Geoscientific
Model Development
Discussions

Interactive comment on "ESMValTool v2.0 – Extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP" *by* Veronika Eyring et al.

Anonymous Referee #1

Received and published: 23 January 2020

General Comments

This paper documents the latest version of the ESMValTool v2.0. The paper and the tool have lots of very good points. For example, one can reproduce IPCC Working Group 1 figures fairly easily. As an IPCC author and also a user of the IPCC, this alone makes the tool very valuable to our community. The ability to quickly assess the new CMIP6 models is equally valuable and so on. I strongly support the development of the tool.

That said, I found the paper very hard to read and review. The authors want to discuss each option (what they call recipes) found in the tool. In describing the recipe, it is good to show a figure or 2 to help the reader appreciate what the tool can do. Given the large number of recipes, this makes the paper very long with many figures.

There is a tension between showing the figures as examples and trying to make certain scientific points/conclusion with those figures. I understand this tension well as I have also tried to produce papers like this one. Scientists want to talk about the science, not just use the figure as an example. The problem is the caveats and sometimes the complete justification for the conclusion is left out because the science is not the point. It is an example.

In this paper, there are conclusion sentences in the figure captions and figure caption text in the main body in many, many places. Cleaning these up would greatly help the readability of the paper. As a reader, I strongly dislike conclusions in the figure caption. This is especially true when the conclusion is not restated in the main text. The technique of giving conclusions in the caption is fine in a talk, but not in a paper. It just adds to the length.

Also, many figure captions are missing units and other details needed to document the figure. Sometimes these details are found in the text, but they belong in the caption. Again, this would help the readability of the paper. I highlight some of these in my specific comment section below. As a suggestion to greatly shorten the paper, one could have the nice summary section as section 3 and move all of what is now section 3 to an appendix. I feel this would help most readers and give the authors room to add more overview type of text. If the readers need the details, they can find them in the appendix.

Another general comment is that there needs to be some documentation of the models used in the paper. Shorthand names are given in most cases. Somewhere these names have to be connected to references. Most likely another table is needed.

Finally, it needs to be made clear that most users will not run the tool themselves. The users mainly interact with a browser that displays information which has been all ready computed. This is addressed in section 4 but needs to be made clearer in the abstract, introduction and summary sections. There is a lot of misunderstanding in the community about this tool. Clearing up this aspect will help.

Specific Comments 1. Line 57 – end-to-end provenance – This is jargon. What does this mean? 2. Line 58 – ensure reproducibility – Of what? Need to mention it is reproducibility of the analysis/figures. 3. Line 72 – broad user community - Who do the authors have in mind? Non-specialists would struggle with most of the analysis presented in the paper. 4. Line 77 – 2.1 and 4.7C for a doubling – By when? Is this transient (then a year is needed along with the time averaging period) or equilibrium? 5. Line 105 – to achieve this – To me this sounds like the tool is the solution. It is only part of the solution. Reword. 6. Line 131 – Reference needed for CMOR tables and definitions. I assume the CMIP6 web address is fine. 7. Line 137 – Section 2 needs to mention that the tool helps address a major CMIP focus – to evaluate model performance by comparing models and observations. It should also mention the OBS4MIPS effort here. 8. Line 163-164 – The caption and the text use different words to describe what is shown: error versus deviation. Make them match. The text in the caption (lines 1022-1026) need moved to the text. An important thing to note in the caption is that the colors will change if models are added or removed. (Defining what is meant by "relative"). 9. Lines 170 – 185 – The high correlation for TAS is likely related to mountains and the land-sea mask. Is there anyway to removed these imposed boundary conditions from the analysis? For precipitation, the observations over the ocean are uncertain. Likely more uncertain than the land obs. Is it possible to weight the land more in this analysis? 10. Figure 2, lines 1035-1036 – The sentence that starts "The figure shows both" belongs in the text. 11. Lines 185-201 – What variables were used to make the index? In the caption the phrase that starts "and in this case . . ." (line 1045) belongs in the text. 12. Line 218 – 220 – Figure 4 shows . . . - This is figure caption text. In the figure caption, only the first and last sentences should

remain. The middle sentences belong in the main body text. What are the units? 13. Lines 237 – 243 – Much of this text belongs in the figure caption. The sentence in the caption "Larger biases can be seen" (lines 1062 – 1064) belongs in the text.

*** Note: I stop commenting on figure caption text in the main body and main body text in the caption below this point. It occurs in most captions. ****

14. Line 253 – peculiar – What is peculiar about these regions? Change to "some". 15. Line 260 – From the caption of figure 8, I have no idea what is plotted in figure 8. What are the units? The main text again describes the figure and belongs in the caption. 16. Line 284, figure 9 – The units for both plots are missing. It is unclear to me what is being plotted. 17. Line 312, figure 10 – Units? The label says %. Percent of what? Occurrence? 18. Line 311 – ZIP – Is this shorthand standard? "Compressed" seems better but less precise. 19. Line 323 – severely impacts – This seems way too strong. It could impact prediction or projections. I have not seen many examples where it does. Change to "could severely impact" or reword. 20. Lines 325 – 365 – I think this discussion could be greatly shortened. Just need to reference Lembo et al. 2019. I do not get much from figure 10. There is a lot of text for little gain. Even the analysis sentence in the figure caption (which belongs in the text) only states a known conclusion. Figure 14 and 15 – what are the units and shadings? 21. Line 385 – Somewhere in this discussion, the point that the observational record in many cases is too short to reliably assess the variability. 22. Line 419 – observed – Change to "reported". 23. Line 421 – unsuitable – This seems too strong. It depends on the questions being asked. 24. Line 410 – Weather regimes – One has to use these with caution. For large climate changes, they may not be useful/reliable. Caveats are needed in this section. 25. Line 436 – Figure 17 – Units? 26. Line 445 – ZIP – See comment #18 above. I have stopped commenting on the use of ZIP. 27. Line 446 – Figure 18 – Units? 28. Line 464 – Figure 19 – Units? 29. Line 492 – strength of northward current – This is incorrect. It is the strength of the overturning circulation – near surface and deep. 30. Line 494 – Figure 22 caption – but it is not clear . . . - This

is a funny statement for this paper. It could be computed using the model spread as an estimate. Also, this phrase belongs in the main body text. 31. Line 561 – Add "of temperature" after "Arctic amplification". 32. Line 570 – Figure 29 caption – A. PHC3 is used in the caption. The label is PHC. Make the labels the same. B. Eurasian Basin – Needs defined in some way. Latitude-longitude? C. Add "in Arctic" after "Eurasian Basin". D. Atlantic water too depth – sentence belongs in text. In the caption, need to say how Atlantic water is defined. 33. Line 577 – Eurasian and Amerasian basins – These basins need defined in some way. 34. Line 589 – 590 – linearly interpolated to climatology levels – Of what? Observations? Need reference. 35. Lines 593 – 597 – Discussion seems to suggest that velocities are interpolated and then transports are computed. If this is the case, this is calculation is wrong. The transport needs to be computed first on the native grid and then interpolated. Doing the velocity first can and will lead to incorrect transports because of issues with the dot product. 36. Lines 600 – 605 – Why use only T to define Atlantic water? It seems like S should be used too. 37. Line 689 – Figure 33 – Relative bias (%) needs better defined. I am not sure what it means. I also do not know what accumulated and averaged bare soil covered area mean. 38. Line 694 – LUC – Defined somewhere? I could not find it. 39. Line 712 – 713 – 5X5 model grid cells – Model grid cells are not 5X5. This is some interpolated grid. 40. Line 728 – Almost only snow-free areas are visible – This makes no sense. 41. Lines 725 – 732 – This result seems suspect to me. It needs checked against models where land cover can change during the integration. Why is the sign of the change the same in both hemispheres given that the data are for July. 42. Line 770 – not well reproduced – This seems too strong given the large observational uncertainty. 43. Line 777 – Figure 36 – The observational uncertainty shading band seems way too narrow. There are many estimates for observed carbon fluxes and they disagree a lot. The internal uncertainty estimates for any given observed data set is typically quite small relative to the disagreement between obs data sets. Therefore the figure and the text are quite misleading. Revise. 44. Line 802 – Figure 39 – Units? 45. Line 807 – but also the spatial distribution . . . - What does this phrase mean? 46. Line 807

– 809 – Figure 40 – Why are there the colored lines on the right side of the figure? The main text refers to a color scale. I do not see a color scale. 47. Line 845 – 858 – Can one access the data plotted in the figures if needed? This would be useful both to IPCC authors and anybody who needs to replot the data. 48. Line 963 - How does the tool handle "bad data"? By bad I mean having the wrong units or the grid is wrong or the data is missing - as examples. How much of the error checking is human and how much is automated? Can other figures be generated if the input data is ok while bad data exists for some other variables? 49. Line 972 – Are the observations available? Saying they are not distributed with the tool does not address this important question.