Responses to the comments of anonymous referee #2

We would like to thank for the comments that helped to improve our manuscript. Please find below your comments in blue, our responses in black and modifications in the revised manuscript related to technical or specific comments in italic and in quotes. Additional text modifications have been done as most figures have been updated following your suggestions. All modifications are highlighted in the revised manuscript.

The authors' present an analysis assessing CAMx model under-predictions of ozone production in Europe, and arrive at the main finding that NOx emissions are likely under-predicted in existing emission inventories. Overall, the paper is well written, clear, scientific methods appropriate, and in general findings/conclusions wellsupported. I have a couple of critiques that hopefully will help strengthen this paper. First, I think the sensitivity analysis adjusting NOx emissions could be more specific to transportation emissions, rather than applied across total anthropogenic emissions. Second, the sensitivity analysis of wind speed seems to be in the opposite direction based on the model bias for this meteorological parameter. With revisions to the model test cases, I believe this manuscript could be considered for publication in Atmospheric Chemistry & Physics.

General Comments

1. Section 2.3 ("Emissions"). The authors present a nice literature summary suggesting that transportation emissions of NOx are uncertain and may be underestimated by a factor of 2-4 (Page 5, Line 2). However, it appears that the authors' scaled all anthropogenic NOx emissions by a factor of 2 (Table 3) in sensitivity tests of the model. Based on Figure 2, scaling up road transportation emissions by a factor of 4 would roughly equal a factor of 2 increase in anthropogenic emissions. I suggest scaling the transportation sector only in the sensitivity analysis rather than all anthropogenic sources. First, it is not clear that point source emissions should exhibit uncertainties as large as the transportation sector. Second, the diurnal and day-of-week cycle in transportation emissions differ (Nassar et al., 2013) from point/area sources, which could affect diurnal and day-of-week patterns in the model and affect NO2 and O3 evaluations (Marr and Harley, 2002). Third, transportation emissions are likely more concentrated in urban cores relative to other sources of NOx (e.g., power generation/industry), which could affect the spatial distribution of emissions and model evaluations performed on rural background monitors (Page 6, Line 13).

Thank you for this suggestion. We included the suggested sensitivity test in our study. The sensitivity scenario is labeled "4traf_NO_x" where we scaled up only the road transport NO_x emissions (SNAP 7) by a factor of 4. Indeed the behavior of the "4traf_NO_x" scenario is very similar to the "2NO_x" scenario, pointing out that probably the uncertainties in the NO_x emissions that are related to the transportation sector are more responsible for the model bias patterns examined in this study. We updated all the respective figures and modified the text to properly describe and discuss the updated results.

2. Section 3.3 ("Sensitivity of ozone to meteorology"). The rationale behind increasing temperature by +4 degrees Celsius in the model (Table 3) makes sense based on systematic underestimates in temperature in the base case (Figure S3, also stated on Page 14, Line 3). However, why is wind speed reduced in the model rather than increased, when the model generally systematically underestimates wind speed in the base case (Table 4/Figure S4/Figure S6)? By increasing wind speed in the model, the ozone under-predictions will likely be worse, and a stronger argument can be made that meteorology is unlikely to explain the model discrepancies in relation to emissions.

We agreed and followed your suggestion by applying a sensitivity test with double the wind speed ("WSx2"), but we also kept the previous sensitivity test of "WS/2" to investigate also the overestimation of the low wind speed (< 2 m s^{-1}) in the regions where the fraction of the low wind speed overestimation compared to the total bias is not very small. We updated all the respective figures and we edited or added some text to properly describe and discuss the updated results.

Specific Comments

3. Section 2.3 (Page 4, Line 13). I think this paragraph could benefit from a description of why anthropogenic VOC emissions are uncertain at a ~50% level, similar to the proceeding discussion of why anthropogenic NOx emissions are uncertain. One sentence here seems too brief.

We discussed some possible sources of uncertainty for the anthropogenic VOC emissions:

"The VOC emission uncertainties can be due to a number of reasons such as: i) the small number of measured vehicles for the transportation sector, since the VOC species resolution rely on measurements, ii) not enough available measurement data for the combustion-, process-, and production-related emissions compared to the much higher number of individual emission sources, iii) the large variety of the VOC compositions in the used solvents, iv) the measurement uncertainties (Theloke and Friedrich, 2007) "

4. Section 3.1 (Page 10, Line 13). "… the overestimation of higher ones …" I believe the authors' mean *underestimation* here.

Corrected.

5. Section 3.2 (Page 12, Second Paragraph). A summary point at the end of the paragraph would be helpful here. It seems that the authors' might want to emphasize that NOx emissions need to be increased across most regions to improve results.

A sentence was added as a summary point at the end of that paragraph:

"Overall, our emission-sensitivity analysis indicates that the NO_x emissions, especially from the transportation sector (SNAP 7) in central, eastern and southern Europe might be too low in the emission inventories."

6. Section 3.2 (Page 12, Lines 21-22). It would help to label the slopes of the dashed grey lines in Figure 9, to help the reader more clearly discern the points made in this paragraph.

We modified the regions in the European domain and it is now even more clear that different regions have different model bias patterns and/or different responses to the sensitivity tests. Therefore, we removed the dashed grey line form Fig. 9, as we believe that this line will represent more the larger regions (ME and BX) and for most of the emissions sensitivity tests these regions have different responses, which can compensate each other. We think that it is better for the reader to extract more consistent information from this figure by examining the position of the region-colored scatter points around the 1:1, 1:2 and 2:1 lines.

7. Section 3.3 ("Temperature"). Figures S3 and S5 seem inconsistent. While Figure S3 shows a general under-prediction of temperature by the base case model, Figure S5 seems to be showing a lot stations being over-predicted in the model (yellow and orange markers). I'm wondering if this related to the under-predicted sites being blocked out by the over-predicted sites in the coloring scheme. Suggest revising the presentation of Figure S5.

We modified the Figures S4 and S6 (i.e. S3 and S5 in the original manuscript) changing the coloring scheme and shrinking the markers in Fig. S6. With the new region separation the comparison of the temperature bias in Eastern Europe (EA region) between the Figures S4 and S6 is now more clear. The temperature overestimation that is observed in Fig. S6 is also depicted in Fig. S4. We believe that Figures S4 and S6 look now consistent.

8. Section 3.3 (Page 14, Lines 25-31). These sentences do not seem to support the sensitivity test performed in the model where wind speeds are decreased, since: (i) the model seems to be doing well already (Line 25), (ii) most observations show a model under-prediction in wind speeds rather than over-prediction (Line 28), and (iii) the low wind speed conditions where the model over-predicts wind speeds comprise a minor fraction of observations (Line 30). Suggest revising this sensitivity test for wind speed, to increase rather than decrease in the model.

We responded to your general comment (2) about doubling the wind speed. Also, we replaced the wind speed normalized mean bias and gross error with the not normalized ones in Fig. S5, as we believe that the non-relative statistical metric for the bias and gross error gives a more robust overview of the model performance. In addition, the model evaluation in Fig. S5 is performed against observational data bins and so a general overview of the model performance in relative terms can also be seen.

9. Section 4 (Last Paragraph). I think this last statement made here could be stronger by performing a sensitivity test of transportation NOx emissions only, which would be in line with the literature suggesting that this sector is consistently underestimated in Europe (Anenberg et al., 2017; Karl et al., 2017). Rather than draw attention to all anthropogenic emission sources, it would be helpful to identify which sectors specifically need the most improvements in emission inventories.

Thank you for your suggestion. As we wrote above, we followed your suggestion and performed the sensitivity test with traffic NO_x emissions.

References

Theloke, J., and Friedrich, R.: Compilation of a database on the composition of anthropogenic VOC emissions for atmospheric modeling in Europe, Atmos. Environ., 41, 4148-4160, doi:10.1016/j.atmosenv.2006.12.026, 2007.