Review comments to manuscript gmd-2019-70:

Trends of inorganic and organic aerosols and precursor gases in Europe: insights from the EURODELTA multi-model experiment over the 1990–2010 period

by Ciarelli et al.

General Comments

The manuscript investigates trends in aerosol concentrations and chemical composition as well as their precursors in Europe over two decades that have experienced significant changes in emissions. The analysis is based on output of a suite of air quality models and observational data. The paper is well written and figures and tables are clear and informative. The paper contributes to the field of models evaluation and sensitivity analysis to different parameterizations and settings and will be of interest to GMD readers. However the following major points and specific comments could be addressed to improve it.

- The EURODELTA experiment includes output from eight air quality models of which five are included in the current manuscript. Although the additional three models do not allow a proper quantification of the trends they could be included as a model intercomparison exercise. Those models may indeed adopt more sophisticated (or computationally demanding) settings whose skills need to be quantified to better inform model users on the optimal model setup (see more details comments below). I would suggest including a new model intercomparison and evaluation section and present multiple statistical metrics of model skill for the years when all eight models have data.
- Despite the paper focuses on the quantification of temporal trends, it is important to quantify the actual model skills vs observations. This could be useful also in the context of informing users on the optimal model setup to be adopted when aiming to simulate aerosol concentration and composition in future studies. It would be also useful to add results of these analyses to the conclusions, where the model/settings with highest skills are highlighted.
- The trend analysis are performed for different aerosol species and over different regions. However there is no clear quantification of the spatial variability of these trends and how the model performance varies by year. It could be interesting to expand the current analyses and include a more detailed investigation on changes of model performance over different years and different sites. This will also contribute to a better quantification of inter-model variability and identification of sources of uncertainty in model output.

Specific Comments

- Despite there is no word limit for GMD abstracts, the current one is very long, so I would suggest to present the results in a more concise way (e.g. by grouping models/species having similar performance/behavior, etc.).
- Page 2, line 7: I would use “evaluate” instead of “validate” here and in other instances in the paper since you are performing a model evaluation (not validation).

- Page 2, line 13: the meaning of relative trend or relative reduction (here) was never defined in the paper. It should be clarified here and/or in the methods since this terminology is used in many instances in the text ad figures.
- Page 4, line 16: the years of the AeroCom experiment could be included
- Page 5, line 24: what is the spatial resolution of the EDT models? Is it the same for all of them? This information should be included.
- Page 7, line 36: LAI from MODIS is available only from 2002. What did the EDT simulations use for prior years?
- Page 8, line 15: it would be interesting to have some more details about the temporal variability of emissions since they are one of the key drivers of changes in aerosol properties. This would also support your statement that trends in emissions are non-linear and would be helpful to interpret the results at finer temporal scales. Some information about the how trends in emissions change in space would be also relevant (you could also refer to literature studies if available).
- Page 9, line 27: this claim could be supported by a supplementary figure or referring to the literature, as suggested in a prior comment.
- Page 10, line 5-16: these claims are not supported by your analyses. Could you show some more analyses and explain how the trends vary in space (or refer to the literature)?
- Page 11, line 20: Period is missing at the end of the sentence. Also it is not clear the number of stations, data availability and frequency of the data used for the analysis.
- Page 12, line 6: is there a way for you to indicate which dots these sites correspond to in Figure 7?
- Page 13, line 6: do you have observations to be used for model evaluation in Figure 9/10?
- Page 13, line 24: this claim could be better supported by referring to the spatial patterns of emissions, as suggested to do in previous comments.

Tables

Table 7: there is no mention of Table 7 in the manuscript but I would keep it, so please add reference to it and comments. Also the statistical metrics should be defined. What is the temporal resolution/frequency of the data used in this evaluation and how many observations do you have at each site?

Figures

Figure 5 and 8: not clear how many observational sites are included and the temporal resolution of the data used.

Figure 14: Rephrase caption to clarify that panel a includes all seasons as in Table S2, while summer and winter campaigns are shown in panels b and c.

Figure S1: why averages above 7 μg m^{-3} are excluded? Explain this also in Table S1.

Technical corrections:

- Page 2, line 34: be consistent between “emission” and “emissions”.
- Page 3, line 10: remove “the formation of”.

- Page 4, line 29: SIA was not defined earlier.
- Page 8, line 35: typo “Weather Research and Forecasting”.
- Page 9, line 11: typo “illustrates”.
- Page 12, line 24: replace “model” with “models”.
- Page 17, line 5: add also spatial resolution (if it is the same).
- Page 17, line 21: SOx missing subscript.
- Figure 2: Missing subscripts on titles for chemical species.