Reply to comments of Anonymous Referee # 2:

Original comments are in *italic*, replies in blue and proposed new text is in normal font.

The paper describes the application and evaluation of the CAMx model with the Volatility Basis Set scheme used for formation of secondary organic aerosols. The study includes several sensitivity simulations varying the volatility and emission parameters of the organic species. In-general, the study goes along the same lines of several existing applications, some quoted by the authors in the introduction. In that sense, I found little new or innovative pieces in the paper. From the other side, such evaluation exercises are useful for collecting experience with the VBS approach. Till now, it falls short of demonstrating a major breakthrough in the models performance as a reward for high complexity and bulkiness.

The paper is comparatively well written except for the results section 3.

I however noticed a few omissions, some with potentially heavy consequences, which should be brought up.

Thank you for your comments and suggestions to improve our manuscript. Our answers to questions follow below:

General comments

1.The most-important omission is the analysis of the emission dataset. A potentially woeful problem, for instance, is seen from Figure 8, the S1 scenario. The concentration map evidently reproduces the emission distribution, which almost completely misses half of the countries. This is a major caveat of the input dataset, which, if confirmed by the explicit emission analysis, would disqualify the whole exercise: the authors would have to switch to another emission dataset.

Thank you for this remark. We improved the color scale as shown below in order to facilitate the visualization of countries where low OA concentrations are predicted, e.g. over Germany, Spain and UK. Emissions used in this study are based on the main European inventories TNO-MACC and EMEP which are widely used in European modelling studies. Details of the construction of emission inventory, improvements and uncertainties are discussed in Kuenen et al. (2014). We believe that the emission inventory used in this study is the best available in Europe, however, one should bear in mind its limitations. Using the same emission inventory, Bergström et al. (2012) showed that wood burning emissions in Sweden were underestimated during winter as also suggested by our results in this study.

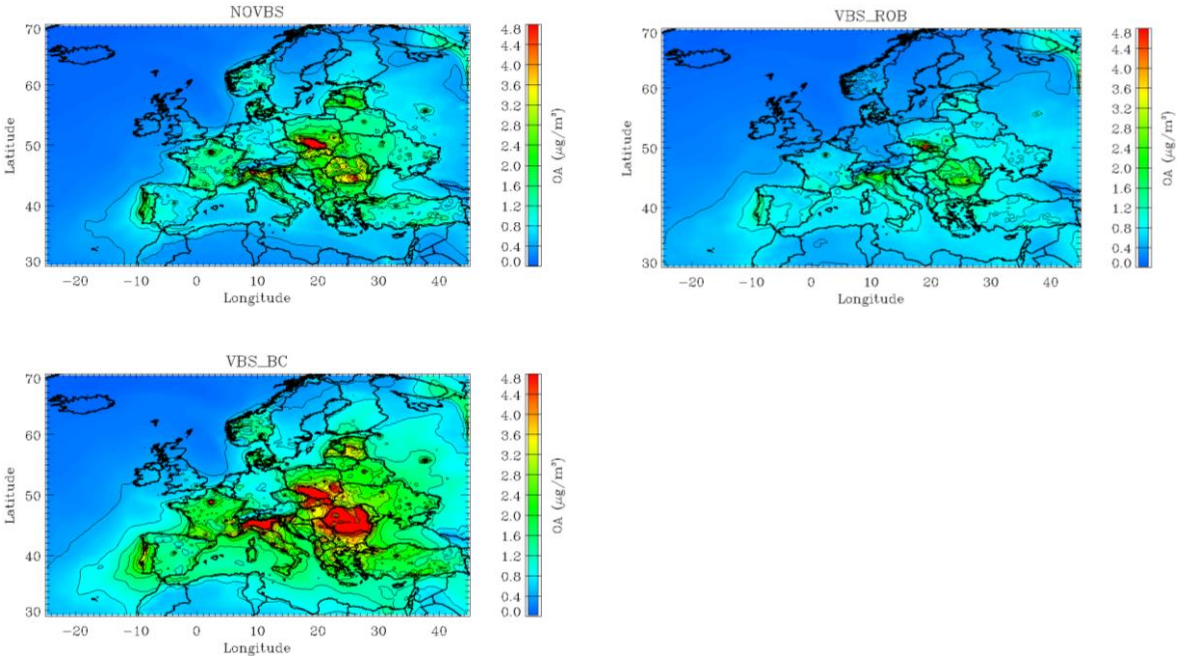


Figure 6. Predicted OA concentrations over Europe for the NOVBS, VBS_ROB and VBS_BC scenario in February-March 2009. Note that the color scale was limited to a maximum of 4.8 µg m^{-3} to facilitate comparison of the panels.

2. From the other side, the authors fell to a frequent modeller's trap of blaming emission for poor model performance, often with thin supporting analysis. Some of these blames may be justified, some may be not. For instance, I found it hard to believe the long discussion in p. 35657, where the authors try to explain the strong systematic NO2 under-estimation – and blamed emission. I found an alternative and much simpler potential explanation: nitrates are strongly over-estimated in most of cases, which would probably make-up for the deficit and suggest problems in the model chemistry rather than emission.

Thank you for this remark. We revised Section 3.1 to make it clear that there might be more than one reason for the discrepancy between modeled and measured NO_2 concentrations. One of them is the difficulty in reproducing the PBL height correctly as discussed in detail by Bessagnet et al. (2016). On the other hand, although NO_x emission estimates in Europe are thought to have an uncertainty of about ± 20%, the complete data set used in the inventories has much higher uncertainty (Kuenen et al., 2014). A recent study identified a significant discrepancy between emission estimates and actual flux measurements, with the highest underestimation being a factor of two in central London mainly due to under-representation of real world road traffic emissions (Vaughan et al., 2016). Most of the models within the EURODELTA III exercise underestimated NO_2 concentrations during the modeled periods (Bessagnet et al., 2016) especially during day time. The referee's potential explanation as over-estimated nitrate making up the deficit for NO_2 might certainly be a sound one. In addition to uncertainties in NO_2 emissions, there might also be too much HNO_3 in the model due to not enough deposition. Wet deposition of oxidized nitrogen was underestimated supporting the hypothesis of insufficient deposition leading to overestimation of nitrate.

We added daily average time series of NO_2 for the period of Feb-Mar 2009 for the stations used in Table2 (Figure S1, below). Moreover we reported daily average time series of NO_2 at stations not exceeding 5 ppb (92% of the stations used in the top panel) in order to remove the influence of polluted areas in the proximity of rural-background stations that might be difficult to resolve (middle panel in Fig. S1). The model performance improved significantly giving more confidence regarding the emissions. One should also bear in mind that measurements might have interferences from other oxidized nitrogen compounds leading to too high NO_2 concentrations (Villena et al., 2012). We also included emission map of NO for 1 March 2009 at 6 AM as an example (Figure S2). Spatial distribution of emissions looks reasonable and in line with other model exercises. High emission of NO are predicted in the Benelux area, Po Valley, Germany and in some of the eastern European countries. High NO emissions due to ship traffic are also visible especially in the Mediterranean Sea.

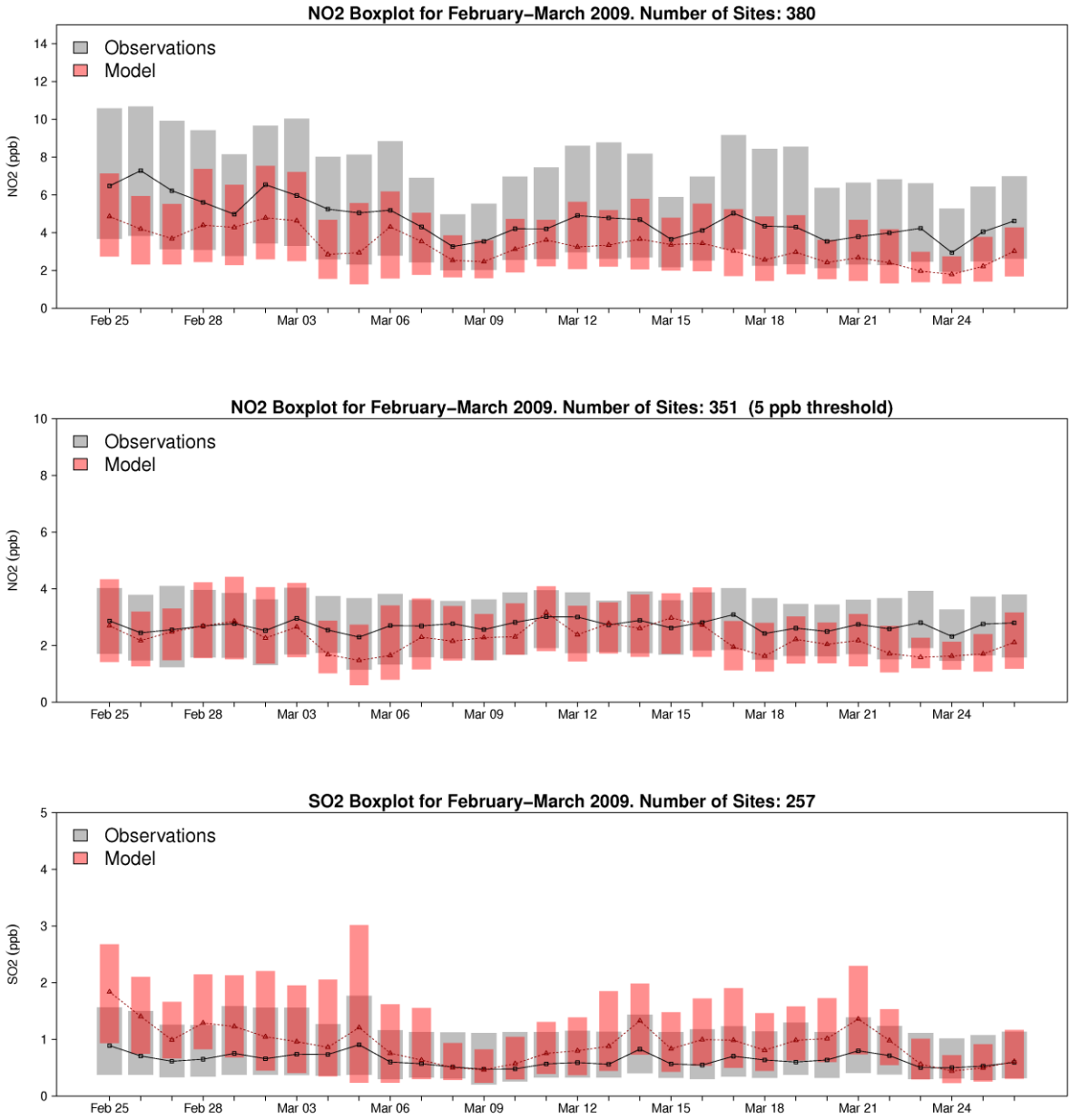


Figure S1. Comparison of modelled (red) and measured (grey) NO_2 and SO_2 concentrations at AirBase rural background sites and station not exceeding 5ppb of NO_2 concentration (central panel). The extent of the bars indicates the 25th and 75th percentile. The black and red lines are observed and modelled median, respectively. Based on base case (VBS_BC).

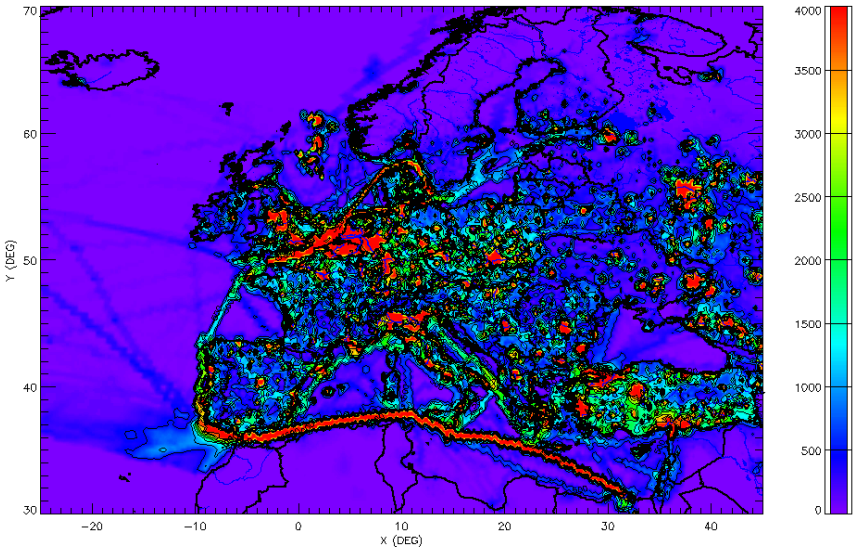


Figure S2. NO emissions in [mol/(h cell)] for 1 March 2009, at 6:00 AM

3. Another weakly presented component is the comparison with other studies. The TNO-MACC emission, EURODELTA, EUCAARI and Airbase archives are usual sources of information for numerous model exercises, not to mention MACC project itself, which covered the considered period with the ensemble of seven models and performed a detailed evaluation against the same Airbase. Numerical results and model scores are available. How does CAMx compare to these? In a couple of places, the authors mention conclusions of other studies but it has to be in a numerical form and made much more systematic.

It is true that comparison with other models was not emphasized strongly in the manuscript. This is because a very detailed evaluation of several models including CAMx for the same periods within EURODELTA III exercise is already available (Bessagnet et al., 2016). We agree, however with the referee's comment and added some more information in the revised manuscript.

Evaluation of the EURODELTA III model inter-comparison exercise showed that all models performed similarly for NO_2 in terms of correlation with values in the range 0.6-0.7, and the spatial correlation was much higher in the range 0.7-0.9 for all models (Bessagnet et al., 2016). There was a general underestimation in the afternoon. Additional analyses showed that NO_2 concentrations within the Paris area were well reproduced by CAMx showing a bias lower than 2 ppb, corresponding to less than 20% of the observed median concentration (Fig. 19 in Bessagnet et al., 2016). At the Po valley sites, on the other hand, NO_2 values were systematically underestimated. These analyses suggest that local emission sources and meteorological conditions such as reconstruction of the PBL diurnal cycle strongly influence NO_2 performance.

4. Among smaller things, I am missing the time correlation coefficient in the list of parameters. It is not only the absolute level that is to be verified, the expensive and complicated VBS mechanism is supposed to deliver better representation of the processes, thus improving the patterns and their evolution. The temporal correlation coefficient is arguably the best parameter to reflect it. Fractional error is good but less straightforward and intuitive parameter, also affected by bias.

We included correlation of determination for the three scenarios (NOVBS, VBS_ROB, VBS_BC) in Figure 5 caption excluding the elevated sites of Puy de Dome and Montseny. R^2 values are improved when the VBS approach is taken into account (VBS_ROB and VBS_BC) with respect to the non-volatile organic scheme (NOVBS). The values are reported in the caption of Figure 5:

Figure 5. OA daily average scatter plots for S1, S2 and S3 scenarios for February-March 2009 for stations in Table 3. Solid lines indicate the 1:1 line. Dotted lines are the 1:2 and 2:1 lines. Boxplots indicate medians, 5th, 25th, 75th and 95th quantiles for observations (black) and sensitivity tests (red). The crosses represent the arithmetic means. R^2 is 0.55 for NOVBS, 0.64 for VBS_ROB and 0.59 for VBS_BC.

5. The naming convention is confusing. The base case is usually number one, from which the sensitivity cases are made. It may look like a small thing but while reading I had to again and again remind myself that S3 is, in fact, the base case.

We changed the naming convention of the different scenarios throughout the manuscript as also suggested by Referee 1. The new nomenclature follows below:

S1: NOVBS (without VBS, two-product organic aerosol scheme based on Strader, 1999)
S2: VBS_ROB (VBS organic scheme with volatility distribution proposed by Robinson et al., 2007)
S3: VBS_BC (Base case, VBS organic scheme with volatility distribution proposed by Tsimpidi et al., 2010 and Shrivastava et al., 2011)
S4: VBS_BC_2xBVOC (Double biogenic emission based on VBS_BC)
S5: VBS_BC_2xBBOA (Double biomass burning emission based on VBS_BC)

6. Section 3 is the problematic one from the presentation standpoint. The text is not structured, subsections are routinely comprised of just one huge paragraph without much logic. I would strongly recommend heavy editing of this section.

We agree and revised Section 3.

Specific comments

1. The title does not reflect the paper content. This is the model evaluation exercise, not the AQ assessment.

We changed the title of the paper to read:

Evaluation of European air quality by CAMx including the volatility basis set scheme

2. p. 35647, l.15. I found it strange to praise the model for PM2.5 score, which, as shown already in the next lines, is a result of error compensation (l.20).

We changed the sentence to the following:

CAMx reproduced both total concentrations and monthly variations of measured PM2.5 for all the four periods with average biases ranging from −2.1 to 1.0 μgm^{-3}

3. p. 35648, l.1-3. No, it does not. The only piece shown is that the model appeared sensitive to scaling of the biogenic emission fluxes in one case and anthropogenic in another. The residential combustion is a hypothesis of the authors not directly supported by the study. It still sounds plausible and can be brought up in discussion but not in the abstract and not in the so categorical form.

Thanks for this remark. We will change the text p. 35647, l.15 of the paper to the following:

Further sensitivity tests with increased biogenic and anthropogenic emissions suggest that OA concentrations in Payerne were affected by changes in emissions from residential heating during the period of February–March 2009 whereas they were more sensitive to biogenic emissions in June 2006.

Moreover, the hypothesis of high contribution to OA from residential combustion in winter is supported by a recent modelling study performed in Europe by Denier van der Gon et al., 2015 for the same winter period (February-March 2009). In this study, an improved inventory with higher emissions from residential combustion by a factor of 2-3 compared to previously used inventories (EUCAARI) was tested in two CTMs with the VBS scheme (PMCAMx and EMEP). Authors concluded that the model performance for OA was improved when using the revisited wood burning emission inventory (Denier van der Gon et al., 2015. Figure8).

4. P. 35651, l.12-13. I did not understand: were the CAMx levels the same as the ones of IFS or not? If they were different, I would challenge the idea of neglecting the interpolation from the IFS levels. The issue should be clarified and explanations provided.

CAMx levels were the same as the ones in IFS. We modified the sentence as follows:

CAMx simulations used 33 terrain-following σ- levels up to about 8000 m a.g.l. as in the original IFS data

5. P. 35652- 35653. The emission discussion is unstructured and difficult to comprehend. Splitting the paragraphs to "main" species available from TNO-MACC, biogenics, etc, would help.

We agree and we split the paragraphs as suggested by the referee.

6. P.35652, l.14-17. How was the split made? As follows from the rest of the paper, amount of organic matter is one of the primary parameters of the study. This vague sentence is part of the most-important weakness of the paper mentioned above: the emission dataset is not analyzed and, as follows from this sentence, is not even presented properly.

Thanks for this remark. We added the following sentence to the manuscript:

$PM_{2.5}$ and PM_{10} emissions were provided by EMEP and they were split to Elemental Carbon and Organic Matter using the fractions given by IIASA (International Institute for Applied Systems Analysis) per each source and country.

7. P.35653, l.1-10. I did not understand: did the authors run MEGAN themselves, including preparation of the land use specifications, emission factors, etc? From the text it seems so ("were prepared for this study") but then, what was wrong in the native MEGAN setup? And how the

changes suggested in this study modified/improved its performance? Did the authors make this analysis?

We regret the misunderstanding concerning biogenic emission calculations. We used the MEGAN model without any change of the model itself. The data needed to run the model (emission factors of relevant vegetation for our model domain) were retrieved from the MEGAN website (http://lar.wsu.edu/megan/), from MODIS satellite data (leaf area index, LAI) and adapted to the model domain resolution. The same meteorological variables as for CAMx simulations were used to run MEGAN.

8. P. 35654, l.4. "Further aging" from what stage? And why was the ageing stopped? Just because then the model over-estimates the SOA, as stated in the paper? But this cannot be the reason, it is artificial and model-dependent. Is there any physical/chemical ground or hypothesis?

Further reactions of gas-phase products from the first oxidation of biogenic and biomass burning precursors were not considered in the original version of the CAMx-VBS model because of some studies showing over-prediction of SOA especially in rural-areas due to ageing of biogenic SOA (Lane et al., 2008 and Murphy and Pandis, 2009). In this study our aim is to validate the original CAMx-VBS model and compare the results with previous studies performed in North America sharing the same setup (Koo et al., 2014). We will include and investigate the effect of ageing of BSOA in a follow-up study.

9. P. 35655, Statistical methods. These formulas are from textbook. One can put them to appendix for the sake of completeness but this sub-section definitely should be eliminated from the main paper.

We agree and we eliminated the formulas from the main text.

10. P. 35656- 35658. Almost two pages of plain unstructured text, all in one (!) paragraph. I tried several times and still had problems in pushing myself through it.

Paragraph 3.1 was extensively revisited by splitting it according to the different investigated species and shortened as also suggested by Referee 1.

11. P. 35656- 35658. It also looks like the authors do not really pay attention to the physical and statistical meaning of the metrics used. As said in the general comments, mean error is heavily controlled by bias when the latter is large. An independent quantity would be correlation coefficient.

We added correlation coefficients in Table 2 and revised the paragraph, as also suggested by Referee 1.

12. P. 35659, l.13. Another praising the model for meeting totals by a mere error compensation. Not sure if this is a big achievement.

We could not find the above comment in the text. We will change the text p. 35659, l.16 to l.18 of the paper to the following:

The modelled average total non-refractory PM2.5 (sum of nitrate, sulfate, ammonium and OA) concentrations match the measurements with a few exceptions (Fig. 5 and Table 3).

13. P. 35659 – 35661. . . and another 2.5 pages in a single-paragraph of unstructured text.

We modified and shorten the Section 3.2 as also suggested by Referee 1.

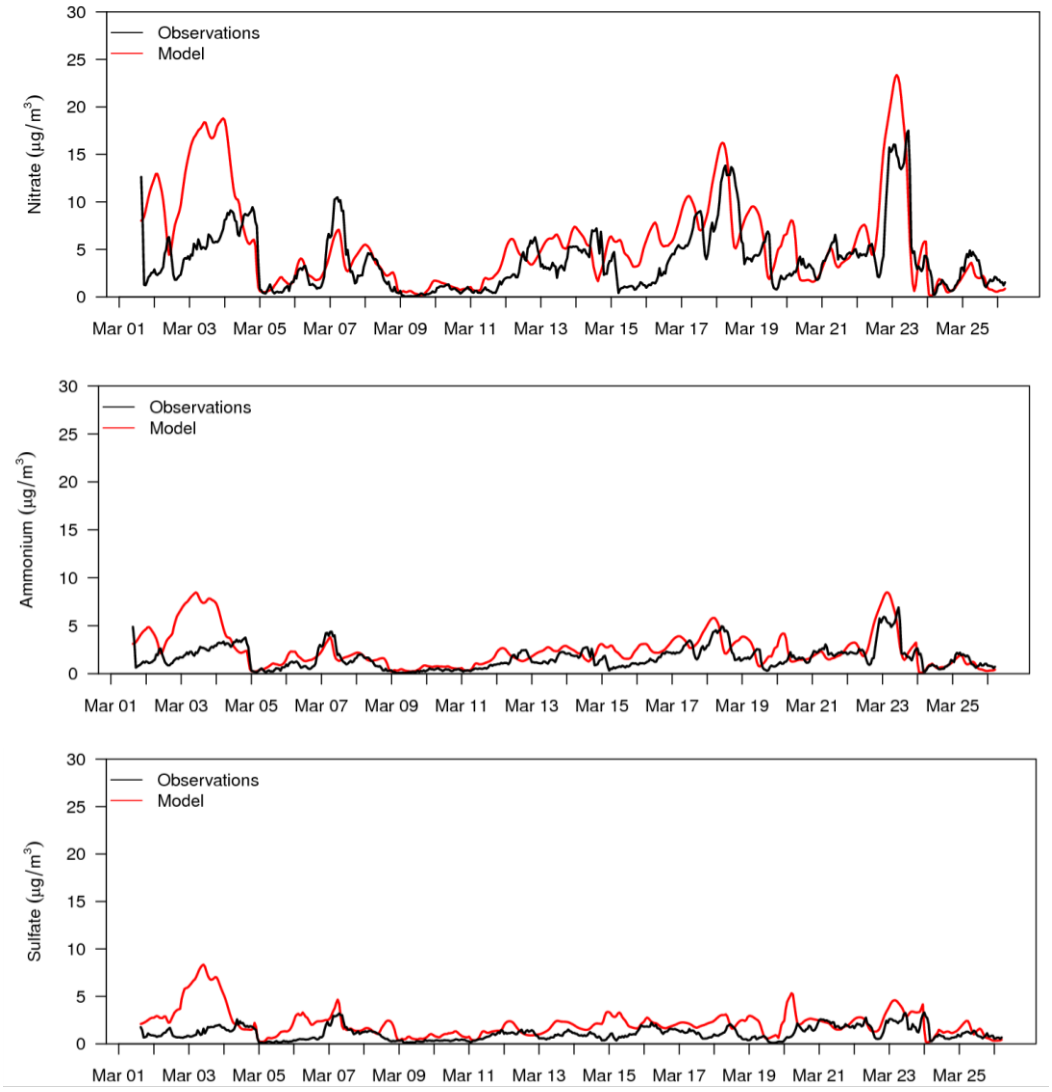
14. P. 35665, l.20. . . .and again "total PM2.5 was modelled very well", for a change without a reference to error compensation. I have strong difficulties with such presentation style.

We revised that part in the Conclusions by merging points 1 and 3 together as shown below:

Although total $PM_{2.5}$ mass concentrations and its variations were well reproduced by the model in all four periods, comparisons with AMS measurements for the February–March 2009 period revealed that the good agreement between model and measurements was most of the time due to overestimation of the inorganic fraction, especially NO_3^-, and underestimation of OA. Sensitivity tests with reduced NH_3 emissions generally reduced the positive bias in NO_3^- suggesting potential uncertainties in NH_3 emissions and their seasonal variability.

15. Figure 6: what panels are for what parameter? The axis font is much too small to figure it out.

We increased the fonts.



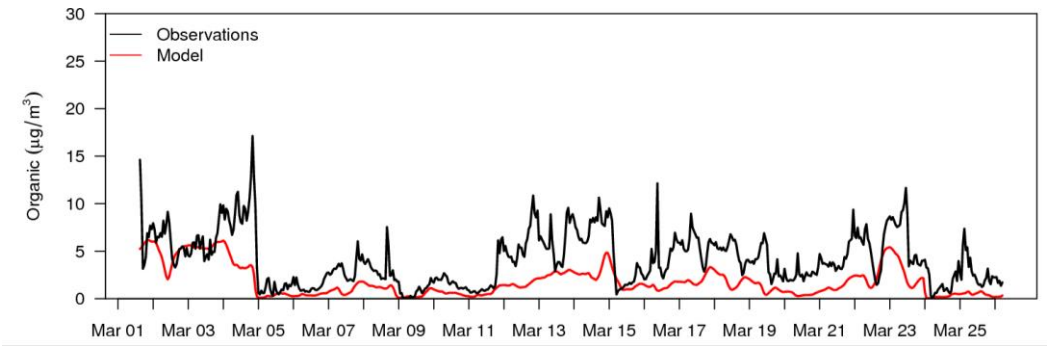


Figure 4. Comparison of observed and modelled nitrate, ammonium, sulfate and organic aerosol at Payerne for March 2009.

References

Bergström, R., Denier van der Gon, H. A. C., Prévôt, A. S. H., Yttri, K. E., and Simpson, D.: Modelling of organic aerosols over Europe (2002–2007) using a volatility basis set (VBS) framework: application of different assumptions regarding the formation of secondary organic aerosol, Atmos. Chem. Phys., 12, 8499–8527, doi:10.5194/acp-12-8499-2012, 2012.

Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M.-T., Colette, A., Couvidat, F., Meleux, F., Rouïl, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isodoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J.-C., Fagerli, H., Gonzalez, L., Menut, L., Prévôt, A. S. H., Roberts, P., and White, L.: Presentation of the EURODELTA III inter-comparison exercise – Evaluation of the chemistry transport models performance on criteria pollutants and joint analysis with meteorology, Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2015-736, in review, 2016.

Koo, B., Knipping, E., Yarwood, G.: 1.5-Dimensional volatility basis set approach for modeling organic aerosol in CAMx and CMAQ, Atmos Environ., 95: 158-164., 2014.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10963-10976, doi:10.5194/acp-14-10963-2014, 2014.

Lane, T. E., Donahue, N. M., and Pandis, S. N.: Simulating secondary organic aerosol formation using the volatility basis-set approach in a chemical transport model, Atmos. Environ., 42, 7439–7451, doi:10.1016/j.atmosenv.2008.06.026, 2008.

Murphy, B. N. and Pandis, S. N.: Simulating the formation of semivolatile primary and secondary organic aerosol in a regional chemical transport model, Environ. Sci. Technol., 43, 4722– 4728, 2009.

Vaughan, A.R., Lee, J.D., Misztal, P.K., Metzger, S., Shaw, M.D., Lewis, A.C., Purvis, R.M., Carslaw, D.C., Goldstein, A.H., Hewitt, C.N., Davison, B., Beevers, S.D., Karl, T.G., 2016. Spatially resolved flux measurements of NOx from London suggest significantly higher emissions than predicted by inventories. Faraday Discuss. doi:10.1039/C5FD00170F

Villena, G., Bejan, I., Kurtenbach, R., Wiesen, P., and Kleffmann, J.: Interferences of commercial NO2 instruments in the urban atmosphere and in a smog chamber, Atmos. Meas. Tech., 5, 149-159, doi:10.5194/amt-5-149-2012, 2012.