

Review of Jaidan et al. "Future changes in surface ozone over the Mediterranean Basin in the framework of ChArMEx"

Re-submission to ACP, 2018

Note: My review is of the revised version of this manuscript, not having seen the original submission.

This manuscript presents a study of the present and projected future of surface ozone over the Mediterranean Basin, as simulated by a range of global chemistry models that took part in the ACCMIP experiment. It is no doubt a useful contribution to ChArMEx and is broadly interesting inasmuch as surface ozone projections from global models are being used in impact studies and international (climate) reports.

Overall, my opinion is that the manuscript needs a further iteration of revisions. Below, I have made some comments on the authors' response to reviewers (Section A), followed by specific comments on the revised manuscript (Section B), and technical corrections after that (Section C).

A. Comments on the authors' responses to reviewers:

1. I agree with the authors' comments that a detailed explanation of the drivers of model biases/differences is not feasible. While this appears unsatisfactory to some, to do this properly in models with 1000s of parameters would require a substantial (albeit necessary) research effort, organised across multiple modelling centres. It's not just the emissions, deposition and chemistry scheme, but also physics parameters in the underlying GCM, including biases in (e.g.) the timing and location of winds, clouds, temperatures, rainfall etc.

There are efforts underway to better understand the interaction of all the biases, but we must recognise that we are dealing with phenomena that emerge from a complex interaction of multiple processes and knowing that models are "right for the right reason" will be a fraught question.

2. There seems to be some confusion about assessing statistical significance, T-test and p-values, at least as written (e.g., bottom of P9 of the response). One does not "calculate the student T test for the 95% confidence level"; rather the Student's t test gives the t-statistic, which - for a given number of degrees of freedom - can then be used to give a p-value (e.g., by using statistical software). See also my specific comments above, related to the graphs.

Additionally, I would not call it a "trend" between 2000 and 2100 as it is really a difference between two time slices.

B. Specific Comments on the manuscript

P2, L33: Some of the ACCMIP models were not chemistry-climate models (e.g., CICERO-CTM2 is a CTM, and MOCAGE and STOCHEM are basically run as CTMs - see the Young et al. ACCMIP paper).

P3, L8 (and for general consideration): There no mention of the hourly ozone output as part of ACCMIP, which might add some further depth (or at least context) to the analysis. It would at least be good to mention the analysis and conclusions of Schnell et al. (2015, ACP, doi: 10.5194/acp-15-10581-2015), who looked at this in the context of AQ in Europe and N America.

P3, L32: See Iglesias-Suarez et al. (2016, ACP, doi: 10.5194/acp-16-343-2016) for a description and evaluation of stratospheric ozone in the ACCMIP models.

P6, L12: How can the mean "simulate appropriately", yet have "a consistent positive bias"?

P7, L4: What types of models did Vautard et al. evaluate? Is their conclusion likely to be valid for ACCMIP?

P9, L4-5: Sentence starts saying "Several studies" and then only references one at the end.

P10, L2-5: The authors mention later, but here it would be good to note that there is considerable variability in the complexity of the VOC scheme (and total emissions of reactive C) between the ACCMIP models. See figure of the emissions in Young et al. (2013).

P10, L30: "We use the Student t test for the 95% confidence interval...". Either the grammar here is wrong, or there's perhaps a misunderstanding about the t test - see comment #2 in Section B.

P12, Section 4.4: I'm afraid I found this section very hard going to understand, and I wonder if it could be re-worded to be clear about what trends are from precursors and what are from climate? (See also my comment about paragraphs below).

For the impact of climate, why did the authors not analyse the subset of ACCMIP models that completed sensitivity studies with fixed emissions? See Stevenson et al. (2013; ACP, doi: 10.5194/acp-13-3063-2013).

P14, Conclusions: This section appears to be rather a laundry list of individual results, with no synthesis and little in the way of outlook. What should people doing impact studies take away from this analysis, for instance?

Figure 2 (and related discussion): Is the seasonal cycle consistent for all the grid squares in this evaluation? Is there any interannual variability in the observations that should be used on the error bars? (The models were not simulating the meteorology for the year 2000, so the comparison needs to be applied fairly, somehow).

Figure 3: Please try and avoid the rainbow colour scale (e.g., see <http://bit.ly/2rN9RjM>; applies to other figures too). Also, what is gained from having so many individual levels? Can anyone tell the difference between 23 shades of blue? Finally, please state whether the standard deviation is the intermodel spread, or temporal. (I guess the former, but it's ambiguous.)

Figure 5: A colour bar for the table might be useful, even if it is just qualitative. ...Is it based on ranking?

Figure 6: Caption starts by saying annual average, when it is a JJA average. ...Also, if you are showing absolute numbers (are you sure you want to do that?), then it would be good to show comparison numbers from (e.g.) a reanalysis product. Climate models are biased for the global mean, so I am sure that they will be so for a smaller region.

Figure 7: Please put (a), (b) etc before the species to which it refers.

Figure 9: This figure is very small, and (similar to my comment on Figure 3), I think the colour bar colours and levels needs revisiting. Furthermore, have the authors considered the "field significance" in their indication of significant (or not) differences? See Wilks (2016, BAMS, doi: 10.1175/BAMS-D-15-00267.1).

Figure 11 and 12: Is a box-whisker plot appropriate for 5 models?

C. Technical corrections to the manuscript

1. There are an awful lot of very long paragraphs. Please split up the text for ease of reading. E.g., P1,L10: new paragraph at "Tropospheric..." (and combine with next shorter paragraph; P1 L31: new paragraph at "A number..." etc.
2. A proof read would help. E.g., P2, L18: "...usually observed in summer period" -> "...usually observed in THE summer period"; Pp, L13: "experience" -> "experiment"
3. Throughout (for consideration): Why write "O3" instead of "ozone"? We say the latter; we don't say "o-3". This helps readability in my view.