We would like to thank the reviewer #2 for his/her helpful comments that improved this manuscript. Below in *italics* please, find our replies to the reviewer's comments. Following their comments, we have thoroughly revised the manuscript as outlined below:

(1) We have added a new figure (and related discussion) to evaluate the extension of aircraft profiles.

(2) We have included a new figure showing summary of comparison at all DISCOVER-AQ sites.

(3) Presentation of the manuscript is improved following the suggestions from both reviewers.

---

## Reviewer: #2

In this manuscript, Lamsal et al. evaluate the most recent operational NO2 product from OMI by comparison to a number of validation measurements (from aircraft, Pandora, MAX-DOAS and surface in-situ instruments) as well as by indirect validation with the US NOx emission data base. Their main result is that the OMI NO2 product is in reasonable to good agreement with all the validation sources used, but that individual retrievals can show large differences for a number of reasons including a priori data used, spatial sampling, and measurement uncertainties in the validation data. The paper is well written, reports on the validation of an important satellite data product and provides a number of interesting and convincing new results. As already stated in my quick review, I think it would have matched the scope of AMT better, but I can also recommend it for publication in ACP. There are however several important points which the authors should consider before submitting a revised version of their manuscript.

### Major Points

1. **Limited geographical coverage:** The main problem of this paper is that it tries to provide an evaluation of the global operational OMI NO2 product but only uses aircraft spirals over 6 sites in Maryland during July 2011, a seasonality of Pandora measurements in Hampton, VA, MAX-DOAS measurements at two sites in Japan, and 2 (arbitrarily?) selected SEARCH surface sites. While this is better than many previous studies, it cannot provide serious constraints on the uncertainty of a product covering most of the globe in different seasons and under widely varying cloud, aerosol, NO2 profile and surface reflectivity conditions. I think the authors have to acknowledge clearly in the abstract, text, conclusions, and if possible also the title of the paper that their results are limited to certain regions, seasons and conditions.

*We agree with the reviewer that the validation study is still limited in scope due to scarce and sporadic $NO_2$ measurements. We have acknowledged this in the abstract, text, and conclusions. For example, we have added "Since validation data sets are scarce and are limited in space and time, validation of the global product is still limited in scope by spatial and temporal coverage and retrieval conditions" in abstract, and "The spatial and temporal coverage of the* *comparisons we have examined in this paper are limited; they may not be representative of other locations and seasons" in the conclusion section.*

2. **Extrapolation of aircraft profiles:** In their analysis, the authors extend the aircraft derived profiles towards the surface using the last measurement point and the gradient of the model profile. As is obvious from Figs. 2 and 3, the NO2 value in the lowest layer has a large impact on the shape of the NO2 profile and thus the column and the AMF derived from it. It is based entirely on the (shape of) the monthly GMI profile as none of the aircraft profiles shows indication for such an increase in NO2 towards the surface.
As I expect most of the spatial and temporal variability of NO2 in the lowest layer, the method used will systematically underestimate the effect of profile assumptions on the AMFs and thereby on the tropospheric columns in Fig. 4.
I think the method used for profile extension and the implications this has on the interpretation of results should be discussed in more detail.

*We have evaluated the aircraft profiles extended to the surface by using coincident in situ (photolytic) measurements at a DISCOVER-AQ site in Padonia. A new figure and relevant discussions are now included in the manuscript. The following text has been added: " We first evaluated the extrapolation scheme by comparing the estimated surface $NO_2$ mixing ratios with $NO_2$ measurements from a photolytic converter instrument at Padonia. Since $NO_2$ measurements at the lowest aircraft altitude are on average 45% lower than the measurements at the ground, extrapolation of aircraft profiles by assuming a constant mixing ratio from the value at the lowest aircraft level will substantially underestimate the true $NO_2$ near the surface. In Figure 3, we show a comparison of our estimates using Eqn. 1 with surface measurements at Padonia. The extrapolated and measured values are well correlated (r = 0.64, N = 14), and generally compare well (mean bias = 23%), although extrapolation could at times overestimate observations when the aircraft encountered elevated plumes with high $NO_2$ concentrations".*

3. **Statistics:** In spite of the large number of spirals flown and Pandora measurements taken during DISCOVER-AQ, there only are around 10 values per location in Figs. 4 and 6. I'm not convinced that computing the correlation makes a lot of sense for data sets having so few points, in particular if they are all from a period of less than 30 days in a limited geographical region. I'd therefore suggest adding two more panels to Figs 4 and 6 each, showing the full data sets in a scatter plot such as in Fig. 8, separately for standard and aircraft a priori profiles.

*We agree. As suggested by the reviewer we have added a summary plot using data from all DISCOVER-AQ sites.*

4. **Model comparison:** I do not see any added value in section 5. Numerous comparisons between OMI NO2 data and different model runs have been published, most of them applying proper data sampling and averaging kernels. I do not see anything in this section that extends upon what is already in the literature. In particular I do not see how this section justifies the statement in the conclusions reading "Finally, we investigated the potential improvement of the retrievals that could be realised using a high resolution model, with updated emission inputs, as a source of a priori profiles." Improvements can only be documented by comparison to independent results and attribution of improvements can only be done if one thing at a time is

changed, not everything (model, resolution, emissions) in one step. I'd therefore suggest removing section 5 and all figures and references linked to it.

*Our main motivation for this section is to demonstrate a proper use of scattering weights that are made available to users to help interpret satellite-model inter-comparison. There have been several user requests to provide documentation on the use of scattering weights, and we strongly believe that this section serves the purpose. To clarify the purpose of this section, we have modified the title of this section as "Use of scattering weights in applications of OMI to evaluate AQ models".*

**Minor Points**

• page 14524, line 11: I don't think this manuscript adds anything new on "objective methods to compare model-simulated NO2 columns with satellite retrievals"

*Please, see our replies to previous comments regarding the section on satellite-model comparison. To our knowledge, this is the first article that provides an alternative approach to compare satellite-retrieved tropospheric $NO_2$ with model results using scattering weights.*

• page 14526, line 3: I know that this is not the topic of this paper but I find the given uncertainty of 2E14 molec cm-2 for the separation between troposphere and stratosphere really optimistic. If the authors believe this number, they should remove all the later statements pointing at this step of the retrieval as one of the possible sources for the differences observed with other data sets, as these differences are all more than one order of magnitude larger.

*As the reviewer implied, we deleted the statement on uncertainty in stratosphere-troposphere separation, but retained statements mentioning the stratosphere-troposphere separation scheme as one of the factors affecting evaluation of satellite tropospheric $NO_2$ retrievals with other measurements.*

• section 2.4 – it would be worthwhile to already mention here how the temperature dependence of the NO2 cross-section is treated in the Pandora retrievals

*We have mentioned this in Section 2.4.*

• page 14531, line 11: As discussed above, the lowest layer in the "measured" profiles is based on model assumptions. I therefore disagree with the statement: "Both the measurements and the model suggest that 20–30% of the tropospheric NO2 column is located near the surface"

*The statement is now entirely based on modeled profiles.*

• section 3.2.4: While Figure 9 looks great, the reader wonders why these two SEARCH sites were selected and how the comparison looks for the other sites. Is there a good reason for this selection and the omission of all the other results?

*Validation of satellite retrievals with point measurements at urban sites is not helpful as the two measurement systems could sample very different air masses. There are further complications when columns measured by satellite instrument are evaluated with in situ surface measurements due to the need for column-to-surface concentration conversion using model. From our previous study (discussed in introduction), we found that only the rural sites are suitable to validate satellite retrievals with surface measurements. Regarding the SEARCH network, there are three rural sites, and one of them is often impacted by urban pollution leaving just two sites for comparison. This is now clarified in the text.*

• section 4: It would be good to make the link between scattering weights and averaging kernels here for readers not familiar with the differences in these two concepts.

*Following Rodger's formalism, averaging kernels (AKs) are defined as the integrating kernels and they provide a weighted sum. Since ideally AK should be a unit vector, one expects real AK to be close to the unit vector. Values of $NO_2$ AK differ considerably from those of Rodger's formalism and do not look like a unit vector. In presence of clouds, elements of AK in the free troposphere can be as high as 10 or even more. Given the difficulty in defining AK for the $NO_2$ algorithm, we prefer to use only the scattering weights, which are easier to interpret. Our main goal here is to provide an alternative to Rodger's formula for comparing OMI $NO_2$ vertical columns with models using scattering weights.*

• page 14540, line 10: While the differences are larger than stated in Boersma et al., they are in line with other estimates of high resolution a priori profile effects (Heckel et al., 2011, Russell et al., 2011).

*That is true. We have modified the sentence and included the two references.*

• page 14544, line 13: What is a factor of 2 change in profile shape? I think a better measure would be the day to day change in tropospheric NO2 AMF

*This sentence is referring to significant spread of $NO_2$ profile shapes observed (in July) from aircraft spirals during DISCOVER-AQ. We have modified the sentence for clarity.*

*Tropospheric $NO_2$ AMF changes with not only profile shapes but also viewing and solar geometries. Here we are interested in day-to-day changes in tropospheric AMF due to profile shapes only.*

• Blond et al. reference: Typo in SCIAMACHY

*Thanks. Done.*

• Crawford et al. reference – this doesn't look like a proper reference to me

*This reference is deleted, and a link to the DISCOVER-AQ site is provided.*

• caption Fig. 3: circles show => circles shows

*Done.*

• figure 5: Please don't use dashed lines for error bars. It would also be nice if you could introduce an x-offset to the Pandora values to avoid overlapping of error bars with the aircraft data

*Figure 5 is modified as the reviewer suggested.*

• figure 8, left: Add 1:1 and 25 (or 30)% lines in the scatter plot

*The figure is modified as suggested by the reviewer.*

We would like to thank the reviewer #3 for his/her helpful comments that are helpful to improve this manuscript. Below in *italics* please, find our replies to the reviewer's comments.

General comments:

This manuscript evaluates the OMI NO2 Standard Product using a variety of data sources, including aircraft, MAX-DOAS, and in-situ measurements and an emissions inventory, allowing a detailed evaluation of several of the factors that govern the uncertainties in the retrieval algorithm. They generally find good agreement between OMI observations and measurements and find that day-to-day variability in NO2 profiles largely influences the retrieved daily columns. Overall, I believe it is a strong and well written paper that provides interesting new insights into the factors governing retrieval validation and accuracy, and feel that the manuscript is suitable for publication in ACP after the following minor comments are addressed.

**Specific comments:**

While the use of different types of datasets (aircraft, in-situ, etc) for validation is clear, it is unclear how the data were chosen. It seems the authors intended to cover a variety of geographical regions and capture seasonal variation but I suggest further discussion of these choices and clearly stating throughout the document and perhaps also in the title that the conclusions were drawn from several specific regions and may not be representative elsewhere or under certain conditions given the limited spatial and temporal coverage.

*We agree with the reviewer that the validation study is still limited in scope due to scarce and sporadic $NO_2$ measurements. We have acknowledged this in the revised manuscript. For example, we have added "Since validation data sets are scarce and are limited in space and time, validation of the global product is still limited in scope by spatial and temporal coverage and retrieval conditions" in abstract, and "The spatial and temporal coverage of the comparisons we have examined in this paper are limited; they may not be representative of other locations and seasons" in the conclusion section.*

It may be worth mentioning the updates to the standard product in Section 2.1 and elsewhere when comparing results to previous studies that used earlier versions of the standard product.

*We added a paragraph as follows: "The OMNO2 retrievals used here, version 2.1 (Bucsela et al., 2013), represent a significant advance over previous version 1.0 (Bucsela et al., 2006; Celarier et al., 2008). The main changes include the use of monthly, rather than annual, mean a priori NO2 profiles, and improvements in the estimates of stratospheric NO2 columns, correction of calibration artifacts (de-striping), and the calculation of scattering weights".*

Please clarify whether all OMI cross-track pixels were used for comparisons.

*We use data from all cross-track positions. This information is included in Section 2.1.*

Page 14531 Line 17-18: Please explain what is meant by "Day-to-day variations in aircraft NO2 shape factors are up to a factor of two".

*The statement is modified as follows: "These measurements also reveal considerable day-to-day variation in NO2 profile shapes within a given month, suggesting that the use of a monthly mean profile in the operational algorithm is potentially a significant source of error in individual retrieved tropospheric NO2 columns".*

Page 14532 Line 10-14: The author suggests that "inaccurate removal of stratospheric NO2 on July 2" may have contributed to the discrepancy between measurements but earlier provide an uncertainty of only 2E14 molecules/cm2 for the stratospheric subtraction step of the retrieval. I suggest omitting the reference to the stratospheric subtraction here.

*We deleted the statement on uncertainty in stratosphere-troposphere separation.*

Page 14543 Line 3-6 and Page 14545 Line 18-20: The author's discuss the importance of surface reflectivity and its potential influence on retrieved NO2 columns. I suggest mentioning results from previous studies that have attempted to reduce AMF uncertainties related to surface reflectivity.

*We have added the statement: "Some previous retrieval studies have used high-resolution MODIS albedo data in an attempt to reduce uncertainty in the tropospheric AMF [Russell et al., 2011, Zhou et al., 2010]".*

**Technical corrections:**

Page 14523 Line 19-20: Is there a typo in "This study takes advantage of state-of-theart NO2 measurement technique: : :"?

*Done.*

Page 14548 Line 23: Typo, should be SCIAMACHY

*Thanks. Corrected.*