**Atmospheric
Measurement
Techniques**
Open Access

Discussions

Interactive
Comment

Interactive comment on "Smoothing error pitfalls" _by_ T. von Clarmann

T. von Clarmann

thomas.clarmann@kit.edu

Received and published: 25 April 2014

I thank the reviewer for spotting several issues which remained unclear in my manuscript. In the following I take the opportunity to reply to some of the comments. Here I have re-ordered some of the comments when it helped to better structure my reply. A full point-by-point reply will be given at the end of the discussion phase.

Comment: _Clarification is also needed that the topic has been addressed in the literature before, in particular by Rodgers (2000)._

Reply: I do quite not understand this comment. In my paper I make very detailed reference to Rodgers (2000). All the passages of the Rodgers book quoted in the review are already referenced in my paper, with explicit page reference. I agree (and have mentioned in my paper!) that Rodgers discusses non-interpolability of the S_a

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

matter but I am not aware that the consequence of this for the smoothing error has been discussed anywhere in the literature.

Comment: *In section 2 the paper presents a misleading definition and discussion of the smoothing error. In particular, Eq. (5) defines the smoothing error as*

$$S_{smooth} = (I - A)S_a(I - A)^T.$$

However, Rodgers (2000, Eq. 3.17) introduced this as

$$S_{smooth} = (I - A)S_e(I - A)^T, with$$

S_e *being the covariance of a real ensemble of states rather than the a priori covariance* S_a...

Reply: Thank you for spotting this. I should indeed have used another notation and will correct this in the revision. However, in Rodgers' Equation 2.23 (p25), S_a is defined as the true variability of the atmospheric state around the a priori assumption. In contrast, in Rodgers [p.49] the possible use of ad hoc choices of the S_a matrix is discussed. In Rodgers [p.58] again the smoothing error is evaluated on the basis of S_a. Anyway, the notation with S_e is much clearer, and in my paper I can simply replace S_a bt S_e because I do not use the regularization term S_a^{-1}. In my notation, I use R for any generic or ad hoc regularization matrix, so there is no ambiguity.

Comment: *... It seems S_e rather than S_a was introduced intentionally by Rodgers (2000), with the following argument (p. 49): "Many remote observing systems cannot see spatial fine structure, the loss of which contributes to the smoothing error. To estimate it correctly, the actual statistics of the fine structure must be known. It is not enough to simply use some ad hoc matrix that has been constructed as a reasonable a priori constraint in the retrieval...*

Reply: I fully agree. This comment is completely in line with my paper: On page 3306, lines 8 ff. I discuss exactly this issue. There I also make reference to page 49 of the Rodgers book.

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Comment: *... I think this is at the heart of the problem raised by the new paper of T. von Clarmann, i.e., that a coarse grid a priori covariance S_a cannot be simply interpolated to obtain an estimate of S_e on a fine grid. This would not be a completely new finding than...*

Reply: In my paper I do not claim to have discovered the non-interpolability of the S_a or S_e matrix. On page 3308, line 16 I attribute this finding explicitly to Rodgers. The new point in my paper is, that this has important implications for the smoothing error. Once discovered, this issue seems trivial, but it has to my best knowledge never discussed before in the literature, and the effect on the error budget of a retrieval is important.

Comment: *...Rodgers (2000, p. 49) also provides a simple solution to this problem: "If the real covariance is not available, it may be better to abandon the estimation of the smoothing error, and consider the retrieval as an estimate of a smoothed version of the state, rather than an estimate of the complete state." T. von Clarmann came to a similar conclusion in his paper, i.e., to better exclude the smoothing error from the error budget and to supply averaging kernels to the data user instead. However, this would also not be a completely new finding than?*

Reply: Also here I have not claimed to have discovered the solution myself but on p 3312 line 13 of my paper I explicitly attribute this suggestion to Rodgers. In the same paragraph I then conclude that "As a result of the discussion above, this approach is not only an option but seems to be the only reasonable choice...".

Comment: *A practical problem that I see with the solution of Rodgers (2000) and the one presented in the paper is that it can become very inconvenient to distribute full averaging kernel matrices along with every retrieval, simply in terms of data size. (I am thinking of current nadir sounders, which provide measurements for millions of footprints per day, or new limb sounders, which may require tomographic retrievals that are associated with large and complex averaging kernel matrices.) Another danger I see is that error budgets excluding the smoothing error could easily be mistaken as*

'total' errors by inexperienced data users? One likely can find good arguments for both, to include or exclude the smoothing error in the error budget. I would say it is a judgement call and suggest to discuss this more balanced in the paper.

Reply: The size of the smoothing error covariance matrix is exactly the same as that of the averaging kernel matrix as long as it is evaluated on the retrieval grid. If it is evaluated on a finer grid which covers all atmospheric variability (whatever this is), then the smoothing error covariance matrix is even larger. Of course one can argue that its diagonal values already contain important information, but the same is true for the averaging kernel matrix.

The trouble with the smoothing error is a classical inconsistent triad (or antilogism) in the sense of Ladd-Franklin: We have three accepted facts: (a) Gaussian error propagation holds for linear transformations of erroneous data; (b) linear interpolation leads, according to Gaussian error propagation, to a reduction of the smoothing error; (c) actually the smoothing error should be larger at interpolated points. None of these three points seems refutable to me, but still these three points together are clearly a contradiction. This contradiction does not involve any merely practical aspect nor faulty actions by inexperienced data users but it is a conceptual contradiction in itself, so the problems found seem to be of another category than those mentioned in the review. So I do not see much room for latitude of judgement.

Further, the solution to characterize the smoothing by the averaging kernel rather than by the smoothing error has the further advantage that the data user can calculate an estimate of the smoothing characteristics according to Eq (5) on any desired grid himself. Thus the 'averaging kernel solution' is superior to the 'smoothing error solution' because the first kind of includes the latter but not vice versa.

In order to better balance the the revised paper, I will take care to restrict the attributes 'questionable' and 'untenable' to the smoothing error as an absolute estimate of the related component of difference between the measurement and the 'true atmosphere'

(whatever the latter is, see below) and to better highlight that the smoothing error concept is a valid and powerful tool in relative terms, i.e., relative to a fixed and pre-defined grid.

Comment: *Another major concern is related to the discussion of the "nature of the retrieved quantities": Section 4 first presents a theoretical solution to the error propagation problem, i.e., to simply evaluate the smoothing error on an infinitesimally fine grid. In principle, this would allow for the atmospheric variability on all spatial scales to be taken into account and the problems related to coarse grid representation would not arise. This solution is then discarded with the argument that the representation of the atmospheric state based on infinitesimal volumes of air itself is meaningless. This is illustrated in the paper in the following way: "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. [...] For number densities and temperature, there are similar problems to define these quantities in any meaningful manner for an infinitesimal point." Based on this proof by contradiction is is argued that the smoothing error concept itself failed. In the conclusion (section 7) the paper reads: "It has further been shown that this problem cannot be solved by representing the atmospheric state on a 'sufficiently fine' grid, because the estimate of the atmospheric state does not converge to a useful value when the grid approaches an infinitesimally find grid." However, the complete section 4 does not convince me at all, I am afraid. Starting with the illustrating example of the mixing ratio at a single point: a point itself is by definition infinitesimal (an "infinitesimal point" or "extensionless point" as referred to in the paper is a tautology). In contrast, a molecule has finite extent. I would agree that it does not make sense to try to define mixing ratios in this context. One would need a finite volume to do this. Likewise, a definition of thermodynamic quantities such as temperature does not make sense on the basis of individual points in space. Temperature is a macroscopic quantity that requires a canonical ensemble with a certain finite volume or system size to be defined..."*

Reply: In the mathematical sense of 'point', an extensionless point is (since Euclid of Alexandria) a tautology. I have used the attribute 'extensionless' just to make clear that I indeed mean the technical mathematical term 'point', as opposed to its physical, everyday, or figurative meaning, and to make clear which particular characteristics of a point I refer to. I do agree that my chain of arguments was sloppy here, because I did not properly distinguish between the molecular scale and the infinitesimal scale. Thanks a lot for pointing this out. I'll think about how to better formulate this issue. On the other hand: isn't my argument valid a fortiori when the problem arises already on the molecular scale, not only on the infinitesimal scale?

Comment:*... However, I think one cannot turn this around and argue that as we cannot define temperature for an infinitesimal volume (a point) the concept of temperature itself is meaningless. This contradicts basic principles of statistical physics. The same would also apply for the smoothing error problem addressed in the paper, I think?*

Here I disagree: Of course we can assign a temperature or vmr to a point (in its mathematical sense). This temperature or vmr, however, (and this is the point where I contradict) does not characterize only the point it is assigned to but a larger air volume around this point. This resolution aspect is essential in this context.

Comment: *Instead, I would think that this is a scale problem. It is certainly true that atmospheric variability occurs on all scales. However, one would really need to know how variability varies from larger to smaller scales, to judge if the smoothing error problem will be relevant on all scales. For example, I would think that "true data" or "real ensemble data" (Rodgers) with a spatial resolution on the order of a few meters would be perfectly fine to estimate smoothing errors for today's satellite remote sensing measurements? There would be no need to go for the infinitesimal limit? I think the problem discussed here arises in practice, because such perfect, high resolution data are usually not available. However, this is not a problem of the theory itself, but an issue that needs to be specifically addressed for individual problems.*

Reply: I do agree that non-availability of the fine-resolved ensemble covariance matrix adds to the problem. But even if we had these data, one satellite group might evaluate their smoothing error on a 10 m grid, the other on a 3 m grid, and again they compare apples and oranges, even in the ideal world where we have access to ensemble co-variance matrices as desired. To do it right, one again has to go back to the 'smoothing difference' which I describe (and support!) in Section 6 of my paper.

Comment: *p3306,l24-28: It is okay to point out that the term "smoothing error" may be misleading in certain situations, as the retrieval may not always provide a smoothed version of the true profile. However, the standard case is that a smooth a priori profile is combined with a more fluctuating true profile, yielding a smooth optimal estimate. So, I think the term "smoothing error" is okay and I do not see a need to refer to it as "so-called" smoothing error (e.g. p3304,l3 and in other places). Mentioning this "standard case"*

Reply: I have added the critical discussion of the term "SMOOTHING error" following a suggestion by the editor Dr Bhartia. I have used the quotes for the term "smoothing ERROR" not because the word 'smoothing' might be misleading, but because this quantity, since not propagated by generalized Gaussian error propagation, should not be called error. The main point of my paper is to show that the so-called (sorry!) smoothing error does not behave as an error should behave.

Comment: *p3307,l17-18: Rather than saying that the error propagation rules are "gen-erally accepted" for all cases except grossly non-linear functions one might say that this is actually the way how the"moderately non-linear" and "grossly non-linear" case are discerned? I thought the difference of the two is that "moderately non-linear" means "linear within the error bars" whereas "grossly non-linear" means that a linear approx-imation is not applicable for error estimation? Also, I guess you actually mean "mod-erately non-linear case" rather than "moderately linear case" in the heading of section 3.1?*

Reply: Thanks a lot for spotting the typo in the section header. I fully agree with your definition of moderately nonlinear and grossly nonlinear via applicability of linear error estimation (which again is borrowed from the Rodgers book, as far as I remember). The reason why I say that these error propagation laws are accepted is that I need this to set up the inconsistent triad (see above): In my paper the nonlinear applications are of no concern, but it is essential to show that the smoothing error is in conflict with this generally accepted law, even in the perfectly linear world.

Comment: *p3310,l20-21: The vertical correlation length of 1 km in this illustrative example seems to be rather short. I am sure it helps to demonstrate the point as it implies significant variability on smaller scales? Real vertical correlation lengths may be larger, though? (Larger values are often used in regularization matrices, I think.)*

We can simply scale the entire example by a factor of three or four, then we might have the typical resolution of a nadir sounder, and a correlation length might be more appropriate for atmospheric problems typically tackled with nadir sounders. The current example may be typical for a limb sounder trying to resolve the hygropause. The point is that it is the contrast between resolution and correlation length which defines the problem. In that sense, the example chosen is representative for more cases than it appears. It may be true that larger values are often used in regularization matrices, but as we have agreed above, it is not the choice of the regularization matrix what counts here but the true ensemble covariance matrix.

Comment: *p3312, l8-13: It would be good to repeat here or in another place that the representation of the continuously varying atmospheric state does not only depend on a fixed number of state variables on a fixed set of grid points, but also on a set of interpolation rules that define the state anywhere in-between. (Isn't that also a reasonable way out of the dilemma that the atmospheric state cannot be defined for an individual point?)*

I agree that this is worthwhile repeating but I do not think that it solves the problem

that the atmospheric state cannot be defined for an individual point. Of course we can assign a number to each point, but this number is determined also by the atmospheric state in around this point. I realize that my wording was sloppy and caused misunderstanding: I should have made clear that 'defined' in this context means to characterize independently of its vicinity. I'll try to find some better wording.

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 3301, 2014.