Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2017-61-RC1, 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.

Anonymous Referee #1

Received and published: 8 August 2017

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.

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.

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.

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.

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?

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.

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

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

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

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.

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).

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?

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.

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”

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