Atmospheric
Chemistry
and Physics
Open Access

Discussions

EGU

ACPD

**Interactive
comment**

Interactive comment on "Comparison of tropospheric NO$_2$ columns from MAX-DOAS retrievals and regional air quality model simulations" *by* Anne-Marlene Blechschmidt et al.

Anonymous Referee #3

Received and published: 16 March 2017

The topic is very relevant, however I have to criticise the approach used since in the current status important open questions remain.

Let me start by asking the author, once more, to improve the language adopted in the manuscript. There are some fixed points on which the community has agreed upon since many years that simply cannot be ignored. For example the term 'validation' should be dropped for the time being in favour of 'evaluation'. This has been clearly stated in a number of important publications that cannot be neglected. Secondly, one cannot talk about 'validatio'n and than start two sections with "Intercomparison method" and "Intercomparison results". A comparison is between two or more things normally. The suffix 'inter' normally refers to a comparison of elements of the same nature, e.g.

model vs model, obs vs obs. If that would not be the case, one would simply talk about "comparison", would he/she not? I think that the natures of observation and model results are already sufficiently different, to complicate further the scene and inferring, with the used of 'intercomparison', that they are not. How about "methodology for the evaluation of the ensemble" and "Results", simple straight forward, clear?

I have serious problems with reading the figures. They are excruciatingly small and the number and nature of the differences between models and models/obs is so crucial to the evaluation of the manuscript quality that I cannot precede in a conclusive way.

The most important objection resides in the ensemble treatment and the fact that the differences among the models are confined in the appendix of the paper. The differences among the models qualify the ensemble and define also the quality of your final results. Once more the figures are too small for me to say something definitive here, but from what I can judge I see small differences among models. This puts in question the necessity for an ensemble treatment especially when based on the median which by definition cuts the outliers contribution to the ensemble result and in this particular case may well make redundant the use of several models that are replicating their results. May be they are different, but this is not visible to me from the figures provided.

I do know the value of ensembles of opportunity but the opportunity should be exploited at maximum making sure that there is an added value within the use of multiple models, that the number of models is adequate, not too many not too few and that the contribution from the model results finally used is original and un biased. This has been demonstrated in a number of works that deserve the attention of the authors.

I think the paper will benefit if the individual relationships among the ensemble members is brought to a higher degree of visibility (not only with larger figures but also conceptually) and analysis. This will increase the scientific significance of this paper which otherwise would look too much like a performance report. The later is useful indeed for the institution/s that use these results but is not at all instructive for the

scientific community.

What I find contradicting a bit in this paper is also the fact that data are used to validate an ensemble, use nature is obscure, and the main message that this brought forward is indirectly that this exercise demonstrates that Max-Doas data are suitable to validate models. So what is validating what and how?

In the present status the manuscript can not, in my view be published in ACP. GMD would be more suitable, but provided that more insight is given into the ensemble workings.

I am prepared to read a new version of this work should the Editor consider it necessary.

———————————————