is less the case in this direction of the retrieval sequence for the “height” H_L of the ground layer and the extinction coefficient k_{L2} at its top. The height H_L is overestimated throughout and for $\phi_{rel} \sim 90^\circ$ the profile information with respect to H_L , and k_{L2} is rather limited for the high AODs in this case. Including relative intensities makes little difference here. For $\phi_{rel} < \sim 90^\circ$ and small θ , however, they improve

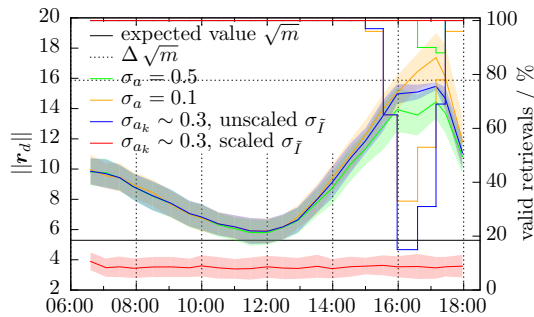


Fig. 15. Data residuals as in Fig. 7, but here for the regularised cases of Fig. 14 (thick lines). Also shown is the percentage of valid retrievals according to Eq. (17) (thin lines).

notably the AOD τ . But for small ϕ_{rel} at the end of the series the last two retrievals fail regardless – this time not due to over-regularisation (the other direction works), but either again due to the inadequacy of the profile representation or due to an interplay of regularisation and a priori. The effect of the latter, carried through from each retrieval to the next, is most obvious when comparing the retrievals in forward and backward direction, especially the profile parameters H_L and k_{L2} . The retrievals (without \tilde{I}) for small relative solar azimuth illustrate that depending on the a priori the AOD can be both persistently overestimated (\rightarrow , green in Fig. 18b, bottom panel) and underestimated (\leftarrow , yellow in Fig. 18b, bottom panel). Similar to the first example – but not as evident here – it would not improve the retrieval of the profile time series if forward and backward direction were to be combined.

4 Conclusions and outlook

In this model study we address the questions of how the information content in the least-squares retrieval of profile shape parameters for aerosol optical extinction profiles $k(z)$ from MAX-DOAS observations can be used more efficiently by choosing appropriate parameters for noise filtering (regularisation) and by exploiting the context of a time series of measurements (empirical a priori). The simulated measurements for two O_4 wavelengths in the UV/VIS (360 and 477 nm) are assumed to take place in an urban, polluted environment with the instrument pointing to the west. Forward calculations are carried out at 470 nm and the retrieval combines both wavelengths using a fixed Ångström exponent a . Other (equally fixed) aerosol optical parameters are expressed in terms of the single-scattering albedo ω_0 , the asymmetry parameter g and a Henyey–Greenstein phase function. The aerosol profile $k(z)$ is parameterised linearly up to a variable height H_L and exponentially above H_L . The either three or four profile parameters are given by extinction coefficients and optical depth for the linear part (k_L and τ_L or k_{L1} , k_{L2} and τ_L) and the optical depth of the exponential (τ_E). The

retrieval uses either differential slant column densities ΔS or ΔS plus relative intensities \tilde{I} for all elevations as fit quantities. Errors for these data are assumed to be completely uncorrelated. The forward model for the fit is involved “online”, i.e. without look-up tables. We use a Tikhonov regularisation scheme with one regularisation parameter chosen to be fixed or to approximate the L-curve criterion.

4.1 Conclusions and further discussion

Summarising our results from both the singular value decomposition (SVD) and Monte Carlo simulations, we first conclude that the inverse MAX-DOAS problem is ill posed in the sense that some of its fit parameters cannot be determined if measurement errors are present. By this we mean that even under ideal conditions (no model errors, low noise, long light paths etc.) the relative error in the retrieved parameter may exceed relative errors of the measurement by 1 or 2 orders of magnitude. For the examples of ground layer profiles containing absorbing aerosols shown here, these parameters correspond to the profile information incorporated in τ_E (and τ_L), while the layer extinction coefficient k_L and to a lesser extent the fully retrieved AOD τ respond moderately. This holds for low and high AODs, but depends, of course, on the signal-to-noise ratio.

Regularisation in form of the Tikhonov scheme with parameter γ (or σ_a^{-1}) proves effective, but the right choice of the regularisation parameter is crucial, as it is shown to depend on the aerosol extinctions themselves. A fixed value of γ does not accommodate a realistic range of AODs. The approximate L-curve criterion employed here works well for low to high AODs, but possibly over-regularises for very small AODs and might just be replaced by an upper bound (for γ) in these cases. Such an upper bound can, in principle, be given by a smallest “possible” measurement error (see the discussion after Eq. 9), but we rather suggest a more refined parameter choice as proposed in the outlook below.

The dynamic update of both the regularisation parameter σ_{a_k} and the a priori \mathbf{x}_{a_k} based on the last valid retrieval within a daytime series of data was demonstrated to work consistently for random measurement noise and strong temporal variation of the aerosol profile. While for the present “over-determined” formulation of the inverse problem we did not find the information content embodied in the signal degrees of freedom d_s helpful for the choice of the regularisation parameter, variations of d_s within the time series (of a day) can be used to establish the starting point \mathbf{x}_{a_0} and order of the retrieval sequence. Choosing \mathbf{x}_{a_0} according to maximal information content allows for reduction of the smoothing error for subsequent retrievals suffering from a lack of information, e.g. for short light paths and high AODs. Results from different retrieval sequences (for example in forward/backward direction) can be combined and used to estimate the actual smoothing error.

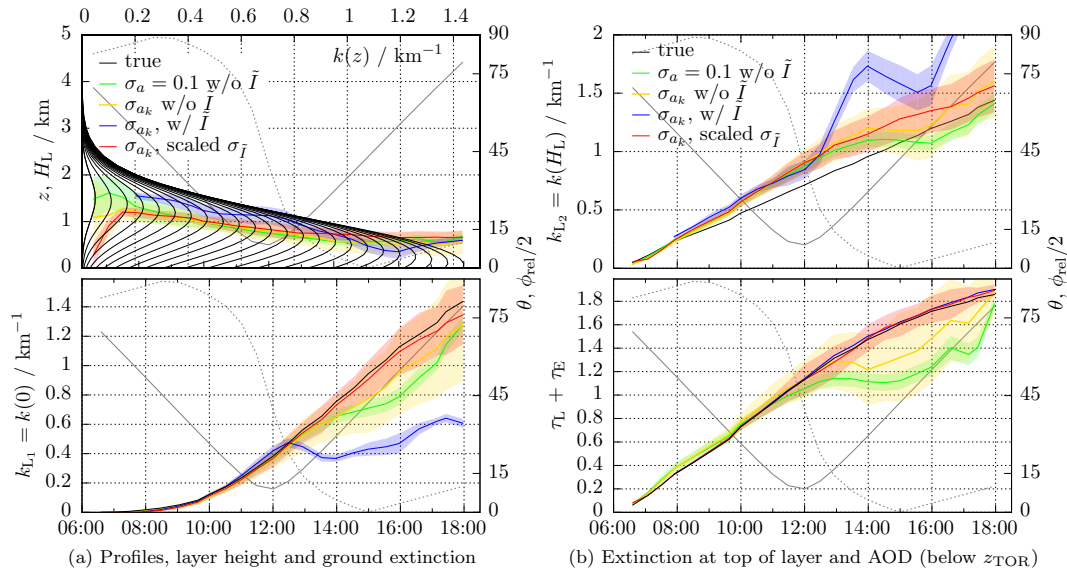


Fig. 16. Retrieval of $\mathbf{x} = (k_{L_1} \cdot 1 \text{ km}, k_{L_2} \cdot 1 \text{ km}, \tau_L, \tau_E)^T$ for the extinction profiles varying in time as shown in (a) (top panel) and different choices of the fit quantities, the regularisation parameter and weights for \tilde{I} . (a) (top panel) “Layer height” $H_L = 2 \tau_L / (k_{L_1} + k_{L_2})$, bottom: ground extinction k_{L_1} . (b) (top panel) Extinction k_{L_2} at height H_L , bottom panel: AOD $\tau = \tau_L + \tau_E$. As in previous figures, lines represent averages, and shaded areas standard deviations for 100 MC runs. The maximum number (100) of valid retrievals is reached except for the first three profiles retrieved with \tilde{I} included explicitly into the fit.

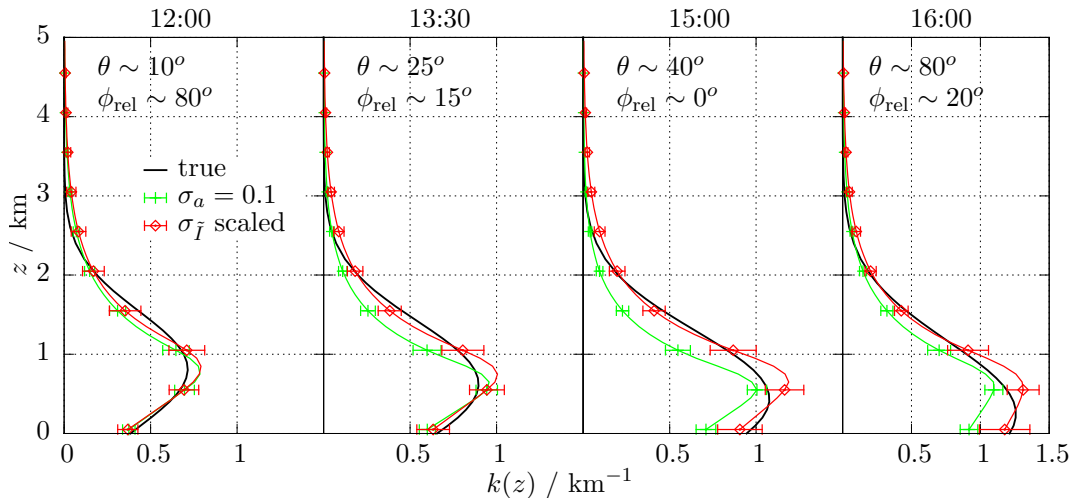


Fig. 17. Example for profiles retrieved as in Fig. 16 for $\phi_{\text{rel}} < 90^\circ$ (same colours).

Explicit addition of relative intensities to the measurement data improves the retrieval of τ in all cases for all viewing geometries. For the reconstruction of profile features, however, we find a certain discrepancy between the error-free analysis using averaging kernels and results from MC simulations. While the former suggest better retrieval of profile shape parameters such as τ_L, τ_E as well, the latter indicate that these retrieval qualities depend to a certain extent on the weighting of \tilde{I} , especially for small relative solar azimuth when the signal-to-noise ratios of \tilde{I} are large. This disagreement is most likely due to neglected correlations between the

ΔS and \tilde{I} , and is here accounted for by giving relative errors of \tilde{I} weights not larger than the ones of ΔS . Alternative approaches will be outlined below.

An immediate consequence of the inverse MAX-DOAS problem being ill posed is the fact that all data errors should be kept as small as possible. On the one hand, large errors do not necessarily mean that all retrieved parameters are equally affected. In our examples of ground layer profiles the layer extinction is a largely decoupled and robust parameter. This implies that the experimental validation of such “safe” parameters related to the largest singular mode in the SVD of

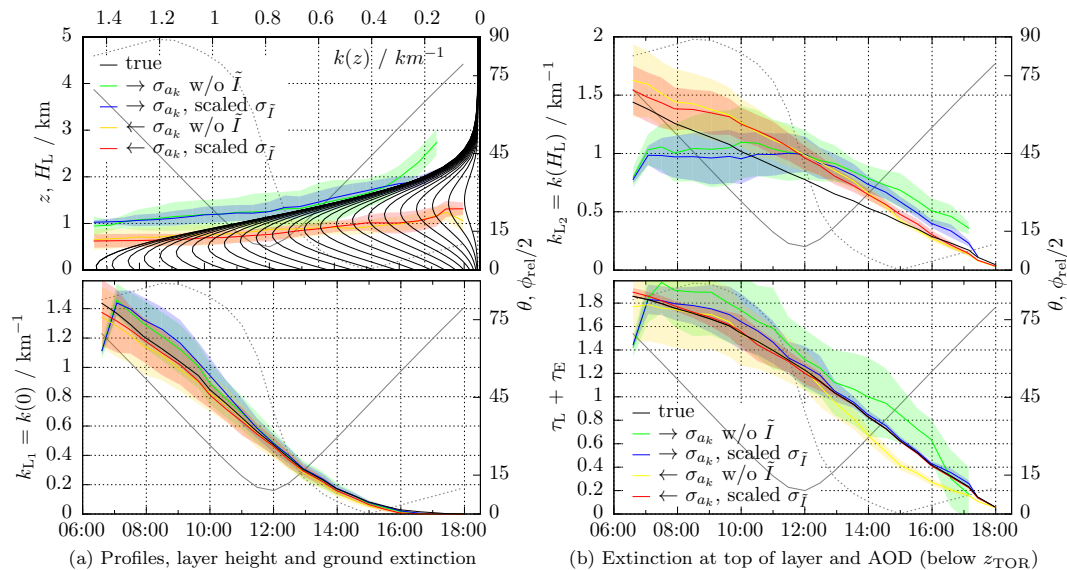


Fig. 18. As Fig. 16, but here with order of profiles reversed in time for retrieval starting with the first (\rightarrow) or last (\leftarrow) set of DSCDs (and RIs).

the forward model's Jacobian does not necessarily mean a validation of the complete profile. On the other hand, the benefit of accurate data can partly be undone by an unsuitable choice of the retrieval parameters, for example by giving the a priori too much weight. Not only the minimisation of measurement errors but also the correct estimate of their (and all the errors') actual size and statistics is of great importance in order to be able to use absolute values of the data residuals $\|\mathbf{r}_d\|$ as indicators for a valid retrieval – obviously vital for the way the empirical a priori is chosen here. Large values of the residual suggest a problem with assumptions on the underlying model or the (systematic) measurement errors. One example discussed was the expected value of $\|\mathbf{r}_d\|$ being too high when \tilde{I} is included into the fit. Another simple example would be a true extinction profile that can only be represented with large errors (h in Eq. 21) by the profile parameterisation used in this study, resulting in high residuals in Eq. (26). Low or consistent residuals, however, do by no means automatically guarantee a good retrieval; see the example of the pure least-squares fit. This might be the point to raise the question of whether coarse look-up tables are able to achieve consistent residuals in a similar manner (compare Fig. 5). As a side remark, we note that the (linear) interpolation involved by look-up tables is one way to implicitly regularise (Groetsch, 1993).

4.2 Outlook

This first attempt to investigate the regularisation behaviour of the inverse MAX-DOAS problem was naturally simplified. An obvious generalisation of the regularisation scheme would be to adjust the parameter γ (or σ_a) during the Levenberg–Marquardt (LM) (or Gauss–Newton (GN))

iteration such that at each step it is chosen to minimise the object function. However, this is computationally expensive and, for example in Doicu et al. (2004), replaced by an empirical relaxation scheme of the regularisation parameter. It might even be possible to carry out the decisive choice of the regularisation parameter at the last step of the LM/GN iteration – that is, for a linear inverse problem. Replacing the regularising contribution in the object function $\gamma\|\mathbf{x} - \mathbf{x}_a\|$ by $\gamma\|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|$ with $\gamma^2\mathbf{L}^T\mathbf{L}$ corresponding to \mathbf{S}_a^{-1} in the optimal estimate (see Sect. 1) allows for parameters to be given different relative weights (e.g. regularise τ_E stronger than τ_L) and to introduce correlations (e.g. for τ_E and τ_L). An alternative viewpoint is that this more general regularisation enables the basis vectors \mathbf{v}_j of the now generalised SVD (GSVD) to be chosen, i.e. to control how individual modes are regularised (Hansen, 1998). If the regularisation can be carried out for the linearised problem, a promising method for the choice of the overall regularisation parameter is the generalised cross validation (GCV, e.g. Hansen, 1998, and references therein) developed for statistical parameter estimation. It is designed to regularise the problem according to the different information content of individual data, and would automatically account for correlation between them. While the Tikhonov regularisation and consequently its filtering behaviour is not intended for any specific class of ill posed problem, certain algorithms originally designed to solve large tomographic problems have been proven quite successful for regularising the inverse problem arising from this kind of remote-sensing measurement. These so-called row acting methods or algebraic reconstruction techniques regularise the least-squares problem iteratively and have been applied to active DOAS measurements

in Laepple et al. (2004) and Hartl et al. (2006) and to stratospheric ozone profile retrieval from limb scatter satellite observations in Degenstein et al. (2009). Since MAX-DOAS observations are essentially tomographic measurements, we believe that their analysis might benefit from this more specialised approach.

The time series retrieval presented in this study requires, strictly speaking, an a posteriori analysis of the information content (d_s), i.e. two retrieval runs. While, in principle, this is not a problem, it would be desirable to perform the time series analysis in a more rigorous probabilistic framework and, for example, to estimate relaxation parameters introduced ad hoc in Frieß et al. (2006) to mimic a model for the temporal evolution used in the application of a Kalman filter.

Relative intensities are used here and in the original work (Frieß et al., 2006) to constrain certain parameters or combinations thereof, namely the integrated extinction coefficients, rather than to add independent profile information. Taking \tilde{I} into account in form of inequalities such as $\tilde{I}_{i_{\min}} \leq f_i(\mathbf{x}) \leq \tilde{I}_{i_{\max}}$ may cause the resulting equality-bound least-squares problem to be less sensitive to uncertainties in the true $\sigma_{\tilde{I}}$ and avoid the problem arising from very different weights of slant column densities and relative intensities in the object function altogether. Alternatively, the forward and inverse problem may be expressed in terms of (relative) intensities alone, which might be a more appropriate formulation to exploit the full wavelength-intensity spectrum measured – for example in a simultaneous retrieval of aerosol and trace gas profiles, or to retrieve other optical parameters such as the single-scattering albedo ω_0 , the asymmetry parameter g etc. Regardless of the formulation, regularisation of the underlying inverse problem remains important in order to avoid losing valuable information.

Acknowledgements. The work described in this paper was partly supported by a grant from City University of Hong Kong (project no. 9041479) and the Guy Carpenter Asia-Pacific Climate Impact Centre (project no. 9360126). A. Hartl would also like to thank the Anhui Institute of Optics and Fine Mechanics and the Chinese Academy of Sciences in Hefei, China, for their support and hospitality.

Edited by: R. Schofield

References

- Brinksma, E. J., Pinardi, G., Volten, H., Braak, R., Richter, A., Schönhardt, A., van Roozendaal, M., Fayt, C., Hermans, C., Dirksen, R. J., Vlemmix, T., Berkhout, A. J. C., Swart, D. P. J., Oetjen, H., Wittrock, F., Wagner, T., Ibrahim, O. W., de Leeuw, G., Moerman, M., Curier, R. L., Celarier, E. A., Cede, A., Veefkind, J. P., Eskes, H. J., Allaart, M., Rothe, R., Pitters, A. J. M., and Levelt, P. F.: The 2005 and 2006 DANDELIONS NO₂ and aerosol intercomparison campaigns, *J. Geophys. Res.*, 113, D16S46, doi:10.1029/2007JD008808, 2008.
- Clémer, K., Van Roozendaal, M., Fayt, C., Hendrick, F., Hermans, C., Pinardi, G., Spurr, R., Wang, P., and De Mazière, M.: Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, *Atmos. Meas. Tech.*, 3, 863–878, doi:10.5194/amt-3-863-2010, 2010.
- Cowell, W. R., ed.: Sources and Development of Mathematical Software, Prentice-Hall Series in Computational Mathematics, Cleve Moler, Advisor, Prentice-Hall, Upper Saddle River, NJ 07458, USA, 1984.
- Degenstein, D. A., Bourassa, A. E., Roth, C. Z., and Llewellyn, E. J.: Limb scatter ozone retrieval from 10 to 60 km using a multiplicative algebraic reconstruction technique, *Atmos. Chem. Phys.*, 9, 6521–6529, doi:10.5194/acp-9-6521-2009, 2009.
- Doicu, A., Schreier, F., and Hess, M.: An iterative regularization method with B-spline approximation for atmospheric temperature and concentration retrievals, *Environ. Modell. Softw.*, 20, 1101–1109, 2004.
- Doicu, A., Trautmann, T., and Schreier, F.: Numerical Regularization for Atmospheric Inverse Problems, 1st Edn., Springer, 2010.
- Frieß, U., Monks, P. S., Remedios, J. J., Rozanov, A., Sinreich, R., Wagner, T., and Platt, U.: MAX-DOAS O₄ measurements: A new technique to derive information on atmospheric aerosols: 2. Modeling studies, *J. Geophys. Res.*, 111, D14203, doi:10.1029/2005JD006618, 2006.
- Frieß, U., Sihler, H., Sander, R., Pöhler, D., Yilmaz, S., and Platt, U.: The vertical distribution of BrO and aerosols in the Arctic: Measurements by active and passive differential optical absorption spectroscopy, *J. Geophys. Res.*, 116, D00R04, doi:10.1029/2011JD015938, 2011.
- Garland, R. M., Yang, H., Schmid, O., Rose, D., Nowak, A., Achtert, P., Wiedensohler, A., Takegawa, N., Kita, K., Miyazaki, Y., Kondo, Y., Hu, M., Shao, M., Zeng, L. M., Zhang, Y. H., Andreae, M. O., and Pöschl, U.: Aerosol optical properties in a rural environment near the mega-city Guangzhou, China: implications for regional air pollution, radiative forcing and remote sensing, *Atmos. Chem. Phys.*, 8, 5161–5186, doi:10.5194/acp-8-5161-2008, 2008.
- Garland, R. M., Schmid, O., Nowak, A., Achtert, P., Wiedensohler, A., Gunthe, S. S., Takegawa, N., Kita, K., Kondo, Y., Hu, M., Shao, M., Zeng, L. M., Zhu, T., Andreae, M. O., and Pöschl, U.: Aerosol optical properties observed during Campaign of Air Quality Research in Beijing 2006 (CAREBeijing-2006): Characteristic differences between the inflow and outflow of Beijing city air, *J. Geophys. Res.*, 114, D00G04, doi:10.1029/2008JD010780, 2009.
- Greenblatt, G. D., Orlando, J. J., Burkholder, J. B., and Ravishankara, A. R.: Absorption measurements of oxygen between 330 and 1140 nm, *J. Geophys. Res.*, 95, 577–582, 1990.

- Groetsch, C. W.: Inverse problems in the mathematical sciences, Vieweg, Braunschweig, 1993.
- Halla, J. D., Wagner, T., Beirle, S., Brook, J. R., Hayden, K. L., O'Brien, J. M., Ng, A., Majonis, D., Wenig, M. O., and McLaren, R.: Determination of tropospheric vertical columns of NO₂ and aerosol optical properties in a rural setting using MAX-DOAS, *Atmos. Chem. Phys.*, 11, 12475–12498, doi:10.5194/acp-11-12475-2011, 2011.
- Hansen, P. C.: Rank-deficient and discrete ill-posed problems, Siam, Philadelphia, 1998.
- Hartl, A., Song, B. C., and Pundt, I.: 2-D reconstruction of atmospheric concentration peaks from horizontal long path DOAS tomographic measurements: parametrisation and geometry within a discrete approach, *Atmos. Chem. Phys.*, 6, 847–861, doi:10.5194/acp-6-847-2006, 2006.
- Heckel, A., Richter, A., Tarsu, T., Wittrock, F., Hak, C., Pundt, I., Junkermann, W., and Burrows, J. P.: MAX-DOAS measurements of formaldehyde in the Po-Valley, *Atmos. Chem. Phys.*, 5, 909–918, doi:10.5194/acp-5-909-2005, 2005.
- Hendrick, F., Van Roozendaal, M., Kylling, A., Petritoli, A., Rozanov, A., Sanghavi, S., Schofield, R., von Friedeburg, C., Wagner, T., Wittrock, F., Fonteyn, D., and De Mazière, M.: Intercomparison exercise between different radiative transfer models used for the interpretation of ground-based zenith-sky and multi-axis DOAS observations, *Atmos. Chem. Phys.*, 6, 93–108, doi:10.5194/acp-6-93-2006, 2006.
- Hönninger, G., von Friedeburg, C., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS), *Atmos. Chem. Phys.*, 4, 231–254, doi:10.5194/acp-4-231-2004, 2004.
- Irie, H., Kanaya, Y., Akimoto, H., Iwabuchi, H., Shimizu, A., and Aoki, K.: First retrieval of tropospheric aerosol profiles using MAX-DOAS and comparison with lidar and sky radiometer measurements, *Atmos. Chem. Phys.*, 8, 341–350, doi:10.5194/acp-8-341-2008, 2008.
- Irie, H., Kanaya, Y., Akimoto, H., Iwabuchi, H., Shimizu, A., and Aoki, K.: Dual-wavelength aerosol vertical profile measurements by MAX-DOAS at Tsukuba, Japan, *Atmos. Chem. Phys.*, 9, 2741–2749, doi:10.5194/acp-9-2741-2009, 2009.
- Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y., and Van Roozendaal, M.: Eight-component retrievals from ground-based MAX-DOAS observations, *Atmos. Meas. Tech.*, 4, 1027–1044, doi:10.5194/amt-4-1027-2011, 2011.
- Kneizys, F. X., Shettle, E. P., Abreu, L. W., Chetwynd, J. H., Anderson, G. P., Gallery, W. O., Selby, J. E. A., and Clough, S. A.: Users Guide to Lowtran 7, Tech. Rep. AFGL-TR-88-0177, Air Force Geophysics Laboratory, Hanscom Air Force Base, MA, 1988.
- Laepple, T., Knab, V., Mettendorf, K.-U., and Pundt, I.: Longpath DOAS tomography on a motorway exhaust gas plume: numerical studies and application to data from the BAB II campaign, *Atmos. Chem. Phys.*, 4, 1323–1342, doi:10.5194/acp-4-1323-2004, 2004.
- Lee, H., Irie, H., Kim, Y., Noh, Y., Lee, C., Kim, Y., and Chun, K.: Retrieval of Aerosol Extinction in the Lower Troposphere Based on UV MAX-DOAS Measurements, *Aerosol Sci. Tech.*, 43, 502–509, doi:10.1080/02786820902769691, 2009.
- Lee, K. H., Li, Z., Wong, M. S., Xin, J., Wang, Y., Hao, W.-M., and Zhao, F.: Aerosol single scattering albedo estimated across China from a combination of ground and satellite measurements, *J. Geophys. Res.*, 112, D22S15, doi:10.1029/2007JD009077, 2007.
- Li, X., Brauers, T., Shao, M., Garland, R. M., Wagner, T., Deutschmann, T., and Wahner, A.: MAX-DOAS measurements in southern China: retrieval of aerosol extinctions and validation using ground-based in-situ data, *Atmos. Chem. Phys.*, 10, 2079–2089, doi:10.5194/acp-10-2079-2010, 2010.
- Moré, J. J., Garbow, B. S., and Hillstom, K. E.: User Guide for MINPACK-1, Tech. Rep. ANL-80-74, Argonne National Laboratory, Argonne, IL, USA, 1980.
- Nocedal, J. and Wright, S.: Numerical Optimization, Springer Series in Operations Research and Financial Engineering, 2nd Edn., Springer, 2006.
- Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy, 1st Edn., Springer, 2008.
- Rodgers, C. D.: Inverse methods for atmospheric sounding: Theory and practice, World Scientific, London, 2000.
- Roscoe, H. K., Van Roozendaal, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Friess, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Peters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO₂ and O₄ by MAX-DOAS and zenith-sky UV and visible spectrometers, *Atmos. Meas. Tech.*, 3, 1629–1646, doi:10.5194/amt-3-1629-2010, 2010.
- Rozanov, A., Rozanov, V.-V., and Burrows, J. P.: Combined differential-integral approach for the radiation field computation in a spherical shell atmosphere: Nonlimb geometry, *J. Geophys. Res.*, 105, 22937–22942, doi:10.1029/2000JD900378, 2000.
- Rozanov, A., Rozanov, V., and Burrows, J.: A numerical radiative transfer model for a spherical planetary atmosphere: combined differential-integral approach involving the Picard iterative approximation, *J. Quant. Spectrosc. Ra.*, 69, 491–512, doi:10.1016/S0022-4073(00)00100-X, 2001.
- Rozanov, A., Rozanov, V.-V., Buchwitz, M., Kokhanovsky, A., and Burrows, J. P.: SCIATRAN 2.0. – A new radiative transfer model for geophysical applications in the 175–2400 nm spectral region, *Adv. Space Res.*, 36, 1015–1019, 2005.
- Shaiganfar, R., Beirle, S., Sharma, M., Chauhan, A., Singh, R. P., and Wagner, T.: Estimation of NO_x emissions from Delhi using Car MAX-DOAS observations and comparison with OMI satellite data, *Atmos. Chem. Phys.*, 11, 10871–10887, doi:10.5194/acp-11-10871-2011, 2011.
- Sinreich, R., Friß, U., Wagner, T., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS) of gas and aerosol distributions, *Faraday Discuss.*, 130, 153–164, doi:10.1039/B419274P, 2005.
- Sinreich, R., Volkamer, R., Filsinger, F., Friß, U., Kern, C., Platt, U., Sebastián, O., and Wagner, T.: MAX-DOAS detection of glyoxal during ICARTT 2004, *Atmos. Chem. Phys.*, 7, 1293–1303, doi:10.5194/acp-7-1293-2007, 2007.

- Tarantola, A.: Inverse Problem Theory, Elsevier, New York, 1987.
- Tarantola, A.: Inverse problem theory and model parameter estimation, SIAM, Philadelphia, 2005.
- Twomey, S.: Introduction to the mathematics of inversion in remote sensing, Dover Publication Inc., New York, 1997.
- Vlemmix, T., PETERS, A. J. M., BERKHOUT, A. J. C., GAST, L. F. L., WANG, P., and LEVELT, P. F.: Ability of the MAX-DOAS method to derive profile information for NO₂: can the boundary layer and free troposphere be separated?, *Atmos. Meas. Tech.*, 4, 2659–2684, doi:10.5194/amt-4-2659-2011, 2011.
- Wagner, T., von Friedeburg, C., Wenig, M., Otten, C., and Platt, U.: UV/vis observations of atmospheric O₄ absorptions using direct moon light and zenith scattered sunlight under clear and cloudy sky conditions, *J. Geophys. Res.*, 107, 4424, doi:10.1029/2001JD001026, 2002.
- Wagner, T., Dix, B., Friedeburg, C. v., Frieß, U., Sanghavi, S., Sinreich, R., and Platt, U.: MAX-DOAS O₄ measurements: A new technique to derive information on atmospheric aerosols – Principles and information content, *J. Geophys. Res.*, 109, D22205, doi:10.1029/2004JD004904, 2004.
- Wagner, T., Burrows, J. P., Deutschmann, T., Dix, B., von Friedeburg, C., Frieß, U., Hendrick, F., Heue, K.-P., Irie, H., Iwabuchi, H., Kanaya, Y., Keller, J., McLinden, C. A., Oetjen, H., Palazzi, E., Petritoli, A., Platt, U., Postlyakov, O., Pukite, J., Richter, A., van Roozendaal, M., Rozanov, A., Rozanov, V., Sinreich, R., Sanghavi, S., and Wittrock, F.: Comparison of box-air-mass-factors and radiances for Multiple-Axis Differential Optical Absorption Spectroscopy (MAX-DOAS) geometries calculated from different UV/visible radiative transfer models, *Atmos. Chem. Phys.*, 7, 1809–1833, doi:10.5194/acp-7-1809-2007, 2007.
- Wagner, T., Deutschmann, T., and Platt, U.: Determination of aerosol properties from MAX-DOAS observations of the Ring effect, *Atmos. Meas. Tech.*, 2, 495–512, doi:10.5194/amt-2-495-2009, 2009.
- Wagner, T., Beirle, S., Brauers, T., Deutschmann, T., Frieß, U., Hak, C., Halla, J. D., Heue, K. P., Junkermann, W., Li, X., Platt, U., and Pundt-Gruber, I.: Inversion of tropospheric profiles of aerosol extinction and HCHO and NO₂ mixing ratios from MAX-DOAS observations in Milano during the summer of 2003 and comparison with independent data sets, *Atmos. Meas. Tech.*, 4, 2685–2715, doi:10.5194/amt-4-2685-2011, 2011.
- Wittrock, F., Oetjen, H., Richter, A., Fietkau, S., Medeke, T., Rozanov, A., and Burrows, J. P.: MAX-DOAS measurements of atmospheric trace gases in Ny-Ålesund – Radiative transfer studies and their application, *Atmos. Chem. Phys.*, 4, 955–966, doi:10.5194/acp-4-955-2004, 2004.
- Zieger, P., Weingartner, E., Henzing, J., Moerman, M., de Leeuw, G., Mikkilä, J., Ehn, M., Petäjä, T., Clémer, K., van Roozendaal, M., Yilmaz, S., Frieß, U., Irie, H., Wagner, T., Shaiganfar, R., Beirle, S., Apituley, A., Wilson, K., and Baltensperger, U.: Comparison of ambient aerosol extinction coefficients obtained from in-situ, MAX-DOAS and LIDAR measurements at Cabauw, *Atmos. Chem. Phys.*, 11, 2603–2624, doi:10.5194/acp-11-2603-2011, 2011.