



Smoothing error pitfalls

T. von Clarmann

Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Karlsruhe, Germany

Correspondence to: T. von Clarmann (thomas.clarmann@kit.edu)

Received: 28 January 2014 – Published in Atmos. Meas. Tech. Discuss.: 1 April 2014

Revised: 2 July 2014 – Accepted: 15 August 2014 – Published: 18 September 2014

Abstract. The difference due to the content of a priori information between a constrained retrieval and the true atmospheric state is usually represented by a diagnostic quantity called smoothing error. In this paper it is shown that, regardless of the usefulness of the smoothing error as a diagnostic tool in its own right, the concept of the smoothing error as a component of the retrieval error budget is questionable because it is not compliant with Gaussian error propagation. The reason for this is that the smoothing error does not represent the expected deviation of the retrieval from the true state but the expected deviation of the retrieval from the atmospheric state sampled on an arbitrary grid, which is itself a smoothed representation of the true state; in other words, to characterize the full loss of information with respect to the true atmosphere, the effect of the representation of the atmospheric state on a finite grid also needs to be considered. The idea of a sufficiently fine sampling of this reference atmospheric state is problematic because atmospheric variability occurs on all scales, implying that there is no limit beyond which the sampling is fine enough. Even the idealization of infinitesimally fine sampling of the reference state does not help, because the smoothing error is applied to quantities which are only defined in a statistical sense, which implies that a finite volume of sufficient spatial extent is needed to meaningfully discuss temperature or concentration. Smoothing differences, however, which play a role when measurements are compared, are still a useful quantity if the covariance matrix involved has been evaluated on the comparison grid rather than resulting from interpolation and if the averaging kernel matrices have been evaluated on a grid fine enough to capture all atmospheric variations that the instruments are sensitive to. This is, under the assumptions stated, because the undefined component of the smoothing error, which is the effect of smoothing implied by the finite grid on which the measurements are compared, cancels out when the difference

is calculated. If the effect of a retrieval constraint is to be diagnosed on a grid finer than the native grid of the retrieval by means of the smoothing error, the latter must be evaluated directly on the fine grid, using an ensemble covariance matrix which includes all variability on the fine grid. Ideally, the averaging kernels needed should be calculated directly on the finer grid, but if the grid of the original averaging kernels allows for representation of all the structures the instrument is sensitive to, then their interpolation can be an adequate approximation.

1 Introduction

The analysis of remotely sensed data of the atmosphere often leads to ill-posed or even underdetermined inverse problems. This is because the measurements do not contain enough information to reconstruct the atmospheric state on a grid as fine as that chosen by the retrieval scientist. A variety of regularization techniques have been proposed to solve such kinds of inverse problems, among them regularization methods by Tikhonov (1963a), Twomey (1963) and Phillips (1962), as well as the maximum a posteriori scheme, which has been systematically investigated by Rodgers (2000) and which had formerly been referred to as optimal estimation (Rodgers, 1976). Any of these regularized retrievals, however, contain formal prior information.

Contrary to its use in analytical philosophy, the term “a priori” does in this context not denote factual (as opposed to logical or analytical) knowledge which is so obviously true that it can be taken for granted (in a Kantian sense). Instead, in remote sensing theory, “prior” or “a priori” are defined only relative to a measurement and denote what is known – or assumed to be known – before the measurement is taken; in other words, these terms are used here in a Bayesian sense.

We call prior information “formal” if it is imported via a formal constraint in the retrieval equation, as opposed to indirect prior assumptions. Indirect a priori assumptions, or indirect constraints, can be applied, for example, by simply using a finite and rather coarse grid for representation of the atmospheric state and an interpolation rule for determination of the atmospheric state between the grid points, or by retrieving a nonlinear function of the target quantity x which constrains the result to positive (e.g. by actually retrieving the logarithm of x) or otherwise bounded (e.g. by actually retrieving the sine or cosine of x values). The interaction of the chosen grid and regularization is discussed, for example, in Haario et al. (2004) and references therein.

With a grid coarse enough, maximum likelihood retrievals which do not require any formal constraint or a priori information are often possible. While the effect of finite resolution is self-evident in the latter case, because nobody reasonably expects the resolution of, for example, a vertical profile be better than the grid on which it is represented, regularized retrievals lead to oversampled profiles, i.e. there are more altitude grid points than independent pieces of information. In this case, it is essential to report the influence of the prior information on the retrieval to the user. Since the constraint can push the retrieval away from the actual true state of the atmosphere towards the prior information, the regularization causes an additional error term. This term is larger when the influence of the prior information is stronger, which is the price to pay for a reduction in the retrieval noise by regularization. This additional error term was initially called “null space error” (Rodgers, 1990) until it was renamed “smoothing error” (Rodgers, 2000).

In this paper it will be shown that this constraint diagnostic has a particular characteristic which makes the related concept questionable in the context of error budget. In Sect. 2 the formal environment will be presented in which the discussion will take place and the notation and terminology will be clarified. In Sect. 3 the error propagation of the smoothing error will be discussed and related problems will be identified. Section 4 is dedicated to the critical discussion of the attempt to save the smoothing error concept by evaluating it on a fine enough grid, and, in Sect. 5, alternative approaches to characterize the impact of prior information on the profile are discussed. In Sect. 6 an application will be identified for which, despite all criticism, a concept closely related to the smoothing error concept is still appropriate. Finally, in Sect. 7, the main lessons learned will be summarized and the implications on the appropriate representation of remotely sensed data will be discussed.

2 Background and notation

For formulation of a constrained retrieval we use the concept and notation of Rodgers (2000) with some minor adjustments by von Clarmann et al. (2003). We minimize a

two-component cost function c

$$c = (\mathbf{y} - F(\mathbf{x}))^T \mathbf{S}_y^{-1} (\mathbf{y} - F(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{R} (\mathbf{x} - \mathbf{x}_a), \quad (1)$$

where \mathbf{y} is the m -dimensional vector of measurements, F the $\mathbb{R}^n \rightarrow \mathbb{R}^m$ signal transfer forward model, \mathbf{x} the n -dimensional vector of the unknown components of the atmospheric state, \mathbf{S}_y the $m \times m$ measurement error covariance matrix, \mathbf{x}_a the n -dimensional a priori information on the atmospheric state and \mathbf{R} an $n \times n$ regularization matrix. This leads, after linear replacement to $F(\mathbf{x})$ by $\mathbf{x}_a + \mathbf{K}(\mathbf{x} - \mathbf{x}_a)$, where \mathbf{K} is the Jacobian matrix with elements $k_{i,j} = \partial y_i / \partial x_j$, to the following retrieval equation:

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{x}_a + (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{R})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} (\mathbf{y} - F(\mathbf{x}_a)) \\ &= \mathbf{x}_a + \mathbf{G}(\mathbf{y} - F(\mathbf{x}_a)), \end{aligned} \quad (2)$$

where the $\hat{\mathbf{x}}$ symbol denotes the estimated profile, and where the so-called gain function \mathbf{G} , which will later be used for brevity, is implicitly defined by the second line of the equation. Various choices of \mathbf{R} are possible: $\mathbf{R} = \mathbf{S}_a^{-1}$, where \mathbf{S}_a is the a priori covariance matrix, leads to a maximum a posteriori retrieval (Rodgers, 2000), while squared and scaled k th-order finite difference matrices have been suggested by Phillips (1962), Tikhonov (1963b, a) and Twomey (1963) and have systematically been investigated for remote sensing applications by, for example, Schimpf and Schreier (1997) or Steck and von Clarmann (2001). Nonlinear variants of these retrieval approaches are common but not relevant to the topic of this paper.

The dependence of the solution on the true state is characterized by the so-called averaging kernel matrix of dimension $n \times n$

$$\mathbf{A} = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{R})^{-1} \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K}. \quad (3)$$

With this we can rewrite Eq. (2) as

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{x} + (\mathbf{I} - \mathbf{A})\mathbf{x}_a, \quad (4)$$

where \mathbf{I} is the $n \times n$ identity matrix. Rodgers (1990, 2000) suggests the application of generalized Gaussian error propagation (cf. next section) to estimate a diagnostic quantity, which is the mapping of the expected deviation of \mathbf{x}_a from the actual \mathbf{x} :

$$\mathbf{S}_{\text{smoothing}} = (\mathbf{I} - \mathbf{A})\mathbf{S}_e(\mathbf{I} - \mathbf{A})^T. \quad (5)$$

\mathbf{S}_e is the covariance matrix of the atmospheric state around the mean state. The diagnostic quantity $\mathbf{S}_{\text{smoothing}}$ is the expected deviation of the retrieval from the true state which is caused by the constraint term in Eq. (2) and is directly comparable to other retrieval errors, e.g. noise. Thus, this constraint diagnostic is called “smoothing error”. It is an intuitive quantity used to characterize the uncertainty due to the difference between the actual atmospheric state and the prior

information. The appropriateness to include this quantity in the error budget, however, requires closer inspection. Before this, some more general caveats in the context of the smoothing error are summarized.

The linear estimate presented in Eq. (5) holds only if indeed $\mathbf{x}_a = \langle \mathbf{x} \rangle$, where $\langle \rangle$ denotes the expectation value. More precisely, it is required that \mathbf{S}_e represents the covariance around $\langle \mathbf{x} \rangle$, and not the covariance around \mathbf{x}_a if the latter happens not to be chosen to equal $\langle \mathbf{x} \rangle$, or around any other arbitrarily chosen a priori state. The use of arbitrarily chosen covariance matrices for the evaluation of the smoothing error is critically discussed in Rodgers (2000, p. 49), while the need to consider a possible bias between the correct expectation value of the atmospheric state and the ad hoc prior chosen to constrain the retrieval is outlined, for example, in von Clarmann and Grabowski (2007). In the latter case the effect of the formal constraint is not only smoothing of the true atmospheric state, and as a consequence the so-called smoothing error has to be complemented by the additional component

$$(\mathbf{I} - \mathbf{A})(\mathbf{x}_a - \langle \mathbf{x} \rangle)(\mathbf{x}_a - \langle \mathbf{x} \rangle)^T (\mathbf{I} - \mathbf{A})^T, \quad (6)$$

which accounts for the bias of \mathbf{x}_a .

Further, it is important that the \mathbf{S}_e matrix includes atmospheric variability on all of the scales which can be represented on the grid on which it is evaluated. \mathbf{S}_e matrices constructed from real data often happen to be singular. This can hint at a situation where the parent data do not resolve atmospheric variability on the small scales corresponding to the grid on which the \mathbf{S}_e is represented. In this case, Eq. (5) will underestimate the smoothing error. The same is, of course, true if the parent data do not fully cover the true spatial and temporal atmospheric variability.

Moreover, the term “smoothing error” can be misleading, because, depending on the retrieval scheme chosen, the retrieved profile is not necessarily a smoothed version of the true profile but can also be a combination of the a priori profile and the profile the unregularized retrieval would tend towards. While in many cases the profile obtained by means of Eq. (2) is smoother than the true profile, there is no reason that this should always be the case. The retrieved profile can also be shifted with respect to the true profile, or, depending on the actual prior information used, it can also have artificial structure.

Examples of error budget estimates including the smoothing error or with the smoothing error as a supplemental diagnostic quantity can be found in Worden et al. (2004), Bowman et al. (2006) and Kramarova et al. (2013).

3 Error propagation

3.1 General linear or moderately nonlinear case

Let

$$\mathbf{v} = f(\mathbf{u}) \quad (7)$$

for any real vectorial argument \mathbf{u} and any real vectorial result \mathbf{v} . The uncertainties of \mathbf{u} map onto the uncertainties of \mathbf{v} as

$$\mathbf{S}_v \approx \mathbf{K} \mathbf{S}_u \mathbf{K}^T, \quad (8)$$

where \mathbf{S}_u and \mathbf{S}_v are the error covariance matrices of vectors \mathbf{u} and \mathbf{v} , respectively, and where \mathbf{K} is the Jacobian matrix of $\mathbf{v} = f(\mathbf{u})$ with elements $\frac{\partial v_j}{\partial u_i}$. Equation (8) is a generalization of the Gaussian error propagation law¹

$$\sigma_{v_j}^2 \approx \sum_i \left(\frac{\partial v_j}{\partial u_i} \right)^2 \sigma_{u_i}^2, \quad (9)$$

where σ_{u_i} and σ_{v_j} are the standard deviations representing the uncertainties of v_j and u_i , respectively. Contrary to the latter equation, which assumes uncorrelated u_i , Eq. (8) is valid also for intercorrelated errors of u_i , which are accounted for by the related off-diagonal elements of covariance matrix \mathbf{S}_u . These error propagation rules are generally accepted in all cases except for grossly nonlinear functions $f(\mathbf{u})$.

Application of this formalism to the mapping of measurement noise onto retrieved atmospheric state variables gives

$$\mathbf{S}_{\text{noise}} \approx \mathbf{G} \mathbf{S}_y \mathbf{G}^T. \quad (10)$$

3.2 Application to retrieved profiles

Typical linear operations performed with retrieved vertical profiles are transformation from one altitude grid to another, e.g. by interpolation from a coarse grid to a finer grid (cf. Rodgers, 2000, p. 162) by

$$\hat{\mathbf{x}}_{\text{fine}} = \mathbf{W} \hat{\mathbf{x}}_{\text{coarse}}, \quad (11)$$

of which a possible inverse operation is

$$\hat{\mathbf{x}}_{\text{coarse}} = \mathbf{V} \hat{\mathbf{x}}_{\text{fine}} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{x}_{\text{fine}}. \quad (12)$$

Here, $\hat{\mathbf{x}}_{\text{coarse}}$ and $\hat{\mathbf{x}}_{\text{fine}}$ are of dimensions n and \tilde{n} , and \mathbf{W} and \mathbf{V} are $\tilde{n} \times n$ - and $n \times \tilde{n}$ -dimensional transformation matrices, respectively. In this context it is important to note that transformation from the coarse to the fine grid is reversible because $\mathbf{V} \mathbf{W} = \mathbf{I}_{\text{coarse}}$, i.e. back transformation from the fine to

¹Although its name may suggest the contrary, this error propagation scheme is independent of the particular error distribution and does not depend on a Gaussian error distribution.

the coarse grid will fully restore the original coarse-grid profile. In contrast, transformation from the fine to the coarse grid implies an irreversible loss of information; because of $\mathbf{W}\mathbf{V} \neq \mathbf{I}_{\text{fine}}$, back transformation to the fine grid will not restore the original profile.

According to Eq. (8), retrieval noise is propagated from the coarse to the fine grid as

$$\mathbf{S}_{\text{noise, fine}} = \mathbf{W}\mathbf{S}_{\text{noise, coarse}}\mathbf{W}^T \quad (13)$$

and from the fine grid to the coarse grid as

$$\mathbf{S}_{\text{noise, coarse}} = \mathbf{V}\mathbf{S}_{\text{noise, fine}}\mathbf{V}^T. \quad (14)$$

The same equations apply to the propagation of the parameter error estimate. The latter is the response of the retrieval to uncertainties in the forward model parameters.

As has already been mentioned by Rodgers (2000), the ensemble covariance matrix \mathbf{S}_e cannot be transformed from a coarser to a finer grid by means of Eq. (13), because it does not represent the variability on any scale finer than that represented on its original grid. It has, however, never been discussed that, as a direct consequence of this, the smoothing error as evaluated using Eq. (5) also cannot be interpolated from its native grid to any finer grid. The smoothing error of $\hat{\mathbf{x}}$ represents smoothing error components only with respect to variability which can be represented on the native grid of \mathbf{S}_e .

The striking consequence of this, which has, to the best knowledge of the author, never been mentioned, is that the generalized Gaussian error propagation does not generally apply to the smoothing error. Even for linear functions $f(\mathbf{x})$, error propagation laws fail when applied to the smoothing error as soon as the linear function involves any kind of interpolation to any grid finer than that on which the smoothing error has been evaluated. Interpolation of retrieved data to grids different from (often: finer than) the initial retrieval grid are a frequent task, e.g. when databases are created in which results of different instruments are represented in a common format and on a common grid (e.g. Sofieva et al., 2013; Hegglin et al., 2013; Tegmeier et al., 2013).

While Gaussian error propagation (Eq. 8) of the smoothing error would give

$$\begin{aligned} \mathbf{S}_{\text{smoothing, fine}} &= \mathbf{W}\mathbf{S}_{\text{smoothing, coarse}}\mathbf{W}^T \\ &= \mathbf{W}(\mathbf{I}_{\text{coarse}} - \mathbf{A})\mathbf{S}_{e, \text{coarse}}(\mathbf{I}_{\text{coarse}} - \mathbf{A})^T\mathbf{W}^T \\ &= (\mathbf{I}_{\text{fine}} - \mathbf{W}\mathbf{A}_{\text{coarse}}\mathbf{V})\mathbf{W}\mathbf{V}\mathbf{S}_{e, \text{fine}}(\mathbf{W}\mathbf{V})^T \\ &\quad (\mathbf{I}_{\text{fine}} - \mathbf{W}\mathbf{A}_{\text{coarse}}\mathbf{V})^T \end{aligned} \quad (15)$$

(cf. Rodgers, 2014, for the representation shown in the third and fourth lines of this equation), the correct linear estimate is

$$\mathbf{S}_{\text{smoothing, fine}} = (\mathbf{I}_{\text{fine}} - \mathbf{A}_{\text{fine}})\mathbf{S}_{e, \text{fine}}(\mathbf{I}_{\text{fine}} - \mathbf{A}_{\text{fine}})^T, \quad (16)$$

with $\mathbf{A}_{\text{fine}} = \mathbf{W}\mathbf{G}_{\text{coarse}}\mathbf{K}_{\text{fine}}$ (Rodgers, 2000, p. 161). Equation (16) cannot be inferred via Eq. (8) from $\mathbf{S}_{\text{smoothing, coarse}}$.

Here, $\mathbf{S}_{e, \text{fine}}$ is the ensemble covariance matrix evaluated on the fine grid and including small-scale variability which cannot be represented on the coarse grid, \mathbf{K}_{fine} is the Jacobian which represents the sensitivities of the measurements to atmospheric variability on the fine grid and $\mathbf{I}_{\text{coarse}}$ and \mathbf{I}_{fine} are the identity matrices on the respective grids.

The problem is caused by the fact that the smoothing error does not characterize the full smoothing effect but instead only that part which is caused by the constraint term in Eq. (2). The additional smoothing caused by the finite grid which cannot resolve all atmospheric variability remains unaccounted for. This representation error term is assumed to be practically zero in the idealized framework by Rodgers (2000), but this assumption will be challenged in the next section.

In order to demonstrate that this difference is not only of academic interest, $\mathbf{S}_{\text{smoothing, fine}}$ has been evaluated both via generalized Gaussian error propagation (Eq. 15) and directly on the fine grid (Eq. 16) (Fig. 1). The grid widths of the fine and the coarse grids have been chosen to be 1 and 3 km, respectively. For simplicity, the coarse grid was chosen to be a subset of the fine grid. The averaging kernels were assumed to be triangular in the fine grid, where the sum over the averaging kernel elements was unity. They were transformed into the coarse grid via

$$\mathbf{A}_{\text{coarse}} = \mathbf{V}\mathbf{A}_{\text{fine}}\mathbf{W} \quad (17)$$

(bottom left panel in Fig. 1). The ensemble covariance matrix $\mathbf{S}_{e, \text{fine}}$ was constructed with diagonal values of 1 (in arbitrary units), and exponentially decreasing all positive off-diagonal values, where the correlation length was varied from values of 1 to 20 km (upper left panel in Fig. 1). Construction of $\mathbf{S}_{e, \text{coarse}}$ relies on the \mathbf{V} matrix (upper right panel in Fig. 1). Averaging kernels and climatological variabilities were chosen to be altitude-independent.

First, a test case with a correlation length of 1 km and a vertical resolution of the retrieval of 6 km is discussed in more detail. The resulting smoothing error on the coarse grid is, in terms of variances, 0.38, and the covariances between adjacent profile points are as negative as -0.24 (dark-blue curve in the lower right panel in Fig. 1, which is hardly discernable because it is overplotted by the central red curve). This anticorrelation is intuitive because smoothing means that if, for example, a profile maximum is smeared, the retrieved values at the maximum will be too low while values at adjacent profile points will be too high. Generalized Gaussian error propagation of the smoothing error to the fine grid according Eq. (15) reproduces the errors at the grid points of the fine grid which are also part of the coarse grid, but at interjacent grid points the propagated smoothing error variances are calculated to be as low as 0.10 (red lines/symbols in the lower right panel in Fig. 1). This is computationally intuitive, because interpolation between values with anticorrelated errors leads to error cancellation; physically, however, this is counterintuitive because interpolation cannot reduce

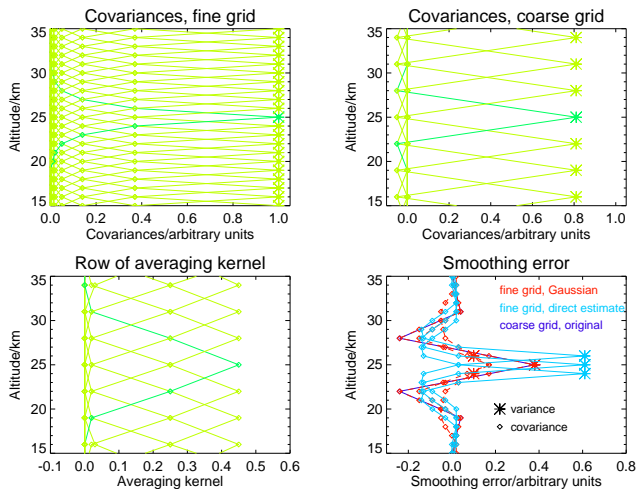


Figure 1. Case study: the upper left panel shows the ensemble covariances on the fine grid (grid spacing 1 km). Only the symbols are significant – the lines are only plotted to guide the eye. The large asterisks are the variances. The variance and covariances referring to 25 km are highlighted for clarity. The top right panel shows the covariances on the coarse (grid width 3 km) grid. The lower left panel shows the averaging kernels on the coarse grid. The lower right panel shows the estimated smoothing errors (in terms of variances/covariances) at 24, 25 and 26 km altitude: the smoothing errors on the fine grid estimated by Gaussian error estimation (red) are largest at 25 km, an altitude which coincides with an altitude of the coarse grid, and are smaller for 24 and 26 km, where the values on the fine grid depend on interpolation. The opposite is true for the direct estimates of the smoothing error on the fine grid (light blue): here the smoothing error is smallest at 25 km and larger at 24 and 26 km. More importantly, the directly estimated smoothing errors are considerably larger. This is because the relevant ensemble covariance matrix contains larger atmospheric variability (cf. top panels). The original smoothing error estimate on the coarse grid (dark blue) is hardly visible because it is identical to that represented on the fine grid.

the smoothing error. The direct evaluation on the fine grid via Eq. (16) gives smoothing error variances of 0.61 (light-blue lines/symbols in the lower right panel). The smoothing errors are larger because they account for the additional variability which can be represented only on the fine grid but which is lost when smoothing errors are evaluated on the coarse grid.

For larger correlation lengths in $S_{e, \text{fine}}$, the smoothing errors decrease but the contrast between the two ways to estimate it on the fine grid remains large. For a correlation length of 20 km and a vertical resolution of the retrieval of 6 km, the correctly calculated smoothing error on the fine grid is still more than 3 times larger than that estimated via Gaussian error propagation. For inferior altitude resolutions, this ratio becomes smaller, but even for a correlation length of 20 km and a vertical resolution of 22 km, the correctly calculated smoothing error is still higher by 37 % compared to the estimate using Gaussian error propagation. Obviously, the

difference between the two ways to estimate the smoothing error does not fully disappear even if the original retrieval has been considerably oversampled (Table 1). Putting theoretical concerns aside, dissemination of diagnostic matrices sampled fine enough to keep the inaccuracy implied by any further interpolation tolerably small can easily be beyond reach for reasons of the pure amount of data to be communicated, and in many real applications the grid on which the diagnostic quantities are provided is defined in a way that the scales which the instrument can measure are resolved (weak gridding criterion) rather than all the scales on which atmospheric variability still occurs (strong gridding criterion).

Therefore either Gaussian error propagation has to be abandoned or the smoothing error problem has to be fixed in a way that the smoothing error concept becomes consistent with the generalized Gaussian error propagation law. Since Gaussian error propagation is an essential part of linear theory and even of quantitative empirical research in general, it might not be acceptable to drop it in favour of the current smoothing error concept. Instead, either a way needs to be determined by which the smoothing error concept can be modified such that it becomes compatible with established error propagation laws, as will be attempted in the next section, or otherwise an alternative way to report the a priori content of the retrieval which makes no use of the smoothing error concept is needed.

4 The nature of the retrieved quantities

Having understood the source of the problem and accepting that there exists natural variability on all physical scales (Richardson, 1920), the natural approach would appear to be to evaluate the smoothing error on an infinitesimally fine grid. This would assure that the smoothing error represents atmospheric variation on all possible scales. Of course, this ideal cannot be reached within finite-dimension algebra, but one could at least try to evaluate the smoothing error on a grid fine enough that further refinement of the grid does not imply additional variability. In other words, the problem should be diagnosed on a grid on which the full variability of the atmosphere can be represented (strong gridding criterion). This approach is based on the assumption that the residual smoothing error not accounted for on a finite grid converges towards zero for a grid spacing approaching zero. In the following it will be shown that this assumption is false.

For an air volume of the size of a molecule, i.e. still much larger than the infinitesimal scale, the mixing ratio of a species is not a meaningful quantity: either, at the given point, there is a target molecule, and thus the mixing ratio is one; there is a molecule of another species, and thus the mixing ratio is zero; or there is no molecule at all and thus the mixing ratio is fully undefined because this would involve division by zero. For number densities and temperature, there are similar problems with the definition of these quantities

Table 1. Ratio of correctly calculated smoothing errors and smoothing errors calculated via Gaussian error propagation.

Correlation length [km]	Resolution [km]										
	4.000	6.000	8.000	10.000	12.000	14.000	16.000	18.000	20.000	22.000	
1000	11.186	5.933	4.377	3.724	3.389	3.184	3.047	2.956	2.894	2.849	
2000	7.751	4.225	3.114	2.636	2.388	2.237	2.139	2.075	2.032	2.001	
3000	6.741	3.698	2.722	2.296	2.072	1.937	1.849	1.792	1.755	1.728	
4000	6.304	3.463	2.545	2.143	1.929	1.800	1.716	1.662	1.627	1.603	
5000	6.074	3.336	2.449	2.059	1.851	1.724	1.642	1.590	1.556	1.533	
6000	5.937	3.260	2.391	2.007	1.802	1.677	1.597	1.546	1.512	1.490	
7000	5.848	3.209	2.352	1.973	1.770	1.646	1.566	1.516	1.483	1.461	
8000	5.788	3.175	2.325	1.949	1.747	1.623	1.544	1.495	1.462	1.440	
9000	5.744	3.150	2.306	1.932	1.730	1.607	1.528	1.479	1.447	1.425	
10000	5.712	3.131	2.291	1.918	1.718	1.595	1.516	1.467	1.435	1.414	
11000	5.688	3.117	2.280	1.908	1.708	1.585	1.507	1.458	1.426	1.405	
12000	5.669	3.105	2.271	1.900	1.700	1.578	1.499	1.451	1.419	1.398	
13000	5.654	3.096	2.264	1.894	1.694	1.572	1.493	1.445	1.413	1.392	
14000	5.641	3.089	2.258	1.889	1.689	1.566	1.488	1.440	1.408	1.387	
15000	5.631	3.083	2.253	1.884	1.685	1.562	1.484	1.436	1.404	1.383	
16000	5.623	3.078	2.249	1.881	1.681	1.559	1.481	1.432	1.401	1.380	
17000	5.616	3.074	2.246	1.877	1.678	1.556	1.478	1.429	1.398	1.377	
18000	5.610	3.070	2.243	1.875	1.676	1.553	1.475	1.427	1.395	1.374	
19000	5.605	3.067	2.240	1.873	1.673	1.551	1.473	1.425	1.393	1.372	
20000	5.601	3.065	2.238	1.871	1.672	1.549	1.471	1.423	1.391	1.370	

in any meaningful manner infinitesimal point on this small scale; quantities which characterize an air parcel in a statistical sense are not applicable any more. The characterization of the atmosphere by statistical terms implies a certain inherent smoothing and thus the true unsmoothed state of the atmosphere is ill-defined. It is not clear with respect to which quantity the expected differences should be characterized by the smoothing error.

Admittedly, the scales discussed here are of no concern in remote sensing. However, it is not the intent here to discuss the state of single molecules but simply to show that there exists no reasonable limit to which mixing ratios, number densities or temperature converge for steadily decreasing scale lengths, i.e. that convergence of the smoothing error cannot safely be expected when the grid spacing approaches zero: for example, mixing of air parcels of different composition range from planetary waves down to the molecular scale. Thus, for any finite grid, there exist sub-grid processes causing their own variability in the atmospheric state not represented by S_e until we reach the molecular scale on which the pathological cases discussed above occur.

In conclusion, the attempt to solve the propagation problem of the smoothing error by use of a grid fine enough that it is guaranteed that interpolation will never occur must be considered as failed. In more practical terms, it is fair to say that if sufficient information is available to construct S_e on a certain fine grid, then there will be scientists who are interested in atmospheric processes on even finer scales which have their own variability.

5 The way out of the dilemma

Since generalized Gaussian error propagation is one of the most essential principles of linear theory, it seems unacceptable to define an error which, even for a linear operation, is not propagated by Eq. (8). The problem can be avoided by changing the notion of what an atmospheric state variable actually represents. All problems discussed above originate from the fact that an ideal measurement of an atmospheric state value represents an ideally resolved, and thus extensionless, point in the atmosphere, and that every measurement of finite spatial resolution is less than ideal and thus affected by a smoothing error representing the expectation of the deviation of the finite-resolution measurement from the fictive true actual value at infinitesimal resolution. Rodgers (2000, p. 48) mentions an alternative understanding of the measurements of the atmospheric state as representing an extended air volume and characterizing the measurement by its measurement and parameter errors (excluding the smoothing error) plus a characterization of the spatial resolution (e.g. via communicating the averaging kernel to the data user). As a result of the discussion above, the dichotomy of understanding the retrieval either as an estimate of the “state smoothed by the averaging kernel” or an “estimate of the true state” Rodgers (2000, p. 48, lines 2–4 in Sect. 3.2.1) does not hold, because any representation of the atmosphere refers to finite air volumes or any other finite representation in both cases. As a consequence of this, the alternative approach of regarding the retrieval as an estimate of the smoothed state is not

only an option but in fact seems to be the only reasonable choice because the concept of the ideally infinitesimally fine resolved atmospheric state has been shown to be untenable. The smoothing error concept which assumes a “true”, i.e. unsmoothed, atmosphere contradicts itself, because the evaluation of the smoothing error on a finite grid with its implicit smoothing through finite representation gives the notion of the retrieval characterizing a finite air volume access through the back door again; in other words it breaks with its own assumption that the “smoothing error” represents the entire smoothing component of the retrieval error relative to the “true” atmosphere in absolute, i.e. grid-independent, terms without grid-dependent representation errors.

The decision to distribute the averaging kernel matrix instead of the smoothing error as the main diagnostic to characterize the impact of the constraint of the retrieval needs further discussion. To compare the effect of the constraint to the effect of measurement errors on a grid finer than that on which the data are distributed, the user might wish to calculate the smoothing error on the finer grid. The user might have an \mathbf{S}_e matrix available or can construct one from known energy cascades between scales or knowledge on relevant small-scale processes. The user can do so because \mathbf{S}_e is not an instrument-dependent quantity, i.e. its construction does not require expert knowledge on the particular instrument. However, the user will have the averaging kernels available only on the original (coarser) grid. In this case the user would use

$$\mathbf{S}_{\text{smoothing, fine}} = (\mathbf{I}_{\text{fine}} - \mathbf{WAV})\mathbf{S}_{e, \text{fine}}(\mathbf{I}_{\text{fine}} - \mathbf{WAV})^T, \quad (18)$$

which is incorrect because $\mathbf{WV} \neq \mathbf{I}_{\text{fine}}$. If, however, the original grid had been chosen fine enough to represent all the atmospheric variability that the instrument is sensitive to (weak gridding criterion), the error caused by the interpolation of the averaging kernel matrix can remain tolerably small. Table 2 shows the ratios of smoothing errors calculated with the correct averaging kernel matrix and those calculated according to Eq. (18) for the series of case studies from Sect. 3.2. In all cases the approximation used leads to an overestimation of the smoothing error, but for cases when the original coarse grid is more than 3 times finer than the resolution of the retrieval, related errors of the estimated smoothing error are smaller than 5%. In all cases, the inaccuracy of the estimated smoothing error due to interpolation of the averaging kernel is orders of magnitude smaller than inaccuracies by application of Gaussian error propagation to the coarse-grid smoothing error. Thus, it seems to be, in agreement with Rodgers (2000, Sect. 11.2.6), preferable to distribute the averaging kernel instead of the smoothing error.

Once having accepted the failure of the smoothing error concept as a grid-independent tool to characterize the full smoothing component of the difference between the retrieved and the true atmospheric state, it is comforting that the finite-resolution concept offers at least three further advantages:

first, the estimate of the error budget for any retrieval involving a given \mathbf{R} (which may or may not be an approximation to, or coarse sampling of, \mathbf{S}_e^{-1}) no longer depends via Eq. (5) on the choice² of the ensemble covariance matrix. Often no reliable estimate of \mathbf{S}_e is available, but any arbitrary choice is in conflict with the smoothing error concept (cf. Rodgers, 2000, p. 48). Second, the averaging kernel is needed for a number of applications of measured data regardless, and to provide it instead of the smoothing error is advantageous for the data user. Third, error budgets of instruments whose retrievals are performed on different grids become intercomparable, which was not the case when the error budget still included the smoothing error. The latter is again related to the core of the problem, viz. that smoothing errors evaluated on different grids actually represent different error components. Although meaningless, it is indeed common practice to compare total error bars (including the smoothing component) of retrievals performed on different grids.

One implication of abandoning the smoothing error concept as part of the error budget is that the usual estimate of the retrieval error covariance matrix shown below is no longer valid, at least not in a general sense where transformation between grids are an issue. Rodgers (1976) states that the retrieval error covariance matrix is

$$\mathbf{S}_x = \left(\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_a^{-1} \right)^{-1}. \quad (19)$$

This covariance matrix which uses the a priori covariance matrix \mathbf{S}_a as an approximation for the ensemble covariance matrix \mathbf{S}_e contains both the measurement noise and the smoothing error component (cf. Rodgers, 2000, p. 58). Thus, all caveats discussed for the smoothing error apply equally to the error estimate of Eq. (19). An error estimate free of smoothing error contributions can be made by direct application of Eq. (10) to the various error sources, viz. noise and parameter errors.

Moreover, Eq. (19) is, regardless of the discussion of the smoothing error in this paper, inapplicable to any choice of the \mathbf{S}_a matrix except for the true climatological a priori covariance matrix \mathbf{S}_e . While reasonable retrievals can be performed with ad hoc choices of the regularization term in Eq. (2), Eq. (19) does not provide a valid error estimate in these cases. The inadequacy of an ad hoc choice of \mathbf{S}_e which has been already highlighted by Rodgers (2000, p. 48) also makes Eq. (19) inadequate for all choices of \mathbf{S}_a except for the true covariance of the atmospheric state under investigation.

²Ideally, contrary to the a priori covariance matrix \mathbf{S}_a , there is no “choice” in the construction of \mathbf{S}_e , because it is not an ad hoc regularization matrix but instead a real ensemble covariance matrix. In practice, however, there remains some ambiguity with respect to the selection of the data source and the definition of the ensemble.

Table 2. Ratio of correctly calculated smoothing errors and smoothing errors calculated using interpolated averaging kernels.

Correlation length [km]	Resolution [km]									
	4.000	6.000	8.000	10.000	12.000	14.000	16.000	18.000	20.000	22.000
1000	0.810	0.927	0.961	0.975	0.983	0.988	0.991	0.993	0.994	0.995
2000	0.782	0.918	0.958	0.974	0.982	0.988	0.991	0.993	0.994	0.995
3000	0.771	0.913	0.956	0.973	0.982	0.987	0.991	0.993	0.994	0.995
4000	0.766	0.911	0.955	0.972	0.981	0.987	0.991	0.994	0.995	0.995
5000	0.763	0.909	0.953	0.971	0.981	0.987	0.991	0.994	0.995	0.995
6000	0.761	0.908	0.953	0.971	0.981	0.987	0.991	0.994	0.995	0.996
7000	0.760	0.907	0.952	0.970	0.980	0.987	0.991	0.994	0.995	0.996
8000	0.759	0.906	0.952	0.970	0.980	0.986	0.991	0.994	0.995	0.996
9000	0.759	0.906	0.951	0.970	0.980	0.986	0.991	0.994	0.995	0.996
10000	0.758	0.906	0.951	0.970	0.980	0.986	0.991	0.994	0.995	0.996
11000	0.758	0.905	0.951	0.970	0.980	0.986	0.991	0.994	0.995	0.996
12000	0.758	0.905	0.951	0.969	0.980	0.986	0.991	0.994	0.995	0.996
13000	0.758	0.905	0.951	0.969	0.979	0.986	0.991	0.994	0.995	0.996
14000	0.758	0.905	0.951	0.969	0.979	0.986	0.991	0.994	0.995	0.996
15000	0.757	0.905	0.951	0.969	0.979	0.986	0.991	0.994	0.995	0.996
16000	0.757	0.905	0.950	0.969	0.979	0.986	0.991	0.994	0.995	0.996
17000	0.757	0.905	0.950	0.969	0.979	0.986	0.991	0.994	0.995	0.996
18000	0.757	0.905	0.950	0.969	0.979	0.986	0.991	0.994	0.995	0.996
19000	0.757	0.905	0.950	0.969	0.979	0.986	0.991	0.994	0.995	0.996
20000	0.757	0.904	0.950	0.969	0.979	0.986	0.991	0.994	0.995	0.996

6 Implication for comparison of retrievals

An exception where a quantity calculated on the basis of a concept closely related to the smoothing error is still a useful and powerful tool is comparison of remotely sensed data according to Rodgers and Connor (2003, their Eqs. 10 to 14). These authors suggest that profiles be validated against each other by testing whether their difference, $\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2$, is significant in terms of χ^2 statistics. The covariance matrix of the difference, \mathbf{S}_δ , needed for this test, however, must not include interdependent components of the smoothing error. Thus, these authors suggest that \mathbf{S}_δ be calculated as

$$\mathbf{S}_\delta = (\mathbf{A}_1 - \mathbf{A}_2)\mathbf{S}_c(\mathbf{A}_1 - \mathbf{A}_2)^T + \mathbf{S}_{x1} + \mathbf{S}_{x2}, \quad (20)$$

where \mathbf{A}_1 , \mathbf{A}_2 , \mathbf{S}_{x1} and \mathbf{S}_{x2} are the respective averaging kernel and retrieval noise covariance matrices and where \mathbf{S}_c is the comparison ensemble covariance matrix. The first term on the right-hand side of this equation characterizes the smoothing difference between both these retrievals.

This estimate of the smoothing difference between two instruments' results is necessary to judge whether the difference between the retrievals is significant or whether it can be attributed to the different smoothing characteristics of the retrievals. In this context, it is not necessary to know the smoothing error relative to the true atmospheric state as it is sufficient to characterize the difference between the smoothing characteristics. Following the Rodgers and Connor (2003) scheme, the difference is calculated on a common so-called "intercomparison grid", which should generally be

at least as fine as the parent grids. When the difference $\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}_2$ between the profiles is calculated on this grid, any degradation of the knowledge of the atmospheric state due to the representation on a finite grid is the same for both profiles and thus cancels out, provided that \mathbf{S}_c has been evaluated on the intercomparison grid or any grid finer than that but is not a result of interpolation, and the grid is fine enough to ensure that the averaging kernels represent all the scales that the instruments are sensitive to (weak gridding criterion) (see Appendix). This implies that, when differences of profiles are considered, the problematic component of the smoothing error, which is the difference between the true atmosphere sampled on the comparison grid and the true atmosphere at "infinite resolution", has no relevance anymore, and the χ^2 analysis is still valid.

The approach of Rodgers and Connor (2003), however, is not without pitfalls: it is essential that the a priori covariance matrix of the comparison ensemble, \mathbf{S}_c , represents all variability of the atmospheric state on the comparison grid. For reasons discussed in Sect. 3.2, the a priori covariance matrix cannot simply be interpolated to the comparison grid.

In summary, the smoothing difference, if calculated correctly, is still a useful quantity, while the parent smoothing errors of the original profiles are affected by the problems discussed in the previous sections and thus should not be part of an error budget.

If the contrast in the vertical resolutions of two measurements is large, then the comparison can be carried out by much simpler means: the better resolved measurement can

then often be regarded as nearly ideal, and x in Eq. (4) can be replaced with the high-resolution profile to yield the high-resolution profile as the poorer resolving instrument would see it. This transformed high-resolution profile can then be directly compared to the poorly resolved profile. This approach has first been proposed and demonstrated by Connor et al. (1994).

7 Conclusions

The following discussion is limited to retrievals using formal a priori information. Recommendations are conditional, assuming that the decision in favour of a constrained retrieval has already been made. Alternatives which avoid the whole problem, such as maximum likelihood retrievals without a formal constraint (cf., for example, Carlotti, 1988), may be worthwhile trying but are beyond the scope of this discussion.

The conclusions are split into two parts, the first of which being theoretical and descriptive, and the second practical and thus prescriptive.

7.1 Theoretical conclusions

It has been shown in this paper that the quantity called smoothing error does not represent an estimate of the regularization-induced difference between the retrieved state and the “true” state of the atmosphere. Instead it characterizes the difference between the retrieved state and an arbitrary representation of the true state, where this arbitrary representation itself, being a representation on a finite discrete grid, has its own representation error which can be understood as an implicit smoothing error. It has further been shown that this problem cannot be solved by representing the atmosphere on a “sufficiently fine” grid with zero representation error, because the estimate of the atmospheric state does not converge to a useful value when the grid approaches an infinitesimally fine grid. This is because the quantities used to characterize the atmosphere in a statistical sense (mixing ratio, concentration, temperature) are not defined at infinitesimal resolution, which would require characterization of extensionless points.

7.2 Practical conclusions

The problem of the smoothing error referring to a finite sampling of the atmospheric state could be considered purely philosophical and practically irrelevant, and the “smoothing error” could be treated as a theoretical term without direct correspondence to the empirical world (e.g. Carnap, 1966, 1974), if the consequence of this problem were not that the quantity called smoothing error does not, contrary to the other retrieval error components, comply with generalized

Gaussian error propagation. While the smoothing error is a valuable and intuitive retrieval diagnostic in its own right, the problems encountered in the context of error propagation cause major reservations against the smoothing error concept in the context of error budget and imply that the quantity calculated according to Eq. (5) thus should not be called an “error” in terms of error propagation. While, if calculated correctly, a smoothing difference of two profiles is still a useful quantity, the inclusion of the smoothing error in the error budget of a retrieval will cause confusion and will lead to inadequate operations by data users. The use of the smoothing error should be restricted to applications where it can safely be excluded that anybody would propagate this quantity to finer grids. An option could be to replace the term “smoothing error” with another term which does not suggest applicability of Gaussian error propagation; for example, one could use the term “constraint diagnostic” instead.

When the quality of two data products is compared, it is important to evaluate the smoothing error on the same grid. Otherwise the retrieval evaluated on the finer grid will erroneously appear to be more affected by the smoothing error than that evaluated on the coarser grid.

A useful and safe way to communicate the smoothing characteristics of the retrieval is to provide the averaging kernel along with the data (Rodgers, 2000, Sect. 11.2.6). While the interpolation of the averaging kernel matrix to finer grids also introduces inaccuracies, these seem to be tolerably small if the original averaging kernel has been evaluated on a grid fine enough that all the scales that the instrument is sensitive to are resolved (weak gridding criterion). A lot of applications of the averaging kernel demonstrate its diagnostic power, and solutions to specific related problems are proposed (e.g. Connor et al., 2008; Stiller et al., 2012; Worden et al., 2013; Neu et al., 2014; Eckert et al., 2014). If for some debatable reason the smoothing error still is to be supplied as part of the error budget, then, at the very least, the native grid on which this error has been evaluated needs to be presented along with the error estimate, and a caveat is needed to warn the data user about the smoothing error pitfalls.

Appendix A: The approximate validity of the smoothing difference

We start from Eq. (20) but, for simplicity, ignore the noise terms which are not relevant here. We formulate Eq. (20) on a grid which is finer than the comparison grid:

$$\mathbf{S}_{\delta, \text{fine}} = (\mathbf{A}_{1, \text{fine}} - \mathbf{A}_{2, \text{fine}}) \mathbf{S}_{c, \text{fine}} \times (\mathbf{A}_{1, \text{fine}} - \mathbf{A}_{2, \text{fine}})^T. \quad (\text{A1})$$

For the fine-grid averaging kernels, the approximation $\mathbf{A}_{\text{fine}} \approx \mathbf{W} \mathbf{A}_{\text{coarse}} \mathbf{V}$ is used: we get

$$\begin{aligned} \mathbf{S}_{\delta, \text{fine}} &\approx (\mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{W} \mathbf{A}_2 \mathbf{V}) \mathbf{S}_{c, \text{fine}} \\ &\times (\mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{W} \mathbf{A}_2 \mathbf{V})^T. \end{aligned} \quad (\text{A2})$$

We are interested to get this on the comparison grid:

$$\begin{aligned} \mathbf{S}_{\delta} &\approx \mathbf{V} (\mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{W} \mathbf{A}_2 \mathbf{V}) \mathbf{S}_{c, \text{fine}} \\ &\times (\mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{W} \mathbf{A}_2 \mathbf{V})^T \mathbf{V}^T \\ &= (\mathbf{V} \mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{V} \mathbf{W} \mathbf{A}_2 \mathbf{V}) \mathbf{S}_{c, \text{fine}} \\ &\times (\mathbf{V} \mathbf{W} \mathbf{A}_1 \mathbf{V} - \mathbf{V} \mathbf{W} \mathbf{A}_2 \mathbf{V})^T \\ &= (\mathbf{A}_1 \mathbf{V} - \mathbf{A}_2 \mathbf{V})^T \mathbf{S}_{c, \text{fine}} (\mathbf{A}_1 \mathbf{V} - \mathbf{A}_2 \mathbf{V}) \\ &= (\mathbf{A}_1 - \mathbf{A}_2) \mathbf{V} \mathbf{S}_{c, \text{fine}} \mathbf{V}^T (\mathbf{A}_1 - \mathbf{A}_2) \\ &= (\mathbf{A}_1 - \mathbf{A}_2) \mathbf{S}_c (\mathbf{A}_1 - \mathbf{A}_2). \end{aligned} \quad (\text{A3})$$

This proves that, except for the additional information contained in the fine-grid averaging kernel, nothing is lost when going from the fine to the coarse grid. The approximation of the averaging kernel used here is considered justifiable if it has been evaluated on a grid fine enough to resolve all the information that the respective instrument is sensitive to (weak gridding criterion).

Acknowledgements. The author thanks T. Shepherd for “interrupting his dogmatic slumber” related to established diagnostic tools and their justification, as well as A. Butz and G. Stiller for helpful discussions. The problem discussed in this paper popped up during the activities SPARC Data Initiative led by M. Hegglin and S. Tegtmeier and was further investigated in the context of the ESA Climate Change Initiative (Ozone CCI) under contract number 4000101683/10/I-NB. The author thanks the eight reviewers, among them B. J. Connor, L. Hoffmann, C. D. Rodgers, J. Tamminen and J. Worden, for an interesting and thorough discussion of the manuscript and numerous helpful comments, which have led to substantial improvements in the paper. The support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of Karlsruhe Institute of Technology are acknowledged.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: P. K. Bhartia

References

- 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 Trans. Geosci. Remote*, 44, 1297–1307, 2006.
- Carlotti, M.: Global-fit approach to the analysis of limb-scanning atmospheric measurements, *Appl. Optics*, 27, 3250–3254, 1988.
- Carnap, R.: Philosophical foundations of physics: An introduction to the philosophy of science, Basic books Inc., New York, 1966.
- Carnap, R.: An introduction to the philosophy of science, Basic books Inc., New York, edited by: Gardner, M., 1974.
- Connor, B. J., Siskind, D. E., Tsou, J. J., Parrish, A., and Remsberg, E. E.: Ground-based microwave observations of ozone in the upper stratosphere and mesosphere, *J. Geophys. Res.*, 99, 16757–16770, doi:10.1029/94JD01153, 1994.
- Connor, B. J., Bösch, H., Toon, G., Sen, B., Miller, C., and Crisp, D.: Orbiting Carbon Observatory: Inverse method and prospective error analysis, *J. Geophys. Res.*, 113, D05305, doi:10.1029/2006JD008336, 2008.
- Eckert, E., von Clarmann, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Leblanc, T., McDermid, S., Pastel, M., Steinbrecht, W., Swart, D. P. J., Walker, K. A., and Bernath, P. F.: Drift-corrected trends and periodic variations in MIPAS-IMK/IAA ozone measurements, *Atmos. Chem. Phys.*, 14, 2571–2589, doi:10.5194/acp-14-2571-2014, 2014.
- Haario, H., Laine, M., Lehtinen, M., Saksman, E., and Tamminen, J.: Markov chain Monte Carlo methods for high dimensional inversion in remote sensing, *J. Roy. Stat. Soc. B*, 66, 591–607, 2004.
- Hegglin, M. I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D., Remsberg, E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, R., and Weigel, K.: SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, 118, 11824–11846, doi:10.1002/jgrd.50752, 2013.
- Kramarova, N. A., Bhartia, P. K., Frith, S. M., McPeters, R. D., and Stolarski, R. S.: Interpreting SBUV smoothing errors: an example using the quasi-biennial oscillation, *Atmos. Meas. Tech.*, 6, 2089–2099, doi:10.5194/amt-6-2089-2013, 2013.
- Neu, J. L., Hegglin, M. I., Tegtmeier, S., Bourassa, A., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Rozanov, A., Toohey, M., von Clarmann, T., Walker, K. A., and Worden, J. R.: The SPARC Data Initiative: Comparison of upper troposphere/lower stratosphere ozone climatologies from limb-viewing instruments and the nadir-viewing Tropospheric Emission Spectrometer, *J. Geophys. Res.-Atmos.*, 119, 6971–6990, doi:10.1002/2013JD020822, 2014.
- Phillips, D.: A Technique for the numerical solution of certain integral equations of first kind, *J. Ass. Comput. Mat.*, 9, 84–97, 1962.
- Richardson, L. F.: The supply of energy from and to atmospheric eddies, *Proc. Roy. Soc. London A*, 87, 354–373, (quoted after R. Benzi, “Lewis Fry Richardson” in: *A Voyage Through Turbulence*, edited by: Davidson, P. A., Kanedea, Y., Moffat, K., and Screenivasan, K. R., Cambridge University Press, 187–208, 2011), 1920.
- Rodgers, C. D.: Retrieval of Atmospheric Temperature and Composition From Remote Measurements of Thermal Radiation, *Rev. Geophys. Space Phys.*, 14, 609–624, 1976.
- Rodgers, C. D.: Characterization and error analysis of profiles retrieved from remote sounding measurements, *J. Geophys. Res.*, 95, 5587–5595, 1990.
- 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.: Interactive comment on “Smoothing error pitfall” by T. von Clarmann, *Atmos. Meas. Tech. Discuss.*, 7, C967–C972, 2014.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 4116, doi:10.1029/2002JD002299, 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, 16037–16055, 1997.
- Sofieva, V. F., Ralpoe, N., Tamminen, J., Kyrölä, E., Kalakoski, N., Weber, M., Rozanov, A., von Savigny, C., Laeng, A., von Clarmann, T., Stiller, G., Lossow, S., Degenstein, D., Bourassa, A., Adams, C., Roth, C., Lloyd, N., Bernath, P., Hargreaves, R. J., Urban, J., Murtagh, D., Hauchecorne, A., Dalaudier, F., van Roozendaal, M., Kalb, N., and Zehner, C.: Harmonized dataset of ozone profiles from satellite limb and occultation measurements, *Earth Syst. Sci. Data*, 5, 349–363, doi:10.5194/essd-5-349-2013, 2013.
- Steck, T. and von Clarmann, T.: Constrained profile retrieval applied to the observation mode of the Michelson Interferometer for Passive Atmospheric Sounding, *Appl. Optics*, 40, 3559–3571, 2001.
- Stiller, G. P., Kiefer, M., Eckert, E., von Clarmann, T., Kellmann, S., García-Comas, M., Funke, B., Leblanc, T., Fetzer, E., Froidevaux, L., Gomez, M., Hall, E., Hurst, D., Jordan, A., Kämpfer, N., Lambert, A., McDermid, I. S., McGee, T., Miloshevich, L., Nedoluha, G., Read, W., Schneider, M., Schwartz, M., Straub, C.,

- Toon, G., Twigg, L. W., Walker, K., and Whiteman, D. N.: Validation of MIPAS IMK/IAA temperature, water vapor, and ozone profiles with MOHAVE-2009 campaign measurements, *Atmos. Meas. Tech.*, 5, 289–320, doi:10.5194/amt-5-289-2012, 2012.
- Tegtmeier, S., Hegglin, M. I., Anderson, J., Bourassa, A., Brohede, S., Degenstein, D., Froidevaux, L., Fuller, R., Funke, B., Gille, J., Jones, A., Krüger, Y. K. K., Kyrölä, E., Lingenfeller, G., Lumpe, J., Nardi, B., Neu, J., Pendlebury, D., Remsberg, E., Rozanov, A., Smith, L., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., and Wang, R. H. J.: SPARC Data Initiative: A comparison of ozone climatologies from international satellite limb sounders, *J. Geophys. Res.-Atmos.*, 118, 12229–12247, doi:10.1002/2013JD019877, 2013.
- Tikhonov, A.: On the solution of incorrectly stated problems and method of regularization, *Dokl. Akad. Nauk. SSSR*, 151, 501–504, 1963a.
- Tikhonov, A.: On the regularization of incorrectly stated problems, *Dokl. Akad. Nauk. SSSR*, 153, 49–52, 1963b.
- Twomey, S.: On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature, *J. ACM*, 10, 97–101, 1963.
- von Clarmann, T. and Grabowski, U.: Elimination of hidden a priori information from remotely sensed profile data, *Atmos. Chem. Phys.*, 7, 397–408, doi:10.5194/acp-7-397-2007, 2007.
- 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, 2003.
- Worden, J., Kulawik, S. S., Shephard, M. W., Clough, S. A., Worden, H., Bowman, K., and Goldman, A.: Predicted errors of tropospheric emission spectrometer nadir retrievals from spectral window selection, *J. Geophys. Res.*, 109, D09308, doi:10.1029/2004JD004522, 2004.
- Worden, J., Wecht, K., Frankenberg, C., Alvarado, M., Bowman, K., Kort, E., Kulawik, S., Lee, M., Payne, V., and Worden, H.: CH₄ and CO distributions over tropical fires during October 2006 as observed by the Aura TES satellite instrument and modeled by GEOS-Chem, *Atmos. Chem. Phys.*, 13, 3679–3692, doi:10.5194/acp-13-3679-2013, 2013.