Atmos. Chem. Phys. Discuss., 9, C8029–C8036, 2009
www.atmos-chem-phys-discuss.net/9/C8029/2009/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Comparison of OMI NO_2 tropospheric columns with an ensemble of global and European regional air quality models" *by* V. Huijnen et al.

Anonymous Referee #1

Received and published: 8 December 2009

In their manuscript 'Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models', V. Huijnen and colleagues report on a comparison of one year of tropospheric NO2 simulations over Europe from 9 regional and two global models. The model fields are compared to each other and to satellite observations from the OMI instrument as well as in-situ surface data in the Netherlands. In addition to columns and surface concentrations and their seasonality, the diurnal variation, the vertical distribution and the effect of application of averaging kernels is discussed.

The paper is well written and provides an interesting overview over the ability of current

European regional models to simulate tropospheric NO2 columns and surface concentrations. However, most of the statements made are rather qualitative, and for the reasons discussed below, it is difficult to draw any firm conclusions from the study. In my opinion, some more work is needed to come to more quantitative conclusions and also many smaller issues need to be addressed before the paper can be accepted for publication in ACP.

General comments:

This paper aims at comparing en ensemble of regional models with satellite columns and surface observations. Unfortunately, a number of problems makes a quantitative comparison difficult if not impossible:

- Contrary to what is stated in the abstract, the models did not use the same emission inventories – the global models used completely different emission data and between the regional models, there are many smaller differences which make interpretation of the results difficult
- the MOZART run used by many (but again not all) regional models was flawed and provided unrealistic boundary conditions
- some of the models did not provide consistent data sets
- the satellite data changed version in the middle of the time period
- only two of the 9 models saved results at all the altitudes necessary to apply averaging kernels for quantitative comparison with the satellite data

This is clearly not a good starting point for a quantitative comparison and is a problem for all results shown here, but I assume that this cannot be fixed in a realistic time frame. I'd therefore suggest to at least clearly state these problems at the beginning of the paper (and not to mention them one by one over the different sections).

ACPD

9, C8029–C8036, 2009

One main point of the paper is the comparison between model fields and satellite data. This comparison is performed by showing monthly averages, discussing some of the features and then providing regional averages and 'spread' of the data from all the models. I think that a more detailed and quantitative discussion is needed here. The difference between global and regional models is their spatial resolution, and OMI data has been used as it provides good resolution. I therefore suggest that the authors show scatter plots between all individual models and the satellite data and provide slope, offset, and correlation coefficient for these comparisons. As the authors suggest that a model ensemble should be used, this should also be compared to the satellite data. With these numbers, the reader can get a better idea on how well the model data really agree with the measurements on the spatial scales relevant for regional models.

Throughout the paper, it is stated that the agreement between the models is satisfactory, and the spread of values is quantified as being 20 – 40%. I find it difficult to reconcile this with the figures which show large differences between the models in all respects. As one example, the largest noon value in Figure 9 is nearly four times the value of the lowest model, and this is already averaged over one month and all of the Netherlands. I think it would be worthwhile to add a section discussing the variability in the models, the most probable reasons, what this implies for application of such models in air quality forecasting and ways forward to improve agreement.

The suggestion of using a model ensemble as a more robust means of predicting NO2 pollution is mentioned briefly in the paper, and it would deserve a bit more discussion. Inclusion of the model mean in figures (as already suggested above) would be useful in several places, and this point could be taken up again in the conclusions.

The discussion of the averaging kernels left me a bit clueless. The paper shows, that for the region and models selected, application of the averaging kernels has a small effect only. This is a surprise as one would expect that regional models better represent the spatial variability in NO2 vertical profiles, leading to a significantly larger variability in the satellite sensitivity than when using the TM4 a priori. This does not appear to be

ACPD
9, C8029–C8036, 2009

Interactive Comment

the case but it is also difficult to judge as only the averages over the western Europe region are shown. In the paper it is argued that the lack of impact of the AK is due to a compensation effect between the boundary layer and the free troposphere. If I understood it correctly, this is because the regional models do not have enough NO2 in the free and upper troposphere, and this has two effects: An underestimation of the tropospheric column and an underestimation of the satellite sensitivity. It would be good if this point could be illustrated a bit more by constructing a 'best guess' profile from a combination of e.g. EURAD and TM5 for a typical situation and then demonstrating the effects.

Abstract

In my opinion, the abstract needs to be rewritten as several of the statements made are misleading or not fully supported by the study:

'The participating models apply principally the same emission inventory'

According to the text, EMEP uses its own emission inventory. In the SILAM model, a mixture of TNO and EMEP inventory is used. For shipping, lightning and aircrft emissions, varying approaches were used. The two global models use yet another inventory (RETRO). Treatment of the diurnal variation also varies between modles. Considering this, this statement is misleading.

'It is also shown that the NO2 concentrations from the upper part of the troposphere (higher than 500 hPa) contribute up to 20% of the total tropospheric NO2 signal observed by OMI'

It should be mentioned that this conclusion is based on model profiles, not measurements, and does depend on the specific model runs used.

'Compared to the global models the RAQ models show a better correlation to the OMI NO2 observations'

This appears to be the case but should be quantified by showing spatial correlation

ACPD
9, C8029–C8036, 2009

Interactive Comment

coefficients between models and measurements for each model individually and for the model average used in some of the comparisons.

‘The spread in the modelled tropospheric NO2 column is on average 20–40%.’

According to the table caption, this number is based on using OMI as the standard. I don’t think this is appropriate and suggest using the model average as baseline for comparison. This will lead to significantly larger spread in summer when model values are low. Figure 9 gives an indication how large the spread of the model results is.

‘These findings suggest that OMI tropospheric columns in summer over polluted regions are biased high by about 40%’

While OMI might well be biased high in summer, I don’t think that comparison to model results is the right way to quantify such a bias. In general, one should use the measurements (with error bars) as a tool to judge the models, not the other way round.

‘The diurnal cycle and profiles in the regional models are well in line’

Considering the spread of values in Figs. 9 and 12, I don’t think that I can agree with this statement, at least not if absolute values are the quantity of interest. If the sentence is referring to relative changes this should be stated and discussed quantitatively in the text.

Detailed comments

P 22274, l7: ‘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.

P 22276, l20: 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.

P 22276, l24. What is an ‘optimum resolution’?

ACPD

9, C8029–C8036, 2009

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

P22280, l9: 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.

P2280, l20: Do all RAQ models use the same diurnal, weekly and monthly emission profiles? If not, please add this information to Table 1

P 22281, l20: 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.

P2283, l14: 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.

P 22287, l5: 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.

P22287, l25: 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?

P22289, l5: 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 underestimation of the tropospheric column.

P22289, l20: 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

ACPD

9, C8029–C8036, 2009

Interactive Comment

so that the readers can judge if this is a good approach.

P22292, l10 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 realise 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.

P22295, l7: 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.

P22295, l18: I would call the agreement qualitative, not quantitative.

P22298, l25: 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?

P22299, l5 earlier, not earier

P22299, l3: 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.

P22299, l7: '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.

P22299, l20: I think that more quantitative statements are needed on the 'good correspondence in spatial patterns and seasonal cycle' here as suggested in the general

ACPD

9, C8029–C8036, 2009

comments

P22300, l23: 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.

P22300, l26: As before, I think this argument is questionable (although I tend to agree that the OMI seasonality appears to be too small).

P22301, l14: height, not hight

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

ACPD

9, C8029–C8036, 2009