

## ***Interactive comment on “Evaluation of version 3.0B of the BEHR OMI NO<sub>2</sub> product” by Joshua L. Laughner et al.***

**Anonymous Referee #1**

Received and published: 19 September 2018

In this manuscript, the authors evaluate the performance of their updated tropospheric NO<sub>2</sub> vertical column product (and the model output needed to derive it) against ground-based vertical column observations and in-situ aircraft profiles. They find that, in general, the updated retrieval is an improvement over the previous version and over the latest NASA standard product. The paper demonstrates the improvement in retrieval performance that is achieved by using daily a-priori NO<sub>2</sub> profiles over monthly averages. This paper also makes a number of pertinent key arguments that will be of interest to the wider community, e.g.: (1) that there is a lack of measurements in the southeast US to constrain sources of uncertainty in the retrieval of NO<sub>2</sub> in that region; (2) the need for further development in dense sensor networks that can evaluate contrast between pollutant plumes and background concentrations; and (3) the value of

C1

using slant column densities to evaluate the spatial distribution of NO<sub>2</sub> predicted by a model.

Overall, this manuscript is detailed and clearly written. It covers subject matter that is certainly relevant to the AMT audience, and should be of interest to the community. I would recommend this paper for publication in AMT, and make some small recommendations that could be considered prior to acceptance.

General Comment:

This manuscript seems to focus largely on how the updated model chemistry/emissions and the daily vs. monthly averaged a-priori profiles impact the performance of the BEHR NO<sub>2</sub> retrieval. But there were other pertinent updates in the retrieval: varying tropopause height, directional surface reflectance, and a new combined surface pressure dataset. Are the differences in retrieval performance against observations as a result of these updates very minor compared to the differences attributed to the updated model chemistry and profile temporal resolution? It seems worthwhile mentioning/commenting on. I wondered whether there were any specific instances/cases where some of these other updates could be relatively more important.

Detailed comments:

1) Abstract: I suggest including the detail that BEHR is focused on retrievals over North America in the abstract.

2) Introduction: The introduction includes some background on previous evaluation efforts for the NASA SP and KNMI DOMINO products. Given the focus of this paper on evaluating this latest version of the BEHR algorithm, some details on how previous versions of this algorithm have been evaluated (and its performance) also seems relevant. This could add more motivation/context for the necessity of updating the algorithm (in addition to the already cited work from Laughner et al. 2016 that focused on the importance of daily profiles).

C2

3) p. 5, l. 29-30: The authors mention using the GEOS-Chem global chemistry model to extend aircraft profiles to the surface. It could be relevant here to include at least the horizontal resolution of the model output used. Is the model output identical to the model experiment run cited in Nault et al. (2017)? It wasn't clear to me whether these authors were directly using output from that experiment.

4) p. 6, l. 18: I wonder whether some summary statistics might more easily advance the authors' argument that model output from v3.0 "show better agreement" than model output from v2.1. At the moment, there is only a description of the more obvious qualitative details.

5) p. 6, l. 31: "since the strongest lighting occurs. . ." Replace "lighting" with "lightning"

6) p. 9, l. 29: I suggest inserting "modeled" between "daily" and "profiles" so the point is very clear.

7) p. 12, l. 2: The word "adequately" is used rather subjectively here. I know it is discussed later in the manuscript, but I wonder whether there is any way to quantitatively evaluate just how well the WRF-Chem NO<sub>2</sub> profiles are indeed capturing the day-to-day variability. If not, perhaps this caveat could be more clear, and you could mention here that you propose some suggestions later in the manuscript.

8) p. 12, l. 12-15: This explanation of negative VCDs in the DISCOVER-CO dataset is unclear to me.

9) p. 12, l. 28-30: The authors introduce the notation of "BEHR v.30 (M)" and "(D)" for the first time here, I think. While it's fair to say it is obvious, the authors could explicitly clarify that "(M)" refers to the product using monthly average profiles, and "(D)" refers to the product using daily profiles.

10) Table 2 and 3: I was wondering whether there would be value in reporting correlation coefficients in addition to slopes. Can the authors explain why they haven't included these in their evaluation of the product performance? For example, I won-

C3

dered whether they can demonstrate that in addition to improving a bias, using daily profiles (D) explain more of the variability than monthly profiles (M) alone.

11) Section 4.2: Can the authors clarify why they have chosen to separate their evaluation by looking at all the Pandora data, vs. just the Pandora data during coincident aircraft spirals? Is this meant to demonstrate how continued long-term monitoring is superior to short-term campaign coverage for evaluation purposes?

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-248, 2018.

C4