**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Evaluation of a regional air quality forecast model for tropospheric NO$_2$ columns using the OMI/AURA satellite tropospheric NO$_2$ product" *by* F. L. Herron-Thorpe et al.

F. L. Herron-Thorpe et al.

farrenthorpe@wsu.edu

Received and published: 9 July 2010

Referee said: The paper discusses a comparison between NO2 column measurements made by the AURA-OMI instrument and AIRPACT air quality model results. A reasonable agreement is found in winter months. In summer an interesting large difference is found related to wildfires, where the model produces much higher NO2 values than OMI.

To my opinion the paper in it's present form is not ready for publication, and a major

revision is needed. The main reasons for this judgement are the following:

1. The authors should be more careful in the formulation of the method and results. My (many) detailed comments given below illustrate this point. Discussions are often unclear, not well formulated and sometimes not based on evidence.

In response to this comment we have drastically revised the methods and results. Furthermore, discussion of data not presented explicitly in the paper has been omitted (i.e. specific daily wildfire biases).

2. The conclusion from comparisons with and without kernels leaves the reader confused. Is it important to account for the kernels? What are the quantitative errors if the kernels are not used? The discussion is mentioning several issues (computationally expensive, unrealistic high values) which seem to be quite specific for the kernel implementation used by the authors and can not be generalised to other NO2 comparisons. Kernels are applied only for one month, and not for the full 18 months. The abstract mentions: "applying the averaging kernel in cloud free conditions has little effect". but fig 3 suggests that the a-priori profile has large limitations, which should influence the cloud-free results. How do these findings match?

In response: The previous kernel discussion was based only on computations derived from Eric Bucsela's code for computing averaging kernels and tropospheric columns. The problems with erroneous values was a known problem with the integration technique used in Eric's code, so we had little confidence in doing a full-scale comparison of those averaging kernel results at the time. Since then we have exported the averaging kernels from his code and computed our own resultant columns and found very large differences (discussed in the newest version of the paper). Results from applying the kernels are now found to be quite significant and are discussed in this newest version. Discussion of computational expense has been deleted.

3. The OMI product from KNMI also contains kernels. It is a pity that these kernels have not been used. For instance it would be interesting to compare the kernels from the two

products. Although many results are also shown for the KNMI product, a description of the two algorithms is not provided in a balanced way. It is not clear what the main differences are between the two algorithms.

In response: At the time this work started, KNMI averaging kernels were not available. In fact, the original paper submitted did not include any KNMI results. We added some KNMI results as a response to the first ACP submission discussion. After this referees comments we went back and computed the averaging kernels, and created related figures for this paper. This newest revision includes using the KNMI averaging kernels. We have also expanded the discussion of the KNMI product and algorithm to balance the NASA description.

4. The discrepancies observed during wildfires are interesting, but are not really discussed. What are the problems in modelling these fire emissions and what is the error estimate on this? Is the model overestimating? The new experiences with "BlueSky" may help to illustrate model uncertainties. What are the issues in the OMI retrievals which may result in a large underestimate of the NO2 from fires? What is the most likely cause for the discrepancy (OMI or model)?

In response: We realize our previous version was not clear on this topic (as also pointed out by the other referee). We have made this section more clear. We are confident that fire emissions in the model were too persistent and not updated quickly enough due to the older version of BlueSky that was used (relying on ICS-209 reports). This is discussed in the newest version of the paper.

5. There is too much discussion about emission estimation. On this topic there are no results presented, and therefore only a brief mention of the inversion plans (and literature on inversions) is justified. The introduction and the end of the conclusions section would be a good place for this. The authors seem to be too optimistic about the inversions: from the uncertainties, the wildfire issues and differences between the retrievals it is not clear to me that one should even attempt to use OMI for emissions

estimates.

In response: We agree and have removed much of this discussion.

Detailed comments: "SCHIAMACY" should be SCIAMACHY (in references and introduction)

In response: Done.

p. 66, l15: "but unfortunately the NRT product does not represent the best calculation provided in the official data collection". Please explain why the NRT product is less accurate.

In response: Done. This has been edited to explain that: "This NRT data is valuable to applications needing access as quickly as possible (i.e. field studies), but is not guaranteed to have uninterrupted data delivery or be at the same quality as the official data collection. Also, users of OMI NRT data are not allowed to distribute the data as it may have problems or may not be a product of the best algorithms"

p 67, l13: "to the height of stratosphere-troposphere exchange". Is it meant in this way? I would expect "to the height of the tropopause".

In response: Done.

p 68, l20: "require the use of radiative transfer models and a geo-referencing scheme to identify areas where there is commonly pollution in the boundary layer." Radiative transfer models are needed in most remote sensing retrievals (rather trivial statement). The geo-referencing refers to the masking applied to estimate the stratosphere, I guess. But this depends on the approach and is not required. For instance a reference sector (over pacific) approach has been widely used which I would not classify as geo-referencing (it is a more simplified assumption).

In response: This has been edited to read: "As discussed in Bucsela et al. (2006), the standard analysis of Earth spectral radiance measurements from OMI makes use of a

radiative transfer model and a geo-referencing scheme to determine trace gas column abundances."

Please motivate the choice for < 35% cloud cover ? Other choices have been made by other groups. Are the results in the paper sensitive to the exact limit for cloud cover?

In response: We have added the following: "The standard cloud-screened tropospheric NO2 product from OMI (level-3) uses a cloud fraction cut-off of 30% (Celarier, 2009), while some studies use pixels with cloud fraction up to 40% (Mijling, 2009) when there are limited cloud-free days. The Pacific Northwest region has frequent cloud cover that significantly limits the number of pixels available with less than 30% cloud fraction. Using OMI pixels with <40% cloud fraction significantly changes the calculated monthly average values because of the large "below cloud" additions. For these reasons, our analysis is only relevant for early afternoon abundances of tropospheric NO2 for pixels with little cloud cover (<35%)."

p 69, l9: "for pixels not dominated by point sources". Why is the analysis not relevant for point sources? Clearly the footprint size of the satellite should be accounted for, but when the appropriate averaging is applied I do not see a problem.

In response: Although this exact phrase is no longer present in the newest version, we still feel that point sources in our domain cannot be individually evaluated. If the point sources in our domain were isolated it could be possible, even accounting for footprint size, but the inclusion of nearby mobile, biogenic, and area sources do not allow this. The averaging kernel accounts for vertical instrument sensitivity, not horizontal resolution. Furthermore, the averaging kernel results suggest that sensitivity of NO2 is near zero at the surface, making it inappropriate to estimate individual point source emissions.

p 69, l10: "Our overall goal is to make future NOx emissions adjustments based on previous OMI/AIRPACT comparisons". This is not the subject of this paper, so irrelevant for the discussion. I suggest to remove this line.

In response: Done. In lieu of the changing direction of this paper this has been removed.

p69, l22: "(http://www.doas-bremen.de/doas glossary.htm)". It is uncommon to refer to such websites.

In response: Done. This reference has been removed.

p69, l23: "An 18 month trend analysis". "Trend" is a word commonly used for changes over time, for periods > 20 years. The use of the word "trend" is confusing.

In response: Done. This wording was found in multiple areas in the paper. One of the authors identified this problem as well and made the appropriate changes.

p70, eq 1: Why is this called "density". It is the total amount of molecules in the layer per 1cm by 1cm surface area.

In response: No Change. This is the common nomenclature used in DOAS based remote sensing. It is not referred to as a "density" but a "column density" which is commonly measured in units of number per area (rather than per volume). It is the number of molecules in the column. The height of the column is factored out of the reported value because the sensitivity of a remote sensing instrument is nominally equal in the x and y dimensions, but much worse in the z. A nadir looking UV/VIS instrument cannot resolve changes in the profile. Essentially, the number of molecules below the instrument are added up, but the instrument does not know the depth of the column, and an assumed profile is used but not "resolved". However, AIRS is able to accomplish this by working in the infrared and "slicing" out pieces of the atmosphere where the temperature is homogeneous. It observes different channels of the spectrum where the molecule cross section has a unique temperature response. So, AIRS is able to report fractional abundance (volume based rather than area based) because the layers have relatively homogeneous temperature and number density.

p70, eq1-2-3-4: These equations are standard, and almost trivial. It would suggest to

remove these equations and to simply mention that the layer concentrations have been converted to layer column amounts, in molecules/cm2.

In response: We prefer to keep this section as it is our intention to provide other CMAQ users with a useful method of calculating VCDs. However, it has been edited to be more clear in that intention.

p70, eq 1-4: It is a bit surprising that the density is calculated by the ideal gas law, eq 2. It would be more accurate to use a total air mass per model grid box. Is this not available in CMAQ/MCIP?

In response: Total air mass is not a variable from CMAQ/MCIP. Not all gases are modeled in CMAQ. Using the ideal gas law is not inaccurate as it is one of the fundamental equations used in CMAQ and CMAQ is known to be mass-inconsistent. Furthermore, VCDs are number density based, not mass density based, so it is more straight forward to use a number density derivation of VCD.

p71, sec 2.2: Since two OMI products are compared in this paper, it would be useful to the reader to provide a summary of the main differences between the two retrievals (and possible impacts of those differences if known). Section 2.2 only describes the NASA approach, and does not discuss the KNMI approach.

In response: This newest version of the paper has been significantly expanded to include the KNMI approach as well.

p71, l6: I believe that these details on the NASA-OMI algorithm are documented. So why do you refer to "Eric Bucsela, private communication"? It is better to cite a refereed paper, in this case Bucsela 2006.

In response: We have changed this reference to Bucsela, 2006. However, not all details discussed in the paper are stated in Bucsela's publications. We have tried to reference peer-reviewed materials as much as possible throughout this paper, but having worked so closely with Eric through this work it is not possible to find everything he

C12428

has shared with us in his journal publications. This is especially true with his averaging kernel code which is not an operational product.

p71, l14: "where corresponding points from AIRPACT are masked from averaging as well.": How is this done? Is this based on spatial collocation, or on the cloud fraction in the AIRPACT model system? Is it done for AIRPACT or for the spatially averaged AIRPACT results?

In response: This wording has been changed to be more clear: "To account for this, our monthly average calculations are limited to using pixels with less than a 35% cloud fraction reported by OMI, with corresponding daily values from AIRPACT masked from averaging."

p71, l15-22: "This can be a relevant source of error when comparing model results to satellite derived columns." This paragraph discusses meteorological variables and their impact on the OMI retrieval, but the conclusion is vague. Are there papers that discuss this effect ? Is number density really needed in the retrievals, and, if so, is it a large source of error (provide order of magnitude)? If the authors can not answer these questions it is better to remove this paragraph.

In response: P and T differences on average would be fairly small - using the ideal gas law for P and T. Say T is off by 10 deg, that is 3% and P off by 50 mb that is 5% so in quadrature it is < 5%, an ignorable error. We have edited the newest version of the paper to read: "Using the ideal gas law, simple calculations of variance in number density due to a few degrees of temperature or tens of millibars suggest this error is small. However, significant retrieval errors may occur when the surface pressure used is not accurate, especially over regions with complex topography as discussed by Boersma (2007)."

p71, sec 2.3: NASA averaging kernels are used. The KNMI product also includes averaging kernels. Is there a reason for not using those as well? It would be interesting to see a comparison between the two kernels.

C12429

In response: At the time this work started, KNMI averaging kernels were not available. In fact, the original paper submitted did not include any KNMI results. We added some KNMI results as a response to the first ACP submission discussion. After this referees comments we went back and computed the KNMI averaging kernels, and created related figures for this paper. We have also expanded the discussion of the KNMI product and algorithm to balance the NASA description.

Fig.2: Please make a clear distiction which boxes have been computed by the authors, and which boxes are computed by NASA (KNMI) as part of the retrieval process. In the averaging kernel computation there are quantities dAMFCld and dAMFClr. Where do those come from? The text mentions "IDL routines and lookup tables". What do these routines require as input, and what do they provide as output? Please provide such links between the main text and the boxes in Fig.2. The authors use quite some text and a figure to explain the processing, but the actual computing steps remain unclear.

In response: This figure has been removed and related text in the paper has been clarified.

p72, l8: "However, the latest version of KNMI data includes the averaging kernels in the daily level 2 data files." Why have not these kernels been used? Apparently these kernels are provided as a matrix, and it would be numerically cheap to apply them!

In response: They were not actually in matrix form and require separate 3D calculations per OMI pixel. However, they were used as requested. Results are discussed in this newest version.

p72, l15: "used by OMI algorithms". This suggests that both the NASA and KNMI retrievals use GEOS-CHEM ?

In response: This has been clarified and now discusses that KNMI uses the TM4 model, not GEOS-CHEM.

p73, l12: "Despite the numerical intensity involved with applying the NASA averaging

C12430

kernel, spatially averaging the model results to the daily OMI swath requires only a simple additional function in scripting as compared to independent comparisons." This is an incoherent sentence: what is the link between the first and second part? Perhaps this line can be removed.

In response: The discussion topics of "numerical intensity" and "computational expense" has been removed from the paper.

p73, l15: "It is a useful and efficient method to adjust model results based on the variance in OMI footprint size throughout the swath." I do not understand this remark. Please explain (or remove).

In response: Done. This sentence has been removed.

p74: Is the kernel applied to the original AIRPACT or to the averaged AIRPACT data? Please provide details how correlations are computed: is it a spatial correlation for the monthly-mean AIRPACT and OMI distributions? What formula is used? Computing correlations is standard, but nevertheless in my experience different expressions are used by different groups (differing especially in the calculation of the means/references).

In response: The kernel is applied to the original AIRPACT (daily). This is described in the new section describing the averaging kernel methods. Note: It is not proper to apply a daily averaging kernel to a monthly averaged model run. Pertaining to correlation, it is a linear correlation (Pearson's r) of spatial distribution of monthly averages. We have edited the introduction of the methods to make this more clear. Sentence now reads: "To properly evaluate the linear correlation (Pearson's r) between OMI and AIRPACT VCD monthly average spatial distributions, we spatially averaged the AIRPACT grid to the pixels within the daily OMI swaths."

p74, Fig 4: It would be nice to include a panel showing the lat-lon distribution of the kernel-applied columns. This will help to appreciate the impact of the application of the

C12431

kernel.

In response: Done. Kernel applied columns are shown in new figures.

p75, l2: "This sometimes leads to very large values if the product of the layer mixing ratio and pressure equals the layer above it." I do not understand how this can be a problem?! Very large values of what? The averaging kernel?

In response: The older version of the paper was trying to explain the problems with Bucsela's averaging kernel code (integration problems). This analysis has been altered to use our own integrations but still discusses the results from his code. The newest version of the paper reads: "When using this method in the summer, the background increase does not occur across the domain, but erroneously large values appear in areas that otherwise had no significant NO2 concentrations, as shown in Figure 9. These values tend to be over 1 x 1017 molecules per square centimeter and occur when the integration technique used overcompensates for successive layers that are numerically similar. The numerical results of the code are less reliable for computing monthly averages because of the erroneously large values. However, it does add another useful perspective when comparing AIRPACT's summer months to the NASA product."

p75, l9: "Overall, the same general trends are found in the modeled columns". Please be more precise. What is meant by "trend" here? As mentioned before, the word "trend" is used in several places and is confusing to me (apparently has a different meaning to the authors).

In response: Done. This wording was found in multiple areas in the paper. One of the authors identified this problem as well and made the appropriate changes.

Fig. 7: Concerning the model range (min-max values): please specify if this is calculated for the non-averaged, averaged or kernel-applied model values.

In response: This figure showing daily results has been omitted.

p75, l25: "In general, the interpretation of bias trends changes much more when we

C12432

spatially average AIRPACT results, as opposed to applying the averaging kernel and masking the erroneous data." I do not understand this statement. Please reformulate.

In response: This discussion has been omitted and is no longer relevant.

p76, l7: " the spatially averaged AIRPACT values give a better representation of what should be directly compared to OMI". This is a statement which is out of place. Spatial (horizontal) averaging is unrelated to the kernel issue (which is related to the vertical profile), and the neglect of the kernels can not be compensated by spatial averaging, as this sentence seems to suggest. It is also not fully clear to me if the spatial averaging is really a better representation of OMI. (The figures seem to suggest it is, but there may be other reasons for a better comparison)

In response: This discussion has been omitted. Any discussion of spatial averaging is limited to that corresponding section.

p76, l20: teem -> team

In response: This sentence is no longer present in the paper.

p76, l20: It is well known that instruments like OMI are able to observe biomass burning signals. For instance the African biomass burning and it's seasonality is well observed and in reasonable agreement with models. Why should it be different for these wild fires? Again, referring to a discussion at the science meeting is not convincing: is there evidence in the literature to confirm this statement that fires can not be observed by OMI.

In response: This section has been edited to be more clear about the problem we found in AIRPACT fire emissions. We were not trying to suggest that OMI's sensitivity to the boundary layer NO2 was to blame for the discrepancy, but rather it was a possibility that was considered. The edited section explains that we see that temporally averaged tropospheric column NO2 densities in wildfire areas display very large discrepancies and in fact AIRPACT is an order of magnitude higher than OMI for many monthly aver-

C12433

ages. This is attributed to the way fire size and progression were estimated in BlueSky which relied only on ICS-209 ground reports.

p77, l7: "This is in contrast to the slow periodicity of high NO2 values retrieved by OMI". What do you mean by "slow periodicity"?

In response: This has been removed.

p77, l 22: "July through December shows much higher OMI". It is useful to mention the difference between the NASA and KNMI product in this case.

In response: We agree and new figures showing their differences have been added as well as discussion.

p78: It is mentioned that the KNMI product has negative values and suffers from "stripes". I was a bit surprised that these effects are not smoothed out in the monthly averages, and have such an impact on the correlation (which are spatial correlations computed for the monthly mean field if I understand correctly).

In response: The pacific northwest has many cloudy days and so the number of values going into a monthly average may be lower than 10. If more than one of these has a negative value it can dominate the monthly average. Most values in NASA and AIRPACT are above 1E14 molec/cm/cm. KNMI averages severely affected by negative pixels can often be lower than 1E12 molec/cm/cm or below zero. We have added further discussion of KNMI negatives and added figures that show the numerical distribution of monthly averages.

p78, l17: "current collection 3 of OMI tropospheric NO2 provided by NASA seems to cause a systematic trend of higher values in the summer." Is this referring to e.g. Fig. 6, or is this taken from the literature (if so, provide a reference)? p78, l22: "However, all long-term NASA timelines show a clear anti-correlation with season." Same comment, please provide reference if available. p79, l7: Same comment: Are these conclusions drawn from this work. If not, is there a reference? (Where do these conclusions come

C12434

from?)

In response: This discussion has been revised and now includes reference to figures in the paper as well as information from Bucsela. The newest version of the paper now reads: "Tropospheric NO2 columns calculated by NASA are lower during the winter (Fig. 5d) and higher during the summer (Fig. 4d) due to the way that the tropospheric column is calculated in the OMI NO2 algorithms. This annual cyclic variance in tropospheric NO2 found in the NASA product (Fig. 6) is not real, but rather an artifact of the known issues with the NASA OMI NO2 stratospheric correction. Bucsela (private communication) has identified these issues as: 1) the small amounts of tropospheric contamination in the data used to derive the stratosphere, 2) failure to account for the diurnal cross-track variation of the stratosphere, 3) issues introduced in the wave-2 interpolation, and 4) the determination of stratospheric AMF from an annual mean profile. Wave-2 interpolation issues include the shapes of continents in the masked regions and the potential to hide planetary-scale structure in the troposphere (e.g. from lightning NOx) and small-scale structure in the stratosphere. Stratospheric NO2 is highest in the mid-latitudes during the summer (Cohen, et al, 2003) and NASA's tropospheric NO2 algorithms do not seem to subtract enough of the stratospheric contribution for those months."

p79, l16: "However, this effect may be expected naturally: less sunlight is incident on the airshed during cloud covered cool months, so less NO2 is photolyzed to NO." This is too simple. The lifetime of NOx, conversion to NOy species and removal through e.g. HNO3 play a crucial role in the difference between summer and winter.

In response: This statement has been revised. It now reads: "In fact, nearly all urban KNMI averages were largest in cooler months. This is matched to a smaller degree in the AIRPACT results and suggests that NO2 levels near urban areas are higher in winter than in summer. The cause for this winter/summer difference is probably a complex balance between less photolysis of NO2 in winter associated with lower sun angles and greater photochemical processing of NOx to HNO3 in summer."

C12435

p80, l1: "wildfire emissions would require a 4-D-var analysis using near real-time NO2 retrievals". Why is near-real time needed? Why is 4D-Var needed? When the fire location and timing is known, it seems that a scaling of the source strength can be applied to match OMI as a cheap way to do the inversion. However, from the earlier discussions on the fire comparisons it seems to me that the biggest problem is the characterisation/understanding of the OMI measurements: which fraction of the fire produced NOx is actually observable by OMI (given the smoke and clouds).

In response: This discussion of targeting emissions by dominant source type has been removed because our results suggested any biases were regionally based and not source-type based. Subsequently, this discussion mentioned by the referee has been removed.

p80, l3: "DOAS satellite retrievals do not readily resolve boundary layer concentrations". I do not understand this remark which contradicts the results presented. My impression is that most of the features observed in e.g. figures 4 and 5 are to be interpreted as originating from the boundary-layer. At least the city hotspots should be interpreted as NO2 in the boundary layer.

In response: This wording has been removed. The averaging kernel figures and discussion now makes the point that sensitivity to concentrations near the surface is very low.

p80, l15: "Ultimately, we have decided it would be best to make adjustments based on the average of the two data sets." This choice is completely arbitrary it seems ...

In response: This line has been omitted.

p80, l22: " long term trends". The word "trend" has been used several times in the paper and is confusing. What is "long term" in this case: 18 months, 5 years, 1 month?

In response: We agree and the word "trend" has been removed from the paper.

p81, l3: "computationally expensive". This is not a general problem, but is quite specific

C12436

for the way the authors derive the kernel for the NASA product it seems.

In response: We have removed this discussion

p81, l3: "erroneously large". I guess the OMI NO2 column cannot be trusted as well in these cases ?!

In response: This statement is specifically in regards to the integrations from Bucsela's code. We have removed this point of the paper as much as possible by redoing the analysis without Bucsela's code.

p81, l9: "significant number of trends". Please use other words, e.g. "conclusions"

In response: Done. The newest version of the paper no longer incorrectly uses the word "trend".

p81, l23: "BlueSky". Are there already indications that the quantitative emissions are substantially smaller in this model ?

In response: The indication is that fire size and progression can be updated more frequently using new BlueSky technology. Due to the infrequency that ICS-209 reports are filed by firemen, we think the modeled fires from older BlueSky technology were allowed to persist at max emissions for too long. The newest version of the paper discusses this problem succinctly.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 27063, 2009.