

Earth Syst. Dynam. Discuss., referee comment RC3
<https://doi.org/10.5194/esd-2021-59-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on esd-2021-59

Anonymous Referee #3

Referee comment on "MESMER-M: an Earth system model emulator for spatially resolved monthly temperature" by Shruti Nath et al., Earth Syst. Dynam. Discuss.,
<https://doi.org/10.5194/esd-2021-59-RC3>, 2021

Review of 'MESMER-M: an Earth System Model emulator for spatially resolved monthly temperatures' by Nath et al.

General summary: This paper extends the existing MESMER approach to include a monthly downscaling module, to enable the generation of large ensembles of spatially explicit monthly temperatures that are representative of ESM behaviors, which could be an useful tool for regional impact assessments. The paper is clearly written for the most part and contributes to the existing literature on ESM emulators. However, I have some concerns that I would like the authors to fully address.

General comments

(1) The training and verification results (such as those described in Lines 218-219, Lines 238-240) show non-trivial dependency on the number of ensemble members available for training, which raises concern not just for the robustness of this method, but also the usefulness of the method presented. If MESMER-M relies on a large number of ensemble members to get robust results, then it defeats its own purpose. A useful component to add into the paper is a sensitivity test to show what is the size of training runs needed to get a robust training result.

(2) The technical details need to be better described/clarified, for the potential users of MESMER-M to fully understand the approach taken, the assumptions made, and the procedure to carry out the training, calibration, validation, and generation of 'super-ensemble' using this method. Please see some of my specific comments below.

(3) My understanding is that the particular MESMER-M presented here can generate a 'super-ensemble' under the SSP5-8.5 scenario. Although the authors stated that 'MESMER offers the perspective to improve our understanding of the likelihood of future impacts under multiple scenarios', a different version of MESMER and MESMER-M has to be developed per scenario. If so, this needs to be clearly stated, and could the authors comment on how straightforward this process would be to expand this work under multiple scenarios.

(4) The fact that the emulator solely relies on annual temperature as input, and the assumption that other forcings have very little impact on the local monthly temperature response, makes the applicability of the monthly temperature probability distributions derived from the emulator limited. There should be more discussions around in what applications would the emulator results be particularly useful in Section 6.

Specific comments

(1) In the Abstract, it should be 'model projection uncertainty' instead of 'model uncertainty'. Model uncertainty comes from incomplete representation of physical processes, uncertain/unknown physical parameterisations, structural uncertainty.

(2) Line 5, it should be 'selected climate variables' instead of 'select climate variables'.

(3) Line 10, what does 'mean response' refer to here? Also this is an odd sentence structure, consider reframe to 'represent the monthly temperature cycle in response to the yearly temperatures'.

(4) Line 50 & in the Abstract, it's important to be clear which one the authors are trying to refer to, internal variability or natural variability, to me these are different things.

(5) Line 64, is there any particular reason why a spatial resolution of 2.5 by 2.5 is used? Could the authors comment on how they expect the results to change if the analysis is done at a lower or higher resolution, and does the training of the emulator have any requirement or restriction when it comes to spatial resolution?

(6) Line 66, please specify why this reference period is chosen.

(7) Line 101, the authors need to explain Bayesian Information Criterion to the readers (the significance of BIC), and why 8? Figure 2 suggests you are using 6 instead of 8.

(8) Line 121, please explain the Yeo-Johnson transformation, for the benefit of the readers who are not familiar with this, which I expect would be the case for many readers.

(9) Line 127, I don't understand how and what fitting is being done using maximum likelihood here?

(10) Line 134, again, for the benefit of the readers, please specify what a Gaspari-Cohn function is here and why you choose this function to apply here.

(11) Line 156, I don't understand the authors' decision to only look at the top 50 highest power spectra. Please elaborate your thinking and reasoning behind this.

(12) Line 166-168, please consider rewriting this bit to clarify what's exactly being done to create these emulated quantiles. This part reads very confusing as it is now. In the following sentences, the quantile comparison description also lacks clarity.

(13) Section 3.4. Please explain why these particular biophysical variables (as listed in Table 1) are considered (chosen over other variables) and used in this study.

(14) Line 195, I don't quite understand how this procedure is done, how did the authors use the physical model to augment the harmonic model results. Please elaborate.

(15) Line 224-225, the authors should consider adding some discussions here on why these two models show such outlier behaviour.

(16) Section 4.2, please explain why these 4 ESMs are presented here, how they are representative (e.g. span across some projection range), and why WNA and WAF regions are chosen here.

(17) Figure 3 & Figure 4. The labelling on the upper left corner should be 4 ESMs.

(18) I would suggest changing the whisker colors in Figure 6 & 7 so that it's easier to see them.

(19) Table A1 clearly shows the split between training/test runs is not always 70/30, as opposed to what's stated in the text. Please check and confirm what's being done.