OMI NO2 retrievals are widely used in the atmospheric science community in order to derive key quantities for our understanding of atmospheric chemistry and climate and to improve air quality assessments. This information includes for example surface NOx emissions, whose uncertainties are poorly quantified in chemical transport models or surface NO2 concentrations. Errors in these parameters are strongly correlated to errors in the retrieved NO2 columns. This work aims at reducing uncertainties in the retrieved NO2 columns by the mean of increasing the spatial and temporal resolution of input parameters used in the retrieval algorithm. This motivation is based on previous studies that identified parameters to which the retrieval algorithm is most sensitive: NO2 shape factor, albedo, terrain pressure and clouds. The authors present a new retrieval (BEHR) using a new dataset for these input parameters. They demonstrate an improvement in the retrieved NO2 column using this product compared to the currently used products (DOMINO and OMI-Standard), by comparing the satellite observations with aircraft in situ measurements from the ARCTAS-CA experiment. For that purpose they propose a new approach to derive in situ NO2 tropospheric column from aircraft measurements using only observations in the boundary layer, which allows a statistically more robust evaluation compared to other methods (full spirals for instance). This latter development is especially useful as validation studies are currently suffering from the lack of NO2 columns data available. However, here and there the methodology used is not entirely clear or well justified, and more explanations are needed. I would recommend publishing this work providing that the authors address the questions/remarks I listed below.

Main remarks:

-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.

-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 occuring during that period? Please justify the use of monthly averaged NO2 profiles rather than daily profiles.

-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.

-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.

Minor remarks:

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.

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

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

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.

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.

Figures:

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).

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

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.

Interactive comment on Atmos. Chem. Phys. Discuss., 11, 12411, 2011.