

Interactive comment on “Comparison of OMI NO₂ tropospheric columns with an ensemble of global and European regional air quality models” by V. Huijnen et al.

V. Huijnen et al.

huijnen@knmi.nl

Received and published: 18 February 2010

AC1-G: We appreciate the critical and constructive comments of the reviewer. We have tried to address most of the points indicated by him/her. However, as acknowledged by both reviewers, not all could be achieved within the current environment. This is inherent to reporting on a semi-operational product like the GEMS-RAQ operational ensemble, in which the contributing models as well as the satellite data are sometimes subject to updates and changes during the operations.

AC2-G: The main concern of rev#1 is that little quantitative conclusions on individual model performance can be drawn from this study due to discrepancies between differ-

C10927

ent model setups and uncertainties in the observations.

1. During the one year of operations changes have occurred in the RAQ ensemble which have to be taken into account and which complicate the analysis. We agree with the reviewer that this can be made more clear in the introduction, and we have listed the updates of the model components during operations in the paper, see section 2: Participating models.

2. The RAQ approach is based on an ensemble of models, with different implementations of processes, emissions, boundary conditions and meteorology. Only part of these have been prescribed, e.g. the meteorological driver (ECMWF) the anthropogenic emissions and boundary conditions, and not all models follow the prescriptions. A basic idea behind ensemble forecasting is the point that the spread in values will reflect model uncertainty.

We believe that the given analysis is a valuable documentation and reference for the current GEMS-RAQ ensemble concerning their performance to model tropospheric NO₂. This point is made more clear in the introduction of the revised manuscript.

AC3-G: The reviewer argues that the analyses of model vs OMI are not sufficiently quantitative. He/she requests for a presentation of scatter plots in the comparison study. We agree that the information from scatter plots helps to quantify the discrepancy between models and observations. However, to spare a multitude of figures we will not show scatter plots for individual models in the revised manuscript. As a response we have included quantitative results for correlation, slope and offset in the revised paper, which summarise the results of the scatter plots.

AC4-G: The reviewer suggests that the use of an ensemble as a reference model helps in the comparison to OMI and in the individual model intercomparison. We agree with the reviewer. We provide extensive new analyses where we construct an ensemble based on the median of the RAQ models. The baseline comparison with OMI is now performed based on the median. The idea to rely on an ensemble of models,

C10928

rather than a single model is also one of key concepts in the GEMS regional air quality project. The rationale for such an ensemble is included in the introduction of the revised manuscript.

AC6-G: The reviewer suggests to include a separate section on the spread between the models, and to quantify better the spread between the models. This is a good suggestion which we have taken into account in the revised manuscript. In this new section we describe differences between models and the ensemble median. Also included is a discussion on the statistics derived from the scatter plots.

Additionally we have moved the section on the intercomparison of the vertical profiles just after the intercomparison of the tropospheric columns, as we believe this fits better in the line of the story.

AC7-G: The reviewer has difficulties to understand the analyses concerning the averaging kernel, see also AC13 below. The reviewer notes that it is surprising that the effect of the kernels applied to the RAQ models is so small compared to the kernel-free comparison, in particular given the larger spatial variability in the RAQ models, compared to the a-priori. The small difference is partly an effect of the regional averaging that is adopted here. Also from figure 11 in the original manuscript we show that there is some effect of local differences, see also AC12 from rev#2. To make this point more clear in the revised manuscript we have included an extra figure to highlight the regional differences between the modeled columns with and without application of the averaging kernel.

Also to make our analyses more clear we have reformulated most of the section on the effect of the averaging kernel, giving more attention to the quantitative ratio between the partial columns in the boundary layer and the free troposphere.

abstract

AC9-G: In response to the reviewer we have rewritten the abstract. The conclusions

C10929

have been phrased more carefully. The spread of values has been re-examined in a new section on the model comparison which is based on the median model as reference. We have also given more attention on the skill of the ensemble median, compared to the individual models.

Detailed comments

RC1: P 22274, I7: 'Studies have shown that a change in emission levels causing changes in climate can counteract ...' I'm not sure what the authors are trying to say here. Please rephrase.

AC1: sentence removed. The given arguments and references are not essential for the manuscript.

RC2: P 22276, I20: The DOMINO data version used is 1.02, the version described in the Boersma et al. reference given is 0.8. Please briefly discuss the differences.

AC2: In this manuscript we present results from the OMI DOMINO Collection 3 version 1.0.2. The main differences with version 0.8 described in Boersma et al. [2007] are the use of level-1 radiance and irradiance spectra with much improved instrument calibration parameters (Collection 3, see Dobber et al. [2008]), and the switching off of the a posteriori stripe correction used in earlier versions (before v1.0.0), leaving some residual stripes. More details can be found in the DOMINO Product Specification Document (http://www.temis.nl/docs/OMI_NO2_HE5_1.0.2.pdf). We include these notes in the revised manuscript.

Dobber, M., Kleipool, Q., Dirksen, R., Levelt, P., Jaross, G., Taylor, S., Kelly, T., Flynn, L., Leppelmeier, G. and Rozemeijer N.: Validation of Ozone Monitoring Instrument level 1b data products, *J. Geophys. Res.*, 113(D15S06), doi:10.1029/2007JD008665, 2008.

RC3: P 22276, I24. What is an 'optimum resolution'?

AC3: This refers to the optimal resolution achieved for OMI at nadir. The sentence is rephrased to: "OMI achieves optimal resolution of 13x24 km² at nadir"

C10930

RC4: P22280, I9: The difference between the RETRO and the TNO inventory (a factor of 2.5) cannot be explained by RETRO being 'outdated for the current evaluation period'. Emission reductions have been achieved since 2000, but not nearly on this level. Also, the TNO inventory is for 2003 and could also be considered to be outdated.

AC4: Exact reasons for the discrepancy and a justification which inventory is better cannot be found from literature. We acknowledge that any such suggestions in the original manuscript cannot be proven, as suggested by rev#1 and rev#2, and have been removed from the text. Also we agree with reviewer #1 that the decrease in emissions of TNO versus RETRO over western Europe cannot be explained by an emission-trend, and also that emissions over eastern Europe remain uncertain in both inventories, despite the fact that they are similar.

RC5: P2280, I20: Do all RAQ models use the same diurnal, weekly and monthly emission profiles? If not, please add this information to Table 1.

AC5: We acknowledge that the imposed NO_x emissions are one of the most important drivers in observed NO₂, while significant differences exist between all models, such as differences in anthropogenic emissions, biogenic emissions, emission heights, distribution of NO_x- emissions over NO and NO₂, choice of diurnal, weekly and seasonal cycle. We are now more clear about this in the introduction, and more explicit with respect to the individual choices for the different models in section 2.3, the description of the emissions. Table 1 in the revised manuscript mentions more explicitly the main differences between the individual RAQ models, such as the anthropogenic NO_x emissions as applied in the EMEP model. The reviewer requests also a discussion of the different assumptions in the diurnal, weekly and seasonal cycles in the different models. Although most RAQ models apply such cycles, the details with respect to their implementation are different, and their effects are difficult to assess.

However, the diurnal cycle of the model NO₂ concentrations was already reported in the original manuscript, and this information documents differences in the diurnal

C10931

variability among the models. Details of these differences can only be investigated in specific sensitivity studies, which is outside the scope of the current analysis. Therefore a more detailed specification of all differences in the individual models does not lead to additional insight, and are therefore omitted. We address these issues shortly in the revised manuscript.

RC6: P 22281, I20: The OMI data product appears to have changed over the period discussed in this paper. As this might have an impact on the comparisons shown e.g. in Fig 7, the possible impact of the processor change needs to be quantified.

AC6: The surface albedo map in the DOMINO retrieval algorithm has been replaced after 17 feb 2009. It has been shown that the improved OMI surface reflectivities [Kleipool et al., 2009] are higher than GOME/TOMS reflectivities by approximately 1%, owing to the minimum Lambertian equivalent reflectance approach used by the latter. Retrievals with the OMI surface reflectivities result in tropospheric columns over Europe that are reduced by 12% compared to retrievals with the GOME/TOMS reflectivities, as shown by Hains et al. [2010], and as indicated in section 3.2 of the original manuscript.

Kleipool, Q. L., Dobber, M. R., de Haan, J. F., and Levelt, P.F.: Earth surface reflectance climatology from 3 years of OMI data, *J. Geophys. Res.*, 113, D18308, doi:10.1029/2008JD010290, 2008.

Hains, J., Boersma, K., Kroon, M., Dirksen, R., Cohen, R., Perring, A., Bucsela, E., Volten, H., Swart, D., Richter, A., Wittrock, F., Schoenhardt, A., Wagner, T., Ibrahim, O., van Roozendaal, M., Pinardi, G., Gleason, J., Veeffkind, P., and Levelt, P.: Testing and Improving OMI DOMINO Tropospheric NO₂ Using Observations from the DANDELIONS and INTEX-B Validation Campaigns, *J. Geophys. Res.*, doi:10.1029/2009JD012399, in press, 2010.

RC7: P2283, I14: Why is the assumed high bias of OMI more present in summer than in winter? I do not see indication for this seasonality in the studies cited.

C10932

AC7: Recent studies using the DOMINO v1.02 data [Boersma et al., 2009; Zhou et al., 2009; Hains et al., 2009; Lamsal et al., 2010] show modest to good agreement between OMI data and independent observations, but OMI appears to be biased high by 0%-40%. Over Europe, Hains et al. [2009]. Zhou et al. [2009] find a high bias of 20% in summer, but only a small bias in winter. Although these results are based on a relatively small set of independent measurements with considerable uncertainties over the Netherlands, Switzerland, and the Po Valley, they are consistent with known systematic errors in retrieval parameters. Apart from the plausible suggestion that the TM4 profile is probably more realistic during winter than during summer, errors in the calculation of the altitude-dependent air mass factors close to the surface, errors in the surface reflectivity, and errors in cloud retrievals contribute to the high bias in the retrievals.

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J.: Validation of urban NO₂ concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, *Atmos. Chem. Phys.*, 9, 3867-3879, 2009.

Lamsal, L., Martin, R.V., van Donkelaar, A., Celarier, E.A., Boersma, K.F., Dirksen, R., Luo, C., and Wang, Y.: Indirect validation of tropospheric nitrogen dioxide retrieved from the OMI satellite instrument: insight into the seasonal variation of nitrogen oxides at northern

Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., and Wang, P.: An improved tropospheric NO₂ retrieval for OMI observations in the vicinity of mountainous terrain, *Atmos. Meas. Tech.*, 2, 401-416, 2009.

RC8: P 22287, I5: How is the model spread defined? And is it relative to the model mean as stated in the text or to OMI as said in the table caption? In my opinion, there is little justification for using OMI as a reference here.

AC8: Metrics for a quantitative evaluation of the model spread in the different regions

C10933

are made more clear throughout the revised manuscript. The spread is quantified as the rms, generally presented as a percentage of the mean of all RAQ models. This is evaluated based on the regional mean from all contributing RAQ both for the winter and summer season. Also the mean of the RAQ model ensemble and the mean observations are reported.

RC9: P22287, I25: The better agreement in June 2009 than in July 2008 could still be related to the change in OMI data version. Or are June 2008 OMI values also a factor of two lower than in July 2008 in Eastern Europe?

AC9: An error in our script to generate plots for the area-averaged seasonal variation of OMI observations over eastern Europe and the Iberian Peninsula led to false conclusions that OMI NO₂ showed significantly smaller values in (late) spring 2009, compared to (early) summer 2008. The revised figures make the analysis more consistent with respect to biases in 2008 and 2009, with generally larger biases in spring-summer compared to winter. As written in AC6, the change in albedo leads to a decrease in observed columns of the order of 12%. In an additional study we also compared biases between model median and OMI observations for July-August 2008 against July-August 2009. On average for the mid/south region the bias decreases from 60% to 40%. Over the Western-Europe region this decreases from 47% to 26% and for the eastern European region from 68% to 54%. But these decreases could mainly be attributed to increases in modeled columns; OMI observations themselves decreased by no more than 5% between summer 2009 and summer 2008. A sensitivity study would be required to show the effect of the albedo only. Summarizing, part of the different bias could indeed be related to the change in the data-product, but the main difference between summer and winter still exists. We have reformulated the text describing the differences between 2009 and 2008 in this respect.

RC11: P22289, I5: As stated above, I don't think that models should be used to validate measurements. Even if the surface concentrations agree well with in-situ observations, the modelled vertical profile might still be systematically wrong leading to an underes-

C10934

timation of the tropospheric column.

AC11: We argue that the surface concentrations from background stations, averaged over a region and a season, can help to constrain the origins from the bias in the tropospheric columns. But the reviewer is right that the biases between the models and surface observations cannot be directly related to the bias in the OMI-observations, due to uncertainties in the profile shape. In subsequent sections we study in more detail the effect of the profile shape on the modeled as well as the observed columns (sec. 8: effect of the averaging kernel, sec. 9: differences between profiles from individual models). At this location we have removed the suggestion that OMI is biased high.

RC12: P22289, I20: Why is this an argument for using a model ensemble? I'd expect that one would identify the model which performs best against validation data and then use it instead of deteriorating the performance by averaging with model results that are not in agreement with observations. If the authors would like to make this point they should elaborate it more and also include the model averaged surface concentration in Fig. 8 so that the readers can judge if this is a good approach.

AC12: In the revised manuscript we have introduced the RAQ model median as the ensemble reference. A measure based on the median does not deteriorate due to model outliers, and it is shown in several studies (e.g. van Loon et al, 2007) to behave better than any of the individual models. However, this approach is still open for improvements, as suggested by the reviewer, when observational data would be available to select a more optimal model ensemble, but this requires a separate study and is outside the scope of our paper. As suggested by the reviewer in the revised manuscript we have included an analysis of the average biases of the ensemble median. It shows that the median performs equally well compared to the best performing individual models. We have reformulated the text accordingly.

van Loon, M., Vautard, R., Schaap, M., Bergstrom, R., Bessagnet, B., Brandt, J., Builtjes, P.J.H., Christensen, J.H., Cuvelier, C., Graff, A., Jonson, J.E., Krol, M., Langner,

C10935

J., Roberts, P., Rouil, L., Stern, R., Tarrason, L., Thunis, P., Vignati, E., White, L., and Wind, P.: Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, *Atmos. Environ.*, 41, 2083-2097, 2007.

RC13: P22292, I10 and Figure 10: using the partial columns in the discussion and the Figure is confusing. I was trying to understand the Figure and it took me some time to realize that this is not proportional to the vertical profile but depends on the thickness of the individual layers in TM4 which is not given. As the figure is linear in pressure, it would look very different if it would have been plotted as a function of altitude. This figure does not show how different altitudes in the atmosphere contribute to the (retrieved) tropospheric NO₂ column but rather how different model layers contribute.

AC13: In the revised manuscript we present a modified version of figure 10 from the original manuscript, to better reflect the essence of the partial columns. The reviewer also suggests to use the altitude on the vertical axis. This implies that visually some more weight is given to the levels in the free troposphere, where gridbox altitude ranges are larger than in the boundary layer. We think that the pressure axis is more appropriate, as in this case it is better reflected that equal weight should be given to all partial columns that are used to construct the full tropospheric column. We have reformulated the text describing the effect of the averaging kernel. In the revised manuscript this section is carefully reformulated, explaining quantitatively the compensating effect that takes place.

RC14: P22295, I7: Before, you explained the difference in summer by OMI retrieval problems. Here it is suggested that this is related to averaging kernel issues which according to Fig. 11 appear to be minor in both summer and winter.

AC14: In the discussion on the averaging kernel we show why modeled total tropospheric columns from the RAQ model do not change much when applying the averaging kernel. In summer this is due to a cancellation of effects in the boundary layer and the free troposphere. In winter we showed that both counteracting effects are of smaller

C10936

importance: the a-priori is more similar to the RAQ model, both in the boundary layer and in the free troposphere. Therefore both in summer and in winter the difference between tropospheric columns with/without AK is small on average.

Still, the large concentrations in the boundary layer in the a-priori profile may contribute to an over-estimation of the OMI product. This is not in contrast with the fact that the on average the regional models do not change when applying the averaging kernel.

RC15: P22295, I18: I would call the agreement qualitative, not quantitative.

AC15: We will replace 'quantitative' by 'qualitative'

RC16: P22298, I25: Maybe I have missed that point in the papers cited, but from where exactly comes the information that OMI is 0-40% high in summer but less in winter? Is that from the Hains et al. paper which was not available to me?

AC16: see AC7.

RC17: P22299, I5 earlier, not earlier

AC17: Replaced.

RC18: P22299, I3: Again, I don't like the validation of measurements by models. Also, you imply that the problem with the TM4 a priori is the reason for the OMI overestimation which is at least partly in contradiction to the results of the Averaging Kernel tests which imply that you would get the same OMI columns when using EURAD or CAM profiles.

AC18: In the revised manuscript we point out that there is no contradiction between the suggestion that OMI NO₂ and the RAQ models show a bias, and the fact that the RAQ models do not change when applying the current AK's.

In the original manuscript we mention that the TM4 a priori has significantly larger surface NO₂ concentrations, as compared to the two other RAQ models, while RAQ models are in line with surface observations. We realize that this is not very clear. In

C10937

the revised manuscript we evaluate the OMI columns using surface observations rather than other models, see also AC19.

RC19: P22299, I7: 'We showed that the TM4 a priori NO₂ concentrations near the surface as used in the retrieval algorithm are significantly larger relative to all contributing global and RAQ models' - I haven't seen this in the paper but it is a good idea. I'd suggest adding the TM4 values to Fig. 8.

AC19: We chose not to include the TM4 surface concentrations in figure 8, as this figure is concerned with the RAQ model validation. This is also the reason why TM4 a-priori profile data is not included in figures 12-13, where the vertical profiles are presented. In the original manuscript we based our conclusions of high TM4 surface concentrations on figure 10, where we show that partial columns at the surface, and hence surface concentrations, are much larger in TM4 compared to the RAQ models, as described in section 8. We realize that this argumentation (see also AR18) is indirect, while for the current analysis, the evaluation of the TM4 retrieval, this analysis plays an important role. Therefore we have included an explicit description of the evaluation of TM4 surface concentrations compared to observations in the section describing the effect of the AK. We now include a note that TM4 surface concentrations are over-estimating NO₂ compared to the presented measurement rural sites, especially in summer. We find that in August 2008, the sampling of a-priori TM4 profile data at the rural stations from the Dutch Air Quality Network give a value that is twice as high as the surface observations, while in winter it is in line with the ensemble median, and underestimates the surface observations by 20%.

RC20: P22299, I20: I think that more quantitative statements are needed on the 'good correspondence in spatial patterns and seasonal cycle' here as suggested in the general

AC20: We have reformulated the conclusions, being more careful with our statements. For instance, we have included quantitative statements based on the analysis of corre-

C10938

lation coefficients.

RC21: P22300, I23: I don't think that changing photolysis rate is a good example for the effects of increased model resolution as the data shown here have been selected for clear sky scenes.

AC21: The reviewer is fully correct. Recent investigations suggest that higher NO₂ in the TM5-zoom version is related to the sensitivity of the model to the time-stepping. The parameterization was originally developed for 1800sec time-steps, but with shorter time steps the vertical transport (mixing) seems more effective, leading to longer NO₂ life times.

RC22: P22300, I26: As before, I think this argument is questionable (although I tend to agree that the OMI seasonality appears to be too small).

AC22: See also AC19. The reviewer is correct that the quality of the modeled vertical profile shapes can be questioned, when not having observational data to compare to. But we also think that on average the analyses of surface observations (the only independent reference here) point at a clear over-estimation of surface concentrations in the a-priori. We have reformulated our conclusions with more care.

Interactive comment on Atmos. Chem. Phys. Discuss., 9, 22271, 2009.