

***Interactive comment on “Averaging kernel prediction from atmospheric and surface state parameters based on multiple regression with MOPITT CO and TES-OMI O<sub>3</sub> multispectral observations” by H. M. Worden et al.***

**Anonymous Referee #1**

Received and published: 22 March 2013

The paper by H. M. Worden et al. presents a method to predict averaging kernels to be used in observation system simulation experiments (OSSE). Since it is not always practicable to use the actual averaging kernels in the OSSE, often an over-simplifying approach is taken. Worden et al. clearly identify the need of a more advanced but still practicable method. Thus their work is highly significant in the context of AMT.

The method proposed by the authors is adequate and convincing, and as far as I can see, the maths behind it is correct.

C433

The presentation is, on the whole, very good but I have several comments how the presentation possibly still can be clarified:

Title: The title is very long, and contains some unnecessary but not all necessary information.

a) it is essential to say that the method is limited to nadir sounding. Prediction of limb averaging kernels has its own difficulties, e.g. the variable mismatch of the retrieval grid and the measurement grid. These difficulties are not tackled in this paper.

b) I think it is not essential to mention the instrument names in the title (because the applicability of the method is more general, and the title is overloaded with acronyms)

c) Is it important to mention that the observations are multispectral? First, TES is hyperspectral rather than multispectral, and I think these days monospectral atmospheric sensors are not very common any more.

I suggest: "Nadir sounder averaging kernel prediction from atmospheric and surface state parameters based on multiple regression".

Abstract I15: The term "training set" is very instructive but it is a technical term in the context of artificial neural networks (where, to my knowledge, coefficients of grossly nonlinear functions are fitted in order to emulate discrete transitions by continuous differentiable functions). Since the meaning of the term "training" is slightly different here, I suggest to define it before using it, because, as far as I can judge, although instructive and pretty self-explaining, it is no established technical term outside neural networks. See also my comment below.

Abstract I26: It took me a while to understand the acronym CONUS. I suggest to delete it in the abstract (since the acronym is never used in the abstract after its definition) and to define it in the body of the paper.

General: The abstract and the body of the paper are quite independent things. Definitions used in the abstract are not to be referred on in the body. Both must be able to

C434

stand alone.

p2754 l. 15: In Rodgers 2000 the method is called "maximum a posteriori". The term "optimal estimation" refers to the paper of Rodgers in 1976 in Rev. Geophys.

Eq 1: Aren't the "x" vectors? Then they should be bold face italic (see

[http://www.atmospheric-measurement-techniques.net/submission/manuscript\\_preparation.h](http://www.atmospheric-measurement-techniques.net/submission/manuscript_preparation.h)

Eq 2: subscripts intentionally in bold face?

Eq 3 and p2755 l4: Isn't F a vector rather than a matrix? Then it should be bold face italic.

p2756 l10: SVD is not defined in the body of the paper, only in the abstract. The definition in the abstract is obsolete, because the abbreviation is not used in the abstract. I suggest to delete the abbreviation in the abstract, and to add the definition of the abbreviation here.

p2758 bottom: I can understand what overlapping AKs are, but what are highly correlated AKs? Is this a sloppy wording for "overlapping AKs cause highly correlated retrieval errors" or something like this? Rewording is appreciated. Of course one can calculate covariances and correlation coefficients between two AKs, but what is their meaning? I find this statement confusing and find the statement on their overlapping nature sufficient.

p2760 top: I suggest to insert: "The averaging kernel prediction scheme used here uses a regression function. In analogy to artificial neural networks terminology, we call the data set used to infer the coefficients of the regression function 'training set'. The training and test ..."

p2762 l2/3: It is correct that the predictors are not independent (because of the error correlations between them, i.e. because the predictors are not an orthogonal system, and possible null spaces in the coefficient fit) but it confuses me that the contributions

C435

of the predictors are not always a "linear combination." Doesn't this statement contradict Eq. 5, which is a linear combination? Perhaps the statement is based on a notion where "linear combinations" are directly linked to a vector space with linearly independent basis vectors. But I think the term "linear combination" is often used in a wider sense. Isn't the definition of "linear dependence" that non-trivial "linear combinations" (sic!) can yield zero? Doesn't this definition imply that "linear combinations" also exist in linearly dependent systems? I might have missed the point here but clarification of this issue would be appreciated, or just delete the statement "are not a linear combination".

Eq 10-12: Somewhere it should be said that the x represent the log<sub>10</sub> of the VMR. It is said above Eq 6 that the MOPITT retrieval is performed in the log<sub>10</sub> space, but this does not imply automatically that the x used in the formalism presented here also is log<sub>10</sub>(VMR) (Experts, of course, know, but...). Either write on p2761 l24: "MOPITT uses  $x = \log_{10}(\text{VMR})$ , or define x on p2764 l17.

---

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 2751, 2013.

C436