

Elimination of hidden a priori information from remotely sensed profile data

T. von Clarmann and U. Grabowski

Forschungszentrum Karlsruhe, Institut für Meteorologie und Klimaforschung, Karlsruhe, Germany

Received: 18 April 2006 – Published in Atmos. Chem. Phys. Discuss.: 18 July 2006

Revised: 13 November 2006 – Accepted: 11 January 2007 – Published: 23 January 2007

Abstract. Profiles of atmospheric state variables retrieved from remote measurements often contain a priori information which causes complication in the statistical use of data and in the comparison with other measured or modeled data. For such applications it often is desirable to remove the a priori information from the data product. If the retrieval involves an ill-posed inversion problem, formal removal of the a priori information requires resampling of the data on a coarser grid, which in some sense, however, is a prior constraint in itself. The fact that the trace of the averaging kernel matrix of a retrieval is equivalent to the number of degrees of freedom of the retrieval is used to define an appropriate information-centered representation of the data where each data point represents one degree of freedom. Since regridding implies further degradation of the data and thus causes additional loss of information, a re-regularization scheme has been developed which allows resampling without additional loss of information. For a typical ClONO₂ profile retrieved from spectra as measured by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), the constrained retrieval has 9.7 degrees of freedom. After application of the proposed transformation to a coarser information-centered altitude grid, there are exactly 9 degrees of freedom left, and the averaging kernel on the coarse grid is unity. Pure resampling on the information-centered grid without re-regularization would reduce the degrees of freedom to 7.1 (6.7) for a staircase (triangular) representation scheme.

recent methods mostly rely on information theory (Rodgers, 2000). It is now common practise to represent the state parameter to be retrieved on an altitude grid which is finer than the altitude resolution of the instrument (see, e.g. von Clarmann et al., 2003b). Since inference of vertical profiles of atmospheric state parameters on such a fine grid otherwise would lead to an ill-posed inversion problem, stochastic (Rodgers, 2000) or other, e.g., smoothness (Tikhonov, 1963; Twomey, 1963; Steck and von Clarmann, 2001) constraints are applied which make the retrieval stable, i.e. regularize the inversion problem.

The disadvantage of regularization is that the elements of the solution of the retrieval problem do carry inter-dependent information, and that the solution does not merely depend on the measurement but includes a certain portion of a priori information. The following problem areas in the use of constrained retrievals have been identified: A meaningful comparison of two profiles of different content of a priori information requires to estimate the uncertainty of the difference of the profiles which is due to their different a priori content (Rodgers and Connor, 2003). The estimation of this uncertainty, the so-called “smoothing error of the difference” relies on the knowledge of the climatological variability in all scales of the quantity to be compared, in order to be able to test the significance of the observed differences. Often reliable climatological variance-covariance information is not available. These artefacts in profile differences do not only affect applications like quantitative validation of remotely sensed data but even visual inspection of measured data. It is virtually impossible to recognize significant structures (which basically are differences of the observed state variable) and to distinguish them from artefacts caused by a priori information in, e.g., time series or 2-D plots of results. No meaningful sums (e.g. total reactive atmospheric nitrogen calculated from its components) or ratios (mainly used for correlation analyses of trace species, e.g., Kondo et al. (1999); Rex et al. (1999); Plumb et al. (2000); Wetzell et al.

1 Introduction

While in early applications of remote sensing to atmospheric research retrieval approaches often were based on ad hoc methods or unconstrained maximum likelihood estimation,

Correspondence to: T. von Clarmann
(thomas.clarmann@imk.fzk.de)

(2002); Ray et al. (2002); Glatthor et al. (2005); Mengistu Tsidu et al. (2005); Muscari et al. (2003); Esler and Waugh (2002); Müller et al. (1999)) of results can be calculated, if the a priori content of the data is significantly different or varying. Furthermore, depending on the nature of the a priori information, profile retrievals at different geolocations are no longer statistically independent. This adds complication to statistical applications such as averaging or data assimilation. Beyond this, there are two further, certainly non-scientific but nevertheless practically relevant problems with regularized profiles: First, the data amount of a complete set of diagnostic data can be enormous. Second, the experience was made that a majority of data users refuses to deal with this diagnostic information and thus runs risk to misinterpret the data.

For applications as discussed above, a representation of the retrieved state parameters which is free of formal a priori information is desirable and often advantageous. Singular value decomposition-based approaches have been developed for similar purposes: Ceccherini et al. (2003) have proposed such an approach for validation purposes targeted at maximum likelihood estimates of the atmospheric state. Joiner and da Silva (1998) proposed two methods, namely null-space filtering of retrievals and partial eigen-decomposition retrievals. All these approaches, however, transform the estimated atmospheric state variables into a space without obvious physical meaning. This paper is targeted at a physically obvious representation of retrieved data which can directly be used without running risk of major misinterpretation, without giving up the major advantage of the regularized retrieval, which is its inherent stability.

2 Theoretical background

With very few exceptions explicitly mentioned, we use the theoretical retrieval concept established by Rodgers (2000). While all MIPAS data processors the authors are aware of perform the retrieval in the framework of Newtonian iteration, for the discussion of the method presented in this paper the linear formulation of the retrieval problem is sufficient and used. The estimate of the n -dimensional atmospheric state vector $\hat{\mathbf{x}}$ is calculated from the m -dimensional vector containing the measured spectral radiances, \mathbf{y} , as

$$\begin{aligned}\hat{\mathbf{x}} &= \mathbf{x}_a + (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{R})^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_a)) \\ &= \mathbf{x}_a + \mathbf{G}(\mathbf{y} - \mathbf{F}(\mathbf{x}_a)),\end{aligned}\quad (1)$$

where \mathbf{x}_a contains the a priori information on the atmospheric state, $\mathbf{F}(\mathbf{x})$ is the radiative transfer forward model which provides the spectral radiances as a function of \mathbf{x} , \mathbf{K} is the $m \times n$ Jacobian matrix, whose elements are $\frac{\partial y_i}{\partial x_j}$; superscript T denotes transposed matrices; \mathbf{S}_ϵ is the measurement covariance matrix. \mathbf{R} is a regularization term, which, in maximum a posteriori (formerly called “optimal estimation”) applications as suggested by Rodgers (2000), is the inverse of

the a priori covariance matrix, while, in other applications, it often is a first order smoothing operator (Twomey, 1963; Tikhonov, 1963). Setting \mathbf{R} zero leads to the unconstrained least squares solution, which is also called maximum likelihood solution. \mathbf{G} is called the gain function. All complications arising from iterative processing to solve problems caused by non-linearity of $\mathbf{F}(\mathbf{x}_a)$ are omitted here since they do not contribute to the aspect of the problem discussed. The retrieval covariance matrix \mathbf{S}_x of $\hat{\mathbf{x}}$ is

$$\mathbf{S}_x = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{R})^{-1}, \quad (2)$$

while the noise error covariance matrix \mathbf{S}_m of $\hat{\mathbf{x}}$ is

$$\mathbf{S}_m = \mathbf{G} \mathbf{S}_\epsilon \mathbf{G}^T. \quad (3)$$

Removal of the effect of the a priori information without rerunning the retrieval, i.e. the transformation of the maximum a posteriori solution $\hat{\mathbf{x}}_{\text{MAP}}$, or any other regularized solution $\hat{\mathbf{x}}_{\text{REG}}$ to the related maximum likelihood solution $\hat{\mathbf{x}}_{\text{ML}}$ according to

$$\hat{\mathbf{x}}_{\text{ML}} = (\mathbf{S}_x^{-1} - \mathbf{R})^{-1} \left[\mathbf{S}_x^{-1} \hat{\mathbf{x}}_{\text{REG}} - \mathbf{R} \mathbf{x}_a \right] \quad (4)$$

usually first needs resampling on a coarser grid to avoid singularity of $\mathbf{S}_x^{-1} - \mathbf{R} = \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}$. It is worthwhile mentioning that the maximum likelihood solution is not really unconstrained but implicitly constrained by the coarse grid on which it is represented and the related interpolation convention.

The $n \times n$ averaging kernel matrix

$$\mathbf{A} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{R})^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} = \mathbf{G} \mathbf{K} \quad (5)$$

is a helpful diagnostic tool, which is useful to rewrite Eq. (1) such that it becomes obvious which part of the solution is controlled by the measurement and which by the a priori information:

$$\hat{\mathbf{x}}_{\text{REG}} = \mathbf{x}_a + \mathbf{A}(\mathbf{x} - \mathbf{x}_a) + \mathbf{G}\epsilon, \quad (6)$$

where \mathbf{x} represents the true atmospheric state and ϵ is the measurement noise characterized by \mathbf{S}_ϵ . In the following we assume $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{\text{REG}}$.

The number of degrees of freedom of the signal was shown to be

$$d\text{g}f_{\text{signal}} = \text{tr}(\mathbf{A}) \quad (7)$$

for optimal estimation applications where $\mathbf{R} = \mathbf{S}_a^{-1}$ (Rodgers, 2000). However, this concept also holds for other regularization methods such as smoothing in the sense of Tikhonov (1963) or Twomey (1963) (see Appendix A). To avoid confusion, we follow the suggestion of Steck (2002) and use the term “number of degrees of freedom of the retrieval ($d\text{g}f_{\text{retrieval}}$)” here rather than “number of degrees of freedom of the signal ($d\text{g}f_{\text{signal}}$)” which is reserved for maximum a posteriori retrievals in a Bayesian sense.

3 Appropriate representation

In the following we first address the question which representation of the retrieved profile is appropriate. To allow the removal of the formal a priori information, the grid on which the profile is represented shall be coarse enough for the inversion of $\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K}$ to be stable, but no coarser, in order to avoid unnecessary loss of information. Then we suggest a scheme to transform the data retrieved by Eq. (1) to the appropriate grid such that the averaging kernel matrix in the coarse grid is zero while the loss of information is minimal. This transformation scheme, however, can also be used to represent the data on any grid coarser than the one found most appropriate for the particular profile. The latter application is particularly useful, if profiles of different state variables, or profiles of one state variable at different times or locations shall be represented in a comparable manner, which generally implies that at least one of the profiles has to be represented on a sub-optimal grid.

We assume that the strength of the regularization in Eq. (1) has been chosen appropriately (in a Bayesian sense or any other optimality criterion, e.g. Steck, 2002) and that the number of degrees of freedom of the regularized solution is the maximum number of degrees of freedom reasonably obtainable from the given measurement.

A linear transformation of the retrieved profile $\hat{\mathbf{x}}$ to a coarser grid uses the transformation matrix

$$\mathbf{W}^* = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \quad (8)$$

where \mathbf{W} is the interpolation matrix which samples the coarse grid profile on the fine grid. The profile on the coarse grid then is

$$\tilde{\mathbf{x}} = \mathbf{W}^* \hat{\mathbf{x}} \quad (9)$$

and the coarse-grid Jacobian is

$$\tilde{\mathbf{K}} = \mathbf{K} \mathbf{W} \quad (10)$$

The averaging kernel matrix on the coarse grid, $\tilde{\mathbf{A}} \in IR^{k \times n}$ is

$$\tilde{\mathbf{A}} = \tilde{\mathbf{G}} \mathbf{K} = \mathbf{W}^* \mathbf{G} \mathbf{K} = \mathbf{W}^* \mathbf{A} \quad (11)$$

The $\tilde{\mathbf{A}} \in IR^{k \times n}$ averaging kernel matrix is understood to represent the response of a retrieval in the k -dimensional grid to a delta perturbation in the finer n -dimensional grid. With $\mathbf{W}^* \mathbf{W} = \tilde{\mathbf{I}}$ where $\tilde{\mathbf{I}}$ is $k \times k$ unity, the averaging kernel $\tilde{\mathbf{A}} \in IR^{k \times k}$ related to a retrieval in the coarser k -dimensional grid is

$$\tilde{\mathbf{A}} = \tilde{\mathbf{G}} \tilde{\mathbf{K}} \quad (12)$$

$$= \mathbf{W}^* \mathbf{G} \mathbf{K} \mathbf{W} \quad (13)$$

$$= \mathbf{W}^* \mathbf{A} \mathbf{W} \quad (14)$$

$$= \left(\mathbf{W}^T \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \mathbf{W} + \mathbf{W}^T \mathbf{R} \mathbf{W} \right)^{-1} \mathbf{W}^T \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \mathbf{W} \quad (15)$$

Averaging kernels of Eqs. (11) and (12) are not directly comparable, because they have delta functions of different widths as a reference: The delta function in the k -dimensional grid, which is the reference to Eq. (12), is wider than the delta function in the n -dimensional grid, which is the reference for Eq. (11).

The $k \times k$ covariance matrix of the retrieval represented in the coarse grid is

$$\mathbf{S}_{\tilde{\mathbf{x}}} = \mathbf{W}^* \mathbf{S}_{\mathbf{x}} \mathbf{W}^{*T} \quad (16)$$

An appropriate representation of the retrieval in a sense as discussed above fulfills the conditions

$$\tilde{\tilde{\mathbf{A}}} = \tilde{\tilde{\mathbf{I}}}, \quad (17)$$

and

$$k = \text{int}(dgf_{\text{retrieval}}) \quad (18)$$

where $dgf_{\text{retrieval}}$ here is the number of degrees of freedom of the retrieval in the fine grid, which is not typically an integer. As follows from Eq. (15), Eq. (17) can be satisfied with

$$\mathbf{W}^T \mathbf{R} \mathbf{W} = \mathbf{0}, \quad (19)$$

for which we have to find a non-trivial solution, i.e. $\mathbf{R} \neq \mathbf{0}$ and $\mathbf{W} \neq \mathbf{0}$. This means that the transformed retrieval shall keep as many degrees of freedom as possible while the state parameters in the k -dimensional grid shall not be constrained to each other. This kind of representation of a vertical profile where each data point represents one degree of freedom we call “information-centered”.

The number of useful gridpoints can be obtained by singular value decomposition, but not the placement of the altitude gridpoints. Therefore, we propose an approach which is based on the number of degrees of freedom of the retrieval. If the number of degrees of freedom, i.e. the number of independent pieces of information in the measurement, is equal to the dimension of the retrieval vector, an averaging kernel can be inverted, i.e. the a priori information can be removed from the retrieval. In the following we discuss two approaches to remove the a priori information from the retrieval and to represent the retrieval on a coarser altitude grid.

3.1 Staircase representation

First, we determine, how many degrees of freedom dgf_c shall be represented by each component of the k -dimensional state vector in the coarse grid. The ideal value would be one, but since dgf generally is not an integer, dgf_c of each altitude step j is calculated as

$$\forall j : dgf_c = \frac{dgf}{\text{int}(dgf)} \quad (20)$$

Certainly, the excess information related to $k \times (dgf_c - dgf)$ is lost through slight undersampling on the coarse k -dimensional grid. This, however, is the price to pay for an easy representation.

The new gridpoints \tilde{h}_j are distributed over altitude such that

$$\sum_{l=l_1(j)}^{l_2(j)} a_{l,l} \approx dgfc \approx 1 \quad \text{where} \quad \mathbf{A} = (a_{i_1 i_2}) \quad (21)$$

where l_1 and l_2 are the lowermost and uppermost gridpoints in the fine grid to be represented by one new coarse-gridpoint \tilde{h}_j . The new gridpoint then has to be placed somewhere between the altitudes $h(l_1)$ and $h(l_2)$. One can even go a step further and determine the placement of the new gridpoint \tilde{h}_j by linear interpolation of the sub-trace $\sum_{l=l_1(j)}^{l_2(j)} a_{l,l(z)}$ in z . Alternatively, the closest altitude in the n -dimensional grid can be chosen, which offers some operational advantages. The level of sophistication of finding the exact information-centered altitudes is not really necessary, because no two subsequent limb measurements will be exactly identical but may be desired to be represented on the same grid, which requires approximations anyway. When two or more profiles are to be represented on a common grid, the gridpoints can be selected such that the criterion defined in Eq. (21) is applied until the respective sums $\sum_{l=l_1(j)}^{l_2(j)} a_{l,l}$ exceed $dgfc$ for each of the profiles.

In summary, we work bottom up along the diagonal of the averaging kernel matrix adding up diagonal elements until the trace of the block of the matrix considered by now exceeds $dgfc$ and assign an altitude gridpoint to this submatrix of the averaging kernel. Since this grid is approximately information-centered, it is considered a kind of natural representation of the retrieved profile $\hat{\mathbf{x}}$.

Resampling of the constrained oversampled fine-grid retrieval, however, degrades the profile and further reduces the information below $dgf_{\text{retrieval}}$. This follows from Eq. (15), because Eq. (19) is not usually satisfied for arbitrarily chosen \mathbf{W} and \mathbf{R} matrices. This leads to $k \times k$ averaging kernels unequal unity, which is equivalent to less than k degrees of freedom. To compensate for this additional loss of information, it seems necessary to first remove the a priori information in the fine-grid retrieval. This, however, is not possible in most cases, since the unconstrained n -grid solution suffers from ill-posedness. Instead, we search for a new constraint \mathbf{R}' , which in effect is equivalent to the resampling of the retrieval to the appropriate coarse grid, i.e. which fulfills Eq. (19). In order to substitute the original constraint \mathbf{R} by the new constraint \mathbf{R}' , we make use of Eq. 10.48 of Rodgers (2000)

$$\hat{\mathbf{x}}' = (\mathbf{S}_x^{-1} - \mathbf{R} + \mathbf{R}')^{-1} [\mathbf{S}_x^{-1} \hat{\mathbf{x}} - \mathbf{R} \mathbf{x}_a + \mathbf{R}' \mathbf{x}'_a], \quad (22)$$

where $\hat{\mathbf{x}}'$ is the transformed profile in the n -dimensional grid, \mathbf{S}_x is the estimated $n \times n$ covariance matrix of $\hat{\mathbf{x}}$, \mathbf{x}_a is the original a priori profile, and \mathbf{x}'_a is the new a priori profile, which is set to zero at all altitude gridpoints for the applica-

tion outlined below. The new averaging kernel matrix after this substitution is

$$\mathbf{A}' = \mathbf{A}(\mathbf{R}') = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{R}')^{-1} \mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} \quad (23)$$

The new constraint term \mathbf{R}' and the particular resampling matrix \mathbf{W}' are chosen to fulfill the following conditions

$$\text{tr}(\mathbf{A}(\mathbf{R})) \approx \text{tr}(\mathbf{A}(\mathbf{R}')), \quad (24)$$

which is not required to be satisfied exactly because of $dgf \neq \text{int}(dgf)$, and

$$\text{tr}(\tilde{\mathbf{A}}') = \text{tr}(\mathbf{W}'^* \mathbf{A}(\mathbf{R}') \mathbf{W}') \Leftrightarrow \tilde{\mathbf{A}}' = \tilde{\mathbf{I}} \Leftrightarrow \mathbf{W}'^T \mathbf{R}' \mathbf{W}' = \mathbf{0}, \quad (25)$$

where $\text{tr}(\tilde{\mathbf{A}}')$ is the number of degrees of freedom of the resampled profile. This is achieved by a constraint \mathbf{R}' in the fine grid which is equivalent to a resampling onto the coarse grid. The formal resampling of an n -grid profile subject to such a constraint onto the k -grid then conserves all information except that contained in the neglected fraction of a degree of freedom.

In the following we present a pair of \mathbf{R}' and \mathbf{W}' matrices which fulfill these two conditions. For convenience, we assume that the coarse grid is a subset of the fine grid. The new constraint \mathbf{R}' is supposed not to constrain values at the gridpoints which also are members of the coarse grid, while it shall produce values between the coarse gridpoints which do not carry additional information but are completely determined by the values at the coarse gridpoints and a chosen interpolation scheme, e.g. a staircase function.

A constraint \mathbf{R}' compliant to this requirement is a block-diagonal $n \times n$ matrix, of the form

$$\mathbf{R}'_{\square} = \begin{bmatrix} \mathbf{R}'_{\square_{\text{sub};1}}, & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}'_{\square_{\text{sub};2}}, & \cdots & \mathbf{0} \\ \vdots & & & \vdots \\ \mathbf{0}, & \mathbf{0} & \cdots & \mathbf{R}'_{\square_{\text{sub};k}} \end{bmatrix}, \quad (26)$$

where each of the disjoint diagonal blocks $\mathbf{R}'_{\square_{\text{sub};j}}$ represents one gridpoint in the coarse grid, representing one degree of freedom. The coarse gridpoints shall be placed where

$$\sum_{l=l_1(j)}^{l_2(j)} a_{l,l} = \frac{dgfc}{2} + (j-1)dgfc, \quad (27)$$

while the block boundaries are placed where

$$\sum_{l=l_1(j)}^{l_2(j)} a_{l,l} = j dgfc. \quad (28)$$

Blocks of the size $i(j) \times i(j)$ are set zero when $i(j)=1$, while blocks equal or larger than 2×2 are set up as first order Tikhonov-type regularization matrices (Tikhonov, 1963) of the type

$$\mathbf{R}'_{\square_{\text{sub}}} = \gamma \mathbf{L}_1 \mathbf{L}_1^T \quad (29)$$

where \mathbf{L}_1 is a local-gridwidth weighted $i(j) \times (i(j)-1)$ first order difference matrix of the type

$$\mathbf{L}_1^T(j) = \begin{pmatrix} (h_l - h_{l+1})^{-1} & 0 & \cdots & 0 \\ 0 & (h_{l+1} - h_{l+2})^{-1} & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & (h_{i_j-1} - h_{i_j})^{-1} \end{pmatrix} \times \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \cdots & 0 & \vdots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \quad (30)$$

for regular fine grid, or a similar matrix which each non-zero matrix element weighted by $h_{l+1} - h_l$ where j indicated the row of \mathbf{L}^T gridwidth-compensated second order difference operators. Such a regularization does not constrain the \hat{x} -values which correspond to coarse-grid-points but, when sufficient regularization strength for each matrix-block is chosen such that no altitude resolution is allowed within each block, values at the additional gridpoints are forced to follow the staircase function between the coarse-grid points. Such large γ values are necessary to avoid singularity of the first term in Eq. (22). Transformation of arbitrary fine gridded profiles to the coarse grid then uses

$$\mathbf{W}'_{\square} = \begin{pmatrix} \frac{1}{i_1} \cdots \frac{1}{i_1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{1}{i_2} \cdots \frac{1}{i_2} & \cdots & 0 & \cdots & 0 \\ & & & \cdots & & & & \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \frac{1}{i_k} \cdots \frac{1}{i_k} \end{pmatrix}, \quad (31)$$

where i_j is the number of fine-grid points to be represented by the j th coarse-grid point. The number of non-zero entries in each row of \mathbf{W}'_{\square} is i_j . For staircase profiles, however, which have been generated with \mathbf{R}'_{\square} and $\gamma \rightarrow \infty$, also a simpler transformation matrix of the type

$$\mathbf{W}'_{\square} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ & & & & & & & \ddots & & & & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \quad (32)$$

will transform the matrix to the coarse grid without loss of information, because due to the regularization applied all values within a block are the same. Reformation of the staircase-profile from the coarse-grid profile uses

$$\mathbf{W}'_{\square} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 1 & \cdots & 0 \\ & \cdots & & \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \quad (33)$$

where the number of non-zero entries in the i th column is n_i . It should be noted that also \mathbf{W}'_{\square} satisfies the condition

$$\mathbf{W}'_{\square} \mathbf{W}'_{\square} = \tilde{\mathbf{I}} \quad (34)$$

The validity of Eq. (19) can easily be verified for regularization and transformation matrices as defined in Eqs. (26–30) and (33). The re-regularized and resampled profile thus contains $k = \text{int}(dgf)$ degrees of freedom.

Certainly, the same solution \hat{x}' which we obtain by Eq. 22 can in principle be found by an unconstrained maximum likelihood retrieval of a staircase profile on the coarse grid. However, this optimal grid is not typically known in advance and thus the fine-grid retrieval is a prerequisite. The availability of the fine-grid retrieval further offers extensive diagnostics, e.g. the $n \times n$ averaging kernel matrix \mathbf{A}' (Eq. 23), which is not available for a direct maximum likelihood retrieval on the coarse grid.

3.2 Triangular representation

Neither is the staircase representation considered a realistic representation of the atmosphere, nor do most forward models support this representation. A representation where atmospheric state parameters are interpolated (e.g. linearly) between the altitude gridpoints is preferred in most applications. As with the staircase representation, we require the information content of the resampled retrieval to be the integer value of the degrees of freedom of the fine-grid retrieval (Eq. 20) and define this to be the number of coarse-grid points. In order to represent the profile from the uppermost to the lowermost end of the original profile, we define

$$\tilde{h}_1 = h_1 \quad (35)$$

and

$$\tilde{h}_k = h_n \quad (36)$$

where $\tilde{\mathbf{h}} = (\tilde{h}_1, \dots, \tilde{h}_k)^T$ is the grid of the coarse representation, while $\mathbf{h} = (h_1, \dots, h_n)^T$ is the original fine altitude grid. In order to satisfy Eq. (20), the other coarse-grid points \tilde{h}_2 to \tilde{h}_{K-1} are the same as for the staircase representation. In order to emulate this coarse grid retrieval in the fine grid retrieval, the following regularization matrix can be constructed:

$$\mathbf{R}'_{\Delta} = \begin{bmatrix} \mathbf{R}'_{\Delta, \text{sub};1} & 0 & \cdots & 0 \\ 0 & \mathbf{R}'_{\Delta, \text{sub};2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}'_{\Delta, \text{sub};2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \mathbf{R}'_{\Delta, \text{sub};k-1} \end{bmatrix} \quad (37)$$

Contrary to \mathbf{R}'_{\square} , the diagonal blocks of \mathbf{R}'_{Δ} are not disjoint but each two adjacent blocks share a common diagonal element. Each block j is of the size

$$i(j) \times i(j) = (l_2(j) - l_1(j) + 1) \times (l_2(j) - l_1(j) + 1).$$

Blocks $\mathbf{R}'_{\Delta_{\text{sub};j}}$ of the size 2×2 are set zero, while blocks equal or larger than 3×3 are set up as second order Tikhonov-type regularization matrices (Tikhonov, 1963) of the type

$$\mathbf{R}'_{\Delta_{\text{sub}}} = \gamma \mathbf{L}_2 \mathbf{L}_2^T \quad (38)$$

where \mathbf{L}_2^T is a local-gridwidth weighted $(i(j)-2) \times i(j)$ second order difference matrix of tridiagonal form with sub-diagonal, diagonal, and super-diagonal elements $\lambda_i \Delta_i$, $-\lambda_i (\Delta_{i-1} + \Delta_i)$, $\lambda_i \Delta_{i-1}$, respectively, where

$$\Delta_l = h_{l+1} - h_l \quad (39)$$

and

$$\lambda_i = \frac{2}{\Delta_{i-1} \Delta_i (\Delta_{i-1} + \Delta_i)}. \quad (40)$$

In the triangular application, the transformation matrix \mathbf{W}'_{Δ} transforms the coarse-grid profile to the fine grid by linear interpolation, with

$$\mathbf{W}'_{\Delta} = \begin{pmatrix} \frac{\tilde{h}_2 - h_1}{h_2 - h_1} & \frac{h_1 - \tilde{h}_1}{h_2 - h_1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\tilde{h}_2 - h_{i_1-1}}{h_2 - h_1} & \frac{h_{i_1-1} - \tilde{h}_1}{h_2 - h_1} & 0 & 0 & 0 \\ 0 & \frac{\tilde{h}_3 - h_{i_1}}{h_3 - h_2} & \frac{h_{i_1} - \tilde{h}_2}{h_3 - h_2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \frac{\tilde{h}_3 - h_{i_1+i_2-1}}{h_3 - h_2} & \frac{h_{i_1+i_2-1} - \tilde{h}_2}{h_3 - h_2} & 0 & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \frac{\tilde{h}_k - h_{n-i_k}}{h_k - h_{k-1}} & \frac{h_{n-i_k} - \tilde{h}_{k-1}}{h_k - h_{k-1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \frac{\tilde{h}_k - h_{n-1}}{h_k - h_{k-1}} & \frac{h_{n-1} - \tilde{h}_{k-1}}{h_k - h_{k-1}} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (41)$$

The transformation of an arbitrary profile from the fine to the coarse grid leads to a lengthy expression for the generalized formulation of related matrix \mathbf{W}^*_{Δ} . Multiplication of any fine-grid profile \mathbf{x} by \mathbf{W}^*_{Δ} returns the footpoints of the contiguous sequence of segments of which each is a regression line constrained such that the regression lines intersect at the pre-defined coarse-grid altitudes. If need be (e.g. for applications as discussed in Sect. 4), \mathbf{W}^*_{Δ} can easily be computed numerically for the actual interpolation scheme \mathbf{W}'_{Δ} according to Eq. (8). For the special case discussed here, where the fine-grid profile has been re-regularized such that all profile values on the fine-grid falling between the coarse-grid points can be generated by linear interpolation and thus carry no independent information, the following simpler matrix can be used for the transformation from the fine grid to the coarse grid:

$$\mathbf{W}^*_{\Delta} = \begin{pmatrix} w^*_{1,1} & \cdots & w^*_{1,n} \\ w^*_{k,1} & \cdots & w^*_{k,n} \end{pmatrix}; \quad w^*_{j,l} = \begin{cases} 1 & : h_l = \tilde{h}_j \\ 0 & : h_l \neq \tilde{h}_j \end{cases} \quad (42)$$

This interpolation matrix just picks out the independent data points of the interpolated profile. Again, for $\gamma \rightarrow \infty$ the re-

regularized retrieval based on the constraint \mathbf{R}'_{Δ} is transformed onto the coarse grid by transformation matrix \mathbf{W}^*_{Δ} without additional loss of information, and we have

$$\mathbf{W}^*_{\Delta} \mathbf{W}'_{\Delta} = \tilde{\mathbf{I}}. \quad (43)$$

4 Smoothing errors of deduced quantities

Various scientific application areas of data involve the analysis of differences of profiles (validation, data assimilation, visual recognition of structures in data) or calculation of sums (e.g. total reactive chlorine) or ratios (correlation analyses) of data. After regridding the data and related averaging kernel matrices and covariance matrices (c.f., Calisesi et al., 2005), and transformation of the data to the same a priori profile (Rodgers and Connor, 2003), the different amount of a priori information in the data still has to be taken into account. There are two options, which we discuss in the following: Either the amount of a priori information in the data is left untouched, and possible artefacts caused by different a priori content are characterized by the smoothing error of the deduced quantity, or the data are made directly comparable.

4.1 Un-matched a priori content

The approach to compare retrieved profiles $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ of the same quantity \mathbf{x} without prior adjustment of their a priori content has been discussed in detail by Rodgers and Connor (2003), who suggest to evaluate the smoothing error covariance matrix $\mathbf{S}_{\text{smooth},\text{diff}}$ of the difference $\hat{\mathbf{x}}_2$ and $\hat{\mathbf{x}}_1$,

$$\mathbf{S}_{\text{smooth},\text{diff}} = (\mathbf{A}_2 - \mathbf{A}_1) \mathbf{S}_e (\mathbf{A}_2 - \mathbf{A}_1)^T, \quad (44)$$

where \mathbf{A}_1 and \mathbf{A}_2 are the related averaging kernel matrices, and where \mathbf{S}_e is the climatological covariance matrix of the state variable $\hat{\mathbf{x}}_1$. $\mathbf{S}_{\text{smooth},\text{diff}}$ is a component of the total error of the difference \mathbf{S}_{diff} , which is used to test the significance of the difference via χ^2 statistics:

$$\chi^2 = (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2)^T \mathbf{S}_{\text{diff}}^{-1} (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2) \quad (45)$$

In cases where the a priori profile \mathbf{x}_a has been chosen different from the climatological profile \mathbf{x}_e , around which the variability is described by \mathbf{S}_e , a mean smoothing error difference

$$\tilde{\Delta} \hat{\mathbf{x}}_{\text{smooth},\text{diff}} = (\mathbf{A}_2 - \mathbf{A}_1) (\mathbf{x}_e - \mathbf{x}_a) \quad (46)$$

has to be considered.

The same basic approach to calculate deduced quantities without prior consideration of different a priori content and characterize related artefacts by the smoothing error of the deduced quantity certainly holds also for other deduced quantities. The smoothing error covariance $\mathbf{S}_{\text{smooth},\text{sum}}$ of, e.g., total $[\text{NO}_x] = [\text{NO}_2 + \text{NO}]$, represented by vectors $\hat{\mathbf{x}}_{1+2}$, $\hat{\mathbf{x}}_1$, and $\hat{\mathbf{x}}_2$, respectively, is

$$\mathbf{S}_{\text{smooth},\text{sum}} = (\mathbf{A}_1 - \mathbf{I}) \mathbf{S}_{a,1} (\mathbf{A}_1 - \mathbf{I})^T + (\mathbf{A}_2 - \mathbf{I}) \mathbf{S}_{a,2} (\mathbf{A}_2 - \mathbf{I})^T, \quad (47)$$

where \mathbf{A}_1 , \mathbf{A}_2 , $\mathbf{S}_{a,1}$, and $\mathbf{S}_{a,2}$ are the related averaging kernel and covariance matrices, respectively. The mean smoothing error is

$$\bar{\Delta}\hat{\mathbf{x}}_{\text{smooth},\text{sum}} = (\mathbf{A}_1 - \mathbf{I})(\mathbf{x}_{e,1} - \mathbf{x}_{a,1}) + (\mathbf{A}_2 - \mathbf{I})(\mathbf{x}_{e,2} - \mathbf{x}_{a,2}) \quad (48)$$

The smoothing error variance $\sigma_{\text{smooth},\text{ratio};n}^2$ of the ratio $\hat{x}_{1;n}/\hat{x}_{2;n}$; n of two profile elements $\hat{x}_{1;n}$ and $\hat{x}_{2;n}$ at altitude gridpoint n is

$$\sigma_{\text{smooth},\text{ratio};n}^2 = \quad (49)$$

$$\frac{1}{\hat{x}_{2;n}^4} \left(\hat{x}_{1;n}^2 \left((\mathbf{A}_2 - \mathbf{I})\mathbf{S}_{e,2}(\mathbf{A}_2 - \mathbf{I})^T \right)_{n,n} + \hat{x}_{2;n}^2 \left((\mathbf{A}_1 - \mathbf{I})\mathbf{S}_{e,1}(\mathbf{A}_1 - \mathbf{I})^T \right)_{n,n} \right), \quad (50)$$

and the mean smoothing error is

$$\bar{\Delta}\hat{\mathbf{x}}_{\text{smooth},\text{sum}} = \left((\mathbf{A}_1 - \mathbf{I})(\mathbf{x}_{e,1} - \mathbf{x}_{a,1}) \right)_{n,n} + \left((\mathbf{A}_2 - \mathbf{I})(\mathbf{x}_{e,2} - \mathbf{x}_{a,2}) \right)_{n,n}. \quad (51)$$

As long as done properly, the approach described above is, of course, valid. The crucial point of this approach, however, is that the smoothing error covariance matrix can only be evaluated when the climatological covariance matrix \mathbf{S}_e is known. This is often not the case. Unfortunately, there are situations when assumptions on \mathbf{S}_e are dominating the calculation of χ^2 .

For an assumed a priori covariance matrix $\tilde{\mathbf{S}}_e$, the deviation of the estimated smoothing error of the difference of the two profiles from the true smoothing error is

$$(\mathbf{A}_2 - \mathbf{A}_1)(\tilde{\mathbf{S}}_e - \mathbf{S}_e)(\mathbf{A}_2 - \mathbf{A}_1)^T. \quad (52)$$

It cannot easily be associated with certain altitude scales of atmospheric variability. The problem of frequent non-availability of a reliable \mathbf{S}_e covariance matrix inferred from representative sufficiently high resolved real data is a major motivation to work with quantities of matched a priori content.

4.2 Matched a priori content

In the following, we shortly discuss various approximations aiming at making profiles comparable such that the smoothing error of the difference of the compared profiles vanishes. Since these approximations are not perfect, there remains a residual smoothing error, which we characterize below. The application to deduced quantities other than differences, as discussed in the previous section, is straightforward and thus not shown here.

An approximation widely used is to consider the better resolved one of the two profiles, $\hat{\mathbf{x}}_2$ to be compared as an ideal measurement and to smooth it by application of the averaging kernel of the coarser resolved measurement before

comparison. This approximation does not rely on an estimate of the climatological covariance \mathbf{S}_e . If the finer resolved profile indeed consists of a large number of point measurements, this approach is close to exact, and our proposed re-regularization/re-sampling method does not offer other than practical advantages. The user who has no fine-grid averaging kernels available would just use transformation matrices as defined in Eqs. (33) or (41) instead.

In a less than ideal case, however, the residual smoothing error to be considered covariance is

$$\mathbf{S}_{\text{smooth},\text{diff}} = \mathbf{A}_1(\mathbf{I} - \mathbf{A}_2)\mathbf{S}_e(\mathbf{I} - \mathbf{A}_2)^T \mathbf{A}_1^T, \quad (53)$$

where \mathbf{I} is $n \times n$ identity. $\mathbf{S}_{\text{smooth},\text{diff}}$ is zero in the ideal case when $\mathbf{A}_2 = \mathbf{I}$. The mean smoothing error is

$$\bar{\Delta}\hat{\mathbf{x}}_{\text{smooth},\text{sum}} = \mathbf{A}_1(\mathbf{I} - \mathbf{A}_2)(\mathbf{x}_e - \mathbf{x}_a) \quad (54)$$

When comparing retrievals whose altitude resolution is too different to be ignored but too similar that the approach mentioned above would be justified, one may wish to cross-wise smooth the retrievals in order to achieve cancellation of smoothing errors. This, however, is not appropriate since the application of averaging kernels to profiles is not commutative. The residual smoothing error covariance is

$$\mathbf{S}_{\text{smooth},\text{diff}} = (\mathbf{A}_1\mathbf{A}_2 - \mathbf{A}_2\mathbf{A}_1)\mathbf{S}_e(\mathbf{A}_1\mathbf{A}_2 - \mathbf{A}_2\mathbf{A}_1)^T, \quad (55)$$

and the mean smoothing error is

$$\bar{\Delta}\hat{\mathbf{x}}_{\text{smooth},\text{sum}} = (\mathbf{A}_1\mathbf{A}_2 - \mathbf{A}_2\mathbf{A}_1)(\mathbf{x}_e - \mathbf{x}_a) \quad (56)$$

Instead, we suggest to first re-regularize and resample the finer resolved one of the two profiles, $\hat{\mathbf{x}}_2$, following the procedure described in Sect. 3. Then the resulting profile $\tilde{\hat{\mathbf{x}}}_2'$ is smoothed by the averaging kernel $\tilde{\mathbf{A}}_1$ of the other profile re-gridded to the k -grid. If covariance matrices, regularization matrices, a priori profiles and averaging kernels – all being quantities needed to carry out the rigorous intercomparison of profiles with unmatched a priori content anyway – it is even possible to re-regularize/re-sample both profiles onto the same grid, namely the one found optimal for the coarser resolved one according to the criterion discussed in Sect. 3.

Using the decomposition

$$\mathbf{S}_e = \mathbf{S}_e^{\text{fine}} + \mathbf{S}_e^{\text{coarse}}, \quad (57)$$

where $\mathbf{S}_e^{\text{coarse}}$ is the part of the variability of the state variable which could completely be represented on the coarse grid, and a component $\mathbf{S}_e^{\text{fine}}$ which is the part of the variability which can only be represented on the fine grid, the residual smoothing error can be written in the k -grid as

$$\begin{aligned} \mathbf{S}_{\text{smooth},\text{diff}} &= \mathbf{W}^{*/k}(\mathbf{A}_1\mathbf{A}_2' - \mathbf{A}_2'\mathbf{A}_1)\mathbf{S}_e^{\text{fine}}(\mathbf{A}_1\mathbf{A}_2' - \mathbf{A}_2'\mathbf{A}_1)^T \mathbf{W}^{*/T} + \\ &\quad (\tilde{\mathbf{A}}_1\tilde{\mathbf{A}}_2' - \tilde{\mathbf{A}}_2'\tilde{\mathbf{A}}_1)\mathbf{W}^{*/k}\mathbf{S}_e^{\text{coarse}}\mathbf{W}^{*/T}(\tilde{\mathbf{A}}_1\tilde{\mathbf{A}}_2' - \tilde{\mathbf{A}}_2'\tilde{\mathbf{A}}_1)^T \quad (58) \\ &= \mathbf{W}^{*/k}(\mathbf{A}_1\mathbf{A}_2' - \mathbf{A}_2'\mathbf{A}_1)\mathbf{S}_e^{\text{fine}}(\mathbf{A}_1\mathbf{A}_2' - \mathbf{A}_2'\mathbf{A}_1)^T \mathbf{W}^{*/T}. \end{aligned}$$

The second term disappears because of $\tilde{\mathbf{A}}_2' = \mathbf{I}$ and all residual smoothing error covariance is associated with subscale variability $\mathbf{S}_e^{\text{fine}}$. The same applies holds for the mean smoothing error. With

$$\mathbf{x}_e = \mathbf{x}_e^{\text{coarse}} + \mathbf{x}_e^{\text{fine}} \quad (59)$$

and $\mathbf{x}_a = \mathbf{0}$ as suggested in Section 3, we get

$$\bar{\Delta} \hat{\mathbf{x}}_{\text{smooth.,sum}} = \mathbf{W}'^* (\mathbf{A}_1 \mathbf{A}_2' - \mathbf{A}_2' \mathbf{A}_1) \mathbf{x}_e^{\text{fine}}. \quad (60)$$

This approximation is exact, i.e. the residual smoothing error disappears, in the case that the variability of the true state is sufficiently well characterized by $\mathbf{S}_e^{\text{coarse}}$, i.e. if the variability can be represented completely in the coarse grid established in Sect. 3.

Otherwise, there remains a residual smoothing error not accounted for (Eqs. 58–60), and our suggested method thus is inferior to the approach of comparing profiles of unmatched a priori content and characterizing the smoothing error of the deduced quantity as described in Sect. 4.2. In real applications, however, there often is no reliable information on the mean true state and its variability available.

In these situations, the rigorous assessment of the smoothing error of the deduced quantity is not feasible and it may be helpful to get rid of smoothing error components related to large scale variability and to be able to relate the ignored part of the smoothing error to a smaller scale.

When data represented as suggested here are to be compared to high-resolution in situ measurements free of a priori information, sampled at $l > n$ gridpoints, the latter are, as usual, smoothed by the averaging kernel of the lower resolved data. If the $k \times n$ averaging kernel matrix $\tilde{\mathbf{A}}' = \mathbf{W}^* \mathbf{A}'$ is available, it should be interpolated to $k \times l$ shape, i.e. to the grid of the high resolution measurement, and used for the smoothing operation. Otherwise, the $k \times k$ averaging kernel matrix $\tilde{\mathbf{A}}'$ is used. This is unity by definition but its interpolation to the $k \times l$ shape is identical to the respective interpolation matrix used to represent the high-resolution measurement on a coarse grid. In other words, the high-resolution profile consisting of l gridpoints is smoothed by a $k \times l$ matrix of the type \mathbf{W}_{\square}^* or \mathbf{W}_{Δ}^* , depending on the representation on the coarse grid chosen.

Besides the intercomparison of two measurements of the same atmospheric profile, the concept of matched a priori content has another wide application area, which is the recognition of patterns and structures in, e.g., 2-D fields of retrieved state parameters, which basically is the detection of differences of data in the time of space domain. The altitude resolution and the a priori content of data retrieved directly with Eq. (1) typically varies with the independent variable, e.g. with time, latitude, etc. This variation of a priori content can be caused because the Jacobians \mathbf{K} , hence also the averaging kernels, depend on the atmospheric state itself. Another possible reason is that a different number of measurements y could be available for the profile retrievals.

The assessment of the significance of the detected structures and patterns in retrievals of unmatched a priori content requires not only the climatological covariance matrix \mathbf{S}_e of a profile but the full multi-dimensional covariance information in the space where the pattern is detected is required. Apparent structures in the fields can then be caused by the varying altitude resolution. After re-regularizing/resampling the profiles on a common altitude grid similar to the one defined in Sect. 3 but constant over the ensemble of profiles under assessment, these artificial structures and patterns are removed. While in retrievals following Eq. (1), a different amount of measurement information partly influences the altitude resolution and partly the noise-induced random retrieval error, the altitude resolution is constant in the re-regularized/resampled data and only the retrieval error varies. Variation of the random retrieval error, however, is easier to handle in most applications than the variation of altitude resolution and a priori content.

5 Case study: MIPAS

To illustrate the behaviour of the proposed re-regularization/resampling method, we show an application to the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) (Fischer and Oelhaf, 1996; Endemann and Fischer, 1993; SAG, 1999), which is an Earth observation instrument onboard the Envisat research satellite. MIPAS records high-resolution limb emission spectra of the Earth's atmosphere, from which vertical profiles of atmospheric constituents' abundances and temperature are retrieved. Operational near-real time data analysis (Ridolfi et al., 2000; Nett et al., 1999), under responsibility of the European Space Agency (ESA), avoids problems with ill-posedness of the inversion by using a grid for representation of the vertical profiles which is defined by the tangent altitudes of the measurements, and by restriction of the altitude coverage of the retrieved profiles to heights where the measurements contain a clear signal of the target quantity, and where no problems due to saturation effects in the measured spectra occur. However, there also exist further MIPAS data processors (von Clarmann et al., 2003a), which aim at an extended MIPAS data product for scientific use. One of these processors is the one operated at Forschungszentrum Karlsruhe, Institute for Meteorology and Climate Research (von Clarmann et al., 2003b). This processor uses constrained retrieval approaches, with all implications mentioned above, in particular mapping of a priori information onto the retrieved profiles and, as a consequence, interdependences of retrieved data.

For this case study, we use simulated limb emission spectral measurements of ClONO₂. First, a vertical profile of ClONO₂ volume mixing ratios is retrieved, using the regular processing setup as described in Höpfner et al. (2004). For radiative transfer modeling, the Karlsruhe Optimized

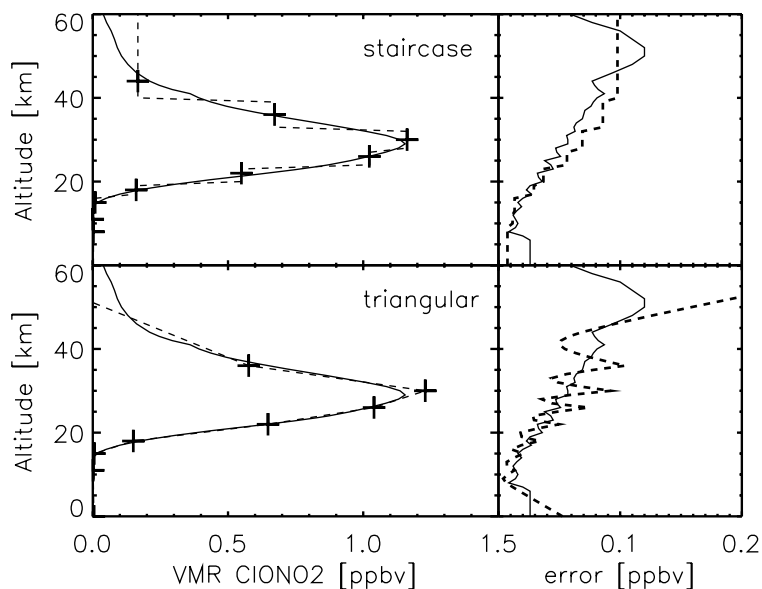


Fig. 1. A retrieval of MIPAS ClONO₂ on a fine altitude grid (solid line) and the related re-regularized profiles (dashed lines) which contain nearly the same number of degrees of freedom, in staircase (upper left panel) and triangular (lower left panel) representation. Re-regularized profiles can be sampled on the coarse grid without loss of information. These resampled profiles are, except for the coarse grid itself, free of a priori information. + signs indicate the selected altitude gridpoints of the coarse grid. The right panels show the related retrieval errors.

and Precise Radiative Transfer Algorithm (KOPRA) (Stiller, 2000) was used. Retrievals were performed with the MIPAS data processor as described in von Clarmann et al. (2003b). Volume mixing ratios are sampled on a 1-km grid from 4–44 km, a 2-km grid from 44–70 km, and at 70, 75, 80, 90, 100 and 120 km. Since this grid is finer than the altitude resolution provided by the measurement geometry, i.e. the vertical distance of adjacent tangent altitudes, and since there is not enough spectral information in the measurement to extract altitude-resolved information on this fine altitude grid, the retrieval needs regularization to be stable. In this case, a constraint has been chosen which minimizes the first order differences of mixing ratios at adjacent altitude gridpoints as discussed in (Steck and von Clarmann, 2001, and references therein). While represented at many more altitude gridpoints, the retrieved profile has only 9.7 degrees of freedom. Our proposed method selects the following 9 altitude gridpoints to represent the profile as a staircase-type profile: 8.0, 11.0, 15.0, 18.0, 22.0, 26.0, 30.0, 36.0, and 44.0 km. Simply re-sampling the profile onto the coarse grid further reduces the degrees of freedom to 7.1 (see discussion in the paragraph above Eq. 22). Contrary to this, our proposed method conserves exactly 9 degrees of freedom, and the averaging kernel matrix is exactly 9×9 unity. The original retrieval and the re-regularized/resampled profile are shown in Fig. 1, upper left panel. The mapping of measurement noise onto the retrieval is shown in Fig. 1, upper right left panel. These retrieval errors are similar for the original and the re-regularized/resampled data sets. Lower left and right panels of Fig. 1 show

the profiles and the error estimates, respectively, for the triangular representation. In this application, resampling alone without prior re-regularization would reduce the numbers of degrees of freedom to 6.7. Also in this application, retrieval errors of the re-regularized profile are, on average, similar to those of the original profile. A characteristic of the triangular representation, however, is, that errors are culminating at the coarse gridpoints. Because errors of adjacent coarse gridpoints are anti-correlated, errors at altitudes in between, which do not carry independent information but where mixing ratios are determined by linear interpolation, are substantially smaller. Large errors at the uppermost and lowermost coarse gridpoints are irrelevant and can be understood as extrapolation error: Any small uncertainty of the volume mixing ratio vertical gradient will trigger large uncertainties at altitudes far off the altitude range where the original retrieval contains information.

Certainly re-regularized/resampled profiles are in some sense also affected by a prior constraint because the coarse grid along with the applicable interpolation scheme is a prior constraint in itself, but this kind of prior constraint is obvious also to data users who are not familiar with the averaging kernel formalism. In consequence, the risk of misinterpretation of data is largely reduced.

6 Conclusions

We have proposed a re-regularization/resampling scheme which allows to represent a retrieval on an appropriate altitude grid such that its averaging kernel becomes unity. This means that, within the newly defined co-ordinate system, the re-regularized/resampled profile is entirely free of a priori information. All remaining a priori is inherent in the grid definition and interpolation scheme which certainly does not allow to represent sub-scale structures of the true atmospheric state. The loss of information is limited to less than one degree of freedom, which usually is tolerable for limb measurements but may limit applicability to nadir instruments where the relative loss of information may become significant. This is, because for nadir sounders the averaging kernels are typically very broad, and the signal measured by the instrument represents a wide altitude range. Our proposed representation of retrieved data simplifies the work with deduced quantities like differences, sums or ratios, guarantees the statistical independence of profiles retrieved with the same a priori information, excludes any bias of a retrieval towards the a priori information, and reduces the amount of data to be transferred to the data user. The re-regularized/resampled profiles certainly represent only a smoothed version of the truth but the related smoothing error is limited to small-scale atmospheric structures variability which cannot be represented on the chosen altitude grid. Further, the remaining a priori information in the re-regularized/resampled data consists only of the grid itself and interpolation function used to represent the data and thus is more obvious to the non-expert data user who neither has the tools available to work with averaging kernel matrices nor wants to work in spaces without obvious physical meaning such as those spanned by the singular vectors of the solution.

While the usual approach to characterize retrievals which include a priori information by their averaging kernels and smoothing errors is, of course, still valid if done properly (which however, assumes availability of the climatological covariance information of the state variable under assessment), our re-regularization/resampling scheme is advantageous in less favourable conditions, i.e. if the required climatological covariance information is not available. Beyond this, in a scientific community which is segmented to an extent into data providers and data users who do not interact directly but communicate their results via databases, as favored by, e.g., the responsible officials in the Global Monitoring for Environment and Security (GMES), data providers will lose control over the correct handling of data and related diagnostics. Thus, the importance of easy-to-use data representation will increase in order to avoid misinterpretation of data.

Appendix A

Here we show that the concept to calculate the number of degrees of freedom of the retrieval as the trace of the averaging kernel is applicable to smoothing constraints, too. This is not self-evident since the inference of this in Rodgers (2000), Eqs. (2.48)–(2.56), involves the non-inverted a priori covariance matrix \mathbf{S}_a . The smoothing constraint \mathbf{R} , which replaces the \mathbf{S}_a term in a Tikhonov-type retrieval is singular due to its rank $n-1$. Smoothing constraints can formally be understood as Bayesian constraints in cases when there is a priori knowledge only on the vertical gradient of the profile but not on the values themselves.

We bypass this problem of singularity of the regularization matrix by decomposing the retrieval which involves a smoothing constraint into an unconstrained part of which the degrees of freedom are trivial to estimate and an optimal estimation/maximum a posteriori-type part, where the arguments of Rodgers (2000) (Eqs. 2.48–2.56) hold.

The retrieval of \mathbf{x} can be rewritten as a retrieval of an integrated quantity, e.g. vertical column of a trace species, altitude-averaged temperature etc., plus the retrieval of $n-1$ first order differences of values at adjacent altitude grid-points. This transformation of the retrieval vector is done by multiplying the vector with matrix \mathbf{C} of the type

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

This transformation of the co-ordinate system does not change the degrees of freedom. The Tikhonov-type first order regularization matrix

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ & & & & \ddots & & \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

is transformed to

$$\mathbf{C}^{-1T} \mathbf{R} \mathbf{C}^{-1} = \mathbf{R}^* = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

This proves that the retrieval of the integrated quantity is not constrained by the first order difference smoothing approach. Therefore, by definition, this quantity adds exactly one degree of freedom to the retrieval, and the related diagonal element of the averaging kernel is unity while all off-diagonal elements are zero. The $n-1$ differences now are

constrained by a diagonal matrix (lower right $(n-1) \times (n-1)$ block in \mathbf{R}^*), which can be understood as an inverse diagonal covariance matrix in the sense of Bayesian maximum a posteriori retrieval, assuming that there is no knowledge on higher order differences (which would cause off-diagonal elements in the relevant block of \mathbf{C}^*). Therefore, the number of degrees of freedom of the retrieval of differences is the trace of the relevant $(n-1) \times (n-1)$ averaging kernel matrix \mathbf{A}^* . Thus the number of degrees of freedom of the retrieval is

$$dgf_{\text{retrieval}} = 1 + \text{tr}(\mathbf{A}^*)$$

We recombine the retrieval of the integrated quantity and the differences retrieval and write the related averaging kernel matrix $\mathbf{A}^{*/} \in \mathbb{R}^{n \times n}$ as

$$\mathbf{A}^{*/} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{A}^* \end{bmatrix}$$

and find that

$$\text{tr}(\mathbf{A}^{*/}) = dgf_{\text{retrieval}}.$$

Since the number of degrees of freedom are conserved when reversibly transforming back to the original coordinate system, we obtain

$$dgf_{\text{retrieval}} = \text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^{*/}).$$

Following this approach the equivalence can also be shown for higher order difference operators as regularization constraint, or linear combinations of these.

Acknowledgements. This work was partly funded by BMBF under contract 50 EE 0512. The authors would like to thank F. Hase and G. Kopp for helpful comments.

Edited by: M. G. Lawrence

References

- Calisesi, Y., Soebijanta, V. T., and van Oss, R.: Regridding of remote soundings: Formulation and application to ozone profile comparison, *J. Geophys. Res.*, 110, D23306, doi:10.1029/2005JD006122, 2005.
- 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. Opt.*, 42, 6465–6473, 2003.
- Endemann, M. and Fischer, H.: Envisat's High-Resolution Limb Sounder: MIPAS, *ESA bulletin*, 76, 47–52, 1993.
- Esler, J. G. and Waugh, D. W.: A method for estimating the extent of denitrification of Arctic polar vortex air from tracer-tracer scatter plots, *J. Geophys. Res.*, 107, 4169, doi:10.1029/2001JD001071, 2002.
- Fischer, H. and Oelhaf, H.: Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb-emission spectrometers, *Appl. Opt.*, 35, 2787–2796, 1996.
- Glatthor, N., von Clarmann, T., Fischer, H., Funke, B., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Milz, M., Steck, T., Stiller, G. P., Mengistu Tsidu, G., and Wang, D. Y.: Mixing processes during the Antarctic vortex split in September/October 2002 as inferred from source gas and ozone distributions from ENVISAT-MIPAS, *J. Atmos. Sci.*, 62, 787–800, 2005.
- Höpfner, M., von Clarmann, T., Fischer, H., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Mengistu Tsidu, G., Milz, M., Steck, T., Stiller, G. P., Wang, D.-Y., and Funke, B.: First spaceborne observations of Antarctic stratospheric ClONO₂ recovery: Austral spring 2002, *J. Geophys. Res.*, 109, D11308, doi:10.1029/2004JD005322, 2004.
- Joiner, J. and da Silva, A. M.: Efficient methods to assimilate remotely sensed data based on information content, *Q. J. R. Meteorol. Soc.*, 124, 1669–1694, 1998.
- Kondo, Y., Koike, M., Engel, A., Schmidt, U., Mueller, M., Sugita, T., Kanzawa, H., Nakazawa, T., Aoki, S., Irie, H., Toriyama, N., Suzuki, T., and Sasano, Y.: NO_y-N₂O correlation observed inside the Arctic vortex in February 1997: Dynamical and chemical effects, *J. Geophys. Res.*, 104, 8215–8224, 1999.
- Mengistu Tsidu, G., Stiller, G. P., von Clarmann, T., Funke, B., Höpfner, M., Fischer, H., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., López-Puertas, M., Milz, M., Steck, T., and Wang, D. Y.: NO_y from Michelson Interferometer for Passive Atmospheric Sounding on Environmental Satellite during the Southern Hemisphere polar vortex split in September/October 2002, *J. Geophys. Res.*, 110, D11301, doi:10.1029/2004JD005322, 2005.
- Müller, R., Groß, J.-U., McKenna, D. S., Crutzen, P. J., Brühl, C., Russell III, J. M., Gordley, L. L., Burrows, J. P., and Tuck, A. F.: Chemical ozone loss in the Arctic vortex in the winter 1995–96: HALOE measurements in conjunction with other observations, *Ann. Geophys.*, 17, 101–114, 1999, <http://www.ann-geophys.net/17/101/1999/>.
- Muscari, G., de Zafra, R. L., and Smyshlyaev, S.: Evolution of the NO_y-N₂O correlation in the Antarctic stratosphere during 1993 and 1995, *J. Geophys. Res.*, 108, 4428, doi:10.1029/2002JD002871, 2003.
- Nett, H., Carli, B., Carlotti, M., Dudhia, A., Fischer, H., Flaud, J.-M., Perron, G., Raspollini, P., and Ridolfi, M.: MIPAS Ground Processor and Data Products, in *Proc. IEEE 1999 International Geoscience and Remote Sensing Symposium*, 28 June–2 July 1999, Hamburg, Germany, 1692–1696, 1999.
- Plumb, R. A., Waugh, D. W., and Chipperfield, M. P.: The effects of mixing on tracer relationships in the polar vortices, *J. Geophys. Res.*, 105, 10 047–10 062, 2000.
- Ray, E. A., Moore, F. L., Elkins, J. W., Hurst, D. F., Romashkin, P. A., Dutton, G. S., and Fahey, D. W.: Descent and mixing in the 1999–2000 northern polar vortex inferred from in situ tracer measurements, *J. Geophys. Res.*, 107, 8285, doi:10.1029/2001JD000961, 2002.
- Rex, M., von der Gathen, P., Braathen, G. O., Harris, N. R. P., Reimer, E., Beck, A., Alfier, R., Krüger-Carstensen, R., Chipperfield, M., de Backer, H., Balis, D., O'Connor, F., Dier, H., Dorokhov, V., Fast, H., Gamma, A., Gil, M., Kyrö, E., Litynska, Z., Mikkelsen, I. S., Molyneux, M., Murphy, G., Reid, S. J., Rummukainen, M., and Zerefos, C.: Chemical Ozone Loss in the Arctic Winter 1994/95 as determined by the Match Technique,

- Atmos. Environ., 32, 35–59, 1999.
- Ridolfi, M., Carli, B., Carlotti, M., von Clarmann, T., Dinelli, B., Dudhia, A., Flaud, J.-M., Höpfner, M., Morris, P. E., Raspollini, P., Stiller, G., and Wells, R. J.: Optimized Forward and Retrieval Scheme for MIPAS Near-Real-Time Data Processing, *Appl. Opt.*, 39, 1323–1340, 2000.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, edited by: Taylor, F. W., World Scientific, 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 4116, doi:10.1029/2002JD002299, 2003.
- SAG: MIPAS, Michelson Interferometer for Passive Atmospheric Sounding, An ENVISAT Instrument for Atmospheric Chemistry and Climate Research, Scientific Objectives, Mission Concept and Feasibility, Instrument Design and Data Products, European Space Agency, 1999.
- Steck, T.: Methods for determining regularization for atmospheric retrieval problems, *Appl. Opt.*, 41, 1788–1797, 2002.
- Steck, T. and von Clarmann, T.: Constrained profile retrieval applied to the observation mode of the Michelson Interferometer for Passive Atmospheric Sounding, *Appl. Opt.*, 40, 3559–3571, 2001.
- Stiller, G. P. (Ed.): *The Karlsruhe Optimized and Precise Radiative Transfer Algorithm (KOPRA)*, vol. FZKA 6487 of Wissenschaftliche Berichte, Forschungszentrum Karlsruhe, 2000.
- Tikhonov, A.: On the solution of incorrectly stated problems and method of regularization, *Dokl. Akad. Nauk. SSSR*, 151, 501–504, 1963.
- Twomey, S.: On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature, *Journal of the ACM*, 10, 97–101, 1963.
- von Clarmann, T., Ceccherini, S., Doicu, A., Dudhia, A., Funke, B., Grabowski, U., Hilgers, S., Jay, V., Linden, A., López-Puertas, M., Martín-Torres, F.-J., Payne, V., Reburn, J., Ridolfi, M., Schreier, F., Schwarz, G., Siddans, R., and Steck, T.: A blind test retrieval experiment for infrared limb emission spectrometry, *J. Geophys. Res.*, 108, 4746, doi:10.1029/2003JD003835, 2003a.
- von Clarmann, T., Glatthor, N., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Mengistu Tsidu, G., Milz, M., Steck, T., Stiller, G. P., Wang, D. Y., Fischer, H., Funke, B., Gil-López, S., and López-Puertas, M.: Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), *J. Geophys. Res.*, 108, 4736, doi:10.1029/2003JD003602, 2003b.
- Wetzel, G., Oelhaf, H., Ruhnke, R., Friedl-Vallon, F., Kleinert, A., Kouker, W., Maucher, G., Reddmann, T., Seefeldner, M., Stowasser, M., Trieschmann, O., von Clarmann, T., and Fischer, H.: NO_y partitioning and budget and its correlation with N₂O in the Arctic vortex and in summer midlatitudes in 1997, *J. Geophys. Res.*, 107, 4280, doi:10.1029/2001JD000916, 2002.