**Atmospheric
Chemistry
and Physics**

Technical Note: Variance-covariance matrix and averaging kernels for the Levenberg-Marquardt solution of the retrieval of atmospheric vertical profiles

S. Ceccherini[1] and M. Ridolfi[2]

[1]Istituto di Fisica Applicata "Nello Carrara" (IFAC) del Consiglio Nazionale delle Ricerche (CNR), Firenze, Italy
[2]Dipartimento di Fisica, Università di Bologna, Bologna, Italy

Abstract. The variance-covariance matrix (VCM) and the averaging kernel matrix (AKM) are widely used tools to characterize atmospheric vertical profiles retrieved from remote sensing measurements. Accurate estimation of these quantities is essential for both the evaluation of the quality of the retrieved profiles and for the correct use of the profiles themselves in subsequent applications such as data comparison, data assimilation and data fusion. We propose a new method to estimate the VCM and AKM of vertical profiles retrieved using the Levenberg-Marquardt iterative technique. We apply the new method to the inversion of simulated limb emission measurements. Then we compare the obtained VCM and AKM with those resulting from other methods already published in the literature and with accurate estimates derived using statistical and numerical estimators. The proposed method accounts for all the iterations done in the inversion and provides the most accurate VCM and AKM. Furthermore, it correctly estimates the VCM and the AKM also if the retrieval iterations are stopped when a physically meaningful convergence criterion is fulfilled, i.e. before achievement of the numerical convergence at machine precision. The method can be easily implemented in any Levenberg-Marquardt iterative retrieval scheme, either constrained or unconstrained, without significant computational overhead.

1 Introduction

The retrieval of the vertical distribution of an atmospheric parameter from remote sensing measurements is generally performed by fitting forward model simulations to the available observations (Rodgers, 2000; Twomey, 1977). The fitting procedure consists in the minimization of a cost function, generally made of the summation of two terms. The first term, usually called "chi-square", is the squared weighted norm of the residuals, i.e. the square of the norm of the differences between observations and simulations, the weight being provided by the inverse variance-covariance matrix (VCM) of the observations. The second term of the cost function, if any, is usually a constraint to the solution. This term can be used for example to penalize solutions that either oscillate beyond acceptable limits (Bowman et al., 2006; Ceccherini, 2005; Doicu et al., 2004; Ridolfi and Sgheri, 2009; Schimpf and Schreier, 1997; Sofieva et al., 2004) or deviate from an a-priori (Rodgers, 2000). Since the forward model is usually a non-linear function of the atmospheric state vector, an iterative method is needed to find the minimum of the cost function. In case of weak non-linearity of the forward model, the Gauss-Newton (GN) technique can be successfully applied (Press et al., 1992). On the other hand, if the forward model is significantly non-linear, a GN step can lead to an increase rather than to a decrease of the cost function. This situation, often occurring in the retrieval of vertical profiles of atmospheric parameters, is generally overcome by applying the Levenberg-Marquardt (LM) modification (Levenberg, 1944; Marquardt, 1963). This modification combines the steepest descent (Press et al., 1992) and the GN methods by damping the GN iteration step. A damping

---

factor is adjusted at each iteration in such a way that if the step leads to a reduction of the cost-function the damping factor is decreased, making the next step closer to the GN step. Whereas, if the step leads to an increase of the cost function the iteration is repeated with an increased damping factor, producing a step closer to the gradient descent direction.

In principle, if a minimum of the cost function has been approached close enough it should be possible to reduce the LM damping factor to a negligibly small value and reach the convergence with a final GN step. This is possible in well-conditioned retrievals but not in both ill-posed and ill-conditioned retrievals. In ill-posed retrievals the minimum of the cost function does not correspond to an unique solution but to a whole class of solutions. In this case it is not possible to perform a final GN step because this involves the inversion of a singular matrix. In ill-conditioned retrievals the matrix to be inverted is not singular but its condition number is too large (Press et al., 1992) and the inversion greatly amplifies numerical, approximation and measurement errors. This large error amplification prevents the GN step from finding the minimum of the cost function. Therefore, in ill-posed and ill-conditioned retrievals it is not possible to finish the iterative procedure with a negligible LM damping factor and the question arises on how to characterize properly the LM solution.

The VCM and the averaging kernel matrix (AKM) are widely used tools to characterize the solution of the retrieval. Accurate estimation of these quantities is essential for both the evaluation of the quality of the retrieved profiles and for their correct use in subsequent applications such as data comparison (Ceccherini et all, 2003; Cortesi et al., 2007; Pougatchev, 2008; Ridolfi et al., 2007; Rodgers and Connor, 2003), data assimilation (Lahoz et al., 2007) and data fusion (Ceccherini et al., 2009, 2010). While the calculation of the VCM and of the AKM is relatively easy in the GN case (see e.g. Rodgers, 2000), in the LM case one must be more cautious.

So far two different methods have been used for the calculation of the VCM and AKM of the LM iterative solution. The first one (Press et al., 1992) calculates these quantities assuming a negligibly small LM damping in the formula used to update the state vector at the last iteration. Of course this method involves an approximation that, as we will show in the following sections, might also be relatively coarse. The second method (Ceccherini et al., 2007) calculates the VCM and the AKM from the formula of the last iteration step, without neglecting the LM damping term. This method is correct if the minimum is reached in only one iteration, while it may also involve a relatively coarse approximation if several iterations are required for the minimization. In this paper we introduce an alternative method for the calculation of the VCM and AKM of the LM solution. The proposed method accounts for both the LM damping term and all the iterations required for the minimization of the cost function.

In this paper the methods of Press et al. (1992) and of Ceccherini et al. (2007) and the new proposed one are applied to the retrieval of the ozone vertical profile from mid-infrared limb emission simulated measurements. These three results are compared to those obtained using, in the case of the VCM, the statistical estimator applied to a set of simulated retrievals and, in the case of AKM, the numerical calculation of the derivatives.

The paper is organized as follows. In Sect. 2 we describe the theory of the new proposed method. In Sect. 3 we apply the new method and the other two methods available from the literature to simulated measurements and compare the resulting VCMs and AKMs to those derived using the statistical estimator and the numerical derivatives, respectively. In Sect. 4 we draw the conclusions of the work.

2   Theory

Consistent with the notations adopted by Rodgers (2000), we represent the observations by the m-dimensional vector $\boldsymbol{y}$ and the unknown atmospheric state by the n-dimensional vector $\boldsymbol{x}$. The forward model $\boldsymbol{f}(\boldsymbol{x})$ provides simulated observations given the atmospheric state $\boldsymbol{x}$. The relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon}$ is the m-dimensional vector containing the experimental errors of the observations, characterized by a VCM $\mathbf{S}_{\mathrm{y}} = \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \rangle$, where the symbol $\langle \cdots \rangle$ denotes the expectation value. The solution of the inverse problem is the state $\boldsymbol{x}$ minimizing a cost function usually defined as:

$$\xi^2 = (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}))^T \mathbf{S}_{\mathrm{y}}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x})) + (\boldsymbol{x}_{\mathrm{a}} - \boldsymbol{x})^T \mathbf{R}(\boldsymbol{x}_{\mathrm{a}} - \boldsymbol{x}), \tag{2}$$

where $\mathbf{R}$ is a $n \times n$ matrix operator that may be used to constrain the solution towards the shape (using e.g. the first and/or the second vertical derivatives, see Ridolfi and Sgheri, 2009) and/or the value of an a-priori estimate $\boldsymbol{x}_{\mathrm{a}}$.

Since $\boldsymbol{f}(\boldsymbol{x})$ is a non-linear and complicated function of $\boldsymbol{x}$, an analytical solution for the minimum of ξ^2 does not exist. Therefore an iterative method is often used to find the minimum. The iterations are stopped when a pre-defined convergence criterion is fulfilled and the final state vector $\boldsymbol{x}_r$ at the last iteration (index r) represents the solution of the retrieval.

If the forward model is moderately non-linear, the GN method can be used for the minimization of ξ^2, with the following iterative formula:

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \left(\mathbf{K}_i^T \mathbf{S}_{\mathrm{y}}^{-1} \mathbf{K}_i + \mathbf{R} \right)^{-1} \cdot$$
$$\left[\mathbf{K}_i^T \mathbf{S}_{\mathrm{y}}^{-1} (\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}_i)) + \mathbf{R}(\boldsymbol{x}_{\mathrm{a}} - \boldsymbol{x}_i) \right], \tag{3}$$

where $\mathbf{K}_i$ is the Jacobian matrix of the forward model calculated at $\boldsymbol{x} = \boldsymbol{x}_i$.

If the forward model is significantly non-linear, the step in Eq. (3) can lead to an increase rather than to a decrease of the cost function ξ^2. In this case it is useful to adopt the LM modification that is based on the following iteration step:

$$
\begin{aligned}
\boldsymbol{x}_{i+1} &= \boldsymbol{x}_i + \left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} + \lambda_i \mathbf{D}_i \right)^{-1} \cdot \\
&\quad \left[\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}_i) \right) + \mathbf{R}(\boldsymbol{x}_\mathrm{a} - \boldsymbol{x}_i) \right] \\
&= \boldsymbol{x}_i + \mathbf{G}_i \left(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{x}_i) \right) + \mathbf{M}_i \mathbf{R}(\boldsymbol{x}_\mathrm{a} - \boldsymbol{x}_i),
\end{aligned}
\tag{4}
$$

where we have defined:

$$
\mathbf{M}_i = \left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} + \lambda_i \mathbf{D}_i \right)^{-1}
\tag{5}
$$

$$
\mathbf{G}_i = \mathbf{M}_i \mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1},
\tag{6}
$$

where $\mathbf{D}_i$ is a diagonal matrix with the diagonal elements equal to the diagonal elements of $\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i$ (see e.g. Pujol, 2007) and λ_i is a scalar damping factor that depends on the iteration index i. The iterations start with a small damping factor λ_0. At each iteration, if the corrected state $\boldsymbol{x}_{i+1}$ provides a reduction of the cost function (i.e. $\xi^2(\boldsymbol{x}_{i+1}) < \xi^2(\boldsymbol{x}_i)$) then the state $\boldsymbol{x}_{i+1}$ is accepted and the damping factor for the next iteration is reduced. If the state $\boldsymbol{x}_{i+1}$ provides an increase of the cost function ($\xi^2(\boldsymbol{x}_{i+1}) > \xi^2(\boldsymbol{x}_i)$) then $\boldsymbol{x}_{i+1}$ is rejected and the iteration is repeated with a larger damping factor.

In order to find the VCM $\mathbf{S}_r$ of the solution $\boldsymbol{x}_r$ we have to propagate the error $\boldsymbol{\epsilon}$ from $\boldsymbol{y}$ onto $\boldsymbol{x}_r$. The error vector $\boldsymbol{\sigma}_r$ on $\boldsymbol{x}_r$ can be written as:

$$
\boldsymbol{\sigma}_r = \mathbf{T}_r \boldsymbol{\epsilon},
\tag{7}
$$

where we have introduced the $n \times m$ matrix $(\mathbf{T}_r)_{jk} = \frac{\partial (\boldsymbol{x}_r)_j}{\partial \boldsymbol{y}_k}$ that can be seen as the r-th matrix of the sequence $(\mathbf{T}_i)_{jk} = \frac{\partial (\boldsymbol{x}_i)_j}{\partial \boldsymbol{y}_k}$ for $i = 0, 1, \cdots r$. Then the VCM $\mathbf{S}_r$ is given by:

$$
\mathbf{S}_r = \left\langle \boldsymbol{\sigma}_r \boldsymbol{\sigma}_r^T \right\rangle = \mathbf{T}_r \left\langle \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \right\rangle \mathbf{T}_r^T = \mathbf{T}_r \mathbf{S}_\mathrm{y} \mathbf{T}_r^T.
\tag{8}
$$

The matrices $\mathbf{T}_i$ can be obtained by calculating the derivative of Eq. (4) with respect to $\boldsymbol{y}$. Neglecting the derivatives of $\mathbf{K}_i$ with respect to $\boldsymbol{x}_i$ (a hypothesis already exploited by the GN approach itself), and consequently with respect to $\boldsymbol{y}$, we get:

$$
\mathbf{T}_{i+1} = \mathbf{T}_i + \mathbf{G}_i (\mathbf{I} - \mathbf{K}_i \mathbf{T}_i) - \mathbf{M}_i \mathbf{R} \mathbf{T}_i.
\tag{9}
$$

Rearranging Eq. (9) and considering that the initial guess $\boldsymbol{x}_0$ does not depend on the observations $\boldsymbol{y}$, we obtain the following recursive formula for the matrices $\mathbf{T}_i$:

$$
\begin{cases}
\mathbf{T}_0 = \mathbf{0} \\
\mathbf{T}_{i+1} = \mathbf{G}_i + (\mathbf{I} - \mathbf{G}_i \mathbf{K}_i - \mathbf{M}_i \mathbf{R}) \mathbf{T}_i.
\end{cases}
\tag{10}
$$

Equation (10) for $i = 0, 1, \cdots, r-1$ determines $\mathbf{T}_r$. This matrix is then used in Eq. (8) to provide the VCM $\mathbf{S}_r$ of the solution $\boldsymbol{x}_r$.

The elements of $\mathbf{A}_r$, the AKM of $\boldsymbol{x}_r$, are defined as the derivatives of the components of the solution $\boldsymbol{x}_r$ with respect to the components of the true profile $\boldsymbol{x}$. Since $\boldsymbol{x}_r$ is a function of the observations $\boldsymbol{y}$ that in turn are functions of the true profile $\boldsymbol{x}$, we can write:

$$
\mathbf{A}_r = \frac{\partial \boldsymbol{x}_r}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}_r}{\partial \boldsymbol{y}} \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \mathbf{T}_r \mathbf{K},
\tag{11}
$$

where $\mathbf{K}$ is the Jacobian matrix of the forward model calculated in the true profile $\boldsymbol{x}$. Equations (8) and (11) show that both the VCM and the AKM depend on $\mathbf{T}_r$ which in turn, as shown by Eq. (10), depends on the path in the parameter space followed by the minimization procedure, from the initial guess to the solution. We note however that, if an iteration step is done with $\lambda_i = 0$ (GN iteration) from Eqs. (5) and (6) we get $\mathbf{G}_i \mathbf{K}_i + \mathbf{M}_i \mathbf{R} = \mathbf{I}$ and from Eq. (10) it follows that $\mathbf{T}_r$ is independent of the steps performed before the considered iteration. Therefore, we can say that a GN iteration resets the memory of the path followed before that iteration.

When exact convergence is reached we get $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i$ and consequently also $\mathbf{T}_{i+1} = \mathbf{T}_i$ (from the definition of $\mathbf{T}_i$). Substituting this condition in Eq. (10) and using Eqs. (5) and (6) we get:

$$
\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} = \left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} \right) \mathbf{T}_{i+1}.
\tag{12}
$$

In well-conditioned retrievals the matrix $\left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} \right)$ can be inverted and we get the solution:

$$
\mathbf{T}_{i+1} = \left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} \right)^{-1} \mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1},
\tag{13}
$$

that is the same solution we get from Eq. (10) when the iteration $i + 1$ is performed with $\lambda_i = 0$. This means that in well-conditioned retrievals that reach exact convergence, the VCM and AKM calculated with Eqs. (8), (10) and (11) coincide with those that would be obtained assuming $\lambda_i = 0$ at the last iteration. When the retrieval is ill-posed, matrix $\left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} \right)$ is singular and cannot be inverted. This means that Eq. (12) does not determine uniquely $\mathbf{T}_{i+1}$, but a whole class of matrices $\mathbf{T}_{i+1}$ fulfills this equation. Among all the possible solutions, Eq. (10) determines a solution that depends on the path followed by the minimization procedure in the parameter space. A similar consideration applies to ill-conditioned retrievals in which matrix $\left(\mathbf{K}_i^T \mathbf{S}_\mathrm{y}^{-1} \mathbf{K}_i + \mathbf{R} \right)$ has a too large condition number (see Sect. 2.6 of Press et al., 1992), and its inversion implies an amplification of measurement errors that is so large that one or more components of retrieval vector are unknowable. This means that, at machine precision, many solutions become compatible with the same minimum of the cost function and the uniqueness of the solution is lost. Therefore, in ill-posed and ill-conditioned retrievals the LM method acts as an external constraint selecting one solution among all the possible ones. Consequently, as any other external constraint, the LM method affects the
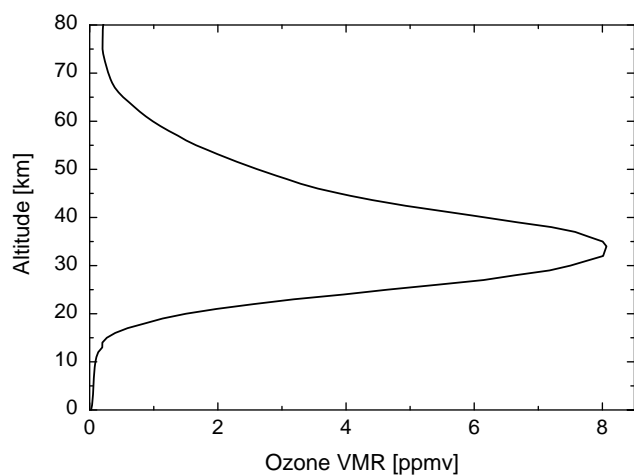
Fig. 1. Ozone profile (ppmv) used to generate the simulated observations. It corresponds to a mid-latitude climatological atmosphere in July (Remedios et al., 2007).

solution, the VCM and the AKM. We note that this latter conclusion applies also to retrievals in which the iterations are stopped before achievement of the numerical convergence at machine precision, using a physically meaningful rule (see Sect. 5.6.3 of Rodgers, 2000).

3 Calculation of the VCM and AKM in a test case

3.1 Set-up of test retrievals

We test the method proposed in Sect. 2 on the basis of ozone volume mixing ratio (VMR) vertical profile retrieval from synthetic limb measurements. In particular, for the simulation of the measurements we consider the specifications of MIPAS (Michelson Interferometer for Passive Atmospheric Sounding). MIPAS measures the middle-infrared atmospheric limb emission spectrum from the polar European satellite ENVISAT (ENVIronmental SATellite). Full details of the instrument are described in Fischer et al. (2008). The simulated observations used here correspond to the MIPAS nominal measurement mode adopted from January 2005 onward, for which the maximum optical path difference is 8 cm. The limb scans include views with tangent altitudes ranging from 6 to 70 km with 1.5 km steps in the upper troposphere – lower stratosphere region and coarser steps (from 2 to 4.5 km) above (Dudhia, 2008).

We obtain the synthetic observations by adding Gaussian random noise to the forward model simulations. The standard deviation of the noise is taken equal to the noise equivalent spectral radiance of the real MIPAS measured spectra. The forward model used is the one internal to our retrieval program, hence permitting us to avoid systematic errors that are not considered in this work.

Our retrieval algorithm is based on the code developed by the European Space Agency for the operational MIPAS retrievals (Ridolfi et al., 2000; Raspollini et al., 2006). It performs a non-linear least-squares fit of the observations in selected spectral intervals (Dudhia et al., 2002) to retrieve at the tangent point grid, pressure and temperature, and then sequentially, the VMR of water vapour, ozone, nitric acid, methane, nitrous oxide and nitrogen dioxide in the altitude range from 6 to 70 km. As additional fitting parameters in each retrieval, the state vector includes also components representing atmospheric continuum absorption cross-sections as a function of altitude and of the selected spectral intervals.

Our tests focus on the retrieval of the ozone VMR profile. The simulated observations are generated assuming a mid-latitude climatological atmosphere in July (Remedios et al., 2007). The corresponding ozone profile is reported in Fig. 1. We minimize the chi-square function (first term of Eq. 2) with the LM iterative technique (Eq. 4). External constraints such as regularization or optimal estimation are disabled ($\mathbf{R=0}$). The damping factor λ_i is calculated as follows. The initial value is equal to 0.1. At each iteration, if the corrected VMRs provide a chi-square smaller than the previous one, the damping factor for the next iteration is reduced by a factor of 4. If the corrected VMRs provide a chi-square larger than the previous one the iteration is repeated with a damping factor increased by a factor of 8. In order to make sure that convergence is reached with good accuracy, the iterations are stopped when one of the two following conservative criteria is fulfilled: (I) the relative variation of chi-square is less than 0.001; (II) the number of 10 iterations is reached. Of course conservative stopping criteria are necessary but not sufficient to ensure the good accuracy of the achieved minimum. For this reason, as we will explain in Sect. 3.2, we also verified a-posteriori the accuracy of the convergence by testing the obtained chi-square values.

As an example in Table 1 we report the reduced chi-square (chi-square divided by $m-n$) and the damping parameter at each of the iterations required for the minimization. Rows of the table with the same iteration number correspond to the fact that the step performed with the input damping factor produces an increase of the chi-square and that the iteration is repeated with an increased damping factor until a decrease of chi-square is achieved.

In order to check whether it is possible to finish the retrieval with a GN iteration, we also tried to repeat iteration 10 with an LM damping factor equal to zero instead of 0.8. We found that such a GN iteration provides a reduced chi-square value equal to 9.04, i.e. much larger than the value of ≈ 1.05 achieved with LM. This large increase is due to the fact that some components of the state vector (representing continuum cross-sections) are not well-determined by the observations and make the retrieval ill-conditioned. This ill-conditioning amplifies the numerical errors occurring in matrix inversion, the small errors in the Jacobian $\mathbf{K}_i$ (due to code optimizations), and the measurement noise errors, to a level

that prevents the GN method from converging. Therefore, the presented test case requires finishing the retrieval with an LM damping factor different from zero, and, as described in Sect. 2, the solution, and its VCM and AKM, depend on the path followed by the minimization procedure in the parameter space.

3.2 Calculation of the VCM

We calculated the VCM of the obtained solution using four different approaches. The first calculation is done as suggested in Press et al. (1992) and is obtained from the LM formula of the last iteration step with damping factor equal to zero (as if the last iteration was a GN iteration). This method provides the following formula for the VCM (we remind the reader that in our algorithm we use $\mathbf{R}=\mathbf{0}$):

$$\mathbf{S}_r^{(1)} = \left(\mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} \right)^{-1}. \tag{14}$$

The second calculation is done as suggested in Ceccherini et al. (2007) and is obtained from the LM formula of the last iteration, with the actual value of the damping factor. This method provides the following formula for the VCM:

$$\mathbf{S}_r^{(2)} = \left(\mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} + \lambda_{r-1} \mathbf{D}_{r-1} \right)^{-1}$$
$$\mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} \left(\mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} + \lambda_{r-1} \mathbf{D}_{r-1} \right)^{-1}. \tag{15}$$

The third calculation provides the VCM $\mathbf{S}_r^{(3)}$ proposed in Sect. 2 and is obtained from Eqs. (8) and (10).

The fourth calculation consists in the a-posteriori calculation of the VCM on the basis of its statistical estimator (Bevington and Robinson, 2003) applied to the profiles resulting from a statistically significant set of simulated retrievals. We carried out 1000 retrievals from 1000 sets of synthetic measurements generated assuming the same atmospheric state and measurement conditions, but with different seeds in the algorithm generating the random noise added to the forward model simulations. Then we estimated the VCM as:

$$\mathbf{S}_r^{(4)} = \overline{(\boldsymbol{x}_r - \overline{\boldsymbol{x}_r})(\boldsymbol{x}_r - \overline{\boldsymbol{x}_r})^T} \tag{16}$$

where the over-bar denotes the arithmetic average of the results of the 1000 retrievals. We consider this estimate of the VCM to be very accurate as it is based on a statistically significant set of simulated retrievals and its evaluation does not involve shortcuts or approximations.

Although the convergence criteria used are very conservative, it could be argued that an occasionally large LM damping factor could produce a small chi-square variation far from convergence, and erroneously trigger a successful convergence check. In order to verify that this condition does not occur in our retrievals we checked a-posteriori the accuracy of the obtained minimum. For this purpose we calculated the average of the final reduced chi-square values obtained in the 1000 retrievals. We got 1.02, to be compared with

Table 1. Details of the iterations required for the minimization of the cost function (chi-square) in our test retrieval. Rows of the table with the same iteration number correspond to the fact that the step performed with the input damping factor produces an increase of chi-square and that the iteration is repeated with increased damping factor until a decrease of chi-square is obtained.

Iteration	Damping factor	Reduced chi-square
0		34.0310
1	0.1	1.2466
2	0.025	1.0862
3	0.00625	1.1607
3	0.05	1.0859
4	0.0125	1.1233
4	0.1	1.0757
5	0.025	1.0972
5	0.2	1.0666
6	0.05	1.0793
6	0.4	1.0590
7	0.1	1.0665
7	0.8	1.0535
8	0.2	1.0577
8	1.6	1.0501
9	0.4	1.0520
9	3.2	1.0481
10	0.8	1.0480

the expected value of 1 (we are using simulated observations, therefore the forward model errors do not affect the chi-square value). If we attribute this difference to the convergence error, then the squared ratio between the convergence error and the retrieval error due to measurement noise is 0.02. Therefore the convergence error is at least a factor of 7 smaller than the retrieval error due to measurement noise.

In Figs. 2–4 we compare the four calculated VCMs. In Fig. 2 we report the square roots of the diagonal elements of the VCMs: they represent the errors on the retrieved VMRs at the different altitudes. The lines in red, black and blue are the average on the 1000 retrievals of the square roots of the diagonal elements of $\mathbf{S}_r^{(1)}$, $\mathbf{S}_r^{(2)}$ and $\mathbf{S}_r^{(3)}$, respectively. The gray areas around the lines represent the standard deviations of the plotted quantities. The gray area around the red line is not visible due to the extremely small fluctuations of $\mathbf{S}_r^{(1)}$ in the test retrievals. The green line reports the square roots of the diagonal elements of $\mathbf{S}_r^{(4)}$. The comparison of the errors estimated by the three analytical methods to the errors evaluated a-posteriori with the statistical estimator of Eq. (16) shows that $\mathbf{S}_r^{(1)}$ overestimates the errors below 30 km, $\mathbf{S}_r^{(2)}$ underestimates the errors in the whole altitude range and $\mathbf{S}_r^{(3)}$ provides a good estimation of the errors in the whole altitude range.
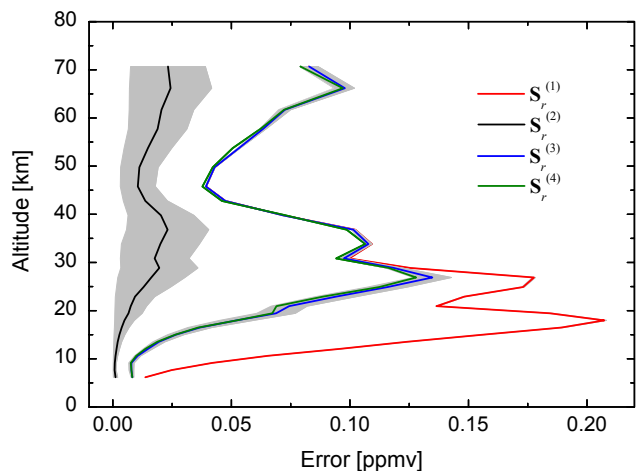
Fig. 2. Average values (ppmv) of the square roots of the diagonal elements of $\mathbf{S}_r^{(1)}$ (red), $\mathbf{S}_r^{(2)}$ (black) and $\mathbf{S}_r^{(3)}$ (blue) on 1000 simulated retrievals. The gray areas around the lines represent the standard deviations. The gray area around the red line is not visible due to the extremely small fluctuations of $\mathbf{S}_r^{(1)}$ in the test retrievals. The green curve represents the square roots of the diagonal elements of $\mathbf{S}_r^{(4)}$. See text for definitions of $\mathbf{S}_r^{(1)}$, $\mathbf{S}_r^{(2)}$, $\mathbf{S}_r^{(3)}$ and $\mathbf{S}_r^{(4)}$.

Figure 3 shows the correlations extracted from the four calculated VCMs. Figure 4 reports the differences between the correlations extracted from $\mathbf{S}_r^{(1)}$, $\mathbf{S}_r^{(2)}$, $\mathbf{S}_r^{(3)}$ and those extracted from $\mathbf{S}_r^{(4)}$. While the correlations calculated from $\mathbf{S}_r^{(1)}$ and $\mathbf{S}_r^{(2)}$ significantly deviate from the correlations calculated from $\mathbf{S}_r^{(4)}$, the correlations calculated from $\mathbf{S}_r^{(3)}$ and $\mathbf{S}_r^{(4)}$ are in very good agreement.

The results reported in Figs. 2–4 show that, globally, the VCM calculated with the method proposed in Sect. 2 ($\mathbf{S}_r^{(3)}$) agrees very well with the VCM estimated a-posteriori ($\mathbf{S}_r^{(4)}$), whereas the VCMs calculated with the approximated methods proposed in Press et al. (1992) ($\mathbf{S}_r^{(1)}$) and Ceccherini et al. (2007) ($\mathbf{S}_r^{(2)}$) significantly deviate from it.

As a further test to show that the retrieval error is properly represented by the VCM $\mathbf{S}_r^{(3)}$, for each of the 1000 retrievals described earlier we calculated the quantity α defined as:

$$\alpha = \left[(\mathbf{x}_r - \mathbf{x}_{\text{true}})^T \left(\mathbf{S}_r^{(3)} \right)^{-1} (\mathbf{x}_r - \mathbf{x}_{\text{true}}) \right] / n \quad (17)$$

where $\mathbf{x}_r$ and $\mathbf{x}_{\text{true}}$ are respectively the retrieved profile and the true profile assumed for the generation of the simulated observations. Then we calculated the average $\overline{\alpha}$ over the 1000 retrievals of the quantity α. By definition, if $\mathbf{S}_r^{(3)}$ represents a correct estimate of the retrieval error, i.e. if $\mathbf{S}_r^{(3)}$ estimates properly the retrieval error due to measurement noise and if this is the dominant error component affecting the retrieved profiles, then we expect $\alpha = 1$. We got $\overline{\alpha} = 0.96$. This value confirms that the convergence error is negligible and

indicates a 2% overestimation of the error given by $\mathbf{S}_r^{(3)}$. This marginal overestimation is consistent with the plot of Fig. 2.

3.3 Calculation of the AKM

According to the scheme adopted for the calculation of the VCM we calculated the AKM of the solution obtained using four different calculations. The first calculation is obtained from the LM formula of the last iteration step with LM damping factor equal to zero. This method provides the identity matrix as AKM:

$$\mathbf{A}_r^{(1)} = \mathbf{I} \quad (18)$$

The second calculation is done as suggested in Ceccherini et al. (2007) and is obtained from the LM formula of the last iteration step with the actual value of the damping factor. This method provides the following formula:

$$\mathbf{A}_r^{(2)} = \left(\mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} + \lambda_{r-1} \mathbf{D}_{r-1} \right)^{-1} \mathbf{K}_{r-1}^T \mathbf{S}_y^{-1} \mathbf{K}_{r-1} \quad (19)$$

The third calculation provides the AKM $\mathbf{A}_r^{(3)}$ proposed in Sect. 2 and is obtained from Eqs. (10) and (11).

The fourth calculation provides the AKM $\mathbf{A}_r^{(4)}$ estimated with the numerical calculation of the derivatives appearing in the definition of AKM contained in Eq. (11). The procedure for this calculation is as follows. We perturb a level of the "true" ozone VMR profile and calculate with the forward model the observations corresponding to the perturbed profile. We perform the retrieval from these observations and calculate the difference between the retrieved profile and the profile retrieved from synthetic observations assuming an unperturbed atmosphere. The ratio between this difference and the input perturbation provides the entries of the column of $\mathbf{A}_r^{(4)}$ that corresponds to the altitude of the perturbation. By applying sequentially the perturbation to all the levels of the "true" profile, we calculate the whole $\mathbf{A}_r^{(4)}$. After several trials we selected a perturbation amplitude of 0.01 ppmv (parts per million by volume). This value turns out to be small enough for a good approximation of the incremental limit of the derivatives and, at the same time, large enough to avoid numerical rounding errors. We consider $\mathbf{A}_r^{(4)}$ the most accurate estimate of the AKM and, therefore, we assume this matrix to be the reference AKM of our retrieval.

Figure 5 shows the averaging kernels in the rows of $\mathbf{A}_r^{(2)}$ (panel a), $\mathbf{A}_r^{(3)}$ (panel b) and $\mathbf{A}_r^{(4)}$ (panel c). The trivial case $\mathbf{A}_r^{(1)} = \mathbf{I}$ is not shown. In Fig. 6 we report the differences between the averaging kernels in the rows of $\mathbf{A}_r^{(1)}$ (panel a), $\mathbf{A}_r^{(2)}$ (panel b), $\mathbf{A}_r^{(3)}$ (panel c) and those in the rows of $\mathbf{A}_r^{(4)}$. This comparison shows that the averaging kernels of $\mathbf{A}_r^{(1)}$ corresponding to the altitudes below 30 km significantly differ from those of $\mathbf{A}_r^{(4)}$. All the averaging kernels of $\mathbf{A}_r^{(2)}$ significantly differ from those of $\mathbf{A}_r^{(4)}$. All the averaging kernels of $\mathbf{A}_r^{(3)}$ (calculated with the method proposed in Sect. 2) agree well with those of $\mathbf{A}_r^{(4)}$.
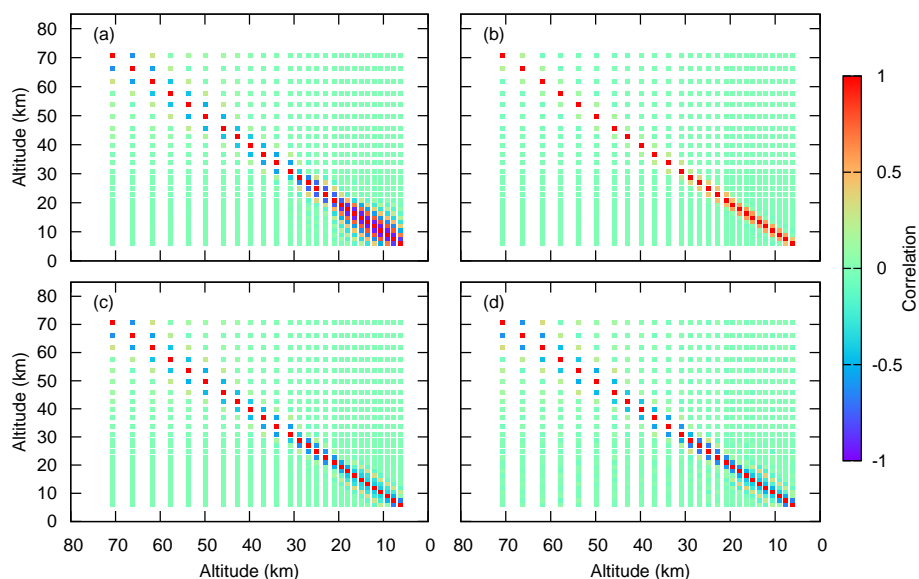
Fig. 3. Correlation matrices derived from $\mathbf{S}_r^{(1)}$ (a), $\mathbf{S}_r^{(2)}$ (b), $\mathbf{S}_r^{(3)}$ (c) and $\mathbf{S}_r^{(4)}$ (d). See text for definitions of $\mathbf{S}_r^{(1)}$, $\mathbf{S}_r^{(2)}$, $\mathbf{S}_r^{(3)}$ and $\mathbf{S}_r^{(4)}$.
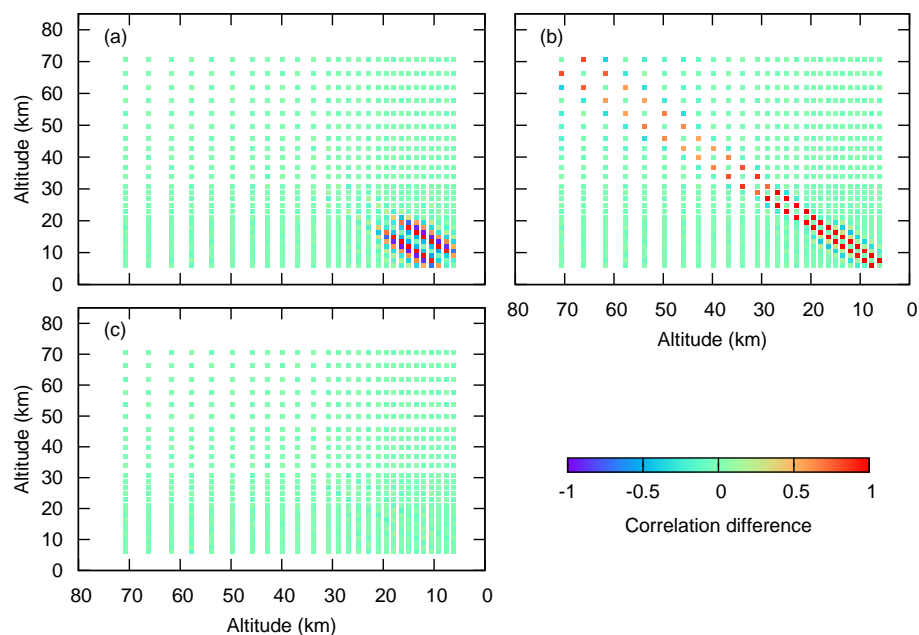

Fig. 4. Differences between the correlations derived from $\mathbf{S}_r^{(1)}$ (a), $\mathbf{S}_r^{(2)}$ (b), $\mathbf{S}_r^{(3)}$ (c) and those derived from $\mathbf{S}_r^{(4)}$. See text for definitions of $\mathbf{S}_r^{(1)}$, $\mathbf{S}_r^{(2)}$, $\mathbf{S}_r^{(3)}$ and $\mathbf{S}_r^{(4)}$.

In summary, we see that both the VCM and the AKM estimated with the method proposed in this paper agree well with the VCM and AKM calculated using the statistical estimator and the numerical derivatives, respectively. Whereas, the two methods reported in the literature significantly deviate from the accurate a-posteriori estimates.

The implementation of the proposed method requires the evaluation of $\mathbf{T}_i$ at each iteration with Eq. (10) and its storage up to the subsequent iteration for the calculation of $\mathbf{T}_{i+1}$. At the end of the iterations the matrix products in Eqs. (8) and (11) need also to be evaluated. In the test case we examined, the additional computing resources required for these operations are not relevant when compared to the overall amount of resources necessary for the retrieval.
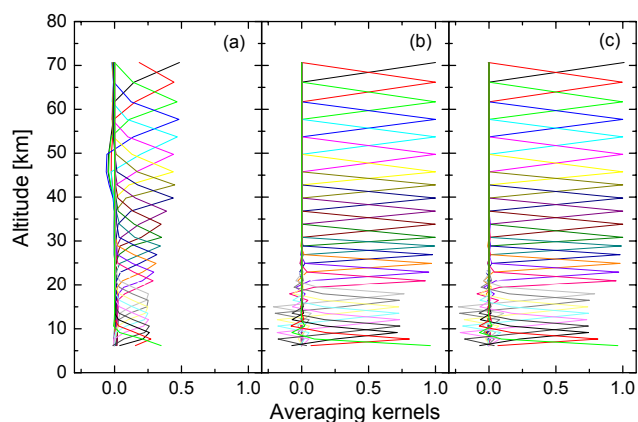
Fig. 5. Averaging kernels in the rows of $\mathbf{A}_r^{(2)}$ **(a)**, $\mathbf{A}_r^{(3)}$ **(b)** and $\mathbf{A}_r^{(4)}$ **(c)**. See text for definitions of $\mathbf{A}_r^{(2)}$, $\mathbf{A}_r^{(3)}$ and $\mathbf{A}_r^{(4)}$.

Fig. 6. Differences between the averaging kernels in the rows of $\mathbf{A}_r^{(1)}$ **(a)**, $\mathbf{A}_r^{(2)}$ **(b)**, $\mathbf{A}_r^{(3)}$ **(c)** and those in the rows of $\mathbf{A}_r^{(4)}$. See text for definitions of $\mathbf{A}_r^{(1)}$, $\mathbf{A}_r^{(2)}$, $\mathbf{A}_r^{(3)}$ and $\mathbf{A}_r^{(4)}$.

4 Conclusions

We studied the problem of accurate evaluation of the VCM and the AKM of vertical atmospheric distribution profiles retrieved using the LM iterative technique. Two methods are reported in literature providing the VCM and AKM on the basis of some approximations. One method assumes the last iteration as a GN iteration (Press et al., 1992), the other accounts for the non-negligible LM damping but assumes a single iteration retrieval (Ceccherini et al., 2007). In this paper we propose a new method. The method uses a rigorous calculation that takes into account all the iterations required by the minimization procedure, from the initial guess to the solution. In well-conditioned retrievals that reach exact convergence the VCM and AKM calculated with the proposed method coincide with those calculated assuming the last iteration as a GN iteration. In ill-conditioned retrievals the LM method acts as an external constraint and the solution depends on the path followed by the minimization procedure in the parameter space. This latter conclusion applies also to retrievals in which the iterations are stopped when a physically meaningful convergence criterion is fulfilled, i.e. before achievement of the numerical convergence at machine precision. In these cases the VCM and AKM calculated with the proposed method differ from those calculated assuming the last iteration as a GN iteration. We perform an ill-conditioned inversion and compare the VCMs and the AKMs derived using the proposed method and the two methods reported in the literature with accurate estimates derived a-posteriori using the statistical estimator for the VCM and the numerical derivatives for the AKM. The VCM and AKM calculated with the two methods reported in the literature significantly deviate from the accurate a-posteriori estimates. Whereas, the new proposed method provides VCM and AKM that are in good agreement with the a-posteriori estimates. The implementation of the method does not require significant additional computing resources.
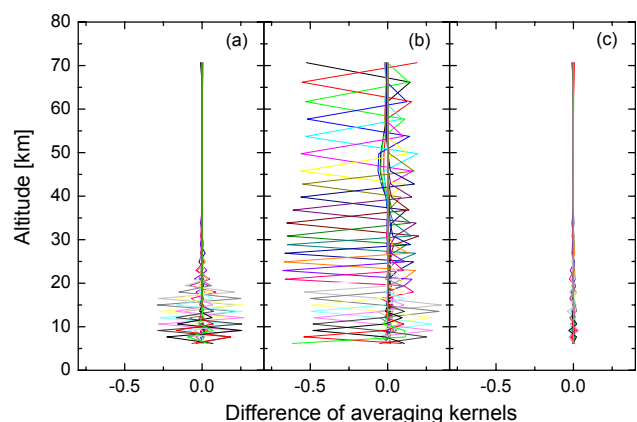
References

Bevington, P. R. and Robinson, D. K.: Data Reduction and Error Analysis for the Physical Sciences, 3rd edition, McGraw-Hill, New York, 2003.

Bowman, K. W., Rodgers, C. D., Kulawik, S. S., Worden, J., Sarkissian, E., Osterman, G., Steck, T., Lou, M., Eldering, A., Shephard, M., Worden, H., Lampel, M., Clough, S., Brown, P., Rinsland, C., Gunson, M., and Beer, R.: Tropospheric emission spectrometer: Retrieval method and error analysis, IEEE T. Geosci. Remote, 44(5), 1297–1307, 2006.

Ceccherini, S., Carli, B., Pascale, E., Prosperi, M., Raspollini, P., and Dinelli, B. M.: Comparison of measurements made with two different instruments of the same atmospheric vertical profile, Appl. Optics, 42(32), 6465–6473, 2003.

Ceccherini, S.: Analytical determination of the regularization parameter in the retrieval of atmospheric vertical profiles, Opt. Lett., 30, 2554–2556, 2005.

Ceccherini, S., Belotti, C., Carli, B., Raspollini, P., and Ridolfi, M.: Technical Note: Regularization performances with the error consistency method in the case of retrieved atmospheric profiles, Atmos. Chem. Phys., 7, 1435–1440, 2007,
http://www.atmos-chem-phys.net/7/1435/2007/.

Ceccherini, S., Raspollini, P., and Carli, B.: Optimal use of the information provided by indirect measurements of atmospheric vertical profiles, Opt. Express, 17(7), 4944–4958, 2009.

Ceccherini, S., Carli, B., Cortesi, U., Del Bianco, S., and Raspollini, P.: Retrieval of the vertical column of an atmospheric constituent from data fusion of remote sensing measurements, J. Quant. Spectrosc. Ra., 111, 507–514, doi:10.1016/j.jqsrt.2009.09.001, 2010.

Cortesi, U., Lambert, J. C., De Clercq, C., Bianchini, G., Blumenstock, T., Bracher, A., Castelli, E., Catoire, V., Chance, K. V., De Mazière, M., Demoulin, P., Godin-Beekmann, S., Jones, N., Jucks, K., Keim, C., Kerzenmacher, T., Kuellmann, H., Kuttippurath, J., Iarlori, M., Liu, G. Y., Liu, Y., McDermid, I. S., Meijer, Y. J., Mencaraglia, F., Mikuteit, S., Oelhaf, H., Piccolo, C., Pirre, M., Raspollini, P., Ravegnani, F., Reburn, W. J., Redaelli, G., Remedios, J. J., Sembhi, H., Smale, D., Steck, T., Taddei, A., Varotsos, C., Vigouroux, C., Waterfall, A., Wetzel, G., and Wood, S.: Geophysical validation of MIPAS-ENVISAT operational ozone data, Atmos. Chem. Phys., 7, 4807–4867, 2007, http://www.atmos-chem-phys.net/7/4807/2007/.

Doicu, A., Schreier, F., and Hess, M.: Iterative regularization methods for atmospheric remote sensing, J. Quant. Spectrosc. Ra., 83, 47–61, 2004.

Dudhia, A., Jay, V. L., and Rodgers, C. D.: Microwindow selection for high-spectral-resolution sounders, Appl. Optics, 41, 3665–3673, 2002.

Dudhia, A.: MIPAS reduced resolution modes, Oxford University: http://www.atm.ox.ac.uk/group/mipas/rrmodes.html, last access: 8 September 2009, 2008.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, 2008, http://www.atmos-chem-phys.net/8/2151/2008/.

Lahoz, W. A., Geer, A. J., Bekki, S., Bormann, N., Ceccherini, S., Elbern, H., Errera, Q., Eskes, H. J., Fonteyn, D., Jackson, D. R., Khattatov, B., Marchand, M., Massart, S., Peuch, V.-H., Rharmili, S., Ridolfi, M., Segers, A., Talagrand, O., Thornton, H. E., Vik, A. F., and von Clarmann, T.: The Assimilation of Envisat data (ASSET) project, Atmos. Chem. Phys., 7, 1773–1796, 2007, http://www.atmos-chem-phys.net/7/1773/2007/.

Levenberg, K.: A method for the solution of certain problems in least squares, Q. Appl. Math., 2, 164–168, 1944.

Marquardt, D. W.: An algorithm for the least-squares estimation of nonlinear parameters, SIAM, J. Appl. Math., 11, 431–441, 1963.

Pougatchev, N.: Validation of atmospheric sounders by correlative measurements, Appl. Optics, 47(26), 4739–4748, 2008.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes in C: The Art of Scientific Computing, 2nd edition, Cambridge Univ. Press, Cambridge, 1992.

Pujol, J.: The solution of nonlinear inverse problems and the Levenberg-Marquardt method, Geophysics, 72(4), W1–W16, doi:10.1190/1.2732552, 2007.

Raspollini, P., Belotti, C., Burgess, A., Carli, B., Carlotti, M., Ceccherini, S., Dinelli, B. M., Dudhia, A., Flaud, J.-M., Funke, B., Höpfner, M., López-Puertas, M., Payne, V., Piccolo, C., Remedios, J. J., Ridolfi, M., and Spang, R.: MIPAS level 2 operational analysis, Atmos. Chem. Phys., 6, 5605–5630, 2006, http://www.atmos-chem-phys.net/6/5605/2006/.

Remedios, J.J., Leigh, R. J., Waterfall, A.M., Moore, D.P., Sembhi, H., Parkes, I., Greenhough, J., Chipperfield, M.P.,Hauglustaine, D.: MIPAS reference atmospheres and comparisons to V4.61/V4.62 MIPAS level 2 geophysical data sets, Atmos. Chem. Phys. Discuss., 7, 9973–10017, 2007, http://www.atmos-chem-phys-discuss.net/7/9973/2007/.

Ridolfi, M., Carli, B., Carlotti, M., von Clarmann, T., Dinelli, B. M., Dudhia, A., Flaud, J.-M., Hoepfner, M., Morris, P. E., Raspollini, P., Stiller, G., and Wells, R. J.: Optimized forward model and retrieval scheme for MIPAS near-real-time data processing, Appl. Optics, 39(8), 1323–1340, 2000.

Ridolfi, M., Blum, U., Carli, B., Catoire, V., Ceccherini, S., Claude, H., De Clercq, C., Fricke, K. H., Friedl-Vallon, F., Iarlori, M., Keckhut, P., Kerridge, B., Lambert, J.-C., Meijer, Y. J., Mona, L., Oelhaf, H., Pappalardo, G., Pirre, M., Rizi, V., Robert, C., Swart, D., von Clarmann, T., Waterfall, A., and Wetzel, G.: Geophysical validation of temperature retrieved by the ESA processor from MIPAS/ENVISAT atmospheric limb-emission measurements, Atmos. Chem. Phys., 7, 4459–4487, 2007, http://www.atmos-chem-phys.net/7/4459/2007/.

Ridolfi, M. and Sgheri, L.: A self-adapting and altitude-dependent regularization method for atmospheric profile retrievals, Atmos. Chem. Phys., 9, 1883–1897, 2009, http://www.atmos-chem-phys.net/9/1883/2009/.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, Vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics World Scientific, Singapore, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 4116–4130, 2003.

Schimpf, B. and Schreier, F.: Robust and efficient inversion of vertical sounding atmospheric high-resolution spectra by means of regularization, J. Geophys. Res., 102(D13), 16037–16055, 1997.

Sofieva, V. F., Tamminen, J., Haario, H., Kyrölä, E., and Lehtinen, M.: Ozone profile smoothness as a priori information in the inversion of limb measurements, Ann. Geophys., 22, 3411–3420, 2004, http://www.ann-geophys.net/22/3411/2004/.

Twomey, S.: Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements, Elsevier, New York, 1977.