We thank both referees for their constructive comments and suggestions and will address them in a revised version of the manuscript. Our responses to the reviewer's comments and detailed changes made to the manuscript follow. All minor editorial changes suggested by the referees have been included in our revision.

Responses to Anonymous Referee #1

Referee: The authors present the Berkeley High-Resolution (BEHR) product as the third OMI NO2 product. Use of high resolution monthly NO2 vertical profiles derived from WRF-Chem is stated to be a major cause of the improvement. As WRF-Chem is an important part of the manuscript, additional details on model are necessary. Little information on WRF-Chem given in Section 3.3 is hidden, but it has to be made visible with more information in a separate section. The BEHR should not be presented as the third product. The manuscript as it stands tells me that a user who is investigating African or Asian emissions using standard or DOMINO NO2 product could switch to the BEHR product. It is available to the user who is interested to investigate ship emissions. That apparently is not the case. Emissions and meteorology determine the shape of NO2 profiles in model. What is the resolution of anthropogenic emissions in WRF-Chem? What are the extensions in latitude and longitude? Do the emissions have diurnal, weekly, seasonal, and annual components? What about other emission sources? Does it have lightning NOx emissions? What is the resolution and domain of the meteorological field? Unless these questions are addressed, we cannot be convinced that NO2 vertical profile shape from WRF-Chem is more representative than those from other global model. Please state clearly in both abstract and conclusion that the product is available just over a domain centered over California. Title of the manuscript needs to be changed.

Response: Our aim is to present the BEHR retrieval as a distinct retrieval algorithm-with accuracy and precision that are characterized with in situ observations. We compare the accuracy and precision of this new retrieval algorithm to that of some commonly used retrievals to give the reader a sense of the advantages of the elements (spatial resolution of albedo, terrain pressure, and tropospheric profiles) we are investigating in this retrieval. With respect to whether this product is available for another part of the world, our intention is that a user investigating the use of the SP or DOMINO NO_2 in some other part of the world could either produce their own retrieval following our procedure or at a minimum evaluate the size of the error and bias in their work if they choose to use the standard or DOMINO products. We agree that or description of the WRF-CHEM calculation was too condensed for a user to repeat our calculations and have added text to the manuscript that would allow the reader to reproduce our results. We've added *"at native 4km × 4km resolution"*, to clarify the resolution of anthropogenic emissions. We've also added *"(30 – 50 N, -124W – -100W)"* to clarify the domain for which the model was evaluated. As stated in the manuscript, emissions are *"for a typical June weekday"*. We've also added *"Emissions from lightning and fires are not included but are expected to have a minimal influence in the urban regions studied here"* to address the exclusion of additional emission sources.

Referee: The methodology is not clear. What is the starting point for your retrieval algorithm? Which radiative transfer model do you use to calculate AMF? What prompted the authors to implement the same method of stratospheric subtraction as for the Standard Product which has been critical in the past? How could you come up with the numbers: – 20% to +20% for terrain pressure, -40% to +40% for albedo, and -75% to +10% for NO2 profile shape? I assume, it will be very difficult to come to the conclusion without using the same algorithm as there are many steps/components that could differ between two independent algorithms. Please describe clearly how the study was carried out.

Response: The starting point for the retrieval algorithm is stated in Section 2.3: *"The Berkeley High Resolution (BEHR) retrieval uses the same method of stratospheric subtraction as that outlined for the Standard Product and AMFs are similarly determined using a lookup table (generated using the TOMRAD radiative transfer code) that depends on viewing parameters and terrain and profile information."* The following was added to clarify which radiative transfer model was used: *"generated using the TOMRAD radiative transfer code"*. The authors agree that an improvement to the method of stratospheric subtraction is important, however, implementation of a new method is beyond the scope of this paper. Our aim is to show the biases in the retrieval that result from using low-resolution terrain and profile inputs. We've added the following to the start of Section 3 to clarify how the analysis was carried out and how biases in terrain pressure, albedo, and NO_2 profile shape were determined: *"In this section we describe each of the new datasets implemented into the BEHR retrieval and then compare the product with each individual parameter implemented with the operational products".*

Referee: You state that MODIS albedo is not available over the ocean. This would mean that the method described here cannot be applied in the operational algorithm. If the main motivation is to obtain high resolution albedo database, wouldn't it be more logical to create high resolution (say 0.1x0.1 deg2) albedo product from OMI than using MODIS-based albedo? OMI-based albedo may be more suited to trace gas retrieval than MODIS BRDF due to retrieval consistency, same measurement time, and more representative spectral bandwidths. Describe why MODIS albedo is chosen for OMI NO2 retrieval algorithm?

Response: Our main motivation is to show that the spatial resolution of the albedo is important for accurately retrieving NO_2 column. We believe this is effectively accomplished using the MODIS albedo and that our evaluation with in situ observations and our demonstration of the reduction of anomalies over salt flats in Nevada confirms that using the MODIS albedo reduces errors induced by the low resolution albedo fields currently in use. We agree that development of a high-resolution OMI based terrain reflectivity might be useful. However development of such a product is beyond the scope of this paper and we are not persuaded that it would be measurably better (for our purposes) than the MODIS albedo.

Referee: I wonder if the retrieval and conclusions drawn here are based on observations in the month of June. The effect of albedo and profile shape could vary seasonally. Would -40% to +40% for albedo and -75% to +10% for NO2 profile shape still be valid for winter?

Response: The conclusions drawn here are indeed based on observations in the month of June. While the exact magnitude of the albedo variability in other seasons is likely to be

different, we note that the 16 day MODIS product will capture snow cover reasonably well—an improvement that might be even larger than the +/-40% we observe in summer. The effect of profile shape will vary seasonally; however, the seasonal issue is not related to the spatial variability that is the focus of this work and therefore, we choose not to address it here.

Referee: Remove redundancies in description of OMI NO2 retrieval algorithms in introduction (Page 12413, line 14-27) and Section 2.

Response: We have removed the following from Section 2: "*(referred to as "operational retrievals" in the remainder of this paper)".*

Referee: Page 12413, line 28: Following my earlier comments, it is probably not a development of new retrieval product, but a kind of sensitivity study.

Response: This has been addressed above.

Referee: Terrain pressure effects: I am surprised to see a large difference in terrain pressure. Terrain pressure in standard and DOMINO product might be based on ETOPO5 or similar, which generally has better spatial resolution than OMI observations. Why the GLOBE topographical database averaged over OMI observations should differ systematically by 5-20% being terrain pressure used in standard and DOMINO product generally higher? I wonder if the differences arise from the method of conversion from terrain height to terrain pressure.

Response: The terrain pressure databases used in the Standard and DOMINO products are indeed at a higher spatial resolution than OMI observations, however, as addressed in the manuscript, these products use the pressure at the center of the OMI pixel instead of averaging over the OMI satellite pixel. We believe that by averaging the pressures from the GLOBE database, we are providing a better representation of the terrain pressure for the AMF calculation. We show that the percent change in terrain pressure between the two methods (Figure 1a) largely depends on the degree of spatial variability of the terrain.

Referee: Page 12417, line 1: What does the "effective terrain pressure" mean? How does it differ from "average terrain pressure"

Response: We've replaced "*to derive effective terrain pressure while NASA and KNMI choose the terrain pressure at the center of the pixel*" with "*instead of using the terrain pressure at the center of the OMI pixel as in the NASA and KNMI products*".

Referee: Section 3.5: Complete new retrieval of OMI NO2: What does it mean? When it has same stratospheric field as in the standard product, how can it become complete new retrieval? Does it treat stripes and temperature correction differently?

Response: We have replaced "*Complete new retrieval of*" with "*BEHR*" for clarification. The only changes are to the terrain pressure, albedo, and NO_2 profile shape.

Referee: Page 12422, line 24: I wonder if the assumption of a constant 40 ppt NO2 would be valid over areas with heterogeneous boundary layer NO2 field. Because of advection, wouldn't the free tropospheric NO2 higher in polluted areas than in clean areas?

Response: Aircraft observations indicate modest spatial variability of upper tropospheric NO_2 (except in regions of intense lightning) and in the polluted regions studied here, this minor variability would have a very small impact on the tropospheric column. We find that a constant 40 ppt is representative of the NO_2 concentration in the upper troposphere in California for the timeframe studied here.

Referee: Page 12418, line 28: Remove comma in front of "mean".

Response: We have replaced *"annual, mean"* with *"annually averaged"*.

<u>Response to Anonymous Referee #2</u>

Referee: Abstract (last line): Saying that much of the variance can be attributed to coarse resolution terrain and profile parameters seems in contradiction with Figure 1 (f, g, h), where we see that the variance in NO2 columns arising from the use of the new parameters is more pronounced for albedo and profile shape than for terrain pressure.

Response: *"Terrain"* here was meant to refer to both albedo and terrain pressure collectively. Replaced *"coarse resolution terrain and profile parameters"* with *"coarse resolution terrain pressure, albedo and profile parameters"*.

Referee: Section 2.3 & 3.3: The main argument of this work is that the new product uses spatially and temporally improved resolutions for the input parameters in the retrieval. While the resolution is clearly improved for albedo and terrain pressure, the NO2 shape profiles used are averaged over a month, which is a coarser temporal resolution than the one used in the DOMINO product. Depending on the type of source considered, daily variability of the NO2 shape profiles can be significant and impact the retrievals. For example, do you know to what extent your data might be impacted by fires occurring during that period? Please justify the use of monthly averaged NO2 profiles rather than daily profiles.

Response: While we agree that daily variability might be important in addition to the spatial variability we investigate in this manuscript, we are not persuaded that WRF-CHEM (or any model) adequately captures the daily variability in PBL height (at 4km spatial resolution)—but have more faith in their representation of the mean behavior on monthly time scales. We chose to use profiles from WRF-Chem because the model provides profile information at a much improved spatial resolution (4km × 4km) compared with the 2° × 2.5° used in the Standard product or the 3° × 2° used in the DOMINO product. We do not know what impact daily variability has on the profile but we suspect that it is smaller than that induced by the coarse resolution of the NO_2 profile sets used in the operational retrievals. That said, we agree that the influence of daily variability would be an interesting

topic of future research. The observations during this period were impacted by fires—however the fires were a considerable distance from the observations and we do not include any obvious fire plumes in the analysis.

Referee: Section 4.1: The authors propose to assume that the boundary layer is well-mixed in order to infer the boundary layer part of the NO2 column from the aircraft observations. They justify this assumption by arguing that it is supported by both model outputs and aircraft measurements. Figure 5 c) shows an in situ NO2 profile over the area of interest (California) that does not exactly correspond to a mixed profile in the boundary layer (which would imply uniform values within it), but rather represents an exponentially decreasing profile shape. Please clarify this point, as the shape of the NO2 profile near the surface might be critical for this analysis.

Response: We agree the figure does not make the point very well and have removed the figure. The figure showed an apparent exponential profile but this was due to combining different altitudes from different locations and was not representative. We have added text describing our assumption of a well mixed profile and indicating that the variance in the comparison to in situ data would be similar if we assumed an exponential profile. Our observations from aircraft do generally support the idea that profiles are well mixed in an average sense—although individual profiles have a variety of shapes—a manuscript detailing these observations is in preparation.

Referee: Section 4.3: This section presents OMI NO2 validation results for different cloud criteria, using both MODIS and OMI cloud products for the thresholds. It is not clear if in the case of the use of MODIS cloud fractions the retrieval is reprocessed using the MODIS-derived cloud parameters or not. In other words, are the MODIS cloud fraction data used only to filter the data or are they also used in the AMF calculation? This is a key point in my opinion, as you need consistency between the cloud fraction used for the threshold and the one used in the radiative transfer calculation of the AMF in order to be able to interpret your results properly. The methodology needs to be clarified here and this section would also benefit from a paragraph with some interpretation of the results.

Response: In the BEHR product, the MODIS cloud fraction is used in the AMF calculation-not just to filter the data. Added following to clarify: *"For the MODIS cloud filtering case, the cloud fraction derived by MODIS and averaged over the OMI satellite pixel is implemented in the AMF calculation to determine NO_2 columns (while the standard OMI cloud product is used for the SP and DOMINO cloud-filtering cases)."*
Also added the following interpretation: *"As discussed in Section 3.4, the poor correlation between aircraft observations and the BEHR product derived using the standard OMI cloud information is expected to result from artifacts in the OMI cloud algorithm related to the coarse terrain reflectivity used in the product."*

Referee: P. 12418, Line 16: " [...] OMI is less sensitive to NO2[...]". This should be removed, as you are not considering the OMI measurement itself here, but only the AMF.

Response: We've removed *"OMI is less sensitive to NO_2,"*.

Referee: P. 12419, L23: Please add some interpretation of this result.

Response: We've added the following to explain the problem with using coarsely resolved profile information as in the current operational products. *"The profiles provided by both models are at a much coarser spatial resolution that the OMI observation (WRF-Chem: 2° × 2.5°; TM4: 3° × 2°). Consequently, an average semi-polluted profile is applied over the large grid cell that contains both urban and rural locations, poorly representing the actual NO_2 profile over both areas and resulting in an underestimation of the NO_2 column in the urban region and an overestimate in the rural region."*

Referee: P 12425, Line 21: "[..]resulting IN differences[...]"

Response: We have replaced *"resulting differences"* with *"resulting in differences"*.

Referee: P 12426, Line 11-13: "[...] verifying the use of the boundary layer method for the validation of satellite products." should be removed, as a good agreement between measurements and retrieved columns is not a criteria to evaluate the methodology of the comparison itself.

Response: We've removed *"verifying the use of the boundary layer method for the validation of satellite products"*.

Referee: P 12426, Line 15: "We interpret this to mean that much of the variance in current retrievals is due not to atmospheric parameters [...]". What do you mean exactly? This sentence is not clear to me.

Response: We've replaced *"atmospheric parameters"* with *"meteorological or chemical variation"* for clarification.

Referee: Fig 4 c): in the color bar, 0 should be at the center (white) for more lisibility (positive values in red and negative values in blue)

Response: The color bar was changed so that zero is white.

Referee: Fig 4 d): the color bar should be removed as it is not related to this figure.

Response: The color bar was removed.

Referee: Fig 5 b): the y-axis should include the altitude in order to be able to read the aircraft altitude as well as the PBL height.

Response: The y-axes have been labeled.