Response to anonymous referee #1

We would like to acknowledge the referee for their helpful and thorough review. We believe that their comments improved the quality of this work.

Some of the measurements and statistics presented in the study have changed as a result of some of the reviewers' suggestions. Moreover, an updated OMI overpass data set has been used, generated by the Aura Validation Data Center (AVDC) (http://avdc.gsfc.nasa.gov) in November 28th 2016.

Our responses (in blue) follow the reviewer's comments (in black italics).

General comments:

One of the major pity is that the air quality model is used only for the OMI results and not for the GOME-2 data. Moreover, the air quality model is used with the shortcoming of a typical OMI nadir pixels centered around each location (p.10, line 7 to 13) instead of the actual pixel size and location. If additional simulations are unavailable to improve the study, the variability of the NO_2 content in space should be estimated in some way (maybe from the NOx emissions) and discussion of the spatial and temporal variability should be enhanced. For GOME-2 comparisons, the fact that the pixels would cover the whole domain of Fig. 1 (p. 7, line 33) can be used to test how the average variability seen by 3 typical stations would compare to the averaged variability seen from satellite. Moreover, the differences between GOME-2A and GOME-2B comparisons results are never mentioned, while Metop-A is in a reduced swath mode since 7/2013 with a pixel half the size of the nominal mode of Metop-B. This should be mentioned and commented (figure 7, table 3 and 4).

The NOx emissions pattern during the campaign period is discussed in section 2.4. GOME-2 observations have been compared to the average of the three stations and the results indicate that this average seems to be affected mostly by the much higher NO_2 concentrations of the urban site. Information on the reduced swath width of GOME2A has been included in section 2.3 and discussed in sections 2.3, 3.2 and 4.

The OMI and GOME-2 tropospheric NO_2 algorithms are only mentioned in Section 2.3, without any detailed description, and conclusions are thrown as if they were consistent among them, while we know that large differences in tropospheric NO_2 exist between different product (applied on a same instrument) due to different algorithm hypothesis. The 2 algorithms should be described in Section 2.3 (maybe with a table of main differences) and impact of the algorithm hypothesis should be mentioned/quantified. Comparison e.g. of the model NO_2 profiles at the 3 stations type wrt to the profiles used as a-priori in the satellites retrievals would be interesting.

Detailed descriptions of the OMI and GOME-2 tropospheric NO_2 algorithms have been added in section 2.3 and a table of the main features has been included. The possible impact of the two different algorithms on the study is mentioned in sections 3.2 and 4. However, it is not the purpose of this study to investigate the impact of the different algorithms. The model and satellite NO_2 a priori profiles have been described in sections 2.2 and 2.3.

More detailed answers are given in the specific comments below.

Specific comments and technical corrections:

Page 1, line 9: consider changing "the main" into "one of the"

The text has been revised accordingly.

Page 1, line 13 (and in the conclusions): why using DOAS/MAX-DOAS while in the rest of the document only "MAX-DOAS" is used?

The text has been revised accordingly.

Page 1, line 18: consider reformulating "mainly due to the higher spatial resolution of OMI" to also mention the different products.

The text has been revised accordingly.

Page 2, line 22: consider changing "is limited… respectively" to "are respectively limited by the …"
Page 3, line 10: "are applied to satellite data …" → Only applied to OMI data!

The text has been revised accordingly.

Page 4, line 18: inconsistency between "a fixed azimuth angle of 255°" and "at azimuth angles of 80° relative to the solar azimuth". Reformulate.

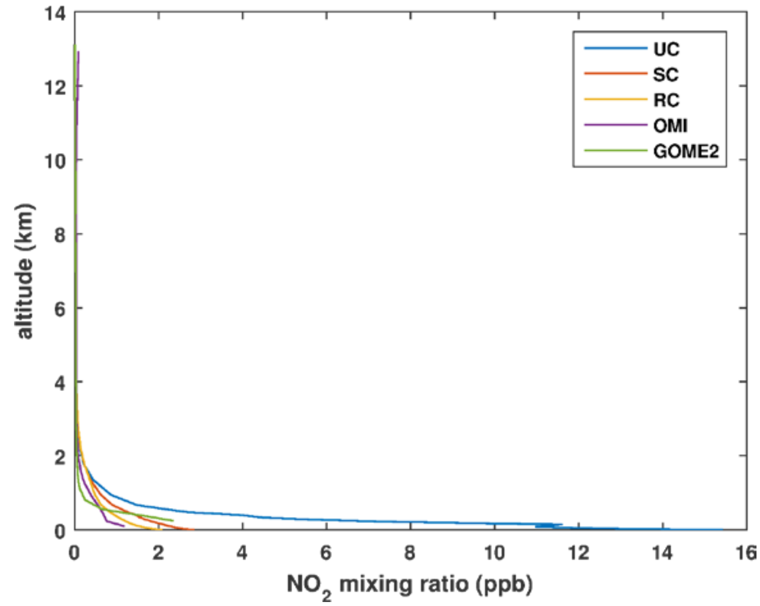
The text has been revised accordingly.

Page 4, line 21: add reference for "only at the elevation angles of 15º and 30º in order to avoid uncertainties introduced due to aerosol loadings at lower elevation angles" and explain what is done with the VCD retrieved at the 2 elevation angles. Is the average value used? Or is a filtering related to the difference of the 2 VCD values used (as in Brinksma et al., 2008)?

Initially, the average of measurements at both elevation angles had been used without any criteria. Following the above comment, an improved approach has been adopted for the revised manuscript. An average of measurements at both elevation angles is used when they agree within 20%, otherwise only results at 30º are compared with satellite retrievals. Text has been revised accordingly.

Page 5, line 7: illustrate the shape of the used "mean vertical profiles from the air quality modelling tool consisting of the photochemical grid model CAMx and the mesoscale weather prediction system WRF" (ideally also showing how they differ from the 2 satellite NO_2 algorithms a-priori profiles)

The model and satellite a priori profiles are similar to exponential decay with altitude. We describe them in the text (sections 2.2 and 2.3) but we have not included any figure in the manuscript. In the following figure the CAMx, OMI and GOME-2 a priori mean profiles are shown. All the a priori profiles are similar except for the CAMx profile of the urban site (UC), in which the NO_2 concentrations at the lowest layers are much higher due to the elevated NOx emissions in the urban area.



Page 5, line 13: the 0.07 and 0.01 difference in AOD between the sites should be referred to the mean value. (0.07 wrt to 0.1 is large while it is small if compared to 1).

This section has been revised. These differences have been calculated for AOD values ranging between 0.1 and 1.1. The long-term (1997-2005) mean AOD over Thessaloniki (UC) is 0.33±0.14 and 0.53±0.17 for winter and summertime, respectively.

Why scaling the LIDAR profile to AOD values (line 14), if the "AOD loading and AOD profile between the three locations is very small" (line 12-13)? How much the choice of this aerosols a-priori profile affects the MAXDOAS NO2 VCD? Give an estimation of the MAXDOAS error.

The same climatological LIDAR profile was used as a priori for all three locations. This profile is scaled during the simulations using AOD values in the range 0-1.5 only for the construction of the AMF LUT. The AMF corresponding to each measurement is extracted from the LUT using AOD measurements from the CIMEL sun-photometer operating in Thessaloniki. We have repeated the AMF simulations using an aerosol profile according to Shettle (1989) for autumn/winter conditions and the average difference in the retrieved NO_2 tropospheric VCDs was < 2%.

Maybe add a subplot on figure 6 with the time-series of the Thessaloniki CIMEL AOD to have an idea of the variability?

We think that this subplot is not necessary and it doesn't improve the discussion because we already take into account the variability in AOD for the NO_2 retrieval using the CIMEL AOD measurements as described above.

Page 5, line 18: it's Fig 4 and not 3. Line 21 it's Fig 2 and not 4.

The text has been corrected.

Page 5, line 24-25: I would put this sentence at the end of the section.

The text has been revised accordingly.

Section 2.3: add a description of the 2 NO_2 algorithms maybe with an overview table) and references to papers (at least Valks et al., 2011 and Bucsela et al., 2013) and some other validation results (here on when discussing the results).

The text has been revised accordingly. Section 2.3 has been completed with references, including the suggested, and a table for the NO_2 algorithms has been added. Also, validation results have been included in sections 2.3, 3.2 and 4.

Page 6, line 17: "a particular direction and path" – how long is the representativeness area of the MAXDOAS? Give references (at least irie et al., 2011) and estimation ranges. In page 7, line 32, an area of 2x2km² is considered representative of the MAXDOAS data, but when using high elevation angles as 15° and 30°, and considering no aerosols, longer distances are obtained. How much the choice of using only one model cell of 2x2km² is affecting the adjustment factor?

Sections 2.4 and 3.3 have been revised. We have estimated the representativeness area of the MAX-DOAS data based on the elevation angle and boundary layer height (BLH) reanalysis data from ECMWF (as the ratio of the BLH to the tangent of the elevation angle). An average of 0.55 km for both elevation angles was calculated as the representative horizontal distance of the MAX-DOAS measurements for the campaign period and OMI overpass time. Only ~2% of the campaign data were found to correspond to horizontal distances larger than 2 km. Thus, we believe that the $2x2km^2$ model cell can be considered representative of the MAX-DOAS data. We also did calculations of the adjustment factors using more than one cell and they seem not to be significantly affected (Table 6 in the revised manuscript).

Page 6, line 19: "the satellite data are adjusted" → only OMI data are adjusted.

Text has been revised accordingly.

Page 6, line 25 to page 7 line 24: a lot of details are given for the model data used here, which is too much compared to the lack of description of the satellite data in Sect 2.3. The goal of the paper is focusing on the satellite data!

Section 2.3 has been revised.

Page 7, line 25: again, only OMI data are adjusted.

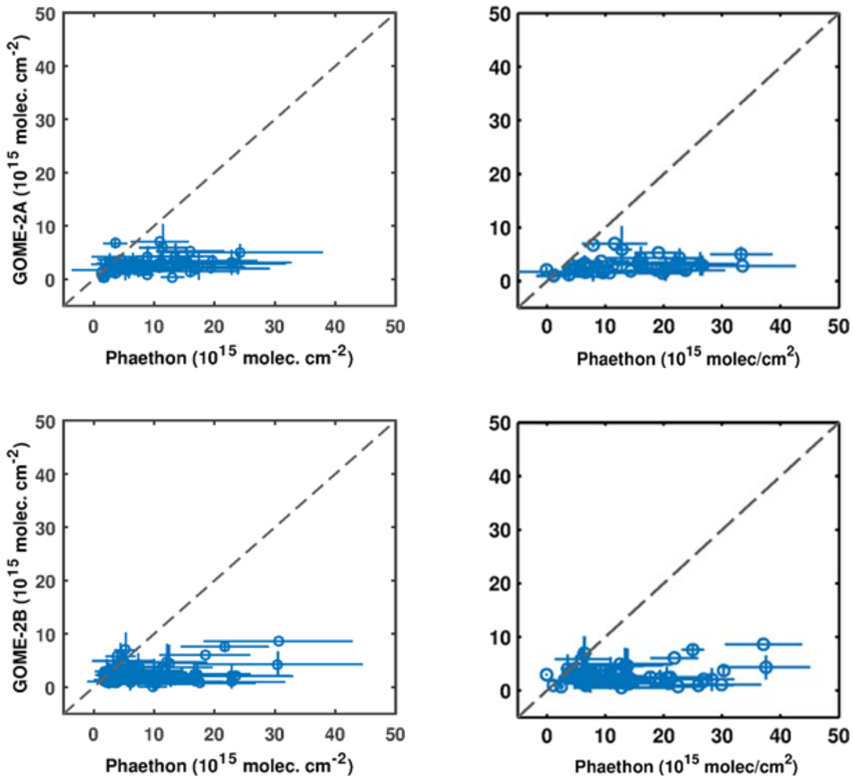
Text has been revised accordingly.

Page 7, line 32: see comment above about the impact of using only one model cell of 2x2km² is affecting the adjustment factor.

Answered in the previous comment about representativeness area.

Page 7, line 33: "the method was applied only to OMI data because for GOME-2 the sub-satellite pixel is very large, covering typically the entire domain of Fig. 1" – see general comment: how the GOME-2 data would compare to an average of the 3 stations?

The average NO_2 VCD values of the 3 stations are mostly affected by the large NO_2 loading of the urban area. Thus, the comparison of GOME-2 overpass data with the average of the 3 stations is similar to the comparison with the urban site observations. In the following figures the GOME2A (upper panels) and GOME2B (lower panels) overpass data are compared with the average of all three stations (left panels) and the urban site (UC) measurements (right panels).



Page 8, line 14 and line 19: again, what columns are used in this study (in the figures, tables, etc)? VCD at 15° or at 30°? Or something else?

Answered in a previous comment. This information is now included in the captions of figures and tables.

Page 8, line 26: the "upper limit for the distance": does this mean that only the closest pixel within this distance is used for the satellite data? Or an average of the pixels values within this distance is used? How much choosing 50km for both instruments would affect the comparison?

Both the pixel selection and the distance criterion are based on the satellite pixel size. In case of OMI the closest pixel within 25 km is used, in order to enhance the possibility that the station would be included in the pixel area, whereas for GOME-2 sensors, the pixels of which are much larger, an average within 50 km seemed to be a better choice. However, the comparison results are only slightly affected when we use a limit of 50 km also for OMI measurements. This is also the case if we choose the closest pixel of GOME-2 instead of the average value. Text has been revised.

Page 8, line 29: typo: "table 3 and 4"

The text has been corrected.

Page 9, line 6: "can be attributed mainly to its smaller pixel size" and smaller distance criterium? "to its higher sensitivity in the boundary layer compared to GOME-2": give a reference. The fact that the 2 algorithms are different should also be mentioned and commented.

The smaller distance limit does not affect the comparison results. There is a detailed answer in a previous comment. Section 3.2 has been revised accordingly and a reference for OMI high sensitivity in the boundary layer has been added as suggested by the reviewer. The differences between the two algorithms are mentioned in the revised manuscript.

Page 9, line 11: "The use of actual satellite geometry for each day is complex and would require a much larger domain for the air-quality simulations, more than double the currently used (120x120 km^2) in order to include all possible pixel sizes and positions for each location. However, such simulations were not available for this study." → This is really a pity. The study would be much more realist…

We agree, but we didn't have such capability.

Page 9, line 19: again, it would be nice to have a figure with the NO_2 model profile at the 3 sites.

We have addressed this comment already and we have provided a figure. However, we have decided not to include this figure in the paper but discuss the shapes and give numbers of the scale height of the different profiles.

Section 4: no discussion to previous validation results is included, neither in Section 3 or 4. This should be added. Discussion of the difference in GOME2A and GOME2B results should also be added. Again importance of the OMI and GOME2 algorithmic differences should be mentioned.

The text has been revised accordingly. Validation results have been included and discussed in sections 2.3, 3.2 and 4. Discussion of the difference in GOME2A and GOME2B results has been included in sections 3.2 and 4. The OMI and GOME2 retrieval algorithms are described in section 2.3 and the importance of the algorithmic differences is mentioned.

Page 18, table 2: mention that the statistics are for tropospheric NO_2. From which angle (12 or 30°)?

The text and table caption have been revised accordingly.

Page 18, table 3: discuss the larger differences in GOME2A wrt GOME2B mean values in RC sites and smaller in SC and UC (not consistent with the pixel size influence, as GOME2A is twice as small than GOME2B). A figure with location of the satellite pixels position/size could be useful.

Sectioned 3.2 has been revised and examples of typical pixel sizes are now illustrated on the maps showing the mean NO_2 spatial distribution over the campaign period as observed by the satellite instruments (Fig. 5 in the revised manuscript).

Page 19, table 4: add slope and intercept values to this table. Again, discuss GOME2A wrt GOME2B differences.

Slope and intercept values have been added in the table (Table 5 in the revised manuscript). Discussion on GOME2A and GOME2B differences has been included in section 3.2.

Page 19, table 5: considering adding a line with values over 2 or more cells in the MAXDOAS pointing direction (related to comment on the MAXDOAS representativeness).

The comment about the representativeness area of the ground-based measurements has been addressed already. The adjustment factors calculated for more than one cells are presented in the table (Table 6 in the revised manuscript) and discussed in section 3.3.

Page 20, figure 1: add a ruler to estimate distances.

Ruler has been added in Fig. 1.

Page 23, caption of figure 4: "and azimuth angle 100º relative to the sun for "→ in the text (page 4, line 19) it's 80° !

The Fig. 4 has been changed and the AMF lines are now referred to an azimuth of 80º relative to the sun.

Page 24, figure 5: it would be good to add the locations of the 3 stations also on the 1rst panel, and add a ruler to estimate distances.

Rulers have been added to both panels and the three stations are shown also on the 1st panel. Also, all CAMx cells have been now included in the 2nd map. Fig. 6 in the revised manuscript.

Page 25, figure 6: why not including the period where the 3 instruments measured at the same site on the figure?

A separate plot has been included in Fig. 2.

Page 26, figure 7: use the same axis limit for the urban scatter plot for the 3 instruments.

The axis limit has been adjusted accordingly (Fig. 8 in the revised manuscript).

References:

Irie, H., Takashima, H., Kanaya, Y., Boersma, K. F., Gast, L., Wittrock, F., Brunner, D., Zhou, Y. and Van Roozendael, M.: Eight-component retrievals from ground-based MAX-DOAS observations, Atmos.

Meas. Tech., 4(1), 1027–1044, doi:10.5194/amtd-4-639-2011, 2011.

Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F., and Pickering, K. E.: A new stratospheric and tropospheric NO2 retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, Atmos. Meas. Tech., 6, 2607– 2626.

Valks, P., Pinardi, G., Richter, A., Lambert, J.-C., Hao, N., Loyola, D., Van Roozendael, M. and Emmadi, S.: Operational total and tropospheric NO2 column retrieval for GOME-2, Atmos. Meas. Tech., 4, 1491–1514, doi:doi:10.5194/amt-4-1491-2011, 2011.

The references suggested by the reviewer have been included in the revised manuscript.