Replies to the comments

General Issues:
I would like to thank the reviewers for the interesting and thorough discussion and the helpful comments which have substantially improved the manuscript. The main critical issues raised by the reviewers were:

- It was suggested to make a stronger link to practical applications. This link has been made in the revised version by (a) including a series of test cases which cover a wide range of possible applications and (b) by splitting the conclusions sections into a theoretical (descriptive) and a practical (prescriptive) part. The latter includes clear practical recommendations. (c) Further the discussion of the interpolability of the averaging kernel seems to me to be of major practical relevance.
- A more balanced description has been required. This has been achieved in revised version by turning all notion-dependent statements into conditional statements, where the condition is the respective notion or definition etc.

In the following reply, the reviews are inserted in *italic face*; the reply is printed in normal face. Changes are tracked in the revised manuscript. The author's reply on the discussion page and the letter to the editor are identical.

1 Anonymous Referee #1

Comment: *Overview: This is a nice thought provoking paper that revisits and challenges established thinking on the quantification and description of error estimates for atmospheric remote sounding measurements. I enjoyed reading the paper; it is well thought out and generally well written. While the discussion is, in places, arguably a more philosophical than is typical in atmospheric science literature, the bottom line message is one the atmospheric remote sounding community would generally benefit from internalizing and acting on in the representation of their results to the wider community.*
Reply: I thank the reviewer for this encouraging comment.

Comment: *As with the other reviewers, I feel that the manuscript would benefit from some more specific examples. Such examples could serve as more concrete "case studies" to which subsequent authors describing their remote sounding datasets can "anchor" a citation. However, I understand that such additions would change the character (and increase the length) of the paper, and go beyond this authors original intent. As such, as, in my opinion, this discussion deserves publication (and where better than in AMT), I would be OK if the author decided not to go down that path.*
Reply: I have now performed a series of tests where I scan a wide range of correlation lengths and vertical resolutions. These should cover most practical applications. Results are now presented in a table and thus do not add too

much length to the paper. I don't discuss in the paper applications of the averaging kernel found in the literature in detail. Instead I have made references to existing literature.

Comment: *I have little to add to the discussion, other than some minor (mostly wording) suggestions, given below. I also appreciated reading the interactive discussion between the other reviewers and the author, and am generally satisfied by the responses given by the author to the thoughtful comments from reviewer #5.*
The writing is very clear and logical indeed. In a few places the choice of phraseology reflects the author's linguistic roots, but it is in all cases perfectly understandable. I've identified a few of these places and suggested potential amendments, but a brief look over by a copy editor would probably not hurt.
Last minute note. I wrote this review before I've had a chance to read over Clive Rodgers comments. Given the deadline is rapidly approaching, I've decided to go ahead and submit this review rather than delay things by trying to consider his points in more detail.
Minor comments:
Page 3302 Line 18: Suggest deleting comma before "because"
Reply: Agreed and deleted.

Comment: *Line 24: "... as fine as* **that** *chosen ..."*
Reply: Agreed and inserted.

Comment: *Page 3306*
Line 1: "... suggest the application of generalized ..."
Reply: Agreed and reworded according to the reviewer's suggestion.

Comment: *Page 3307 Line 15: Consider adding the point that this is more complex in the non-linear case, but that that is beyond the scope of this paper.*
Reply: I've inserted 'linear estimate' at the beginning of the paragraph to make clear that all the following is linear approximation.

Comment: *Page 3307 Lines 15-18: I wonder if, somewhere in this discussion, it might make sense to point out to the less familiar reader that "Gaussian error propagation" applies equally to random variables regardless of whether they have a "Gaussian" probability distribution function or not. It is perhaps unfortunate that the same mathematician's name is attributed to these two cornerstones of statistics, as many people's natural assumption may be that one requires the other.*
Reply: Agreed; footnote added.

Comment: *Page 3308*
Equation 12: Is there any value in discussing the significance of cases where $(\mathbf{W}^T\mathbf{W})^{-1}$ *is singular here?*
Reply: I prefer to avoid this discussion. In practical applications I have never encoutered this case except if the altitude ranges of the two representations do not match, and this would add another branch to the discussion such that the it may be hard for the reader to follow the main train of thoughts.

Comment: *Line 15: As pointed out by others, would be good to define "parameter error" here.*
Reply: Agreed; definition added.

Comment: *Page 3309*
Line 10: Consider citing equation 8 after "propagation"?
Reply: Agreed and cited.

Comment: *Line 22: "... have been chosen* **to be** *1km and ..."*
Reply: Agreed and inserted.

Comment: *Page 3310*
Line 2: "... were chosen **to be** *altitude ..."*
Reply: Agreed and inserted.

Comment: *Line 21: "both" → "the two"*
Reply: Agreed and changed.

Comment: *Line 27: "how" → "by which"*
Reply: Agreed and changed.

Comment: *Line 29: "which will be tried" → "as will be attempted"?*
Reply: agreed and reworded according to suggestion.

Comment: *Page 3311*
Line 4: "Having understood the source of the problem, the natural approach would appear to be to evaluate ..."?
Reply: Agreed and reworded according to the suggestion.

Comment: *Line 19: "... problems with the definition of these quantities ..."*
Reply: Agreed and reworded according to the suggestion.

Comment: *Line 21: would: "it is not our intent to discuss the state ..." be better?*
Reply: Agreed and reworded according to the suggestion.

Comment: *Page 3312*
Line 6: "not acceptable" → "unacceptable"?
Reply: Agreed and changed.

Comment: *Line 14: "an alternative understanding of the meaurements of the atmospheric state as characterizing an extended ..."?*
Reply: Agreed and reworded according to the suggestion.

Comment: *Lines 20-24: I have to confess to not having understood how this "back door" differs from the front one that was discussed in sections 3 and 4. Apologies if I've missed something.*
Reply: Here the front door was meant to be the notion that the truth is understood to be a representation of the atmosphere without any smoothing and that all deviation from this ideal atmosphere by finite resolution being a 'smoothing

error'. By analyzing the smoothing error on a finite grid, one breaks with the assumption of the unsmoothed true atmosphere and goes back to the notion of finite air volumes. I have reworded this part.

Comment: *Page 3313*
Line 24: "notion with respect to the soothing error concept, inapplicable to ..." → "discussion of the smoothing error in this paper, inapplicable to ..."?
Reply: Agreed and reworded according to the suggestion.

Comment: *Page 3314*
Line 1: "also turns" → "also makes"
Reply: Agreed and changed.

Comment: *Line 8: "suggest in their paper that profiles be validated against each other by ..."*
Reply: Agreed and reworded according to the suggestion

Comment: *Line 11: "... these authors suggest that* $\mathbf{S}_{n\Delta}$ *be calculated as"*
Reply: Agreed and reworded according to the suggestion

Comment: *Page 3315*
Line 18: Any citations of papers giving examples of those "retrievals without a formal constraint"?
Reply: Citation added.

Comment: *Page 3316*
Line 6: Perhaps consider having a new sentence and "However, " before "wouldn't". Also this sentence needs to end with a question mark. However, the phrasing of this thought as rhetorical question is out of character with the rest of the paper. Consider rewording?
Reply: Apparently my grammar was incorrect and misleading. It was meant 'if the consequence wasn't that...'; I have reworded this.

Comment: *Figure 1*
Consider adding a legend to the plots. Having to read through the caption to get the meaning of colors etc. is a little tedious. Minor point, consider postscript fonts rather than Hershey fonts for clarity (!p.font=0 in IDL).
Reply: Agreed but postponed because I am fighting with the deadline. I will do this along with other technical corrections once the other revisions will have been fixed.

Comment: *Caption: Suggest "gridwidth" → "grid spacing"*
Reply: Agreed and changed.

Comment: *"More important" → "More importantly"*
Reply: Agreed and corrected.

Many thanks for all these suggestions!

2 Reviews by John Worden

2.1 First Review of Paper

Comment: *Thomas,*
Please consider this as "iteration 1" on my review as I wanted to make a suggestion first based on my experience and get your response before formally going through the paper and making comments (so for now please ignore the odd ratings I gave your paper as there does not appear to be a way for me to make a comment without rating your paper)
You mentioned in the acknowledgements that you were inspired to write this paper as a result of the SPARC intercomparison activities (which I was also involved in via Jessica Neu's inter-comparison paper).
Reply: My mentioning of the SPARC Data Initiative refers to the first project meeting at ISSI, Berne, where it was discussed if and how the different content of prior information of the different climatologies should be taken into account.

Comment: *I think this paper would be quite a bit more useful if you added a few examples on why this subtle aspect of the smoothing error is important.*
Reply: A large list of further examples has been included.

Comment: *I agree that most of these issues are resolved by simply supplying an averaging kernel associated with the measurement.*
Reply: The non-interpolability of the averaging kernel adds some further complication to this issue. See reply to Clive Rodgers' review for further dertails. This issue is now discussed in the paper and numerical tests have been included.

Comment: *That said, a bound on the smoothing error is still important when encountering logarithmically large differences between a remotely sensed measurement and a model...*
Reply: It is not quite clear to me how the smoothing error concept solves the problems of the logarithmically large differences. Aren't those uncertainties which are caused by the fact that the averaging kernel is not representative due to these large differences just propagated onto the smoothing error, which then would be very inaccurate itself?

Comment: *...or comparing measurements with very different sensitivities.*
Reply: In this case, the 'smoothing error of the difference' is applicable (c.f. Section 6 of my paper).

Comment: *For example, in my recent methane papers using TES methane and CO (see Worden et al. ACP 2013), we could not directly use GEOS-Chem profiles through the stratosphere because the differences between GEOS-Chem and our a priori were orders of magnitude different in the stratosphere and these differences propagate into the altitude region of interest (in our case the troposphere) via the averaging kernel.*
Reply: In this case the model atmosphere is apparently outside of the range of values where linear diagnostics still can be applied. I see the problem, but I do not see how the smoothing error would help here since it also depends on the state-dependent averaging kernel.

Comment: *The approach we took to address this issue is to truncate at the tropopause which introduces an error equivalent to the cross-term A_strat_trop##-S_strat_strat##transpose(A_strat_trop), where A_strat_trop is the influence of stratosphere on troposphere.*
Reply: I see the point: the averaging kernel is used only at altitudes where the model atmosphere is in the range of values which is accessible to linear diagnostics. But I do not quite see how the smoothing error concept helps here.

Comment: *A similar problem occurs in the SPARC inter-comparison problem in which limb measurements are being compared to TES nadir profiles but the stratospheric limb measurements are not sensitive to the mid and lower troposphere. Using the approach above introduces an error that is too large so in that case Jessica has to scale an a priori in the troposphere so that it matches the limb estimate.*
Reply: Also here the problem is solved by a variant of the observation operator method but not by application of the smoothing error concept.

Comment: *In both of these cases an error associated with limited vertical resolution (or lack thereof) affects the comparison and hence it is useful to at least bound the smoothing error to determine if the estimates are consistent to within the observation error plus this component of the smoothing error.*
Reply: For the examples you mention, I think that it is not the 'smoothing error' in absolute terms (viz. the expected a priori induced difference between the true atmospheric state and the retrieval) what you actually need but the 'smoothing difference' as discussed in Section 6 of my paper. In my paper I do not criticize the latter application but I confirm that it is a powerful tool for profile intercomparison.
The smoothing error is discussed in the methodical framework of linear theory, and its refutation by a reductio ad absurdum takes place within the framework of linear theory. I do agree that there are issues with the averaging kernel which are not trivial to solve. Most of these problems, however, have to do with the fact that linear methods are pushed towards (or beyond) the limits of linear theory. Log retrievals or prior information orders of magnitude off the measured state raise problems due to the state dependence of the Jacobian K (otherwise it would be easy to transform the retrieval to another a priori). A thorough discussion of these issues, however, would force me to replace the framework of linear theory (in the narrow sense of this term) by a theoretical framework allowing state-dependent Jacobians where the linearization point has to be considered. I do not think that it is adequate to add that additional complication to the discussion since all this is not needed for the title issue of the paper. Workarounds to cope with such problems have already been developed and reported in the literature, and I do not think I should review these here. Instead, I have just added some related references.

Comment: *For these reasons, I would recommend that you augment your paper with some practical examples of this nature so that the paper becomes more useful for the remote sensing community.*
Reply: In the revised version I make reference to applications of the averaging kernel concept in the conclusions section.

2.2 Formal Review of Paper

Comment: *This response represents my "formal" review of the paper. In general the paper is well written and easy to understand and has a point that has not yet been formally stated in the literature, i.e., that the smoothing error cannot directly be linearly related (or mapped) from one gridding to another. In that respect, I can say that the paper is publishable with some minor changes (e.g., add references to the Bowman et al. 2006 and Worden et al. 2004 papers where we discuss the effects of mapping on the calculated errors)*
Reply: The References have been included in the revised version.

Comment: *On the other hand, in the papers current state, its unclear how often the paper would be referenced ...*
Reply: Indeed I assume that the paper will be referenced more often in reviews or rebuttals than in journal publications.

Comment: *...because most well characterized retrievals have already addressed the issue discussed in this paper by supplying the instrument operators (averaging kernelsand constraint vector), as well as the observation errors.*
Reply: The reviewer claims that for most well characterized satellite data products the problem I have identified is already solved because in these cases the observation operator is provided. I do agree that in a majority of cases (not always, but this is another story!) the observation operator solves the problem of data application/comparison. I have now referenced some examples. By my discovery that the smoothing error is not an error in terms of error propagation I provide the theoretical justification that in the context of constrained retrievals this is the only known way to go, not just an option. Still a lot of groups publish the total error budget (with the smoothing error included).
However, there are situations where traditionally it is argued with the total error budget: This is typically done when the data quality (not the data themselves!) are intercompared in order to find out which instrument or which data processor is superior over another. In these cases, the inclusion of the smoothing error would lead to absurd results: Let there be two retrievals, one on an 1-km grid, another on a 3-km grid. Let the vertical resolution of both retrievals be 3 km, and let the retrieval noise be the same for both. In this case, traditional smoothing error analysis would assign a larger total error to the retrieval on the 1-km grid and will thus rate the retrieval on the 3-km grid better although both retrievals carry exactly the same information. Now, at the decline of the golden age of Earth observation pre-flight studies in the context of evaluation of future space missions gain more importance while the work with existing data will (unfortunately) be reduced. In this context a correct understanding of the smoothing error and its pitfalls is essential.

Comment: *It is for this reason that I would suggest adding some "real-life" examples (e.g., from the SPARC comparison) that I think would make this paper more relevant to the remote sensing community.*
Reply: Many more case studies have been added and references to papers where the application of the averaging kernel is demonstrated have been added,

however without changing the main focus of the paper.

3 Review by Clive Rodgers

General Remarks: We agree in more points than it appears at first sight. We seem to agree

- that the averaging kernel is more helpful but sometimes less intuitive than the smoothing error;
- that it is often possible to provide a reasonable estimate of $\mathbf{S}_{e,fine}$;
- that the smoothing error should not be linearly mapped to a grid finer than that on which it has been evaluated (which is the main focus of my paper).

There still seems to be disagreement about

- if it is possible to routinely provide the smoothing error on a grid fine enought that either interpolation to an even finer grid will not be attempted or causes no additional error;
- if it is acceptable to treat the smoothing error as an error while Gaussian error propagation does not yield the smoothing error on a finer grid;
- how serious the errors by interpolation of the averaging kernel are compared of those related with linear mapping of the smoothing error.

3.1 Initial Review

Comment: *Thomas,*
My apologies for lateness in commenting on this paper. Here are some quick comments - it will be a couple of weeks before I can do a detailed job. I don't want to rate the paper yet.
*I agree that smoothing error covariance is difficult to evaluate correctly, because $\mathbf{S}_e$ is usually not well enough known. I regard it as a qualitative measure, giving an indication of the magnitude of the difference between the retrieval and the true state due to the finite width of the averaging kernel. The averaging kernel itself is more helpful. (**GK**, not **GKV**)*
Smoothing error, like the averaging kernel, is properly only defined on a 'fine' grid. Any attempt to evaluate it on a coarse retrieval grid is doomed to failure, for the reasons given in the paper. However the expression for smoothing error covariance given in eqn (16) appears to be in error. Rodgers (2000) eqn 10.3 leads to

$$(\mathbf{W}\mathbf{G}_z\mathbf{K}-\mathbf{I})\mathbf{S}_e(\mathbf{W}\mathbf{G}_z\mathbf{K}-\mathbf{I})^T$$

whereas (16) would give

$$(\mathbf{W}\mathbf{G}_z\mathbf{K}\mathbf{W}\mathbf{V}-\mathbf{I})\mathbf{S}_e(\mathbf{W}\mathbf{G}_z\mathbf{K}\mathbf{W}\mathbf{V}-\mathbf{I})^T$$

*but **WV** is not a unit matrix.*
Reply: I agree that the analytically correct evaluation of the smoothing error requires the averaging kernel which represents the response of the retrieval to

the variation of the true atmosphere on the fine grid. I used in the original manuscript the averaging kernel which represents the response of the retrieval to the variation of the true atmosphere on the coarse grid and have then interpolated this coarse-grid averaging kernel to the fine grid. Thanks a lot for pointing out! Eq. 16 has been corrected in the revised manuscript. The case study has been redone with correct Eq. 16. In the following I will discuss the implications of this error for the paper.

- Implication for the main result of the paper: The main result of the paper is that Gaussian error propagation is not compatible with the smoothing error concept. This conclusion is even more supported when the fine-grid averaging kernel is used in Eq. 16. The conclusion holds a fortiori.
- Implications for the case study: The new results differ only marginally from the old ones.
- Implications for the recommendation to supply the averaging kernel instead of the smoothing error: Here we have a serious problem. Facing the fact that the averaging kernel cannot be interpolated, the user **cannot** calculate the exact smoothing error on his preferred grid himself. This issue is discussed in more detail in the context of the second review.

Comment: *Effectively (16) strips out the fine scale variation from* $\mathbf{S}_e$*.*
Reply: This comment is discussed in more detail in the context of the second review, which comments also my initial reply to this comment.

Comment: *The discussion about the possibility of arbitrarily large variation in the state vector at fine scales seems to be a red herring. Agreed it is conceivable, but no evidence is given that it actually happens, and experimental evidence from gravity waves seems to show the opposite.*
Reply: Here I feel a little misunderstood. I do not claim that this variation is large. I just claim that it does not disappear, even at fine scales. What I intend to say is that, if one has evaluated the smoothing error on a fine grid, there always is a finer grid with its own small-scale variability, and Gaussian error propagation will fail when the smoothing error is propagated to this finer grid. There is no limit beyond which there is no more additional variability and where Gaussian error propagation would be fully applicable to the smoothing error.

Comment: *I haven't done an exhaustive search of the literature, but the first paper I turned up (http://www.atmos-meas-tech.net/4/1627/2011/amt-4-1627-2011.pdf) fig 5 shows a decrease with wavenumber in the region of* k^{-3} *towards small scales, which I think is typical of the scale of variation of atmospheric quantities. Incidentally, this can be used as a basis for extrapolating* $\mathbf{S}_e$ *to finer scales than have been measured, by considering a fourier representation.*
Reply: I agree that an educated guess (as you call it on p 163 of your book, however in the context of $\mathbf{S}_a$) is often possible, although there is a large variety of processes which can lead to large small-scale variability (e.g. photochemical species beneath scattered clouds; intermittently emitting sources, etc.). In my original manuscript I did not at all mention difficulties to evaluate $\mathbf{S}_e$ on a fine grid. The discussions during the access review phase and during the first

part of the official discussion phase have pushed me somewhat towards this issue although my main interest has been the conflict between the smoothing error and Gaussian error propagation. My concern is that the data user who has no access to instrument-specific data cannot transform the smoothing error to a finer grid than that used for the calculation of the original smoothing error. In other words: I do agree with you that the smoothing error can be calculated (directly, from $\mathbf{S}_{e;fine}$) on the fine grid. We also agree that the smoothing error on the fine grid cannot be estimated from a smoothing error on a coarse grid via Gaussian error propagation. We seem not to agree if the term "error" is inadequate or not in a context of a quantity which is in conflict with the established error propagation laws. Although somewhat pushed into this direction, I have not challenged in my paper the possibility to provide some $\mathbf{S}_{e;fine}$. I only mention that often no reliable estimate is available, and I do not think that this statement is too strong.

3.2 Final Review

Comment: *General Comments*
The aim of this paper is to show that the concept of smoothing error leads to problems of interpretation that make it not a useful quantity in characterising retrieval errors. I agree that smoothing error is a tricky concept, and should not form part of the error analysis supplied to data users. However it is still a useful concept if handled correctly. The error analysis is best described in terms of the retrieval noise covariance (uncorrelated between separate retrievals), the systematic error covariance (may be correlated between retrievals), and the averaging kernels on clearly specified representation of the state. (See Rodgers (2000) 11.2.6.) Smoothing error is useful in certain circumstances, but is for the data users to compute based on their own assessment of the state ensemble covariance, and the spatial resolution needed for their particular applications.
Reply: I do agree; in the section of the paper where the smoothing error is presented (after Eq. 5) I have now included a statement on the benefits of the smoothing error.

Comment: *Detailed Comments*
3303 L1: An optimal estimation scheme is one which optimises something. It covers all sorts of possibilities, of which MAP is just one. I did not retitle the MAP approach, I simply used the name which had been normally used in the literature for many years.
Reply: I have reworded this paragraph accordingly.

Comment: *3303 L5: I don't think the use of the term 'a priori' in analytic philosophy is relevant.*
Reply: But I do not think that it is harmful either. Many remote sensing scientists think that its meaning in remote sensing and information theory is equivalent to its general meaning, which is not true.

Comment: *3303 L12: "indirect prior knowledge" does not sound like "knowledge" at all. It is simply a representation, used as a practical neccessity (we can only deal with a finite number of rational numbers), which has the effect of constraining the solution to a particular region of state space.*

Reply: Agreed; I have replaced the term by "indirect prior assumptions". This seems appropriate to me, because a finite number of points without some assumption how the atmosphere behaves in between is of no value.

Comment: *3306 L4: As other reviewers have pointed out,* $\mathbf{S}_a$ *should be* $\mathbf{S}_e$*. Incidentally the error analysis in Rodgers (2000) chapter 3 applies to any retrieval method....*
Reply: Agreed and corrected. This caused no major further complication because the (real) $\mathbf{S}_a$ matrix has played no major role in this paper.

Comment: *...That analysis only uses* $\mathbf{S}_a$ *as a linearisation point.*
Reply: I assume it uses $\mathbf{x}_a$, not $\mathbf{S}_a$ as a linearisation point.

Comment: *3306 L1820: The choice of grid is important. It needs to be chosen so that it can represent all scales of variability that are important to the problem in hand, and so that information present in the raw measurement is not lost...*
Reply: I agree in part. I would go a step further: For the retrieval it is sufficient to define the grid in a way that no information present in the raw measurement is lost. For the smoothing error, however, the grid must be chosen in a way that no variability of the real atmosphere is lost. Related difficulties are the main topic of my paper.

Comment: *...*$\mathbf{S}_e$ *matrices from real data should be constructed in such a way that the are not singular. And, of course, due allowance should be made for the errors in the 'real' data. See Rodgers (2000) section 10.3.3.*
Reply: Agreed.

Comment: *3306 L248: If you can think of a better term, please suggest it!...*
Reply: In the paper I will occasionally use now the term 'constraint diagnostic'. Admittedly this term is less specific than 'smoothing error' but at least it will not raise wrong expectations. The exclusive use of the new term in the paper would be confusing because the reader would lose track what I am talking about; thus I still use the term 'smoothing error' in most instances.

Comment: *If you think of eqn (4) in terms of a fourier expansion or in terms of the eigenvectors of* $\mathbf{A}$*, you will generally find that smaller scale components are suppressed more than large scale components, hence 'smoothing'. If the jacobians are, as is commonly the case, more or less single peaked smooth functions of finite width, you would expect a smoothing effect.*
Reply: The small scale components of the difference $\hat{\mathbf{x}} - \mathbf{x}_a$ are suppressed. This, however, does not necessarily mean that also the small scale components of $\hat{\mathbf{x}}$ are suppressed. It appears to me that you assume $\mathbf{x}_a = \mathbf{x}_e$. In this particular case I agree.

Comment: *3308 L57: The statement in parentheses could be omitted without loss.*
Reply: I have removed it because I am no longer sure if it indeed holds for indirect measurements.

Comment: *3308 L9: You should remind the reader here that* $\mathbf{VW} = I_{coarse}$

but $\mathbf{WV} \neq \mathbf{I}_{fine}$.
Reply: Agreed and added.

Comment: *3308 L19: What makes you think it has remained unnoticed? Isn't it obviously the case?*
Reply: I may have misinterpreted the fact that this issue had not be discussed. I have now changed this statement accordingly.

Comment: *The same arguments apply as for* $\mathbf{S_e}$. *Smoothing error should not in principle be calculated on a grid which is too coarse to capture all relevant atmospheric variability. Or, if you do so, you must be aware that you have introduced another source of error – representation error, describing variability in the null space of the representation you are using.*
Reply: Here I disagree. Error propagation is a generic concept, and the usual laws have to be applicable without expert knowledge about the specific kind of error. Application of generalized Gaussian error propagation to interpolation should yield the expected difference between the estimated and the true state also in the new grid, regardless of the nature of the error, as long as the initial covariance matrix is correct (including the off-diagonal elements). This holds for noise and parameter error but not for the smoothing error. The problem of the smoothing error comes from the fact that the initial smoothing error does not describe the difference of the estimate and the true atmosphere but the difference of the estimate and one specific finite representation of the true atmosphere. In case of the smoothing error, Gaussian error propagation yields a quantity which can be interpreted only with expert knowledge and which has a physical meaning only to people who know for which specific case this error has been calculated. The error budget should be understandable without detailed instrument specific knowledge.

Comment: *3309 L1: What precisely do you mean when you say "Gaussian error propagation is not compatible with the smoothing error concept". What goes wrong? If I understand you properly, then what you mean is that the smoothing error calculated correctly on a fine grid does not equal the smoothing error calculated on a coarse grid, and then interpolated to a fine grid. Why would anyone expect otherwise? It is not that the smoothing error concept or Gaussian error propagation has gone wrong, it is that you are not comparing like with like.*
Reply: We fully agree about what goes wrong but we seem to disagree if this is acceptable (see above). Error bars on different grids are not comparable. Thus, in case of grid-transformation Gaussian error propagation has to be applied to avoid an apples-to-oranges comparison. For all other errors Gaussian error propagation serves this purpose except for the smoothing error. In the latter case, Gaussian error propagation yields a quantity which is non understandable without knowledge on which grid the smoothing error has initially been evaluated, and – contrary to the other error types – the magnitude of the smoothing error depends on this grid.
You argue that the initial grid must be chosen fine enough to include all variability. I object that (a, theoretically) this representation does not exist, and (b, practically) that, if it existed, distributing these large amounts of data would be beyond reach.

Comment: *Gaussian error propagation applies, but it gives you the error of the coarse smoothing error when interpolated to the fine grid, not the smoothing error calculated on the fine grid.*
Reply: But, as said above, it is just the purpose of Gaussian error propagation applied to grid transformations to make initially uncomparable error bars comparable and to remove the dependence on the grid chosen. This works for all other error types except of the smoothing error.

Comment: *That is given by the corrected version of eqn (16). Incidentally, it can be shown (see appendix) that eqn (15) is related to corrected eqn (16) by replacing $\mathbf{S}_{e;fine}$ by $\mathbf{WVS}_{e;fine}(\mathbf{WV})^T$, i.e. $\mathbf{S}_{e;fine}$ with the finer scales removed.*
Reply: We agree about what is going on mathematically but we seem to disagree what expect Gaussian error propagation to produce.

Comment: *3310 L22–24: Gaussian error propagation does not have to be abandoned, it just has to be used appropriately. Smoothing error should be calculated on the fine grid, and then transformed to the coarse grid, if required.*
Reply: Gaussian error propagation is a general concept which is applicable to any linear operation, without restriction. Transformation to a finer grid is a linear operation. Thus the smoothing error is a diagnostic quantity which does not have the properties an error usually has. The smoothing error is a valuable diagnostic in its own right (I have made this clearer in the revised version) but to me it does not have the nature of an error.

Comment: *3311 section 4: There is no doubt variability on all scales, and we may not be able to estimate it from independent measurement or from theory (though turbulence theory may be helpful). However there will be a scale below which we can be reasonably sure that our instrument has effectively a zero sensitivity. We should choose a representation such that scales larger than this can be represented. (See Rodgers (2000) 10.3.1.2 and 10.3.1.2.) A data user who wishes to extend the smoothing error to smaller scales (if he has a suitable $\mathbf{S}_{e;fine}$!) can then assume that all variation at smaller scales is in the smoothing error.*
Reply: I agree that this is the right way to go. However, Gaussian error propagation is a generic concept where exceptions as required for the smoothing error are not foreseen. Thus I suggest not to treat the smoothing error like an error and not to include it in the error budget (I think we agree at least on the latter, c.f. Sect 11.2.6 in Rodgers 2000, and first lines of this review). The statement that, with a suitable $\mathbf{S}_{e;fine}$ available, the user can calculate the smoothing error on his preferred grid needs some discussion. We have $\mathbf{WV} \neq \mathbf{I}_{fine}$, and the user does not typically have $\mathbf{A}_{fine}$ on the desired grid available, because this is an instrument-specific quantity. Thus the user will have to apply my old (incorrect) version of Eq. 16. But we have the additional assumption that the initial grid is fine enough that all scales the instrument is sensitive to are represented. In this case the inaccuracy of old Eq. 16 turns out to be tolerably small. This is important, because otherwise the data user who wants to go to a finer grid would not be helped with providing the original grid averaging kernel. My test calculations corroborate what one might intuitively expect: The differences due to the use of the interpolated averaging kernel are small when the initial sam-

pling is fine compared to the scales of variability and resolution. This saves the recommendation to provide the averaging kernel as one main diagnostic, which otherwise would be of limited use. Of course my test calculations must not be over-generalized because results will depend on the actual shape of the averaging kernels but the test indicates that the harm done by interpolating averaging kernels is orders of magnitude smaller than the harm done by applying Gaussian error propagation to the coarse grid smoothing error in a straight forward way. I think that this result considerably strengthens the practical usefulness of my paper.

Comment: *3312 L5–7: It is simply wrong to attempt to do this, as discussed above.*
Reply: Why? This approach would be fully in agreement with Gaussian error propagation. If we accept Gaussian error propagation but get absurd results, I conclude that the quantity propagated is not an error in the sense of error propagation.

Comment: *3312 L8–13: We do not attempt to retrieve at dimensionless points. The state vector is always a set of numbers which define a continuous function. It is the continuous function that we retrieve, and which is required to evaluate the forward model.*
Reply: I do not mean extensionless points with gaps in between. But if the grid attempts to catch all natural variability, it will approach towards an infinitesimal grid, which means that each gridpoint will represent an infinitesimal extension. In reality, absurdities will be encountered long before the infinitesimal scale will be reached, i.a. already when the molecular scale is reached, which strengthens my case. All this is only about demonstrating that convergence is not achieved. I have reworded this paragraph accordingly.

Comment: *3312 L13–20: As mentioned above, this is exactly what I recommend, e.g. in Rodgers (2000) 11.2.6. (Though very briefly – I was hurrying to finish the book!)*
Reply: ok, here we agree; I have added a reference to this section of your book.

Comment: *3312 L20-24, I don't understand this sentence.*
Reply: This sentence has been reworded.

Comment: *3313: Much of this will need a different emphasis in the light of the difference between $\mathbf{S}_e$ and $\mathbf{S}_a$.*
Reply: I agree and have corrected this paragraph accordingly.

Comment:*3314 L6: useful*
Reply: corrected.

Comment: *3315 L17-19: The formalism can be applied ML retrievals. All you need is the representation function (which is the same as $\mathbf{W}$), and an $\mathbf{S}_{e;fine}$ matrix. The smoothing error can be evaluated on the fine grid.*
Reply: Agreed but not added to the paper because the discussion of maximum likelihood retrievals is beyond the scope.

Comment: *3315 L25: for "implicit smoothing error" I would use "representation error" as a more descriptive term.*
Reply: Agreed that "representation error" is more descriptive. I use now both terms.

Comment: *3315 next sentence: I don't think it has been shown adequately in a practical sense. Extensionless points are not relevant. You could e.g. use a fourier series.*
Reply: As said above, I don't use the term "extensionless points" in the context of points with gaps in between but as points as the limit when inifinitesimally fine resolution is attempted.

Comment: *3316 L4 onward: I do think the problem is purely philosophical and practically irrelevant!*
Reply: My new series of tests shows that interpolation to a finer than the original grid leads to considerable mis-estimates of the smoothing error even in cases when the grid is several factors finer than the correlation length of natural variability and much finer than the resolution of the retrieval. Philosophical or theoretical issues aside, distribution of diagnostics on grids fine enough that problems are practically (while not theoretically) solved will easily be beyond reach for practical reasons. I have added a comment on this. However, I am still not sure if there isn't a misunderstanding. I do NOT challenge that the smoothing error can in principle be evaluated on any fine grid. I only claim that Eq 15 (which is the correct application of Gaussian error propagation) in the context of interpolation to an even finer grid will fail to produce the correct smoothing error on this finer grid.

Comment: *Smoothing error does comply with Gaussian error propagation if done correctly.*
Reply: Here I disagree. Eq. 15 is correct in terms of Gaussian error propagation, and we cannot just redefine the rules of error propagation just because the quantity to be propagated behaved differently than errors usually do.

Comment: *3316 L15: I agree strongly that the averaging kernel should be supplied. I also feel that the smoothing error should not be supplied...*
Reply: I have included another reference to Rodgers 2000 Sect 11.2.6 here.

Comment: *... The users can evaluate that themselves if they want it.*
Reply: Agreed with the caveat that also the averaging kernels cannot interpolated from the fine to the coarse grid. But if the original grid has been chosen fine enough, related inaccuracies turn out to be tolerable.

Comment: *Comments on Thomas's reply:*
C797 Implications for sect 5, L1-6: These are exactly the reasons that TES chose to provide averaging kernels with their data set.
Reply: Agreed

Comment: *C798 2. Possibility...: But there will always be a scale beyond which it doesn't matter for practical purposes.*
Reply: I assume that, if the data user decides to interpolate data on a finer

grid, there is some additional variability on this grid. Otherwise the user had no interest in using this finer grid. And as soon as interpolation to a finer grid occurs, the discussed problems will occur.

Comment: *Appendix: Eqns 15 and 16*
The correct version of eqn 16 requires $\mathbf{A}_{fine} = \mathbf{WGK}_{fine}$*:*
Reply: Agreed and corrected.

Comment:
$\mathbf{S}_{sm;fine} = (\mathbf{I}_{fine} - \mathbf{WGK}_{fine})\mathbf{S}_{e;fine}(\mathbf{I}_{fine} - \mathbf{WGK}_{fine})^T$

Equation 15 uses $\mathbf{A}_{coarse} = \mathbf{GK}_{coarse} = \mathbf{GK}_{fine}\mathbf{W}$.
$\mathbf{S}_{sm;fine} = \mathbf{W}(\mathbf{I}_{coarse} - \mathbf{GK}_{fine}\mathbf{W})\mathbf{S}_{e;coarse}(\mathbf{I}_{coarse} - \mathbf{GK}_{fine}\mathbf{W})^T\mathbf{W}^T$

Substitute $\mathbf{S}_{e;coarse} = \mathbf{VS}_{e;fine}\mathbf{V}^T$ *and bring the* $\mathbf{W}$*'s inside the brackets:*

$\mathbf{S}_{sm;fine} = (\mathbf{W} - \mathbf{WA}_{coarse})\mathbf{VS}_{e;fine}\mathbf{V}^T(\mathbf{W} - \mathbf{WA}_{coarse})^T$

Substitute $\mathbf{A}_{coarse} = \mathbf{GK}_{fine}\mathbf{W}$ *and move the right hand* $\mathbf{W}$*'s:*

$\mathbf{S}_{sm;fine} = (\mathbf{I}_{fine} - \mathbf{WGK}_{fine})\mathbf{WVS}_{e;fine}(\mathbf{WV})^T(\mathbf{I}_{fine} - \mathbf{WGK}_{fine})^T$

This is the same as eqn 16, but with $\mathbf{S}_{e;fine}$ *replaced by* $\mathbf{WVS}_{e;fine}(\mathbf{WV})^T$ *, which is* $\mathbf{S}_{e;fine}$ *projected onto the coarse grid, and interpolated back on to the fine grid, thus removing fine scale information.*
Reply: Agreed that Eq. 15 can be written in the structure of Eq 16. This explains what goes wrong with Eq. 15 (and by the way, answers the related question by Reviewer #5. But since it is still different from (old) Equation 16, this does not prove that old Eq 16 (with the fine-grid $\mathbf{S}_e$ matrix but interpolated averaging kernel) destroys the small-scale information as Eq 15 does. Tests have shown that old Eq 16 may be a justifiable approximation for cases when the small-scale averaging kernel is not available and the grid on which the original averaging kernel has been evaluated is still fine enough to resolve what the instrument can see.

Thank you, Clive; your comments have helped me much to improve the paper. I think that the aspect of interpolability of the averaging kernel, which resulted from our discussion, has added considerable practical value to the paper.

4 Review by Brian Connor

Comment: *The paper presents a detailed discussion of a subject which is important to the interpretation of remote sensing measurements, but is not well understood in the community. The subject of smoothing error has by-and-large been approached correctly by investigators responsible for retrievals from remote measurements, but its implications and limitations are often not understood by data users.*
The current paper will help alleviate this shortcoming. It also presents a quantitative aspect of the subject for the first time, namely its relation to error propaga-

tion. The paper is generally well reasoned and well written, with a few exceptions already discussed by other reviewers. I feel it makes an important contribution and should be published with minor changes. One recommendation made by others which I support is to include a few specific examples. They needn't be lengthy. A sentence or two and a reference could specify how a particular experiment dealt with the smoothing error issue.
Reply: More examples (case studies) have been added and summarized in a table. References how measurements are characterized by the smoothing error and/or by the averaging kernel have been added.

Comment: *Both the concepts of smoothing error and averaging kernels are intended to cope with the limited information inherent in remote measurements. The following discussion is relevant to the author's recommendations regarding their use.*
The concept of 'null space' error was first introduced in Rodgers, 1990. (It was subsequently renamed 'smoothing error' after A. O'Neil pointed out that 'null space' was not strictly correct.) Clive and I discussed that paper when it was in draft form, and we both recognized smoothing error as a problematic (but unavoidable) concept. However the principal difficulty we saw with it was its dependence on the covariance matrix of variability of the true atmospheric state, which is poorly known at best. To the best of my knowledge, the paper being reviewed is the first to recognize the difficulty with Gaussian error propagation applied to smoothing error.
It emerged from those early discussions that one could eliminate smoothing error from the comparison of atmospheric profiles from different sources (measurements or models) if the vertical resolution of one profile was much finer than the other. The smoothing error is removed by smoothing the high resolution profile with the low resolution averaging kernels. As far as I know, this was used first in Connor et al 1994, who called the result the 'convolved profile.' This approach also removes the a priori profile, used by the low resolution measurement, from the comparison.
References:
Connor, B.J., D.E. Siskind, J.J. Tsou, A. Parrish, A.E.E. Remsberg, Ground-based microwave observations of ozone in the upper stratosphere and mesosphere. J. Geophysical Res., 99, 16,757-16,770, 1994.
Rodgers, C. D.: Characterization and error analysis of profiles retrieved from remote sounding measurements, J. Geophys. Res., 95, 5587–5595, 1990.
Reply: Since this 'convolution' method is widely used, I have added it to the paper, along with the reference. I avoid, however, the term 'convolution' because I am not quite sure if a matrix product where A can have rows which differ by more than only a shift really is the discrete representation of a convolution.

Comment: *Comparing profiles with comparable vertical resolution is more difficult. It was a focus of work around the end of the century and eventually led to Rodgers & Connor, 2003. Here again, as the author of the current paper points out, smoothing can be eliminated as a source of difference between measured profiles, by application of the measurement averaging kernels.*
My own approach to the problem is given in Connor et al, 2008, for the OCO mission. It is to provide an estimate of smoothing error as a qualitative guide for the data user, and also to provide the a priori profiles and the averaging

kernels. In addition, I have recommended the use of Rodgers & Connor (2003) in quantitative profile comparisons. Thus, I wholeheartedly agree with the author's recommendation for the provision and use of the averaging kernels, and urge him to accept smoothing error as a simpler, if less precise, aide to understanding.

References:

Connor, B. J., H. Bösch, G. Toon, B. Sen, C. Miller, and D. Crisp (2008), Orbiting Carbon Observatory: Inverse method and prospective error analysis, J. Geophys. Res., 113, D05305, doi:10.1029/2006JD008336.

Rodgers, C.D.; Connor, B.J. (2003). Intercomparison of remote sounding instruments. Journal of Geophysical Research 108(D3), 4116, doi:10.1029/2002JD002299.

Reply: I have added the references and have rewritten the paper in a way that the smoothing error is no longer challenged as such but only in the context of error budgets.

5 Anonymous Referee #5

Comment: *General Comments The paper "Smoothing error pitfalls" by T. von Clarmann discusses implications and limitations of the smoothing error concept in retrieval theory. The "smoothing error" is usually considered being an important diagnostic quantity of an optimal estimation retrieval. It measures the difference between the optimal estimate and the unknown true atmospheric state due to the influence of explicit a priori information. A major finding of the paper is that the smoothing errors for a coarse retrieval grid are not compliant with smoothing errors estimated for a finer grid, if the coarse grid data was simply interpolated to the fine grid and the corresponding smoothing error was determined by means of Gaussian error propagation. Solutions to this problems are presented.*

In general, I found the paper well-written and interesting to read. I would also think it is of interest for many readers of AMT. However, right now I am not convinced that all key statements hold and I would dispute the conclusion that the smoothing error concept is "questionable", "untenable", or "failed"...

Reply: The entire text has been reworded in a way that attributes as listed above are only used with in conditional statements or within a very specific context and no more in absolute or general statements. This is hoped to make the paper more balanced.

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

Reply: I do not understand this comment. In my paper I make very detailed reference to Rodgers (2000). All the passages of the Rodgers book quoted by 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_e matrix but I am not aware that the consequeence of this for the smoothing error has been discussed anywhere in the literature.

Comment: *I will try to express my concerns in more detail in the following general comments. These concerns need to be carefully addressed and the manuscripts needs to be revised before it can be published.*

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

$$\mathbf{S}_{smooth} = (\mathbf{I} - \mathbf{A})\mathbf{S}_a(\mathbf{I} - \mathbf{A})^T.$$

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

$$\mathbf{S}_{smooth} = (\mathbf{I} - \mathbf{A})\mathbf{S}_e(\mathbf{I} - \mathbf{A})^T,$$

with $\mathbf{S}_e$ *being the covariance of a real ensemble of states rather than the a priori covariance* $\mathbf{S}_a$*...*

Reply: Thank you for spotting this. I should indeed have used another notation and have corrected this in the revision. However, in Rodgers' Equation 2.23 (p25), $\mathbf{S}_a$ 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 $\mathbf{S}_a$ matrix is discussed. In Rodgers [p.58] again the smoothing error is evaluated on the basis of $\mathbf{S}_a$. Anyway, the notation with $\mathbf{S}_e$ is much clearer, and in my paper I have replace $\mathbf{S}_a$ by $\mathbf{S}_e$ because I do not use the regularization term $\mathbf{S}_a^{-1}$ except for a few instances where I do now carefully distinguish between $\mathbf{S}_a$ and $\mathbf{S}_e$. In my notation, I use $\mathbf{R}$ for any generic or ad hoc regularization matrix, so there is no ambiguity.

Comment: *...It seems* $\mathbf{S}_e$ *rather than* $\mathbf{S}_a$ *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 (of the AMTD version) I discuss exactly this issue. There I also make reference to page 49 of the Rodgers book.

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* $\mathbf{S}_a$ *cannot be simply interpolated to obtain an estimate of* $\mathbf{S}_e$ *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 $\mathbf{S}_a$ or $\mathbf{S}_e$ matrix. On page 3308, line 16 of the AMTD version 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 (AMTD version) 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 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 (subject to interpolability of the averaging kernel, see C. Rodgers' comments and my replies to them). Thus the 'averaging kernel solution' is superior to the 'smoothing error solution' because the first kind of includes the latter but not vice versa. This recommendation is in full agreement with Rodgers' recommendation in Section 11.2.6. of his book.

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 infinites-*

imal 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)...
Reply: In the mathematical sense of 'point', an extensionless point is (since the definition of Euclid of Alexandria) a term with a redundant attribute (if it is indeed a tautology is another question). I had 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. Anyway, I have rewritten the respective paragraph and avoid the questionable terms now.

Comment: *.... 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: 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 have reworded this paragraph accordingly. 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?*
Reply: 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. The smoothing error shall give an estimate between the smooth estimate and the true atmosphere at ideal resolution but the latter is undefined, which leads to the problems discussed in the paper. In the revised version of the paper, I do not claim that the smoothing error is a meaningless diagnostic tool; I just claim that it does not behave like an error should behave

and thus should not be treated as an error.

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 can add to the problem. But even if we had these data on a 3-m grid, one satellite group might evaluate their smoothing error on a 1-m grid, and again they compare apples and oranges, even in the ideal world where we have access to ensemble covariance 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. I would like to state again that I do **not** challenge that the smoothing error can in principle be evaluated on a very fine grid. Instead my point is that interpolation of the smoothing error to an even finer grid fails. And in an ideal world where we have variability data available on a 3-m grid, there will be people who are interested in processes on a 1-m grid.

Comment: *...Based on that I would also support the concern raised by referee John Worden that the paper needs more illustrative and real-world examples to demonstrate the relevance of the problem. I learned from the online discussion that such practical problems came up in a comparison of MIPAS and TES data. I was also wondering, if the problem would not be of even larger relevance for nadir measurements? Limb measurements, as discussed in the paper, are often not affected by a priori and regularization as much? In its current form the paper is focused on limb applications.*
Reply: I have included a whole series of test calculations in order to make the test applicable to a wider range of situations. An important practical aspect which came up during the discussion with C. Rodgers is the interpolability of the averaging kernel. The fact that in a rigorous sense the averaging kernel matrix is not interpolable prima facie seems to rule out the option that the users can evaluate their own smoothing errors directly on the fine grid, because the averaging kernel matrix is (as opposed to the $\mathbf{S}_e$ matrix) an instrument-deendent quantity which the users cannot construct themselves. That means that we cannot take it for granted that the recommendation to dirtribute the averaging kernels instead of the smoothing errors improves anything. My new series of tests, however', demonstrates that the error by interpolation of the averaging kernel remains tolerably small as long as the original averaging kernel has been evaluated on a grid which resolves all scales the instrument can see. I think that this is an important new result of practical relevance.

Comment: *Finally, I noticed that Rodgers (2000, chap. 10.3.1) suggests another practical solution for scaling an a priori covariance from coarser to finer*

grid: "If we only have an estimate of $\mathbf{S}_{za}$ *[coarse grid] we cannot assume that* $\mathbf{S}_a = \mathbf{W}\mathbf{S}_{z,a}\mathbf{W}^T$ *[fine grid], as this will be singular [...] and will carry the implication that any structure not represented by the interpolation has zero variance. If we really have no other information, we should make and educated guess, for example by changing the zero eigenvalues of the eigenvector decomposition of* $\mathbf{W}\mathbf{S}_{za}\mathbf{W}^T$ *to values conservatively extrapolated form the non-zero eigenvalues." This sounds like a very reasonable approach to me. Its applicability could be directly tested with the idealized example presented in section 3 of the paper, for instance. This possible solution should be mentioned in the paper at least.*
Reply: Attention, this is a recommendation for $\mathbf{S}_a$, not for $\mathbf{S}_e$! What is suitable to construct a well working regularization matrix is not necessarily suitable to estimate the true small-scale variability. With regard to estimating variability on small scales I find the method suggested by C. Rodgers in his review, i.e. to take advantage of the known energy cascade between scales more promising. However, even this approach cannot take into account that at smaller scales completely different physical processes may become relevant. While my main statement is not the challenge of the possibility of the smoothing error on a fine grid but the problem that it cannot be propagated to finer grids by Gaussian error propagation, I still have included a remark in the revised version that a fine-grid $\mathbf{S_e}$ matrix can be constructed.

Comment: *Specific Comments*
p3303,l5-7: I think it is fine to explain what we mean by "a priori" in the context of retrieval theory. I think there is no need to explain its meaning in analytical philosophy in AMT.
Reply: I think an explication is less ambiguous when the intended use of a term is contrasted to and demarcated from other existing uses of the term.

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" in the paper would provide a more balanced view.*
Reply: As soon as the regularization matrix is a regular covariance matrix (i.e. in maximum a posteriori retrievals but not in retrievals using a Twomey-Tikhonov-Phillipps-type finite differences constraint), the constraint will push the retrieval away from the true state towards $\mathbf{x}_a$. Even if $\mathbf{x}_a$ is perfectly chosen, it describes only the mean profile, not the true actual profile. This becomes very clear if we apply the maximum a posteriori approach to a one-dimensional retrieval (i.e. the retrieval of a scalar). In this case smoothing plays no role at all, but the 'smoothing error' still will not be zero, except if either $\mathbf{S}_a^{-1} = 0$ or if the true value happens to be identical to the priori value. This proves that there generally is something else beyond smoothing. I think this is why editor Dr Bhartia suggested to me to add a critical discussion of the term "SMOOTHING error". I have, however, 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 this diagnostic quantity does not behave as an error should behave. In the revised version I have rewritten the part about the terminology; I now restrict the attribute "so-called" to implicit definitions of new terms, and restrict the quotes to distinguish metalanguage (i.e. text about terminology) from object language (text about the content).

Comment: *p3307,l17-18: Rather than saying that the error propagation rules are "generally 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 approximation is not applicable for error estimation? Also, I guess you actually mean "moderately 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 moderatly 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. If, as suggested, I used in my paper applicability of Gaussian error propagation as definition of non-grossly non-linear cases, than I had to conclude (as a result of my paper, because Gaussian error propagation is shown to fail here) that linear interpolation is a grossly non-linear operation, which seems absurd to me. Thus I hesitate to use the suggested definition.

Comment: *p3308,l5-14: It would be nice to adopt the notation of Rodgers (2000, chap. 10.3.1.1). This is basically the same discussion and it would be easier to relate your paper and the discussion of Rodgers (2000) to each other.*
Reply: I prefer the self-explaining indices.

Comment: *p3308,l15: Maybe briefly introduce what is meant by "parameter error" here. The term was not introduced in the paper before.*
Reply: Agreed, definition inserted.

Comment: *p3309,l10-15: I was wondering if it is possible to rewrite Eq. (15) in a way that it corresponds to Eq. (16) plus an additional term, which represents the difference of the two equations? This could be helpful to further understand the differences between Gaussian error propagation and direct linear analysis?*
Reply: This transformed representation has been included. A derivation is shown, e.g. in the appendix of C. Rodgers' second review.

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.)*

Reply: Tests with longer correlation lengths have been added. Further, the examples can easily be scaled. The original example scaled by a factor of three or four might give the typical resolution of a nadir sounder, and a correlation length might be more appropriate for atmospheric problems typically tackled with nadir sounders. The original example may have been typical for a limb sounder trying to resolve the hygropause. The point is the contrast between resolution and correlation length which defines the problem. In that sense, the examples presented are representative for more cases than it appears. It may be true that larger values than those used in the initial example are often used in regularization matrices, but as we have agreed above, it is not the choice of the regularization matrix $\mathbf{S}_a$ what counts here but the true ensemble covariance matrix $\mathbf{S}_e$.

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?)*
Reply: 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 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 have added "at infinitesimal resolution" to make this issue clearer.

Comment: *p3315, l1-10: I think that these statements require their mathematical derivation to be shown. I could not directly see that these are true.*
Reply: For this derivation I need the approximation $\mathbf{A}_{fine} \approx \mathbf{W}\mathbf{A}_{coarse}\mathbf{V}$. It is now shown in the Appendix.

Comment: *Technical Corrections*
p3309,l19-p3310, l21: This paragraph is rather long. Better split in 2-3 paragraphs.
Reply: Because of the new test cases this paragraph has been rewritten anyway. Care has been taken to avoid excessively long paragraphs.

Comment: *Technical Corrections*
p3314, l5: Suggest to remove "in their paper".
Reply: Agreed an deleted.