Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-61-AC1, 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 address the comments of reviewer #1 and questions raised during the open discussion of the paper "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 time and effort reviewing the paper. We feel it has improved thanks to the constructive comments. We have listed all reviewer comments below and our answers are provided in blue.

Printer-friendly version

Discussion paper

Anonymous Referee #1

This manuscript by Dr. Lauer and colleagues examines the performance of four updated earth system models (ESMs) in simulating a selection of processes and climate phenomena for which ESMs have demonstrated a need for improvement. Specifically, this study focuses on ESMs that have been updated during the EMBRACE project. The authors focus on phenomena relating to the coupled tropical and Southern Ocean climate. Overall, the authors find that the updated models have improved over their CMIP5 counterparts in several respects, including simulation of mean tropical precipitation, synoptic-scale variability of tropical precipitation, equatorial sea surface temperatures (SSTs), and Southern Ocean cloud cover. The authors also note, however, that the simulation of some processes demonstrates little improvement relative to the CMIP5 versions, and some process-level performance diagnoses are strongly constrained by observational data uncertainties.

Overall, I feel that this is a well written manuscript that provides useful diagnoses for the earth system modeling community. The study is well structured, and the analyses target important earth system processes and shed light on ESM development. The areas that could use improvement, in my opinion, relate to the synthesis of model performance. There are several areas where the message is not quite clear to me owing to conflicting evidence, observational uncertainty not clearly stated, or lack of specific evaluation criteria. I believe that this article will be a useful contribution to the literature if the authors can clarify several points raised below.

Specific Comments

1) The authors present the results under the presumption that all the ESMs being evaluated are improved (e.g. first line of the abstract). As I read this manuscript,

I am left wondering if some of these ESMs are actually improved overall relative to their predecessor versions. Of the three fully coupled ESMs being evaluated, only EC-Earth seems to be consistently improved. If we use RMSE as an overall performance metric, HadGEM is approximately the same or worse in global surface temperature (Fig. 1), global precipitation (Fig. 2), South Asian monsoon climate (Figs. 3-4), and West African monsoon climate (Figs. 7-9 and 11). Similarly, MPI-ESM performance is approximately the same or worse in all of the same categories except global precipitation (Fig. 2) and African JJAS temperature (Fig. 8). With that in mind, are we confident that all the models are improved overall? The authors acknowledge that there are processes that have not shown improvement, but I believe that they could make some additional effort to synthesize the performance changes in the ESMs.

The word ”improved” in the context of comparing the EMBRACE models to their CMIP5 predecessors was referring to the assumption that changes, modification and extensions applied to the models are either done in order to include more or more detailed processes or to improve the performance or skill level of the models. We agree with the reviewer that the word ”improved” might be mistaken for referring to the model performance only and will replace ”improved” with ”updated” in the revised manuscript.

The reviewer is right that, when looking at global average RMSE values, only EC-EARTH has improved consistently. As we think that such global averages are only a starting point for a model evaluation, we looked at several regions and processes in more detail. As stated in the text, the models do show some improvements (in terms of better performance) in some regions, while they perform similarly or sometime even worse in other regions. It is therefore hard to define whether or not the models' overall performance is improved as this strongly depends on the focus of interest / region / process.

In order to address the legitimate criticisms of the reviewer, we will put more emphasis on synthesizing the performance changes by adding more discussion to each section in the revised manuscript on whether we think that the EMBRACE models show an improved performance compared to their CMIP5 predecessors. We will also make it clearer that the EMBRACE model versions were prototypes, not yet fully tuned, calibrated or developed with one aim of this paper to document the long and sometimes difficult pathway of model development and the challenges of improving large model biases.

2) In several places, the authors claim that the models are ”significantly” improved or have ”significant” biases (p1, line 20; p8, line 1; p10, line 14; p13, line 2, etc.). It is difficult to see how significance is determined. There are no formal significance tests performed. The analyses indicate large observational uncertainties, so it is difficult to determine by eye which model changes or biases are meaningful. I recommend that the authors specify the criteria used to determine meaningful improvements and biases. I also recommend omitting the word ”significant” unless a formal significance test is indicated.

We agree with the reviewer that the word ”significant” in this context is potentially misleading and will rephrase these sentences in the revised version.

3) p. 4, lines 15-16: I do not understand what is meant by ”medium range climate sensitivity.” The phrase ”medium range” typically refers to a period of time.

”Medium range” was referring to the climate sensitivity of the new MPI-ESM model being in the middle of the range of climate sensitivities spanned by the ensemble of

CMIP models. The sentence will be clarified and rephrased.

4) Section 3.1.1: This relates to my first comment, but the text does not seem to convey all the information evident in Fig. 1. First, I recommend that the authors indicate in the caption what the ”bias” and ”rmsd” numbers above each plot represent – presumably it is the global mean bias and root-mean-square deviation from ERA-Interim reanalysis. I also believe it is worth mentioning in the text that (1) there is large uncertainty in the near-surface air temperature climatology in reanalysis data (evident in the bottom right of Fig. 2), and (2) overall CNRM and EC-Earth show evidence of improved surface temperature climatology but HadGEM and MPI are not improved overall.

The reviewer is right, the numbers given above each panel in figure 1 are global average values for the deviation of the model results from the ERA-Interim reanalysis. This will be clarified in the revised captions. The regions with particularly large uncertainties in the ERA-Interim data will be discussed in more detail and a panel showing Met Office Hadley Centre observations ”HadCRUT” as an alternative reference will be added. We agree with the reviewer that the MPI model performance has not changed much (see p. 5, l. 14-15 and p. 6, l. 4-5). The inferior performance of the EMBRACE version of the HadGEM model in reproducing the ERA-Interim surface temperatures is stated in the text (p. 5, l. 30-31): ”In these regions, the near-surface temperature biases in the EMBRACE version [of HadGEM] are up to 2°C larger than in the predecessor version.”.

5) p. 8, lines 3-4: In addition to the general issue raised in my second comment, the bottom right figure seems to indicate that the amplitude of the tropical precipitation biases may be strongly dependent on the choice of observational dataset. How do the wet biases change with a CMAP reference climatology?

When using the CMAP dataset as reference, the model biases in the tropics are typically between 0.5-1 mm day^{-1} smaller, in the mid-latitude storm track regions, the model bias changes sign from too dry (GPCP) to too wet (CMAP). The reviewer has a valid point and we will include the difference plots shown in figure 2 using CMAP as reference dataset in the supplementary material (that will be created along with the revised version).

6) Fig. 3 and later figures: What are the ”corr” values above each plot? Are those pattern correlations with the reference field? Again, these types of metrics should be described in the figure caption. In addition, the ”Reference” label in the top left panel is a bit obstructive in some of the plots. This is especially true in Fig. 4, where the label obstructs the region of strongest horizontal temperature gradients.

The reviewer is right, ”corr” refers to the linear pattern correlation coefficients with the reference dataset. This will be added to the respective figure captions. The ”reference” label will be reduced in size in the revised figures.

7) Fig. 4: The top label in the second and fourth columns indicates ”NCEP” but the figure caption indicates the CRU dataset.

The figure is indeed showing results from NCEP. The figure caption will be corrected. Thanks for spotting this.

8) Figs. 6 and 10: I don’t understand what ”yrs: inconsistent” means at the top of each panel.

The labels ”yrs: inconsistent” in figures 6 and 10 refer to the fact that the different datasets do not cover the same years when averaged (as can be seen from table 3). The label will be removed in the revised figures to avoid confusion and a reference to table 3 will be added to the figure caption.

9) p. 15, line 17: I don’t follow what is meant by ”natural and forced modes of variability” in that context.

This sentence refers to natural variability and a forced response to increased greenhouse gases driving precipitation changes. This will be clarified by rephrasing the sentence.

10) Fig. 8: The top right panel indicates a CRU RMSE of 5.29. Is that a typo? If not, then the model changes illustrated in the figure are very difficult to interpret because they would be within the noise of observational uncertainty. The top right panel, however, does not look consistent with an RMSE that high.

The RMSE given above the CRU panel is not matching the plot shown by mistake. Thanks for spotting this. This will be corrected in the revised version of the figure. The correct RMSE of the CRU dataset compared with ERA-Interim is 1.3 K, which is smaller than the RMSE of the models analyzed.

11) Synoptic precipitation variability (pages 21-26): Overall, I found this section difficult to interpret. First, based on the plots and metrics presented in Fig. 11, it looks like all the EMBRACE models perform worse in simulating band-pass filtered precipitation variance over the illustrated domain – this point is not brought out in the text. (Also, I do not know what the ”sahel” metric is above each plot – again, all metrics presented

in the plots should be explained.) In addition, the authors argue that increasing atmospheric resolution improves the representation of precipitation variability. I am not sure how this evaluation is made. Figs. 11-13 indicate large observational uncertainty, depending on the use of TRMM or GPCP data. If GPCP is used as the baseline, it is not clear that increasing horizontal resolution leads to improvement. Are the authors assuming that TRMM is the more appropriate baseline? If so, why? And why do the EMBRACE models look much worse than their CMIP5 counterparts relative to TRMM (Fig. 11)?
This is a more minor issue, but I think it's worth mentioning that increasing horizontal resolution does not seem to improve deficiencies in the general pattern of elevated variance over Africa, which is too narrow and does not extend far enough south in the model (Fig. 12).

We agree with the reviewer that the EMBRACE models show less variability in the band-pass filtered daily precipitation fields than their CMIP5 counterparts. This will be explicitly mentioned in the revised version. The metrics "sahel" given above the individual panels in figure 11 is the band-pass filtered precipitation variance averaged over the rectangular region "Sahel" defined as 10°W-10°E, 10°N-20°N. This will be added to the caption of figure 11.
We also agree with the reviewer that the precipitation observations are subject to large uncertainties, in particular in the region investigated as there is only a sparse coverage with rain gauges. We use TRMM satellite observations as our reference dataset as the data have the highest horizontal resolution (0.25°x0.25°) of the three precipitation climatologies used (GPCP, TRMM, CMAP). We found that when regridding the TRMM data to a coarser grid such as 1°x1°, the band-pass filtered variability decreases. This can be seen when comparing the panels "TRMM" in figures 11 and 12: the TRMM data in figure 11 have been regridded to 1°x1° while the data are shown at full resolution (0.25°x0.25°) in figure 12 (note the different color scales). In order to improve the comparability, we will regrid all data (models and observations) to a common 2.5°x2.5°

grid when preparing the revised version of figure 11.
The evaluation performed here aims mainly at comparing CMIP5 version and EMBRACE version rather than comparing the model results with the reference dataset. This will be clarified in the revised version of the manuscript by adding an additional discussion.
We will also add a brief discussion of the fact that the band of elevated variance is too narrow in the models compared with both observational datasets.

12) Figs. 17-19: I do not see any gray dots in the EC-EARTH plots. Is that because there is almost perfect overlap between EC-Earth and EC-Earth3?

Unfortunately, no radiation data from EC-EARTH are available, i.e. only the fractional occurrence of cloud cover can be shown for EC-EARTH (middle rows in figures 17-19). This will be added to the figure caption.

13) Table 1 caption: I do not understand the use of the phrases ”can be close” and ”can include.” It is trivial that the models can be close to the CMIP6 versions, but it is more important to know if they are.

The caption of table 1 will be shortened to ”List of models analyzed.”

Technical Corrections

1) p.1, line 15: ”earth” → ”Earth”

2) p. 1, line 29: ”over the last years” → ”in the recent past” or ”over the past few years”

3) p. 2, line 3: ”ITCZ” acronym should be expanded.

4) p.2, line 19: ”is particularly focusing on” → ”has a particular focus on”

5) p. 12, line 1: ”varies” → ”vary”

6) p. 15, line 24: ”that one of is” → ”that is one of”

7) p. 19, line 2: ”Further” → ”Farther”

8) p. 37, line 28: ”further” → ”farther”

All technical corrections will be applied as suggested.

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