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 #2

Received and published: 24 August 2017

Process-level improvements in CMIP5 models and their impact on tropical variability, Southern Ocean and monsoons. Lauer et al.

Submitted to Earth Systems Dynamics

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?).

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.

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 mega-droughts. 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).

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)

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)

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

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

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.

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.

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.

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?

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

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