

## ***Interactive comment on* “Elimination of hidden a priori information from remotely sensed profile data” by T. von Clarmann and U. Grabowski**

**Anonymous Referee #1**

Received and published: 15 September 2006

Referee report Elimination of hidden a priori information from remotely sensed profile data T. von Clarmann and U. Grabowski

In this paper the authors discuss a simplified representation of atmospheric sounding profile retrievals, namely one which maintains most of the information content, eliminates prior information and leads to a unity-matrix averaging kernel (eliminates the kernel).

What I appreciate especially in this paper is the discussion on how a coarse-grid a-priori free representation is linked to a fine-grid constrained problem, chapter 2 and 3. The basic theoretical framework for removing a-priori information is given in Rodgers book, chapter 10 in particular, and the discussion in this paper by von Clarmann and

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Grabowski closely follows (copies) Rodgers derivations. Nevertheless, the topic is discussed only very briefly in Rodgers, and the extension of this discussion provided by the authors is interesting and a valuable addition to the literature on retrieval theory. As such I am in favour of publication in ACP.

However, to my opinion the present text does not yet meet up with the scientific standards of ACP. Before being published I would urge the authors to seriously address my basic remarks below, and adjust the text of the paper accordingly.

First remark:

The advantages of not having to deal with averaging kernels is very much overstated, misleading and sometimes wrong. Whether or not people have conceptual problems to use "sophisticated diagnostic metadata" is not a scientific issue and should not be the topic of a scientific paper, and I would urge the authors to remove many of the statements made on this point.

A fundamental aspect of remote sensing is the fact that the information derived comes from an extended region in longitude, latitude and altitude. The retrieval information (e.g. ClONO<sub>2</sub> profile) therefore ALWAYS has to be supplemented by a description of the spatial extent (the atmospheric box: latitude, longitude and altitude range) where the information, condensed into a retrieved quantity, comes from. The averaging kernels (as the name suggests) do exactly this. It is important to realise that this fundamental aspect is by no means changed when the a-priori is removed. Also in the a-priori free case the retrieval still has to provide a recipe how the retrieved numbers are to be related to the real-world high-resolution vertical profile of the chemical tracer. What happens is that the smoother kernel is replaced by e.g. box functions. The choice of the thick layers is a-priori information by itself, as correctly mentioned by the authors. The simplification in data use is only minor, or basically unchanged!

To my impression the "fear" for averaging kernels comes from the fact that people would like to compare point observations with point satellite retrievals (e.g. the mixing ratio at

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

10.35 km altitude). This is clearly unrealistic since satellite retrievals can not provide this, independent of the fact weather or not a-priori information has been eliminated. People who are scared of kernels are still scared when they are confronted with e.g. figure 1!

Abstract: "For such applications it often is desirable to remove the a priori information from the data product." This is a very vague statement for which there is no backup based on scientific arguments, so the statement should be removed. As discussed above there is no real fundamental difference between constrained and unconstrained retrievals.

Abstract:".. which causes complication in the use of data for .. comparison with models, or data assimilation." I disagree and suggest to remove this line. The comparison with model and application in data assimilation is completely straightforward, is discussed at length by Rodgers and is given in eq. 6. One should take the model profile  $x$ , multiply with the kernel matrix  $A$ , and compare with the retrieved number  $x_{\text{hat}}$ . As simple as that. The retrieval covariance describes the corresponding measurement/retrieval error. A corresponding chi-square test will tell if the retrieval and model agree within error bars. When the a-priori error is chosen to be large, this a-priori information will have almost no impact on the assimilation analysis, and the whole discussion about the a-priori is not very important. The covariance matrix and kernel take care of all this.

Abstract: ".. remove the a priori information .. vs .. is a prior constraint in itself" There is some "friction" here with the strong statements made on the benefits of the a-priori free representation, admitting that the choice of layers is a-priori as well ...

Intro: I7 ..severe complications .. This is a non-scientific statement ! I would suggest to remove it.

Intro, p1, I20: "where the data user does not need to consider sophisticated diagnostic metadata but can directly use the data as they are without running risk of major misinterpretation " I disagree - see my discussion above. I would claim that the interpretation

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

of the retrieval is just as difficult as when kernels are provided. Please remove this line.

In chapter 4 the authors also try to convince the reader that there are many advantages to retrievals provided without averaging kernels. Personally I did not find much new information in this rather long section, and I suggest to remove most of this section and add one or two lines to the discussion. For instance the reduction of the amount of information to the data user is generally not a major issue. In my experience with remote sensing the data volume is for most satellite instruments well within the capabilities of modern data storage and internet links. Averaging kernels are quite easily included in the observation operator in data assimilation, as mentioned. The accuracy of the retrieval product is certainly a much more important aspect, and I would have liked the authors to focus more on this. In particular: - Section 4.1: Validation is an important issue, and a discussion based on Rodgers-Conner justifies a separate section. This section should focus on the coarse grid representation and how to compare this with other data. - Section 4.2: This brings forward a topic completely unrelated to the rest of the paper, and ignores the fact that for the coarse-grid retrieval there is still an interpretation issue related to the thickness of the layers. I would suggest to remove this section. - Section 4.3: Same as for section 4.2. I would suggest to remove this section. - Section 4.4: I find the whole discussion concerning data volume long and boring. One single line added to the conclusion, mentioning the reduction of the data volume, will be enough to make this point. I would suggest to remove this section as well.

After all the work done by pioneers like Rodgers the users of remote sensing products should be familiar with averaging kernels and how to use these!

Second remark:

The authors base their discussion fully on the averaging kernel trace, and show that this trace is equal to 9 in the unconstrained case, compared to 7.1 in the normal case with regularisation. This result can be misleading. When a very weak constraint is

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

introduced the trace of the kernel can be artificially high, but the solutions will be influenced by the noise of the instrument. A discussion of the singular values of the problem is crucial, but completely missing in the paper.

In this respect I wonder if the new regularisation (without constraints) is really performing as well in all cases as the original regularisation as far as the singular values are concerned. The strong constraints imposed on the sub-layers within one big layer are artificial. I can imagine that this constraint, imposed by the coarse-grid representation, will (in general) introduce considerable errors to the forward model (which is the ability to accurately describe the radiation levels seen by the satellite instrument) and an actual loss of information in going from the fine grid to the coarse grid. As such I am not convinced that meaningful retrievals can be obtained on the coarse grid in general. The authors introduce a clever choice of layers, based on the trace of the kernel matrix. This will help, but does not guarantee a meaningful retrieval result. MIPAS is a limb sounder and rather special in this respect (see e.g. Ceccherini 2003), but for e.g. nadir sounding it may perhaps fail? The authors should explicitly discuss other measures of information content and these forward model errors. They should also discuss the conditions for which meaningful results are obtained. Do the authors claim that the approach is applicable to all remote sensing observations, or is it limited to MIPAS and/or limb sounders ?

Third remark:

A recipe on how the coarse-grid information should be used is missing in the paper. The paper is not complete without a detailed discussion of this. I would suggest to add a separate section.

Suppose a user has a very fine grid balloon in-situ measurement available, and would like to compare with the data presented in figure 1. Then he still needs to know how the thick layer quantities are related to his measurements (in the same way as kernels provide this information). Is this a simple box average over the thick layer? An average

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

of number density or mixing ratio? How can a single concentration measurement (eg at 16.8 km altitude) be compared with the thick layers (provided at, say, 10, 15, 20, 25 km). Should this be compared with the 15 km concentration, or should an interpolation be applied. What is the representation error? How does this comparison depend on the regularisation chosen (staircase vs triangular, sec 3.1, 3.2)? What is the meaning of the dashed line in figure 1?

Fourth remark:

It took me a while to realise that the authors in fact describe an unconstrained retrieval approach. This is the fundamental meaning of equation 19: for the coarse grid there are no constraints on the thick layers ( $R=0$ ). With this the method reduces to a least-square maximum likelihood approach on the coarse grid. As a result clearly no averaging kernel concept is needed. The reader has to go through quite some mathematics to arrive at this result. The whole discussion could be reversed: start from the results aimed for, namely the coarse grid representation, and show how this may be related to a fine grid problem with a special constraint. Alternatively, and more practical, the authors could add a paragraph at the end of the introduction to explain in words what sections 2 and 3 are about.

Fifth remark:

Please show (figure 1) and discuss the retrieval results for both the staircase and triangular representation. What are the differences?

---

Interactive comment on Atmos. Chem. Phys. Discuss., 6, 6723, 2006.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)