Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-61-AC2, 2017

Earth System Dynamics
Discussions

Interactive comment on "Process-level improvements in CMIP5 models and their impact on tropical variability, Southern Ocean and monsoons" *by* Axel Lauer et al.

Axel Lauer et al.

axel.lauer@dlr.de

Received and published: 20 September 2017

Below we reply to the anonymous referee #2's comments and questions on our ESDD manuscript "Process-level improvements in CMIP5 models and their impact on tropical variability, Southern Ocean and monsoons". We would like to thank the reviewer for the constructive comments helping us to improve the paper. We have listed all reviewer comments below and answers are provided in blue.

Printer-friendly version

Discussion paper

Anonymous Referee #2

The paper is a very thorough evaluation of several aspects of model performance between two generations of three climate models. There is a lot of useful evaluation material, with high quality analysis and presentation. I think the paper will be useful for the climate modelling community, where other modellers can take lessons from the model development that they can then apply to other models. Modellers may also benefit from seeing an application of the evaluation software, and see the utility for their analysis. However, I think the usefulness of the evaluation for modellers and also a wider audience could be improved by drawing more links between the model performance and the differences between model versions (see major comments), and a few minor edits as well. I recommend publishing once reviews have been addressed

Scientific:

1. There needs to be more lessons about model development drawn out, so that the evaluation can be used in a constructive way by others. The analysis of resolution (3.2.3) is useful to identify the influence of this factor, what it provides and what it doesn't (e.g. improvement in moist processes but no improvement in AEW). But I feel the paper needs more links back to the cause of differences not just from resolution but in terms of model schemes and other model improvements. Comments and conclusion about what has led to improvements, what didn't, and what is still required would be useful. For example on page 37, line 16-31 when commenting on the remaining cloud and convection biases, then further comments about the model would be useful – e.g. what improvements were expected, what model components actually did contribute to model improvements, what these changes didn't achieve, and what are the remaining issues. Page 38 line 16-19 offers some insights into what caused some improvements in the Sahel, but the evidence supporting these claims is not laid

out clearly, and is not taken to the level of modelling decisions (i.e. what caused the improvement in stratocumulus clouds, which components/s?).

The model simulations evaluated here were performed within the project EMBRACE, which was aiming at improving the models in preparation for CMIP6. Besides the targets for improvements in the representation of key variables and processes discussed in the manuscript, the model development also included biogeochemical mixing in the Southern ocean, soil hydrology, the carbon cycle, and a more realistic treatment of climate-vegetation interaction.
Model simulations with one individual component changed at a time suitable for an evaluation and comparison with their CMIP5 counterparts are not available because of computational constraints. As the new models are compared with their CMIP5 counterparts, only variables and derived quantities also available from CMIP5 can be included in the evaluation. This makes identification of the exact causes of differences between the CMIP5 and EMBRACE models quite challenging. In some cases possible reasons for the differences seen can be given but not in all cases. This also makes it very difficult to give clear advice for future model development. We will, however, elaborate more on possible reasons for model improvements and non-improvements where possible.

2. For the benefit of the users of model outputs, I would like to see a discussion of where performance is 'good enough' for using the models for projections, and in what ways. I know this topic is difficult, but it is important context to judge the evaluation – at the moment, the reader is at a loss to know whether the old versions were not suitable for making particular climate projections but the new versions are suitable, or if both versions are still not useful, or both are good enough for a given purpose. This should be covered briefly throughout, in regard to each purpose, i.e. the simulation of mean rainfall to projections of mean regional rainfall, and so on. There are some

inferences made – e.g. page 21, line 6-10 suggests that models are not suitable to use for projections of changes to intra-seasonal variability in WAM rainfall in the old and the new version, and more like this would be useful.

The reviewer has a very good point. The usability of model results for supporting policy relevant decisions is an important aspect for users of the model data. The question of ”how good is good enough” for using model results for any kind of application depends strongly on the process of interest. This includes details such as geographical region, simulated quantity, natural variability, time-scales and time range or metric (e.g. mean, extreme values, probability density function, etc.). We feel that statements on the usability or usefulness of the model results are beyond the scope of this study because of the mentioned complexity of the problem and also because of the high sensitivity of this topic. Similar to the model evaluation chapter (chapter 9) of the fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR5), we regard ”the need for climate models to represent the observed behaviour of past climate” only as a ”necessary condition to be considered a viable tool for future projections”. This does not answer the ”much more difficult question of determining how well a model must agree with observations before projections made with it can be deemed reliable.” (IPCC-AR5)
We will, however, add a brief general discussion of this topic to the introduction of the revised version noting that even a "decent" fit does not necessarily guarantee a correct model behaviour in future (and changing) climate predictions, which is one of the reasons why ensemble based methods are used.

3. Throughout – for evaluation, why use just 20 years? I guess this is the IPCC baseline so is common, but it is short enough to be strongly affected by variability such as megadroughts. A longer period that still has satellite coverage (e.g. 1979-2016) would be better, then sub-periods within this could be also covered. If it is too expensive to

run models for longer periods, then I understand this limitation is unavoidable and this should be mentioned. Also, figure captions need to note the time period in the figure captions (1986-2005).

The CMIP5 simulations analyzed (AMIP, historical) typically only cover the time period up to the year 2005. In order to maximize the comparability of the EMBRACE and CMIP5 simulations and with results from the model evaluation chapter (chapter 9) of IPCC-AR5, we decided to use the time period 1986-2005 as a common denominator even though sometimes more years are available. The figure captions will be updated to include the time periods shown.

4. The bias plots nicely show the differences at high values, but I think bias plots are more useful if they include a middle section of blank/white where the bias is negligible rather than having colours go to zero – this helps interpreting areas where biases are small and avoid over interpreting differences between positive and negative when in fact they are not meaningful (goes for all figures)

The color scales are similar to the ones used in chapter 9 of IPCC-AR5. We would therefore prefer to keep the color scales as they are as this allows for an easier comparison with the multi-model mean results shown in the IPCC-AR5. This will be clarified in the revised manuscript.

5. Pg 26 line 8-16 – this is one school of thought about ENSO and the tropical Pacific, but others would disagree – it is important to cover a range of ideas here. E.g. work by Felicity Graham: Graham et al. 2014 Effectiveness of the Bjerknes stability index in representing ocean dynamics (Climate Dynamics), and Graham et al. 2015 Reassessing conceptual models of ENSO (Journal of Climate)

We will add the alternative conceptional model for ENSO suggested by the reviewer to the introduction of section 3.3 (coupled tropical ocean climate).

6. Specify the version of NCEP used in each case – NCEP1 or 2?

All NCEP data used are NCEP 1 data (Kalnay et al., 1996). This will be clarified in the revised version.

7. Pg 2, Line 15 – why IPCC 2007 not 2013?

We will add IPCC 2013 as a reference.

8. Fig 1 – I think an observations-based dataset should be used here, at least as the comparison panel, rather than two reanalyses. Also, are you reporting the difference between ERAint and NCEP as the observed uncertainty? If so, why not show panels only where the bias is larger than this observed uncertainty (blank out other regions)? Could do this in other figures too.

We are showing differences compared to ERA-Interim to allow for easier comparison with the CMIP5 multi-model mean shown in the IPCC-AR5. The same is true for the color scale. We would therefore prefer to keep ERA-Interim as the reference dataset and the color scales as they are. The comparison with NCEP is shown to highlight the regions with particularly large uncertainties in the reanalyses. In order to strengthen this point, we will add a panel showing the Met Office Hadley Centre observations "HadCRUT" to the figure.

9. Section 3.1.2 – absolute errors in rainfall appear higher where the mean is higher (i.e. the tropics), so reduce the appearance of biases in drier areas. The paper needs a figure showing proportional bias (%), even in additional material, to give some perspective on rainfall biases outside the tropics – for example the biases in Canada, Australia and Siberia look small and almost indistinguishable in the different panels, but important differences could be seen in a % bias plot. If % biases are extremely large due to extremely low rainfall, then these areas could be masked out or else identified and discussed.

We will add a figure showing the relative bias in precipitation to the supplementary material (that will be newly created).

10. Page 9, line 4 – I don't think projections will ever be 'accurate' in the sense they won't give a single, correct answer, so I think this word should be changed to "reliable", "robust" or similar – projections that give useful information but are not a single 'accurate' answer.

We will replace accurate by "reliable" as suggested.

11. Figures 14-15 – I was expecting to see SST and rainfall bias map plots for the coupled versions (to see the shape of the warm pool, the extent of the cold tongue bias, the shape of the double ITCZ etc.) – one can see the temperature bias in Figure 1 somewhat, but it is not very clear. Perhaps a Pacific SST and rainfall bias map here or in additional material?

Following the suggestion of the reviewer, we will add SST and rainfall bias maps zoomed in over the Pacific to the supplementary material (that will be newly created).

12. Page 38, line 20-24 – I think this conclusion is incomplete, yes the lack of improvement with finer resolution certainly indicates that there are problems at the coarse resolution and apparent good performance is probably related to compensating errors. But it also shows that these issues are not solved by finer resolution – thus indicating that either some critical threshold of resolution has not been reached (perhaps we need resolution somewhere $<$14 km) or else some other non-resolution factor is involved (e.g. parameterisations are not working well enough and the improvement to the cloud scheme didn't fix the problem). See major comment 1, I think this type of discussion and conclusion is needed more generally

We agree with the reviewer that the results suggest that there are most likely also factors other than horizontal resolution involved. We will extend the discussion in the revised manuscript to include this conclusion.

Minor

1. Pg 2, Line 3 – ITCZ not defined on first usage in text

2. Pg 6, line 6 – Similar not similarly

3. Figure 2 – caption notes lower right panel is noted as data from CMIP, this should be CMAP

4. Figure 3 – the box with 'Reference' in it obscures an important high rainfall region in Nepal, suggest making it smaller and moving it (perhaps top left corner?), same in many other figures.

5. Page 15, line 7-8 – 1970s and 1980s don't need apostrophes,

6. Page 15, line 24 "that one of is the main" typo

7. Figure 10 caption and/or legend needs to explain the grey shading

8. Page 31, line 10 – a paper from 2010 can't show the 'current GCMs', note that this paper is about CMIP3

All minor suggestions / corrections will be applied as suggested.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2017-61, 2017.

Printer-friendly version

Discussion paper