Response to anonymous referee #2:

We thank referee #2 for constructive and helpful review comments, to which we hope to have responded appropriately. A list of comments including our response is given below.

In "Comparison of tropospheric NO2 columns from MAX-DOAS retrievals and regional air quality model simulations," the authors provide a nice overview of 1) long-term MAX-DOAS records of NO2 in northwest and southwest Europe, 2) a description of regional air quality models used in the CAMS ensemble, and 3) a description of past comparisons of regional CTMs and MAX-DOAS with in situ and satellite data. The comparison of the model ensemble and the four MAX-DOAS NO2 datasets showed general agreement in a broad sense. The authors highlight when and where there are discrepancies between ensemble median model results and MAX-DOAS observations (e.g., seasonal cycle, diurnal cycle), but do not offer ideas on potential approaches for disentangling the causes of these discrepancies.

I felt that the paper lacked a final synthesis, written in more general language, of how future simulations and MAX-DOAS deployments like these can isolate effects from individual processes. I hope that the authors consider adding a broader synthesis of their results to the end of section 4, offering possible paths forward for future analyses: what common and distinct attributes of these four sites share? How might these differences and similarities be exploited to investigate chemistry? Emissions? Meteorology? Where might the authors propose future MAX-DOAS instruments be located? Should one expect an ensemble median to capture hourly NO2 variations? Monthly averages? What is the native scale of NO2 spatial variations at the MAX-DOAS sites inferred from the time scale of NO2 variation and wind speed?

Many changes have been applied to the summary and conclusions section in the revised version including for example a discussion of model resolution and averaging volume of MAX-DOAS measurements, suggestions for sites to investigate in future studies (i.e. stations affected by different meteorological and pollution conditions for example at pollution hotspots in the Mediterranean with strong smog conditions especially during summer and clean mountain sites), a paragraph on OMI satellite comparisons with similar results as in Huijnen et al. (2010), as well as further suggestions on how to track down reasons for differences between model runs and MAX-DOAS retrievals (please see Section 5 of revised manuscript for further details).

Huijnen, V., Eskes, H. J., Poupkou, A., Elbern, H., Boersma, K. F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., and Zerefos, C.: Comparison of OMI NO_2 tropospheric columns with an ensemble of global and European regional air quality models, Atmos. Chem. Phys., 10, 3273-3296, doi:10.5194/acp-10-3273-2010, 2010.

Comments: P2, L10-11: NO2 lifetime is much longer in the upper troposphere, primarily because its chemical family, NOx, is mostly present as NO at high altitudes, which has far fewer permanent sinks.

Changed to (p 2 l 14-16): "The lifetime of NOx is only a few hours in the boundary layer but a few days in the upper troposphere, where less OH radicals are present (Ehhalt et al., 1992) to react with NO_2 and more NOx is present as NO which has fewer permanent sinks than NO_2."

P3, L29: "focusses" – typo

This sentence has been deleted in response to a comment by referee #3.

P4: There is no discussion of model resolution. The NO2 lifetime is a function of model resolution. Also, median values may be biased towards coarser models as those with finer resolution may produce highs when a plume passes and lows when not.

In response to this question, the following text has been added to p 17 l 19 - p 18 l 9 of the revised manuscript (as response to a comment by referee #1, this is combined with a description on how the horizontal sensitivity range of MAX-DOAS compares to model resolution):

"The horizontal grid spacing (Table 1) differs for the 6 model runs evaluated in the present study, with a resolution of approximately 9x7 km^2 for the highest resolution run (LOTOS-EUROS) and 50x50 km^2 for the coarsest one (EMEP). The resolution of the remaining model runs is approximately 20x20 km^2. As described in Section 2.2, the horizontal averaging volume of MAX-DOAS retrievals strongly depends on aerosol loading, viewing direction and wavelength (Richter et al., 2013). As a rough estimate, it ranges from 5 to 10 km for the stations used in the present study. Therefore, the horizontal averaging volume is (apart from the coarsest resolution run) expected to be either on the same spatial scale as the horizontal model resolution or by a factor of 1 to 4 smaller. From the latter (i.e. horizontal averaging volume of MAX-DOAS smaller than model resolution) one would expect an underestimation of enhancements in tropospheric columns observed by MAX-DOAS in case of horizontal changes in tropospheric NO_2 columns below the model resolution and, similarly, an overestimation of local minima in tropospheric NO_2 columns. However, in reality, the comparison between horizontal averaging volume of MAX-DOAS and horizontal resolution of the models is much more complicated, as MAX-DOAS instruments usually measure in one azimuthal pointing direction meaning that measurements are performed only on a specific line of sight whereas model simulations are performed for three dimensional grid boxes. This could for example mean that a pollution plume with a horizontal extent on the order of the model resolution and hence showing up in the simulations is missed by the line of sight of the MAX-DOAS instrument. It would therefore be desirable to perform multiple MAX-DOAS measurements over a range of different azimuthal angles for each station and use these in future model to MAX-DOAS comparison studies.

A pollution plume and related increase in the time series of tropospheric NO_2 VCDs observed by MAX-DOAS would be expected to be reproduced better by model runs with higher horizontal resolution compared to lower resolution runs. The lifetime of NO_2 is also expected to increase with model resolution. However, in the present study, the LOTOS-EUROS run with significantly higher horizontal resolution than the other runs in general did not perform better than lower resolution runs which can probably be explained by its low number of vertical layers. Similarly, the EMEP run with significantly lower horizontal resolution did not perform worse than higher resolution runs, which as expected shows that other differences between the models such as chemistry schemes and treatment of emissions strongly impact on comparison results. It would be interesting to investigate the ability of the models to predict the scales of NO_2 spatial variations derived from time scales of NO_2 variations and wind speeds in the context of model resolution in a future study."

Richter, A., Godin, S., Gomez, L., Hendrick, F., Hocke, K., Langerock, B., van Roozendael, M., Wagner, T.: Spatial Representativeness of NORS observations, NORS project deliverable, available online at: http://nors.aeronomie.be/projectdir/PDF/D4.4_NORS_SR.pdf, 2013.

P5, L7: These sites, with exception of OHP, seem to be in very similar physical settings, with likely similar meteorology (e.g., vertical mixing characteristics). If so, this fact should be mentioned.

This is now mentioned in the summary and conclusions section together with suggestions for MAX-DOAS sites to be incorporated in future comparison studies (p 18 l 16-19):

"As the stations investigated in the present study have, apart from the rural background station OHP, rather similar meteorological and pollution conditions, investigation of stations over a broader range of different conditions would be desirable. Further comparison studies could for instance include stations at pollution hotspots in the Mediterranean such as Athens with strong smog conditions especially during summer and clean mountain sites."

Please also consider including a map of the region with sites indicated on a backdrop of satellite-based tropospheric NO2 column measurements.

The location of the MAX-DOAS stations is now shown in Figure 1 of the revised version, plotted on top of mean tropospheric columns of NO_2 from OMI for February 2011 as well as on top of TNO/MACC-II anthropogenic NOx emissions as an indicator of pollution levels in these and surrounding regions. The spatial distribution of NOx emissions agrees well with pollution hotpots and cleaner areas identified by OMI. Corresponding text has been added on p 4 l 1-4 of the revised version. The latter shows that the spatial distribution of emissions does not seem to be a likely reason for differences between simulations and MAX-DOAS retrievals.

Minor comment: I did not see Lat/Lon values reported for Uccle.

Added to revised version on p 6 l 33

Page 6, Line 29: Has there been any side-by-side operation and comparison of these two instruments? If so, please provide the reference.

The Uccle and OHP MAXDOAS instruments are a commercial mini-MAX-DOAS from Hoffmann Messtechnik GmbH and a BIRA research-grade spectrometer, respectively. Although there has not been formal side-by-side operation of both instruments for verification purpose, a good overall agreement has been obtained between the mini-DOAS and other BIRA research-grade spectrometers similar to the one operated at OHP, e.g. like during the CINDI campaign (see Roscoe et al., 2010). The last sentence has been added to p 7 l 14-17 of the revised manuscript.

Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Frieß, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, Atmos. Meas. Tech., 3, 1629–1646, doi:10.5194/amt-3-1629-2010, 2010.

P9, L5: "As the typical error on MAX-DOAS retrieved VCDs is around 20%" – please describe this statement in more detail: at what time scale? Random or systematic uncertainty? Based on measurement intercomparisons or fitting statistics?

Uncertainty discussion of MAX-DOAS measurements is complex but has been done in previous studies (e.g. Hendrick et al., 2014; Wang et al., 2014; Franco et al., 2015). Briefly, uncertainties are a combination of small systematic errors (for example from the cross-sections used), random errors resulting from the DOAS retrieval, errors introduced by the profile retrieval and a priori assumptions made. In particular the latter contribution can vary depending on aerosol loading, vertical NO_2 profile and cloud contamination. In polluted conditions, uncertainties from profiling dominate. In clean situations, random errors from the fit can become significant. In general, uncertainties can be considered as random or pseudo-random, but systematic errors can result from, for example, the presence of elevated aerosol layers.

Quantification of uncertainties not only from error propagation but also from validation with independent measurements would be desirable, but very few suitable validation measurements are available, and differences are usually dominated by differences in measurement volume. Intercomparisons of different DOAS instruments show excellent (a few percent deviations) agreement on the level of slant columns (e.g. Roscoe et al., 2010) but substantial (20% - 50%) differences at the level of profiles.

Here, a simplified and conservative estimate of 30% uncertainty on all MAX-DOAS measurements has been assumed. Data products with more detailed uncertainty information are currently in development for example in the framework of the FRM4DOAS project (http://frm4doas.aeronomie.be/), and once available, this data and related uncertainty information should be used in future comparison studies.

The last sentence of the previous paragraph has been added on p 10 l 1-3 of the revised version.

Franco, B., Hendrick, F., Van Roozendael, M., Müller, J.-F., Stavrakou, T., Marais, E. A., Bovy, B., Bader, W., Fayt, C., Hermans, C., Lejeune, B., Pinardi, G., Servais, C., and Mahieu, E.: Retrievals of formaldehyde from ground-based FTIR and MAX-DOAS observations at the Jungfraujoch station and comparisons with GEOS-Chem and IMAGES model simulations, Atmos. Meas. Tech., 8, 1733-1756, doi:10.5194/amt-8-1733-2015, 2015.

Hendrick, F., Müller, J.-F., Clémer, K., Wang, P., De Mazière, M., Fayt, C., Gielen, C., Hermans, C., Ma, J. Z., Pinardi, G., Stavrakou, T., Vlemmix, T., and Van Roozendael, M.: Four years of ground-based MAX-DOAS observations of HONO and NO_2 in the Beijing area, Atmos. Chem. Phys., 14, 765–781, doi:10.5194/acp-14-765-2014, 2014.

Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clémer, K., Frieß, U., Gil Ojeda, M., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F., Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro, M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schönhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, Atmos. Meas. Tech., 3, 1629–1646, doi:10.5194/amt-3-1629-2010, 2010.

Wang, T., Hendrick, F., Wang, P., Tang, G., Clémer, K., Yu, H., Fayt, C., Hermans, C., Gielen, C., Pinardi, G., Theys, N., Brenot, H., and Van Roozendael, M.: Evaluation of tropospheric SO2 retrieved from MAX-DOAS measurements in Xianghe, China, Atmos. Chem. Phys. Discuss., 14, 6501-6536, doi:10.5194/acpd-14-6501-2014, 2014.

Page 9, L17-21: See comment on "page 4" above. NOx lifetime depends on model resolution, and NO2 maxima will be diluted in coarser models. Model resolution needs to be better reported.

See reply above.

P10, L5-17: I have a hard time following the language and reasoning behind this conclusion. Please consider clarifying. Is this because the a priori profiles are generated from similar models

as those included in the comparison? Are any systematic effects buried below random sources of uncertainty?

Multiplying simulated NO_2 partial columns by column AVKs of the retrievals prior to summing up partial columns in the vertical does not have a big impact on derived tropospheric NO_2 VCDs. One of the reasons for this is that (as shown by Figure 5 and A1, revised version), AVKs are close to 1 around the boundary layer where MAX-DOAS instruments have the highest sensitivity (generally a bit larger than one close to the surface and smaller than one higher up which has a balancing effect) and that the vertical shape of the column AVK curve is in principal agreement with the shape of simulated NO_2 partial columns. At altitudes above roughly 1 km, AVKs are on average for some stations significantly smaller than one, but simulated NO_2 partial columns are also significantly smaller at these altitudes compared to lower levels, so that the contribution to the tropospheric column is limited. At higher altitudes, MAX-DOAS retrievals tend to follow the a-priori, while retrievals in the boundary layer are not much influenced by the a-priori in general. This is in contrast to the situation for satellite observations of tropospheric NO_2, which usually have a minimum of the AVK in the boundary layer, i.e. where the largest fraction of NO_2 is usually located in polluted situations. A-priori profiles used within the MAX-DOAS retrievals (see Section 2.2) are in principal agreement with the ones simulated by the models. The vertical weighting caused by application of AVKs to partial columns does therefore not significantly impact on derived tropospheric NO_2 VCDs.

The information given in the paragraph above has been added to the results section and the corresponding text changed accordingly (see p 11 l 19 - p 12 l 2, revised version). Note that no profile retrievals are performed at De Bilt, which is therefore not shown in Figure 5.

Information on how a-priori profiles were derived for each station has been added to section 2.2. For Uccle and OHP, exponentially decreasing a-priori profiles were constructed based on an estimation of vertical column densities derived by so-called geometrical approximation (Hönninger et al., 2004; Brinksma et al., 2008) using scaling heights of 1 km and 0.5 km, respectively. For Bremen, an a-priori profile which is constant with height has been assumed in the retrieval. For De Bilt, a-priori profiles of NO_2 are based on a block-profile with NO_2 present the boundary layer, boundary layer heights were taken from a climatology based on ECMWF data.

Brinksma, E.J., Pinardi, G. J., Braak, R., Volten, H., Richter, A., Dirksen, R. J., Vlemmix, T., Swart, D. P. J., Knap, W. H., Veefkind, J. P., Eskes, H. J., Allaart, M., Rothe, R., Piters, A. J. M., and Levelt, P.F.: The 2005 and 2006 DANDELIONS NO_2 and Aerosol Intercomparison Campaigns. J. Geophys. Res.,113, D16S46, doi:10.1029/2007JD008808, 2008.

Hönninger, G., von Friedeburg, C., and Platt, U.: Multi axis differential optical absorption spectroscopy (MAX-DOAS), Atmos. Chem. Phys., 4, 231-254, doi:10.5194/acp-4-231-2004, 2004.

Page 10, L21-31: This analysis and discussion is tangential to the broader scope of the paper and should be removed, as earlier noted by the authors "The impact of clouds on MAX-DOAS retrievals is described in detail by Vlemmix et al. (2015)" I do consider the comparison of model and MAX-DOAS NO2 columns under different cloud conditions to be an interesting topic for its own manuscript.

The discussion and analysis of the impact of clouds on comparison results has been removed from the results section as suggested and is regarded as a topic for future studies, which is now mentioned on p 7 l 34 and p 18 l 21 of the revised manuscript.

P10-11, L34-11: How much of the correlation is determined by seasonal and weekly cycle? Consider isolating correlation at one time of day, one season and one set of weekdays (e.g., M-F)

In response to this comment and comment b) by referee #1 three Tables have been added to the manuscript (note that in these Tables also results of individual model runs are summarized, in response to the requests by the other two referees to put more weight on individual model results in the main part of the manuscript):

-Table 3 shows statistical values of AVK-weighted tropospheric NO_2 VCDs for the four stations for the ensemble and individual model runs

-Table 4 shows the same as Table 3, but for surface partial columns of NO_2

-Table 5 shows the same as Table 3, but for seasonal, diurnal and weekly cycles of AVK-weighted tropospheric NO_2 VCDs

The following text has been added on p 15 l 22-24 of the revised version:

“Comparing Table 3 and 5 shows, that the overall correlations reached at all stations are mainly driven by seasonal and weekly cycles, while significantly lower and in many cases negative correlations are found for diurnal cycles which decreases overall correlations. An exception for the latter is Uccle, where good correlations are also found for diurnal cycles. ”

P12, L35: Consider a reference to Beirle et al. (2003). I think that this paragraph could be expanded. Day-of-week effects, over the long-term, are independent of meteorology and driven entirely by variations of emissions and chemistry. Future day-of-week comparisons would be one means of providing systematic approaches to quantify the many processes affecting NO2 (emissions, meteorology uncertainty, chemistry, observational uncertainty)

A reference to Beirle et al. (2003) has been added to p 15 l 19-21 of revised version:

“Beirle et al. (2003) investigated weekly cycles of tropospheric NO_2 based on GOME satellite observations and found a decrease in values of up to about 50 % towards Sundays over polluted regions and cities in Europe. This is in principal agreement with results of the present study, although the choice of the cities is different.”

Differences in diurnal cycles derived for weekdays and derived for weekends only are now presented and discussed in the revised version (see p 14 l 27 – p 15 l 10, p 16 l 20-27) and a corresponding Figure showing diurnal cycles for weekends only has been added (Figure 10, revised version). Note that results for weekdays only look similar to results based on all days of the week and are therefore not shown in the manuscript. As expected, diurnal cycles retrieved from MAX-DOAS based on weekends only in general show a rather flat shape for the urban stations. However, the shape of model simulated diurnal cycles looks very similar for weekdays compared to weekends, meaning that simulations fail to reproduce the observed changes towards the weekend. It should be checked in future studies if switching off diurnal scalings of emissions during weekends leads to an improvement in model performance compared to MAX-DOAS. A note on these results has also been added to the Abstract (p 1 l 14 – p 2 l 2, revised version).