Response to Anonymous Referee #2 for "EURODELTA-Trends, a multi-model experiment of air quality hindcast in Europe over 1990–2010" submitted to GMDD by Colette et al. 2016 as gmd-2016-309

Note: Referee comments are indicated in bold, answers are in regular font and changes highlighted in yellow in the revised manuscript

General comment:

The paper describes a multi-model experiment including 12 one year long air quality simulations for Europe and two simulations with a length of twenty years. The volume of work presented in this model experiment description paper is really impressive and the paper should be published after some revision as a basis for the hopefully following numerous papers going more into the details.

In its currents form the paper starts quite nicely but it is weakening toward the end. In particular, it is disappointing that not even a single result is presented. At least some overview results have to be included (which is also required according to the requirements for model experiment description papers in GMD), although a more in depth discussion must of course be left to more specific papers.

> In excluding model results from this experiment description paper, we followed the example of several recent articles of this category published in GMD for the Climate Model Intercomparison Projects (CMIPs). After further discussion with the topical editor, we understand that those were exceptional, and since the Eurodelta-Trend exercise is mature enough to include such sample results in the experiment description stage, we are pleased to address the referee's concern (also shared by referee #1) by adding a new section "Sample results". As explained in that section, we present some quickviews of the model ensemble. The evaluation of the ensemble and its ability to capture air pollution trends is one of the stated objectives of the experiment, it requires however a substantial analysis which will be the focus of forthcoming papers.

Specific comments:

According to the (too brief) section 4 WRF-Chem is run on a 0.25 x 0.4 lat-long grid. Was a rotated grid used for the WRF-Chem runs?

Line 34 on page 6 and line 1 on page 7 indicate that WRF-Chem was run for a Lambert conformal grid at 25 km resolution ('A similar strategy [as for the meteorological driver for CMAQ] was used for WRF-Chem'). This is contradictory to the statement above.

> WRF-Chem used the same lat-lon grid as all of the other models except CMAQB (the original text in Section 4 was correct). A revised explanation regarding the how meteorology was driven in the WRF-Chem runs has been added to Section 5:

"For WRF-Chem, an online model that simulates meteorology and chemistry simultaneously ("online"), the meteorology from the WRF-Eurocordex runs (Stegehuis et al 2015) was used as initial and lateral boundary conditions and for applying four-dimensional data assimilation (FDDA), with coefficients as described in Mar et al. 2016."

According to Tab. 3 the horizontal grid width is approximately 25 km whereas the grid width of HIRLAM is approximately 22 km. However, in section 4 it is mentioned that all simulations were performed for the same lat-lon projection with 0.25 x 0.4, with CMAQ being the only exception. Does this mean that MATCH has a different grid width than its meteorological driver? Please explain in more detail.

We have added such a clarification. We now explain that the original EURO4M meteorology was interpolated from the original 0.2 degree horizontal resolution on rotated lat-lon grid (corresponding to ca 22km resolution) to the grid of the CTM simulations.

Including an additional line with the name and the applied resolution of the meteorological driver in Table S1 might be helpful.

We have added this information in Table S1

Tab. 3 implies that the boundary conditions for the WRF-Chem simulation were derived from WRF simulations with 0.4x 0.4 grid width. Is this true, or does 'WRF-0.44 simulation used by other EDT models' just mean that the WRF-Chem run and WRF-0.44 use the same ERA-Interim data for deriving the meteorological boundary conditions? If this was the case: Was meteorology nudging applied during the WRF-Chem run?

The first option is true, WRF-0.44 data used by other models were used to force WRF-Chem at the boundaries, and for applying four-dimensional data assimilation (FDDA), with coefficients as described in Mar et al. 2016.

Was nudging applied for RACMO and HIRLAM? This information could be added to Table 3.

The HIRLAM reanalysis (EURO4M) uses data assimilation in 3 dimensions in the upper air (as explained in Table 3) and optimal interpolation for the surface fields. An initial analysis is conducted every 6 hours and three hourly forecast step saved and used by MATCH. We have added a clarification on this in the text. Thus, nudging was not used in producing the EURO4M data set.

The RACMO simulations are part of the EuroCordex ensemble documented in Jacob et al., (2013) and Kotlarski et al., (2014). We clarified in the text that it excludes nudging.

Page 9, lines 21-22: Please add some more details and also remarks concerning the quality of these data.

The following text has been added at the end of Section 7.1:

"The description of EMEP parameterization for sea spray and windblown dust can be found in Simpson et al. (2012). The accuracy of the model results for sea salt and mineral dust is regularly evaluated with available observations over Europe and documented in EMEP reports (www.emep.int). Model evaluation for mineral dust is limited due to the scarcity of dust in-situ measurements (see EMEP Status Report 1/2014), therefore also AOD/extinction measurements from satellite, Aeronet and Earlinet has recently been used for model evaluation within AeroCom aerocom.met.no)."

Section 8.2: Please add some more details here (resolution, etc.)

The following text has been added:

"The model uses a full tropospheric and stratospheric chemistry scheme (Lamarque et al., 2012) based on MOZART (Model for Ozone and Related chemical Tracers) version 4 (Emmons et al., 2010). CAM4-chem considers 56 vertical levels from the surface to about 40 km with 1.9° x 2.5° horizontal resolution. The simulation used in this analysis was performed in nudging the model to meteorological fields from the MERRA GEOS-5 (Modern Era Retrospective Analysis for Research and Application Goddard Earth Observing System Data Assimilation System Version 5) reanalysis provided by the Global Modelling and Assimilation Office (GMAO)."

Section 9: The 'Addition al diagnostics' part looks just like copy paste from the modelling protocol. A table might be more useful. Eventually, this type of list might be moved to an appendix. Finally, although the reader can guess the meaning of all abbreviations e.g. O3_HL, it should be explained.

Following the suggestion of the referee, we moved this section to a Table, also improving the language to avoid the style of modelling protocol as much as possible. All abbreviation should now be defined.

A few overview results including all contributing models (e.g. Taylor diagrams, box plots, for examples see Solazzo et al., 2012 (doi:10.1016/j.atmosenv.2012.02.045) or Im et al., 2015, (http://dx.doi.org/10.1016/j.atmosenv.2014.09.042) or a table) should be added.

See the answer to the general comment. We have added quickviews of the model ensemble but left out of the present article any comparison with observations. In following this approach, we opted for a similar strategy as the AQMEII model experiment description recently published as an ACP Tech. Note: http://www.atmos-chem-phys.net/17/1543/2017/

In their current form, the conclusions are more like a summary and outlook. But even the current outlook needs to be enhanced. Is there any concept concerning the further analysis of the results and future papers? Which detailed analysis is under work by the members of the consortium?

The ongoing analysis work is by nature an evolving object and it is difficult to write in an article to be published analysis plans that are still changing, but the main topics are, as stated in the conclusion:

- " … assess the capability of these state-of-the-art chemistry-transport models to reproduce the observed changes in the concentrations of the main pollutants …"
- " … assessment of the capability in reproducing the actual trends over the 21yr in the 1990-2010 period …"
- " … attribution of air quality trends to emission changes, to influx at the boundaries of the European domain, and to interannual meteorological variability … "
- "… serve for in depth analyses to scientific communities working on the impacts of air pollution on health, ecosystems or aerosol radiative forcing…"

Minor Points

Further previous multi-model studies should also be mentioned in the introduction (at least multi-model studies with a minimum length of one year of simulation).

We added a few references to the AQMEII project, in addition to the one-year multi-model publications from earlier phases of Eurodelta that were already cited. The following text has been added in the introduction:

"Over the recent past, several multi-model projects covering a time period of one year or less were undertaken such as the earlier phases of Eurodelta cited above but also the various phases of the AQMEII project (Galmarini et al., 2012;Galmarini et al., 2017;Rao et al., 2011;Im et al., 2015)."

Page 5 line 9: Do the authors mean chemistry boundary conditions here?

Yes, it has been corrected

Page 5: Is the WRF version (v3.3.1 according to Tab. 3) the same as described in Stegehius et al, 2015.?

Yes, the version number has been added in Section 5

Page 7: Section 6 is a bit meager. It should be either enhanced or incorporated into a section, which dedicated to all types of emissions.

We followed this suggestion by moving that part to a single "emission" section

Figure 2: The red dots as well as the blue dots are not well resolved in the figure, i.e most of the blue dots look more like a line.

The resolution of the figure has been improved

Other updates

The table of contribution of each modelling group to the various tiers has been updated to include new deliveries since the date of submission of the first manuscript. A problem has also been uncovered in WRF-Chem simulations, so that only an update of tier 1 is available for the updated contribution, while other tiers are now indicated as “planned”.