Geosci. Model Dev. Discuss., 8, C3403–C3409, 2015
www.geosci-model-dev-discuss.net/8/C3403/2015/

Geoscientific Model Development Discussions
Open Access
EGU

GMDD
8, C3403–C3409, 2015

Interactive Comment

Interactive comment on "Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV/Vis tropospheric column retrievals" *by* K. F. Boersma et al.

K. F. Boersma et al.

boersma@knmi.nl

Received and published: 18 December 2015

We thank Reviewer1 for the constructive and positive comments. The reviewer's comments are indicated in italic font below.

This manuscript presents a methodology for properly comparing models for atmospheric composition with Level-2 retrievals from satellites in the UV/VIS. It does so by identifying three potential issues, testing the potential introduced errors for these issues and recommend procedures to avoid the errors. The manuscript is generally

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

well written (apart from Section 2; see detailed comments below) and the scientific methods are sound. However, the presented concepts and methods are not really new and the manuscript is in that sense a bit disappointing. Sampling issues and the use of averaging kernels have been discussed in the literature for some time now. Nevertheless, I support the publishing of this manuscript as a model assessment methods paper, subject to minor revision, because it provides a good summary of the issues and guidelines, which will hopefully encourage scientists in the field to make better use of these concepts.

The concepts and methods used in our study indeed exist for some 20 years now. But whereas the data assimilation community may generally be well aware of the issues at hand, we feel that the modeling community using tropospheric column retrievals of NO2, HCHCO, and SO2 from UV/Vis measurements is still struggling with representativeness errors, and generally lacks guidelines on how to perform a proper model-column comparison. This is partly because the UV/Vis retrievals do not always provide averaging kernels in their data products. Another reason is that the use of UV/Vis retrievals is relatively recent compared with the use of temperature and O3 profiles from satellite instruments in formal data assimilation systems, and new users lack a body of literature to guide them how to actually do that. Our paper indeed has the potential to encourage proper use of the UV/Vis satellite retrievals in the community.

Detailed comments:

Section 1: I am missing some important earlier work, such as the studies by Rodgers and Connor, JGR, 2003 and Migliorini et al., Monthly Weather Review, 2008. These could maybe be added in lines 16/17 of page 7827?

We added these relevant references in section 1.

P7827, line 11: The authors go a bit fast here. While the statement is correct, I had to read it several times to understand the logic. I would advise to add a sentence that explains that the contribution of the prior profile to the final solution increases with *decreasing sensitivity of the measurement.*

Thanks for the suggestion. We now added a sentence along these lines.

P7827, line 22-25: Assuming xa=0 makes the problem non-Gaussian, because xa cannot be negative. How does this assumption affect the discussion?

The DOAS retrieval is not an optimal estimation technique based on Bayes theorem. So this is not a problem. We refer to Eskes an Boersma [2003] for an extensive discussion of the retrieval theory for DOAS columns.

P7828, line 23: 'whereas satellite measurements provide "snapshots" at a particular local time': this is not a description of spatial error but of temporal error. Please remove.

We removed "at a particular local time" because this may be confusing.

Section 2.3 is a slightly odd section. In section 1 the authors say that section 2 will introduce the issues, while section 2.3 is actually used to describe the proposed solution. There is therefore overlap with text later in the manuscript, which does not help the reader. I suggest to either remove section 2.3 or make it part of text later in the document (e.g., Section 3.2). Also, the end of section 2.2 describes conclusions, which should not be part of this introduction.

Thanks for this excellent suggestion. We have moved section 2.3 to 3.2 and streamlined the text as necessary.

P7830, line 3: Should you not do the opposite in this case, i.e., average over all the grid boxes that fall within the satellite foot print?

That is in principle possible, but at the cost of no longer evaluating the model on its own grid, or a regular grid at all. Such evaluations are sometimes done on a daily basis (e.g. Huijnen et al., ACP, 2010). Ultimately, they still re-grid the satellite footprints to a coarser grid for final seasonal mean comparisons (maps, regional averages). Regridding the model-retrieval pairs from the irregular, daily varying satellite grid to a

GMDD
8, C3403–C3409, 2015

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

coarser common grid, brings the total number of manipulations on three (1. model-to-satellite, 2. model-at-satellite back to model grid, and 3. satellite to model grid), whereas the recipe in Eq. (3) only requires one such manipulation.

Section 2.4 is slightly strange as well. Of course, one has to be aware of clouds and their impact on the retrieval. But wasn't the whole point of recommending the use of averaging kernels exactly to deal with these kind of issues. Maybe, the authors could mention the impact of clouds in a shortened paragraph earlier in the manuscript and then present the use of averaging kernels later in the manuscript both in the context of vertical sensitivity in general and in the context of clouds.

Thanks for the suggestion. We have now merged the section on the influence of clouds with section 2.1. In that section we also state that it is in principle possible to use the cloudy retrievals as long as the kernels are being applied.

P7832, last paragraph: Treating systematic errors as random is a strange assumption considering the fact that the authors acknowledge later on that the systematic model errors can be as large as 50 percent. This needs some further clarification.

We admit that this was formulated inadequately. It is better to interpret Eq. (6) as providing an overall envelop of the model-retrieval comparison error, where the individual (model, retrieval, representativeness) error contributions may have substantial systematic components. While some of these systematic contributions may cancel out in an area-averaged, monthly mean model-retrieval comparison such as presented in Section 6, robust systematic error will persist. We therefore removed the sentence 'Here and in Sect. 6 we treat them as random errors, an approach usually followed in retrieval and data assimilation studies (Rodgers, 2000)'.

P7833, lines 10 – 12: I am not sure if I understand this argument. Averaging will remove the random component from the error budget, so systematic errors will actually become more dominant. Or do I miss something?

In general, a priori data that is always and everywhere biased in the same way will then indeed become dominant. But a priori data can also suffer from systematic errors that are variable in space and time. Our argument has its origin in the use of a priori data in the retrieval (e.g. surface albedo) that may be systematically biased high for pixel 1, but biased low for pixel 2. Averaging retrievals over a large area, will then tend to cancel the effect of systematic errors in the a priori data, as long as the mean systematic error (in the albedo) is zero.

Section 3.2: Maybe the text of Section 2.3 could be used here.

Done.

Section 4.1: It is unclear what the relationship is between the model profile used for the AMF and the a priori profile (set to zero under the weak absorbing assumption) used in the retrieval. Both prior assumptions play a role in the definition of the end product, but it is not clear how this is accounted for in the averaging kernel.

The a priori profile xa used in the air mass factor calculation is typically taken from a CTM. The a priori profile is used in the calculation of the AMF, and also in the averaging kernel, as discussed already in section 2.1, right below Eq. (1). The model that provides xa may or may not be the same as the model under evaluation. In our study, the a priori profiles xa in the OMI NO2 retrieval are from the TM4 model, and the models that are being evaluated are TM5 and GEOS-Chem. We decided to not go into much detail here, since this has been extensively described in many retrieval papers including our own (Eskes and Boersma [2004], Boersma et al., [2003; 2007; 2011]).

Page 7837, eq. 8: This was already described in Section 2.3

Good point. We removed Eq. (8) and instead refer to Eq. (3) now in the text.

Page 7838, lines 2 – 4: Does this mean that the DOAS error estimates are not correct? If they would be, they should be taken into account. I am not convinced by the argument that the method drives the statistical interpretation of the results.

The DOAS error, calculated from formal error propagation, is our best estimate of the uncertainty in a single retrieval. We stress that here we calculate an area average of individual retrievals that is the best spatial representation of the retrieved column over a large, grid-cell area. We take the individual errors (and their spatial correlation) into account when computing a superobservation error as in Eq. (9) in the revised manuscript.

Page 7841, line 13: Are the TM5 cloud fractions indeed simulated by the model or do they come from ERA-Interim. If the former is the case, it would be worth mentioning how these cloud fractions are simulated. If it is the latter, this should be mentioned as well.

The TM5 cloud information comes from ERA-Interim, as stated already in Section 4.2 and in 5.2. The details of the cloud fraction simulation are described in Dee et al. [2011].

Section 5.3: Although the presented comparisons do indicate a better agreement between model and observations, it is not rock-solid proof. There remains the possibility that the use of averaging kernels is masking/compensating other errors in the model. This is probably worth mentioning for completeness.

It depends on what is understood as 'proof'. Our point is not that the model is doing better if the kernel is applied, but rather that if the model is sampled as if the satellite would observe the modeled state, the model and retrieval can be properly compared according to the retrieval's capabilities. Such a comparison allows for a better-constrained evaluation of the model. Of course some model errors may still go undetected following such an evaluation, but the evaluation is no longer biased because of the difference in vertical sensitivity of the retrievals vs. the model.

Appendix D is not referenced in the text. However, it contains interesting content that could maybe be used in Section 4.1? See also my earlier comment about Section 4.1.

GMDD

8, C3403–C3409, 2015

Thanks for this nice suggestion. We don't think it would be useful to include it in Section 4.1 that is about the retrieval settings. Still, we think referencing Appendix D is a good idea, and we do that now in Section 5.3, because this is where the actual application of the kernel is discussed.

Interactive comment on Geosci. Model Dev. Discuss., 8, 7821, 2015.