Atmos. Meas. Tech. Discuss., 7, 3301–3319, 2014
www.atmos-meas-tech-discuss.net/7/3301/2014/
doi:10.5194/amtd-7-3301-2014

Atmospheric Measurement Techniques
Open Access
Discussions

Smoothing error pitfalls

T. von Clarmann

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

Received: 28 January 2014 – Accepted: 22 March 2014 – Published: 1 April 2014

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The difference due to the content of a priori information between a constrained retrieval and the true atmospheric state is usually represented by the so-called smoothing error. In this paper it is shown that the concept of the smoothing error 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. The idea of a sufficiently fine sampling of this reference atmospheric state is untenable 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 talk about temperature or concentration. Smoothing differences, however, which play a role when measurements are compared, are still a useful quantity if the involved a priori covariance matrix has been evaluated on the comparison grid rather than resulting from interpolation. This is, 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.

1 Introduction

Often the analysis of remotely sensed data of the atmosphere 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 chosen by the retrieval scientist. A variety of regularization techniques have been proposed to solve such kind of inverse problems, among them regularization methods by Tikhonov

(1963a), Twomey (1963) and Phillips (1962) as well as the optimal estimation scheme, which has been systematically investigated by Rodgers (1976) and which has been retitled maximum a posteriori retrieval by Rodgers (2000). 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, i.e. 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 knowledge. Indirect a priori knowledge, or indirect constraints, can be applied e.g. by just using a finite more or less coarse grid for representation of the atmospheric state and some interpolation rule for determination of the atmospheric state between the gridpoints, 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.

With a grid coarse enough often maximum likelihood retrievals are possible which do not require any formal constraint or a priori information. While the effect of finite resolution is self-evident in the latter case, because nobody reasonably expects the resolution of an, e.g., vertical profile be better than the grid on which it is represented, regularized retrievals lead to oversampled profiles, i.e. there are more altitude gridpoints 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 of

AMTD
7, 3301–3319, 2014

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 the so-called smoothing error has a particular characteristics which makes the related concept questionable. 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 an (admittedly: failed) 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, where, despite all criticism, a concept closely related to the smoothing error concept still is 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). Minimization of a two-component cost function c

$$c = (\boldsymbol{y} - F(\boldsymbol{x}))^{\mathrm{T}} \mathbf{S}_y^{-1} (\boldsymbol{y} - F(\boldsymbol{x})) + (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{a}})^{\mathrm{T}} \mathbf{R} (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{a}}) \tag{1}$$

where $\boldsymbol{y}$ is the m-dimensional vector of measurements,
F the $\mathbb{R}^n \rightarrow \mathbb{R}^m$ signal transfer forward model,
$\boldsymbol{x}$ the n-dimensional vector of the unknown components of the atmospheric state,

AMTD
7, 3301–3319, 2014

$\mathbf{S}_y$ the $m \times m$ measurement error covariance matrix,
$\boldsymbol{x}_\text{a}$ the n-dimensional a priori information on the atmospheric state, and
$\mathbf{R}$ an $n \times n$ regularization matrix,

leads, after linear replacement to $F(\boldsymbol{x})$ by $\boldsymbol{x}_\text{a} + \mathbf{K}(\boldsymbol{x} - \boldsymbol{x}_\text{a})$ where $\mathbf{K}$ is the Jacobian matrix with elements $k_{i,j} = \partial y_i / \partial x_j$ to the following retrieval equation:

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

where the $\hat{}$ 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: $\mathbf{R} = \mathbf{S}_\text{a}^{-1}$ where $\mathbf{S}_\text{a}$ is the a priori covariance matrix leads to a maximum a posteriori retrieval (Rodgers, 2000) while squared and scaled kth order finite difference matrices have been suggested by Phillips (1962), Tikhonov (1963a, b), or Twomey (1963) and which have systematically been investigated for remote sensing applications by, e.g., Steck and von Clarmann (2001). Nonlinear variants of this approach 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 $n \times n$

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

With this we can rewrite Eq. (2) as

$$\hat{\boldsymbol{x}} = \mathbf{A}\boldsymbol{x} + (\mathbf{I} - \mathbf{A})\boldsymbol{x}_\text{a}, \tag{4}$$

AMTD
7, 3301–3319, 2014

where $\mathbf{I}$ is the $n \times n$ identity matrix. Rodgers (1990, 2000) suggests to apply generalized Gaussian error propagation (c.f. Sect. 3) to estimate the smoothing error, which is the mapping of the expected deviation of $\boldsymbol{x}_\mathrm{a}$ from the actual $\boldsymbol{x}$:

$$\mathbf{S}_\mathrm{smoothing} = (\mathbf{I} - \mathbf{A})\mathbf{S}_\mathrm{a}(\mathbf{I} - \mathbf{A})^\mathrm{T} \quad (5)$$

This holds only if indeed $\boldsymbol{x}_\mathrm{a} =< \boldsymbol{x} >$, where $<>$ denotes the expectation value. More precisely, it is required that $\mathbf{S}_\mathrm{a}$ represents the covariance around $< \boldsymbol{x} >$, and not the covariance around $\boldsymbol{x}_\mathrm{a}$ if the latter happens not to be chosen to equal $< \boldsymbol{x} >$, or around any other arbitrarily chosen a priori covariance matrix. 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 evaluate the smoothing error around the correct expectation value of the atmospheric state is outlined, e.g., 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 in consequence the so-called smoothing error has to be modified by adding the term

$$(\mathbf{I} - \mathbf{A})(\boldsymbol{x}_\mathrm{a} - < \boldsymbol{x} >)(\boldsymbol{x}_\mathrm{a} - < \mathbf{x} >)^\mathrm{T}(\mathbf{I} - \mathbf{A})^\mathrm{T}, \quad (6)$$

which accounts for the bias of $\boldsymbol{x}_\mathrm{a}$.

Further, it is important, that the $\mathbf{S}_\mathrm{a}$ matrix includes atmospheric variability on all scales which can be represented on the grid on which it is evaluated. $\mathbf{S}_\mathrm{a}$ 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 $\mathbf{S}_\mathrm{a}$ 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.

By the way, 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 a mixture of the a priori profile and the profile the unregularized retrieval would tend towards. There is no reason that the profile obtained by means of Eq. (2) should always be smoother than the true profile.

3 Error Propagation

3.1 General linear or moderately linear case

Let

$$\boldsymbol{v} = f(\boldsymbol{u}) \tag{7}$$

for any real vectorial argument $\boldsymbol{u}$ and any real vectorial result $\boldsymbol{v}$ (note that, while in the empirical world restriction to vectors of rational numbers would be adequate because transcendent values usually cannot be measured, the concept of derivatives requires the space of real numbers). The uncertainties of $\boldsymbol{u}$ map onto the uncertainties of $\boldsymbol{v}$ as

$$\mathbf{S}_v \approx \mathbf{K}\mathbf{S}_u\mathbf{K}^{\mathrm{T}} \tag{8}$$

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

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

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 $\mathbf{S}_u$. This error propagation rules are generally accepted for all cases except grossly nonlinear functions $f(\boldsymbol{u})$.

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

$$\mathbf{S}_{\text{noise}} \approx \mathbf{G}\mathbf{S}_y\mathbf{G}^{\mathrm{T}}. \tag{10}$$

AMTD
7, 3301–3319, 2014

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 (c.f. Rodgers, 2000, p. 162) by

$$\hat{\boldsymbol{x}}_{\text{fine}} = \mathbf{W}\hat{\boldsymbol{x}}_{\text{coarse}} \tag{11}$$

of which a possible inverse operation is

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

Here, $\hat{\boldsymbol{x}}_{\text{coarse}}$ and $\hat{\boldsymbol{x}}_{\text{fine}}$ are of dimension n and $\tilde{n}$, and $\mathbf{W}$ and $\mathbf{V}$ are $\tilde{n} \times n$ and $n \times \tilde{n}$ dimensional transformation matrices, respectively.

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

$$\mathbf{S}_{\text{noise, fine}} = \mathbf{W}\mathbf{S}_{\text{noise, coarse}}\mathbf{W}^{\mathrm{T}} \tag{13}$$

and from the fine grid to the coarse grid as

$$\mathbf{S}_{\text{noise, coarse}} = \mathbf{V}\mathbf{S}_{\text{noise, fine}}\mathbf{V}^{\mathrm{T}}. \tag{14}$$

The same equations apply to the propagation of the parameter error.

As already mentioned by Rodgers (2000, p. 163), the a priori covariance matrix $\mathbf{S}_{\mathrm{a}}$ 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 seems, however, to have remained unnoticed that as a direct consequence of this, also the so-called smoothing error cannot be interpolated from its native grid to any finer grid. The smoothing error of $\hat{\boldsymbol{x}}$ represents smoothing error components only with respect to variability which can be represented on the retrieval grid, viz. the grid of $\mathbf{S}_{\mathrm{a}}$.

AMTD
7, 3301–3319, 2014

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 so-called smoothing error. Even for linear functions $f(\boldsymbol{x})$, 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 data bases are created were 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; Tegtmeier et al., 2013).

While Gaussian error propagation of the quantity called smoothing error would give

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

the correct linear estimate is

$$\mathbf{S}_{\text{smoothing, fine}} = (\mathbf{I}_{\text{fine}} - \mathbf{WAV})\mathbf{S}_{\text{a, fine}}(\mathbf{I}_{\text{fine}} - \mathbf{WAV})^{\mathrm{T}}, \tag{16}$$

which cannot be inferred via Eq. (8) from $\mathbf{S}_{\text{smoothing, coarse}}$. Here, $\mathbf{S}_{\text{a, fine}}$ is the a priori covariance matrix evaluated on the fine grid and including small-scale variability which cannot be represented on the coarse grid, and $\mathbf{I}_{\text{coarse}}$ and $\mathbf{I}_{\text{fine}}$ are the identity matrices on the respective grids.

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 gridwidths of the fine and the coarse grids have been chosen 1 km 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, with halfwidths of 6 km, where the sum over the averaging kernel elements was unity (bottom left panel in Fig. 1). The a priori covariance matrix $\mathbf{S}_{\text{a, fine}}$ was constructed with diagonal values of one (in arbitrary units), and exponentially decreasing all positive off-diagonal values, where the correlation length was set

AMTD
7, 3301–3319, 2014

Smoothing error pitfalls

T. von Clarmann

1 km (upper left panel in Fig. 1). Construction of $\mathbf{S}_{\text{a, coarse}}$ relies on the $\mathbf{V}$ matrix (upper right panel in Fig. 1). Averaging kernels and errors were chosen altitude-independent. The resulting smoothing error on the coarse grid is, in terms of variances, 0.33, and the covariances between adjacent profile points are as negative as −0.23 (dark blue curve in the lower right panel in Fig. 1, hardly discernable because overplotted by the central red curve). This anticorrelation is intuitive because smoothing means that if, e.g., 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 gridpoints of the fine grid which are also part of the coarse grid, but at interjacent gridpoints the propagated smoothing error variances are calculated to be as low as 0.08 (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 the smoothing error. The direct evaluation on the fine grid via Eq. (16) gives smoothing error variances of 0.55 at the points belonging to both grids and 0.64 for the interjacent gridpoints (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 $\mathbf{S}_{\text{a, fine}}$, the smoothing errors decrease but the contrast between both ways to estimate it on the fine grid remains large.

So 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, how the smoothing error concept can be modified such that it becomes compatible with established error propagation laws, which will be tried in the next section, or 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 the source of the problem understood, it seems to suggest itself 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 at least one could 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 shall be diagnosed on a grid on which the full variability of the atmosphere can be represented. This approach is based on the assumption that the residual smoothing error not accounted for on a finite grid converges towards zero for finer grids, once the grid is fine enough. In the following it will be shown that this assumption is false.

For a single extensionless point in the atmosphere, the mixing ratio of a species is not a meaningful quantity: either, at the given point, there is a target molecule; then the mixing ratio is one. Or there is a molecule of another species, then the mixing ratio is zero. Or there is no molecule at all. Then the mixing ratio is fully undefined because this would involve division by zero. For number densities and temperature, there are similar problems to define these quantities in any meaningful manner for an infinitesimal point. Admittedly, the scales discussed here are of no concern in remote sensing. However, it is not intended 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. 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 of the atmospheric state not represented by $\mathbf{S}_\mathrm{a}$, until we reach the molecular scale on which the pathological cases discussed above occur.

AMTD
7, 3301–3319, 2014

In conclusion, the attempt to solve propagation problem of the smoothing error by a grid fine enough that it is guaranteed that interpolation will never occur must be considered as failed.

5 The way out of the dilemma

Since generalized Gaussian error propagation is one of the most essential principles of linear theory, it seems not acceptable 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 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 "true actual value at an extensionless point". Already Rodgers (2000, p. 48) mentions the alternative to understand a measurement of the atmospheric state to characterize an extended air volume and to characterize the measurement by its 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, this approach is not only an option but 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 contradicts itself, because the evaluation of the smoothing error on a finite grid gives the notion of the retrieval characterizing a finite air volume access through the back door again, i.e., it breaks with its own assumption that the "smoothing error" represents the smoothing component of the retrieval error in absolute terms.

Once having accepted the failure of the smoothing error concept as a tool to characterize the 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

AMTD
7, 3301–3319, 2014

further advantages: first, the estimate of the error budget for any retrieval involving a given **R** (which may or may not be an approximation to $\mathbf{S}_a^{-1}$) no longer depends via Eq. (5) on the choice of the a priori covariance matrix. Often no reliable estimate of $\mathbf{S}_a$ is available, but any arbitrary choice is in conflict with the smoothing error concept (c.f. Rodgers, 2000, p. 48). Second, the averaging kernel is needed anyway for a number of applications of measured data, and to provide it instead of the smoothing error is advantageous for the data user. And 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 is that the usual estimate of the retrieval error covariance matrix 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

$$\mathbf{S}_x = \left(\mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1}. \tag{17}$$

This covariance matrix contains both the measurement noise and the smoothing error component (c.f. Rodgers, 2000, p. 58). Thus, all caveats discussed for the smoothing error apply equally to the error estimate of Eq. (17). 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.

By the way, Eq. (17) is, regardless of the notion with respect to the smoothing error concept, inapplicable to any choice of the $\mathbf{S}_a$ matrix except the true climatological a priori covariance matrix. While reasonable retrievals can be performed with ad hoc choices of $\mathbf{S}_a$ or with replacement of its inverse by other regularization matrices, Eq. (17) does not provide a valid error estimate in these cases. The inadequacy of an

AMTD
7, 3301–3319, 2014

Smoothing error pitfalls

T. von Clarmann

ad hoc choice of $\mathbf{S}_a$ already highlighted by (Rodgers, 2000, p. 48) also turns Eq. (17) inadequate for all choices of $\mathbf{S}_a$ except the true covariance of the atmospheric state under investigation.

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 usefull and powerful tool, is comparison of remotely sensed data according to Rodgers and Connor (2003, their Eqs. 10–14). These authors suggest in their paper to validate profiles against each other by testing if their difference $\hat{\boldsymbol{x}}_1 - \hat{\boldsymbol{x}}_2$ is significant in terms of χ^2 statistics. The covariance matrix of the difference, $\mathbf{S}_\delta$, needed for this test, however, must not include interdependent components of the smoothing error. Thus, these authors suggest to calculate $\mathbf{S}_\delta$ as

$$\mathbf{S}_\delta = (\mathbf{A}_1 - \mathbf{A}_2)^\mathrm{T} \mathbf{S}_\mathrm{c} (\mathbf{A}_1 - \mathbf{A}_2) + \mathbf{S}_{x1} + \mathbf{S}_{x2}, \tag{18}$$

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 $\mathbf{S}_\mathrm{c}$ is the comparison ensemble covariance matrix. The first term of 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 if the difference between the retrievals is different or if 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 but 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 shall generally be at least as fine as the parent grids. When the difference $\hat{\boldsymbol{x}}_1 - \hat{\boldsymbol{x}}_2$ 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 $\mathbf{S}_\mathrm{c}$

has been evaluated on the intercomparison grid or any grid finer than that but is not a result of interpolation. That 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. The a priori covariance matrix cannot simply be interpolated to the comparison grid, for reasons discussed in Sect. 3.2.

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.

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 may be worthwhile trying but are beyond the scope of this discussion.

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 implicit smoothing error. It has further been shown that this problem cannot be solved by representing the atmosphere on a "sufficiently fine" grid, because the estimate of the atmospheric state does not converge to a useful value when the grid

AMTD
7, 3301–3319, 2014

approaches an infinitesimally fine grid. This is because the quantities used to characterize the atmosphere (mixing ratio, concentration, temperature) are not defined for extensionless points.

This problem 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), wouldn't the consequence of this problem be, that the quantity called "smoothing error" does, contrary to the other retrieval error components, not comply with generalized Gaussian error propagation. This fact causes major reservations against the smoothing error concept and implies that the quantity calculated according 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 so-called smoothing error in the error budget of a retrieval will cause confusion and will lead to inadequate operations by data users.

A useful and safe way to communicate the smoothing characteristics of the retrieval is to provide the averaging kernel along with the data. If for some debatable reason the smoothing error still is to be supplied, 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.

Acknowledgements. The author thanks T. Shepherd for "interrupting his dogmatic slumber" related to established diagnostic tools and their justification, and A. Butz as well as Gabriele Stiller for helpful discussions. The problem discussed in this paper popped up during the activities SPARC Data Initiative led by Michaela Hegglin and Susann Tegtmeier and was further investigated in the context of the ESA Climate Change Initiative (Ozone CCI) under contract number 4000101683/10/I-NB.

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

AMTD
7, 3301–3319, 2014

References

Carnap, R.: Philosophical Foundations of Physics: an Introduction to the Philosophy of Science, Basic Books Inc., New York, 1966. 3316

Carnap, R.: An Introduction to the Philosophy of Science, edited by: Gardner, M., Basic Books Inc., New York, 1974. 3316

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. 3309

Phillips, D.: A Technique for the numerical solution of certain integral equations of first kind, J. Ass. Comput. Mat., 9, 84–97, 1962. 3303, 3305

Rodgers, C. D.: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys. Space Ge., 14, 609–624, 1976. 3303, 3313

Rodgers, C. D.: Characterization and error analysis of profiles retrieved from remote sounding measurements, J. Geophys. Res., 95, 5587–5595, 1990. 3304, 3306

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, Series on Atmospheric, Oceanic and Planetary Physics, Vol. 2, edited by: Taylor, F. W., World Scientific, 2000. 3303, 3304, 3305, 3306, 3308, 3312, 3313, 3314

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 4116, doi:10.1029/2002JD002299, 2003. 3314, 3315

Sofieva, V. F., Rahpoe, 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 Roozendael, 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. 3309

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. 3305

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., Lingenfelser, 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. 3309

Tikhonov, A.: On the solution of incorrectly stated problems and method of regularization, Dokl. Akad. Nauk. SSSR+, 151, 501–504, 1963a. 3302, 3305

Tikhonov, A.: On the regularization of incorrectly stated problems, Dokl. Akad. Nauk. SSSR+, 153, 49–52, 1963b. 3305

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. 3303, 3305

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. 3306

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. 3304

AMTD

7, 3301–3319, 2014

Smoothing error pitfalls

T. von Clarmann

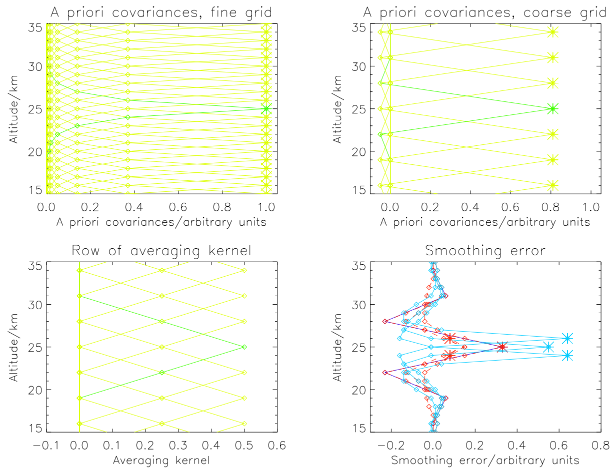

Fig. 1. Case study: the upper left panel shows the a priori covariances on the fine grid (gridwidth 1 km). Only the symbols are significant, the lines are only plotted to guide the eye. The large asterisks are the variances. The variance and covariances referring to 25 km are highlighted for clarity. The top right panel shows the covariances on the coarse (gridwidth 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 important, the directly estimated smoothing errors are considerably larger. This is because the relevant a priori covariance matrix contains larger atmospheric variability (c.f. 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.

AMTD
7, 3301–3319, 2014

Smoothing error pitfalls

T. von Clarmann