Interactive comment on "Presentation of theEURODELTA III inter-comparison exercise – Evaluation of the chemistry transport models performance on criteria pollutants and joint analysis with meteorology" by B. Bessagnet et al.

Anonymous Referee #4

Received and published: 16 June 2016

We thank a lot the reviewer for this review of our paper. The answers are written here below in bold characters after each comment.

General comments

This manuscript is a thorough description of an international model inter-comparison exercise. It can in my view be published in ACP, provided that the comments and concerns below will be taken into account.

The article contains interesting and useful results. However, in my view the discussion of results should focus much more on the results that have some general interest, and on the more general insights and conclusions, and the amount of small details should be substantially reduced. By small details I mean e.g. discussion on how each individual model has performed for each pollutant and each campaign. The amount of figures and tables is also very large; I would advise the authors to reduce these.

In the revised manuscript we have separated the discussion and the conclusion. The conclusion section is then more general and shorter. We understand the concern of the reviewer on the number of figures, however, we did an important effort to keep the most essential figures, we prefer to keep all of the current figures in the manuscript that bring a lot in the discussion, except Fig. 1 we have removed.

However, the figures that that make it possible to draw general conclusions should be included. I suggest that the authors would add to conclusions a discussion on the most important improvements of the models, and areas of improvement for the CTM's in general in the future, based on their findings. The terminology also should be more precise, and some of the conclusion more cautious, taking into account the limitations of the data; details are discussed below.

We have modified the last section of the conclusions as follows : « *The study stresses the importance of emission sources particularly in wintertime, wood burning emissions are likely the most underestimated source, through the missing species called semi-volatile organic compounds. Road traffic emissions could also be underestimated, gasoline and diesel vehicles are both concerned, and more generally all activity sectors involving combustion processes can be concerned. In this study, the importance of meteorological data is highlighted, the difficulties for meteorological models to simulate meteorological variables like wind speed and PBL height during stable conditions can lead to dramatic consequences on air quality modelling. Developments in air quality modelling have not only to focus on processes but also on emissions and meteorological*

input data. To complement the analysis, companion papers will focus on depositions of sulphur/nitrogen compounds and on the behaviour of models for particulate matter species. This ensemble of analyses will help to prioritize the improvement of air quality models used in the frame of the CLRTAP»

Specific comments

Abstract. Explain which experimental datasets were used, and how many stations were included, please. 'Background stations', specify which background; probably regional background, not urban or global background. The discussion would be in my view more clear, if the evaluation of met parameters would be presented first, then evaluation of concentrations. 'performances were good', specify what is meant with 'performance', do you mean e.g. bias or correlations, or both ? PM, specify which PM fraction.

Here is the new abstract :

" The EURODELTA III exercise has facilitated a comprehensive inter-comparison and evaluation of chemistry transport model performance. Participating models performed calculations for four one-month periods in different seasons in the years 2006 to 2009, allowing the influence of different meteorological conditions on model performances to be evaluated. The exercise was performed with strict requirements for the input data, with few exceptions. As a consequence, most of differences in the outputs will be attributed to the differences in model formulations of chemical and physical processes. The models were evaluated mainly for background rural stations in Europe. The performance was assessed in terms of bias, root mean square error and correlation with respect to the concentrations of air pollutants (NO_2, O_3, SO_2, PM10 and PM2.5), as well as key meteorological variables. Though most of meteorological parameters were prescribed, some variables like the planetary boundary layer (PBL) height and the vertical diffusion coefficient were derived in the model pre-processors and can partly explain the spread in model results. In general the day time PBL height is underestimated by all models. The largest variability of predicted PBL is observed over the ocean and seas. For ozone, this study shows the importance of proper boundary conditions for accurate model calculations and then on the regime of the gas and particle chemistry. The models show similar and quite good performance for nitrogen dioxide, whereas they struggle to accurately reproduce measured sulphur dioxide concentrations (for which the agreement with observations is the poorest). In general, the models provide a close-to-observations map of particulate matter (PM2.5 and PM10) concentrations over Europe with rather correlations in the range 0.4 – 0.7 and a systematic underestimation reaching -10 µg m^{-3} for PM10. The highest concentrations are much more underestimated particularly in wintertime. Further evaluation of the mean diurnal cycles of PM reveals a general model tendency to overestimate the effect of the PBL height rise on PM levels in the morning, while the intensity of afternoon chemistry leading to formation of secondary species to be underestimated. This results in larger modelled PM diurnal variations than the observations show and this is so for all seasons. The models tend to be too sensitive to the daily variation of the PBL. All in all, in most cases model performances are more influenced by the model set-up than the season. The good representation of temporal evolution of wind speed is most responsible for models' skillfulness in reproducing the daily variability of pollutant concentrations (e.g. the development of peak episodes), while the reconstruction of the PBL diurnal cycle seems to play a larger role in driving the corresponding

pollutant diurnal cycle and hence determine the presence of systematic positive and negative biases detectable on daily basis."

Introduction. In discussing model inter-comparisons, refer also to the most recent relevant ones, especially Prank et al, 2016, ACP (16, 6041–6070). ". . . showed better performance but higher uncertainty. . .' define what is meant with 'performance' and what you mean with 'uncertainty'. The institutes participating. . . this sentence should be deleted; not scientifically relevant information. 'criteria pollutants': define concept (which criteria ? defined by whom ?); probably the authors refer to the latest EU directives or limit values (?); but that should then be specified.

Yes, we have included the references and taken into account all the remarks. Criteria pollutants refer to the Air quality directives, we have modified accordingly. Uncertainties are related to the model formulation (parameterization) and input data. We modified it as :" ***The objective of this paper is twofold, (i) to present the exercise, the input data and the participating models, and (ii) to analyse the behaviour of models in the four campaigns focussing on the criteria pollutants PM10, PM2.5, O3, NO2 and SO2 as defined in the EU directive on air quality 2008/50/EC (EC, 2008), and relevant meteorological variables."***

Methods. p 8 'lowest levels of emissions': emissions of which pollutant ?

They refer to $PM_{2.5}$ for the residential sector, we clarified it in the revised version.

Discussion. p 23: 'model formulation and setup . . . more influencing than met conditions'.

Define what is meant with 'model formulation and set-up' (is it the setup of input data, which ones ? set-up of model parameters and submodels, which ones ?;or selection of CTM's themselves ?). This statement is also over-interpretation; it has only been shown to be valid for the range of met parameters that were included in the selected conditions, which was not especially wide. Please re-write this, allowing for the limitations of the data used.

We replaced this statement by « ***This confirms once again that on average and for the limited dataset used in this exercise, the model formulation (parameterization of chemical / physical processes, calculation of meteorological diagnosed variables) and set-up (number of vertical levels, value of key parameters, etc...) are more influencing than raw meteorological conditions on model performance.*** ».

p. 23. 'highest errors': which stat. model evaluation parameter is meant by 'error' ?

We have replaced by RMSE instead of error.

Technical corrections

I would also suggest that the whole text and the language will be checked, and the fairly numerous misprints and language mistakes will be corrected.

Yes, we have revised the language; there were several misprints and mistakes.