Referee report regarding the manuscript:

Presentation of the EURODELTA III inter-comparison exercise - Evaluation of the chemistry transport models performance on criteria pollutants and joint analysis with meteorology

Authors: B. Bessagnet et al.

General comments

In my opinion this paper is not suitable for ACP — it is essentially a technical report from the EURODELTA project, and I think it should be published as such; a (rather confusingly presented) model intercomparison, for a limited set of “standard” atmospheric components, is not interesting enough for publication in ACP. Many similar model intercomparisons have been published and, as written, this one does not contribute anything new.

The paper only documents model results for seven different chemical transport models without enough detail to be able to draw any useful conclusions for the general scientific community. I can not see what this paper contributes to the understanding about the atmospheric chemistry or physics, or any new information that aids in improving modelling of the atmospheric composition.

I also think that the authors are fragmenting their research (over several papers). This should be avoided (see the ACP Obligations for authors). In order to better understand model performance you need to take into account all important processes — including deposition and chemistry. If the authors want to publish this material in a scientific journal I think it has to be combined with the information about deposition and chemical composition of particulate matter. Splitting the model evaluation into three different papers is not appropriate.

The paper could have been acceptable for Geoscientific Model Development (GMD, which accepts model evaluation papers) — if the presentation had been better — but I think that a much more scientific approach is needed to make the material presented in this manuscript interesting enough for publication in a peer-reviewed journal.

I am sure that there are a number of interesting scientific questions that the EURODELTA project can answer, and I suggest that the authors focus mainly on that, and keep this kind of model intercomparison/evaluation documentation to technical reports. This paper in itself has little significance for the ACP community.

The paper has a very long author list with 36 authors! However, the brief statement on page 3, about what seven of the participating institutes (and I think NILU is missing in this list) have contributed to the project, is not motivation enough for the inclusion of so many authors. Considering the very long author list, please give a brief explanation of what each individual has contributed to this paper in the reply to this referee comment (the statement of contributions can be added to the supplement of the paper). Please note Point 9 under the General Obligations for Authors for ACP (my highlighting):

"To protect the integrity of authorship, **only persons who have significantly contributed to the research and paper preparation should be listed as authors**. The corresponding author attests to the fact that any others named as authors have seen the final version of the paper and have agreed to its submission for publication." … "The author who submits a manuscript for publication accepts the responsibility of having included as co-authors all persons that are appropriate and none that are inappropriate."

Specific comments

Page 3, lines 34–35: "As a consequence, there were very limited differences in the models set up, representing a sort of sensitivity analysis to several aspects of the modelling chains."

- I do not understand what you mean by this sentence! What do you mean by "a sort of sensitivity analysis" and in what sense is this study a sensitivity analysis? I suggest that this sentence is removed.

Page 4, lines 3–4: "Complementary analyses of depositions fluxes and PM composition data at high temporal resolution will be discussed in companion papers in order to better understand the behaviour of models."

- In my opinion this is fragmentation of research papers, and the consequence is that the present paper becomes uninteresting. As mentioned in the General Comments, I do not think that splitting the information this way in three different papers is useful. If knowledge about the deposition fluxes and PM composition data are important for understanding the behaviour of the models (which I certainly expect them to be) this information is needed in the present paper!

Page 4, lines 25–26: "In CMAQ additional anthropogenic dust is calculated as 90% of unspecified PM coarse emissions and attributed to fugitive dust"

- What is the motivation for adding extra anthropogenic dust? Was this just a modelling mistake?

Or do you have good reasons to believe that the emission inventory used in the present study lacks a substantial amount of fugitive dust? And if this is the case, why did you not increase the emissions in all models?

Page 4, line 27: "CAMx did not activate the sea salts parameterisation in this exercise."

- Why not? Was this a modelling mistake? Or are there problems with the sea salt emissions in CAMx?

Page 4, lines 30–31: Why was CAMx not included in the ENSEMBLE for O_3, NO_2, and SO_2? I would guess that the lack of sea salt would hardly have any impact on these three gaseous species.

Page 8 — Emissions: Are the emissions used in the EURODELTA project available for use by scientists outside the project? If they are, please specify this and where they can be found. If they are not available, more details are needed regarding the emissions in order for others to be able to evaluate/compare this work to other studies using other emission data. Without more detailed information the work presented in this manuscript can not be considered reproducible.

Page 8, lines 16–22: "EMEP national emissions were kept except for..." 14 countries, for which GAINS emissions were used.

- This seems a bit strange - why did you change emissions for these 14 countries and not for the other countries? Please give a motivation.

"Additional factors were applied on two Polish regions (x4 or x8) for $PM_{2.5}$ and PM_{10} emissions"

- For which Polish regions? They need to be specified in detail to make this work reproducible. It is also unclear if the x4 factor applies to $PM_{2.5}$ and x8 to PM_{10} or if the same factors were used for $PM_{2.5}$ and PM_{10}?

Page 8, line 27: What do you mean by "artificial area"?

Page 8, lines 28 and 31: EPER data are only available for the EU-countries + Norway — how did you treat industrial emissions in the other countries?

Page 8, line 31: What is "artificial landuse"?

Page 8, lines 36–38: Considering the great uncertainties in the residential combustion emissions I suggest that you give some more details about the emissions you have used in EURODELTA. The statement that "Germany, Sweden, and Spain clearly have the lowest (levels of) emissions" is not clear enough. Do you mean the lowest emission per capita? Or per square km? In order for the results from the EURODELTA modelling to be comparable to other studies I suggest that you add a table to the Supplement specifying annual total national residential combustion emissions assumed in the EURODELTA inventory.

Page 9, line 3: What are "the usual default profiles"?

Page 9, lines 4–6: "a PM speciation profile provided by IIASA (Personal Communication from IIASA) was used to estimate the fraction of Non-carbonaceous species, Elemental Carbon and Organic Matter per activity sectors and country"

- This PM speciation profile must be provided with the article. Personal communication with an organisation (IIASA) is not a reference that makes it possible for readers to find

the relevant information to be able to reproduce the work. A table specifying the three $PM_{2.5}$ and coarse PM fractions for each emission sector and country should be added to the Supplement of the paper.

Page 9, lines 22–23: There is no reference to a description of the SMOKE system.

Page 10, lines 10–15, Wildfire emissions: Which emitted species were included for wildfires? What gases and which particulate species were included (include information about how the PM-emissions were split between organics, BC, and other PM-components)?

Page 10, lines 22–23: Why were the agricultural and road dust PM sources not activated in the LOTOS-EUROS model?

Page 15, lines 4–6, Regarding the PBL and the LOTOS-EUROS and EMEP models: "LOTOS-EUROS and EMEP that should adopt IFS PBL too, show partially different performance, suggesting that the latter models partially recomputed boundary layer height."

- This is too unclear! You have to be able to describe how these models handle the PBL! In what way do they "partially recompute" the BL height?

Page 16, lines 5–6: "The large positive bias in 2007 and negative in 2009 are largely explained by the boundary conditions that are biased respectively of +8 and -20 μg m^{-3}"

- I agree about the negative bias in the 2009 campaign but the bias of +8 μg m^{-3} in the 2007 campaign can hardly be considered to "largely explain" the very large positive bias (21–23 μg m^{-3}) for CAMx, CMAQ and Chimere — I guess there must be other factors that are more important than the boundary conditions to explain the poor performance of these three models?

Page 16, lines 7–8: "For the summertime campaign 2006 CHIMERE and CMAQ display the lowest correlation for daily averaged concentrations"

- Can you explain the very poor correlation for Chimere and CMAQ for this summer period?

Page 16, lines 11–13: "All models simulate high ozone concentrations over the Mediterranean sea, most of them behaves satisfactorily in Malta and Cyprus stations confirming the ozone concentrations pattern over the seas for the "ensemble" shown in Fig. 6."

- What do you mean by "confirming the ozone pattern over the seas"? Do you mean that a "satisfactory" behaviour at two sites in the Mediterranean region proves that the model ensemble gives good ozone concentrations over all sea areas? Also, in Fig. 6 I see no observation data from Cyprus so for this summer period it is really only one site you base your statement on?

Page 16, lines 20–21: "This result confirms that during stable conditions the pollutant concentration is influenced not only by the PBL height, but also by the overall reconstruction of vertical dispersion."

- What do you mean by “the overall reconstruction of vertical dispersion”? And could the differences of the results not also be due to differences in dry deposition and chemistry?

Page 16, lines 26–28: “Not only the bias is affected by global boundary conditions, but also this result indicates that biased ozone boundary conditions globally impair the normalized statistics confirming the non linearity of ozone chemistry.”

- This sentence hardly makes any sense at all to me. I think it is unclear what you mean and it seems like just speculation to me.
- What do you mean by “globally impairing normalized statistics” and how does this “confirm the non linearity of ozone chemistry”?
- As mentioned above I do not think that you have shown that the global boundary conditions is the main reason for the model problems for the 2007 campaign! Of the four ENSEMBLE models Chimere performs very poorly for 2007 (or at least very differently than the other three models) and this can not be explained by the global boundary conditions.

Page 17, lines 8–9: “This underestimation of NO_2 concentrations is certainly related to rather high ozone concentrations.”

- Can you explain why CAMx behaves differently than the other models (e.g. CMAQ also has high ozone concentrations)?

Page 17, lines 16–17: “Over lands the NO_2 chemistry and the different biogenic NO emissions explain a large part of the differences far from urban areas.”

- How does this explain the differences between the models — be specific.

Page 17, lines 19–20: “It should be pointed out that the observed NO_2 concentrations can be slightly overestimated because of sampling artefact (evaporation of nitric acid).”

- What do you mean by slightly? Give some number/estimate! How large overestimation of NO_2 could you possibly get from the evaporation of HNO_3?
- Provide a reference for this sampling artefact.

Page 17, lines 32–33: “Differently, differences in diurnal temperature between CMAQ and other models seem less relevant with respect to pollutant concentration.”

- How do you know that the temperature differences are less relevant? And does this statement only refer to the NO_2-concentrations or to all pollutants?

Page 18, Sect 6.3 Sulphur dioxide

- General comment: This section is very short and essentially only states that the model results for SO_2 are quite poor with hardly any explanation why. I think a much more detailed investigation of the differences in deposition and chemistry are needed here.

Page 18, lines 6–7: “The overestimation of the first group of models could be explained as follows for MINNI which has the lowest PBL and RCG having the lowest wind speed.”

- The sentence is strangely formulated — perhaps it could have been written something like: “The overestimation in the MINNI model could possibly be partially explained by the low model PBL height”
- However, I do not think that the “explanations” are very satisfying — in my opinion they are not really explanations at all:
    - For 2006 the EMEP model also severely underestimate the PBL height without overestimating SO_2.
    - The wind speed in CMAQ is as low as in the RCG model, without overestimation of SO_2, and these models actually have the smallest bias for U10 for the 2009 period.

Page 18, lines 20–21, Regarding the CMAQ-results:

- I do not think that the CMAQ results are very different for “at least three campaigns” — it strongly deviates for 2006 and deviates somewhat for 2008 but for the other two campaigns the CMAQ results look “similar” to the other models (at least for the RMSE, which is what was discussed here).

Page 22, lines 6–8, Regarding the NO_2 results at the German sites; only meteorological aspects are discussed here, but other things can also lead to modelling problems:

- How do the model results for ozone look at the same sites?
- Could NO_2 emissions be underestimated?

Page 22, lines 8–14, regarding the NO_2 results in the Po Valley

- Are you sure that you are not having problems with underestimated NOx emissions in this region?

Page 22, lines 26–30, the discussion about the correlation between the performances of the ensemble (RMSE) with the variability of the models is a bit confusing

- What values are you correlating?
- Can low correlation coefficients (-0.2 to -0.3) for only three of four campaigns and only two species be considered significant? What the correlation coefficients for the other species?
- Providing a table with the correlation coefficients for the different species and seasons may could probably make this easier to understand.

Page 23, line 21: What do you mean by “a relevant spatial variability”?

Page 23, lines 25–26: “Such spread can be considered as a measure of the uncertainty related do vertical mixing and qualitatively correspond to 80-100% of the observed mean concentration.”

- I do not understand how the model spread can be considered a measure of the uncertainty related to vertical mixing. Could there not be other differences between the models that are important?

Page 23, lines 31–32: As pointed out above I do not think that you have shown that the “lower PBL heights (for MINNI) and wind speed (for RCG)” really **explain** the errors. Also the CMAQ wind speed seems to be as low as the RCG wind speed (according to S0).

Page 24, line 12: “while EMEP seems more able to capture the evolution of the single PM compound.”

- Which single PM compound?

Page 24, lines 21–22: “The analysis of individual compounds of PM will bring more detailed, it will be investigated in a companion paper.”

- Excluding this detailed information from the present paper makes the whole discussion of PM totally uninteresting.

Language

The manuscript is not very well written, which makes it tedious to read. Large parts of the manuscript needs language editing/corrections. It is not the job of the referees of a paper to correct the language — so I only give some examples below, in the Technical corrections section. Some of the 36 authors of the paper are likely very good at English and, since all authors must have seen the manuscript before submission (according to the obligations for authors), I am surprised that they have accepted the submission without helping to improve the language before the paper was submitted. Please make sure that the whole manuscript is checked carefully if it is resubmitted.

Technical corrections

Page 1 line 37: “period” → “periods”

Page 1 line 38: “allowing evaluating the influence” → “allowing evaluation of the influence”

Page 2 line 5: “good very similar” do you mean “good and very similar”?

Page 2 line 18: replace “modelling, techniques” by “modelling techniques”

Page 2 line 19: “calculation uncertainty” do you mean “model (or perhaps modelling) uncertainty”?

Page 3 line 7: “exercise” → “exercises”

Page 3 lines 23–24: I guess the list of “non-model” institutes should include NILU as well (since W. Aas is included in the author list)?

Page 3 line 28: replace “join analysis” by “joint analysis”

Page 8, line 36: replace “most of countries” with “most countries” or “most of the countries”

Page 9, line 32: The first sentence of the “Sea salt emissions” paragraph is strange. As formulated it does not make sense.

Page 11, lines 1–2: “was diagnosed in ECMWF was made available” should probably be “as diagnosed in the IFS-ECMWF model was made available”

Page 12, line 12: “most of species” → “most of the species”

Page 12, line 19: “at some EMEP.” → “at some EMEP sites.”

Page 12, line 27: “converted in m/s” → “converted to m/s”

Page 13, line 1: “Being the boundary layer height a concept valid only for convective” → “Since the boundary layer height is a concept valid only for convective”

Page 13, line 21: “compare” → “compared”

Page 13, line 22: “is” → “was”

Page 13, line 22: “characterized by windy conditions in Europe with cool temperature above average everywhere in Europe” — strange formulation; what do you mean by “cool temperature above average”?

Page 13, line 24–25: “Precipitation were low over the Mediterranean basin but above the climatic average compare to 1961-1990 base period in the rest of Europe.” could be changed to “Precipitation was small over the Mediterranean basin but above the climate average, compared to the 1961-1990 period, in the rest of Europe.”

Page 13, line 28: “spells end” → “spells in the end”

Page 13, line 28–29: “After some cold spells end of February, March 2009 turned cooler with on average warmer temperatures compare to the 1961-1990 base period” — strange formulation; did March 2009 turn *cooler* than the cold spells in the end of February but it was still *warmer* than the climate average?

Page 14, line 3: “whatever the model” → “for all models”

Page 14, line 6: “this bias exceed” → “this bias exceeds” (or “these biases exceed”) and “whatever the campaign” → “for all campaigns”

Page 14, lines 25–26: “In the IFS only 10m winds are used from ships over the oceans for data assimilation (problem of station representativeness for inland stations).” — awkward formulation

— I would suggest something like: “In the IFS only 10m winds from ocean going ships are used in the data assimilation due to problems with station representativity for inland sites.”

Page 14, lines 27–29: “For the lowest winds generally observed during nightime the comparison of the predicted diurnal cycle with observations show a largest positive bias at night than during the afternoon (Fig. 2), this behaviour could lead to an overestimation of the advection process.” This is a very strange sentence that I do not understand. It needs to be reformulated.

Page 15, line 13: “convention” → “convection”

Page 15, line 17: “use the PBL from ECMWF PBL” → “use the PBL from IFS”

Page 15, line 22: “the negative bias of MINNI has the same order of magnitude as the other models” → “the negative bias of MINNI is of the same order of magnitude as those of the other models”

Page 15, line 23: “are still lower” → “are somewhat lower”

Page 15, line 25: “model” → “models”

Page 15, line 28: “on emission areas” → “in emission areas” and “Besides of urban areas” → “Besides in urban areas” (or perhaps “Besides urban areas”)

Page 15, line 29: “that are related to the differences of PBL predicted” → “which is related to the differences in the PBL predicted”

Page 17, lines 14–15: “the mixing of close to emissions is responsible for model output differences” — I think the whole sentence is a bit awkwardly formulated, perhaps this part could be changed to something like: “variations in the PBL height between different models may lead to large differences in modelled concentrations in high-emission areas”

Page 17, lines 32–33: “Differently, differences in diurnal temperature...” — strangely formulated sentence.

Page 18, line 8: “in-deep” → “in-depth”

Page 18, line 9: “This involves a positive bias” → “This leads to a positive bias”

Page 19, line 2: “of the seas” → “over the seas”

Page 19, line 16: “and a few” → “and a little” (or perhaps “and some”)

Page 19, line 27: “all models underestimate” → “all other models tend to underestimate”

Page 19, line 31: “Whatever the campaign” → “For all campaigns”

Page 20, line 13: “are coherent with the completeness of our inventory” — I think a better formulation could be “are consistent with our incomplete inventory”

Page 21, line 22: “smaller areas” → “limited areas”

Page 21, lines 25–26: Remove the sentence: “Finally, as already mentioned, PBL heights derived at SIRTA site has been included too.” — this manuscript is too long to state this twice within the same paragraph.

Page 22, line 27: “close between” → “close to”

Page 23, line 8: “mainly driven by a relevant underestimation” → “at least partly driven by a major underestimation”

Page 23, line 9: “CTMs are affordable in reproducing ozone” → “CTMs are able to reproduce ozone”

Page 23, line 11: “nigh-time” → “night-time”

Page 23, line 22: “Likewise ozone” → “Similar to ozone” or “As for ozone”

Page 24, line 1: “rely in chemistry” → “be due to chemistry”

Page 24, line 7: “Differently, the RMSE rises up 15 μg m^{-3}, representing more than 80% of the observed mean.” — incomplete sentence; I guess you mean “rises up to 15 μg m^{-3} for the campaign XXXX...”?

Page 24, line 27: “are still missing in state of art CTM” → “are still missing in some state of the art CTMs”