<p style="text-align:center">Responses to reviewer 1</p>

Future changes in surface ozone over the Mediterranean basin in the framework of the Chemistry-Aerosol Mediterranean Experiment (ChArMEx)

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).

We thank the reviewer for his/her valuable comments. We answer point to point to the comments in blue:

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 different between two time slices.

We agree with the reviewer that the student T test gives us a p-value from which we can determine if the test is significant or not according to a chosen confidence level.

In addition, as suggested by the reviewer, we changed the term "trend" as necessary throughout the paper.

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).

We replace the following sentence:

"The assessment of the future changes in annual tropospheric O3 at global scale has been done by Young et al. (2013) using a set of chemistry-climate models"

by (see P3.L4 in section 1):

"The assessment of the future changes in annual tropospheric ozone at the global scale has been done by Young et al. (2013) using a set of chemistry models."

To clarify this, we added this point in the new version of the paper (See P3.L27-30 in Section 2.1):

"Most of the models we used are chemistry climate models (CCMs) except three models: MOCAGE which is a chemical transport model (CTM), using off-line meteorological fields from an appropriate simulation of a climate model; STOC-HadAM3 and UM-CAM, referred as chemistry-general circulation models (CGCMs), which produce their own meteorology without any interaction with climate."

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.

We agree with the reviewer, we updated the new version of the paper by

including the following sentence (see L31+ in section 3.2):

"Schnell et al. (2015) evaluated a set of ACCMIP models against hourly surface ozone from 4217 ground based stations in North America and Europe. They found that models are generally biased high during all hours of the day and in all regions. Moreover, they also found that most models well simulate the shape of regional summertime diurnal and annual cycles. They concluded that the skill of the ACCMIP models provides confidence in their projections of future surface ozone."

In addition we added the reference Schnell et al (2015) as suggested by the review

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.

We updated the new version of the paper by including the following sentence (see P4.L6-L9 in section 2.1):

"Iglesias-Suarez et al. (2015) evaluated the stratospheric ozone and associated climate impacts using the ACCMIP simulations in the recent past (1980–2000). They showed that ACCMIP multi-model mean total column ozone trends compare favorably against observations. They also demonstrated how changes in stratospheric ozone are intrinsically linked to climate changes".

In addition, we added the reference Iglesias-Suarez et al. (2016) as suggested by the review.

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

The sentence is misleading and we changed it by (see P6.L19-21 in section 3.1): " The behavior of the annual cycle of surface ozone from ACCMIP models averaged over the period 1990-2010 over the Mediterranean basin is quite similar to the one observed. The bias between the ACCMIP and the observed annual cycle is positive with values between 6.10 and 12.47 ppbv."

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

Vautard evaluated six different chemistry transport models over a full year (1999). Three models are used both at large-scale (typically 50 km) and small-scale resolution (5 km).

The results from Vautard indicate the importance of the meteorological forcings that induce a difference between the model results in the region of Po-Valley. We just used this result to provide a possible reason for the disagreement of the ACCMIP models in this specifically sensitive region.

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

We replace the following sentence:

"Several studies have shown that humidity is the most important meteorological factor affecting OH and CH4 lifetimes (Spivakovsky et al., 2000), which are involved in the chemical production of O3."

by (see P9.L18-19 in section 4.1):

"Spivakovsky et al. (2000) showed that humidity is the most important meteorological factor affecting the lifetimes of OH and CH4 which are involved in the chemical production of ozone."

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).

To clarify this, we added this sentence in the new version of the paper (See P10.L15-16 in Section 4.2):

"Note that there is considerable variability in the complexity of the chemical schemes, in particular for the VOC schemes between the ACCMIP models."

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.

We corrected the sentence by "We use a Student's T-test with a 95% confidence interval..." see also our answer for the 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).

We reworded parts of the section 4.4 as suggested by the reviewer. Our purpose was to focus on the effects of ozone precursors in the context of climate change. We also changed the title to clarify this point.

We agree with the reviewer, it is interesting to study the impact of the climate change by using these sensitivity simulations. However, our goal was to compare the results between the different scenarios by keeping the largest number of models (and only 6 models have provided outputs for the sensitivity study following the RCP8.5 scenario).

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?

We rearrange the conclusion to highlight the message of our paper. We added this paragraph in the end of the conclusion:

"The surface ozone decrease over the MB for the scenarios RCP2.6, RCP4.5 and RCP6.0 is much more pronounced than the relative changes of the global tropospheric ozone burden. This reflects the fact that the surface ozone over the MB is more controlled by reductions of its precursor emissions, water vapor represented by the increase in the specific humidity and the NOx-limited regime over the MB. In this region, for the RCP8.5 scenario, we showed how the future climate change and in particular the increase in methane concentrations can offset the benefit of the reduction in emissions of ozone precursors. Future modeling studies should quantify the sensitivity of the future surface ozone to climate change and methane concentrations changes over the MB"

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).

1) This is a difficult point to answer because we combined spatial and inter-model mean. We have preferred to focus our study on the ozone variability between the ACCMIP models rather than on the ozone variability in the very small domain only covering the MB.

2) We added in Figure 2, the standard deviation of the observations that show the variability of the observations.

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.)

We changed the rainbow colorbar by another colorbar taking into account the small number of the different levels as suggested by the reviewer.

We changed the term "standard deviation" by "inter-model standard deviation" to clarify the sentence.

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

Yes it is based on ranking; the colorbar goes from close to the observation to far from the observation for each metric. We added a qualitative colobar in Figure 5.

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.

We corrected the sentence by changing "annual average" to "summer (JJA) average".

We are not quite sure to understand the point of view from the reviewer but we only could use reanalysis for the contemporary period. From this figure, we can compare the different box plots to the reference (REF). In addition, our study is not focused on the meteorological parameters. We plotted the absolute values to have an idea of the amplitude of ACCMIP models for each parameter. However, we are interested in the parameter difference for two periods (2030 and 2100) in the future to put into evidence the link

between the meteorological parameters evolution and the one of ozone for ACCMIP models.

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

Corrected

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).

We changed the colorbar. We do not consider the field of significance, but we use local tests to have an idea on the statistical significance of surface ozone changes, as we mentioned on page P11.L13-14.

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

We agree with the reviewer that it is more appropriate to use a box plot when the number of models is relatively high. However, we find that we have additional information such as the mean and median. As well as the figure reading is easier to understand with the colors.

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.

Done

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"

We reread the paper and corrected the paper as much as possible. Concerning this example, we changed "usually observed in summer period" by "usually observed in the summer period" and "experience" by "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.

We changed O3 by ozone throughout the paper.

Responses to reviewer 2

Future changes in surface ozone over the Mediterranean basin in the framework of the Chemistry-Aerosol Mediterranean Experiment (ChArMEx)

The authors start to address my previous comments. However, like the other reviewer, I still think that this paper needs further work before it is suitable for publication in ACP. One thing the authors should do is improve the readability of the text by doing the following. (i) Reduce the size of the paragraphs (which are very long) – perhaps by splitting them into smaller units. (ii) Look at the english – in particular, there are quite a few typos. (iii) Rewrite the conclusions, so that they are less of a summary of the results, and more of a reflection of the strengths and weaknesses of the models, in this case for studying climate change in the Mediterranean Basin.

The authors should also address the specific comments below.

We thank the reviewer for his/her valuable comments. We answer point to point to the specific comments in blue:

Specific comments

P. 1 L. 8-10: Maybe I am missing something, but the text suggests the model ensemble mean simulates well the annual cycle of surface ozone, but that a majority of the models overestimate the surface ozone during the period 2000-2010 and for summer. Is this behaviour consistent? Perhaps you need a clarification here and elsewhere in the paper.

We mean that the models simulate well the behavior of the annual surface ozone cycle. We have clarified this in the updated paper (see P6.L19-21 in section 3.1).

L. 15: Where do these increases in CH4 come from?

CH_4 emissions evolution was specified for each of the RCPs. For the RCP8.5, CH_4 emissions will increase between 2000 and 2100 (Van vuuren et al., 2011). In addition, increases in life-stock population, rice production, and enteric fermentation processes drive emissions of methane (Riahi et al., 2011).

P. 2 L. 9: Perhaps include references for these sinks.

As suggested by the reviewer, we added the following reference: Jacob (2000)

P. 3 L. 3: I understand from the list of models that not all are climate-chemistry models. Please clarify. The other referee also made this comment.

We replace the following sentence:

"The assessment of the future changes in annual tropospheric O3 at global scale has been done by Young et al. (2013) using a set of chemistry-climate models"

by (see P3.L4 in section 1):

"The assessment of the future changes in annual tropospheric ozone at the global scale has been done by Young et al. (2013) using a set of chemistry models."

To clarify this, we added this sentence in the new version of the paper (See P3.L27-30 in Section 2.1):

"Most of the models we used are chemistry climate models (CCMs) except three models: MOCAGE which is a chemical transport model (CTM), using off-line meteorological fields from an appropriate simulation of a climate model; STOC-HadAM3 and UM-CAM, referred as chemistry-general circulation models (CGCMs), which produce their own meteorology without any interaction with climate."

P. 5 L. 28: A style point: consider not starting a sentence with an acronym.

Done

P. 8 L. 4: Do you need "obviously"? Omit needless words. Same for "in fact" on P. 9, L. 29.

We corrected the different sentences.

P. 9 L. 6: You mention several studies, but only provide one reference.

We replace the following sentence:

"Several studies have shown that humidity is the most important meteorological factor affecting OH and CH4 lifetimes (Spivakovsky et al., 2000), which are involved in the chemical production of O3."

by (see P9.L19-20 in section 4.1):

"Spivakovsky et al. (2000) showed that humidity is the most important meteorological factor affecting the lifetimes of OH and CH4 which are involved in the chemical production of ozone."

L. 12: I am not sure I understand your statement "according to the radiative forcing". You mean there is a direct relationship between the temperature increases and the different radiative forcings? Please clarify.

We agree with the reviewer that the relationship between the temperature increase and the different radiative forcings are not direct. We clarified this by

P.10 L. 15-19: Quantify the statements made. I suggest you do not use words like "drastically". Please avoid hyperbolic language. Do this elsewhere in the paper.

Done

P. 11 L. 30+: Quantify the statements made.

Done

P. 12 L. 29: I suggest you use "complicated" rather than "complex".

Done

P. 13 L. 6-10: Quantify these statements.

Done

L. 14: Perhaps remind the reader of these scenarios and periods.

Done

L. 24-25: What do you mean by increasing chemical terms? Which chemical terms?

What we called the chemical terms are the chemical production (P) the chemical loss (L) and the chemical budget (P-L). We have specified this in the updated version.

P. 14 Section 5: This is too long and it is difficult to see what inferences one can make about the capability of the models to simulate climate change over the Mediterranean basin. There is just one line at the end of this section. The authors should address this.

We have clarified Section 5

Fig. 12: This is not quite the same as Fig. 11. Please reword. ??

We reworded the figure 11 caption by:

"Future relative change in surface ozone budget over the MB domain for JJA period and for the RCP8.5: (a) chemical production (P), (b) chemical loss (L) and (c) chemical budget (P-L) of surface ozone, (d) dry deposition of ozone (D)

The median is indicated by the thick horizontal black line, the multi models mean by a filled diamond, the (25-75%) range by the colored box and minimum/maximum excluding outliers by whisker. Each point represents a single model. The dashed horizontal line represents the mean for the REF period (2000) considered as a reference."

Table 3: Identify which are chemistry-climate models and which are chemistry-transport models.

In Table 3, we have added a column named "type" that shows the category of each model. In addition, we also have added this point in the new version of the paper (See P3.L27-30 in Section 2.1):

"Most of the models we used are chemistry climate models (CCMs) except three models: MOCAGE which is a chemical transport model (CTM), using off-line meteorological fields from an appropriate simulation of a climate model; STOC-HadAM3 and UM-CAM, referred as chemistry-general circulation models (CGCMs), which produce their own meteorology without any interaction with climate."