Earth System Dynamics Discussions

Earth Syst. Dynam. Discuss., referee comment RC1
https://doi.org/10.5194/esd-2021-59-RC1, 2021

Comment on esd-2021-59

Anonymous Referee #1

Referee comment on "MESMER-M: an Earth System Model emulator for spatially resolved monthly temperatures" by Shruti Nath et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2021-59-RC1, 2021

The authors extend an annual temperature ESM emulator to simulate monthly temperatures that account for short timescale temporal correlation, spatial correlation, and intra-annual changes in distribution (e.g., variability and skewness). I find the paper to be overall well-written and a good contribution to the literature on ESM emulation, and therefore of interest to ESD readers, although there are a number of improvements that I believe could be made to both the content and clarity of presentation. Comments and suggestions follow:

High-level Comments:

1. I typically think of an emulator as something that can be used to produce synthetic simulations under a new scenario that the ESM model has not been run under (or, alternatively, a new parameterization). The meaning here, by contrast, is something that can be used to generate additional synthetic simulations from the same scenario that the ESM was originally run under (SSP5-8.5 in your case), in order to better explore internal variability. That is fine, but I think it would be important to emphasize that you should not expect this emulator to perform as well under scenarios that produce very different GMST responses compared to the SSP5-8.5 scenario. That is, I think the pattern scaling model in Equation (1) is likely reasonable when applied to a single scenario, but would perform poorly if you tried to use the same pattern across very different scenarios.

2. Your emulator allows for changes in variability and skewness within a year, but not across years (unless I misunderstand). There is evidence in model runs of changes in interannual variability over time, as well as changes in other aspects of the temperature distribution (like skewness and tail behavior). Did you look into this at all?

Specific Comments:

1. Equation (1), how is the smoothed GMST calculated?

2. Equation (2), the notation f_s(T_{s,y}) is confusing: you should indicate that this is also a function of month (m).

3. L101, isn't the maximum value of n = 6 (not n = 8)?

3. L101, when you say that you use the BIC to determine the number of harmonic terms, what likelihood function are you using? Is this just the BIC assuming independent Gaussian noise (rather than your full model that accounts for spatiotemporal correlation and non-Gaussian behavior)?

4. Equation (4) specifies a deterministic relationship; I believe you are missing a noise term and you need to specify its distribution.

5. L118, it is not true that inferring correlation requires the data to be Gaussian. You could simply rephrase this to something like "We use a trans-Gaussian process relying on power transformations to account for the fact that temperatures may be non-Gaussian."

6. L121, I'd recommend that you define what the Yeo-Johnson transformations are, because I expect many readers will be unfamiliar with this.

7. L121ff, I don't understand how the power transformations are being applied here. The typical approach would be to transform temperatures to approximate Gaussianity, then apply the additive model that you are using, then back-transform to the original scale. Is that what you are doing? If so, then I think the equation given in Figure 1 is incorrect. If not, can you please clarify where the transformation is happening and why you are using the approach that you are using (which would seem nonstandard to me)?

8. L127, again what likelihood are you using when you say that you fit using maximum likelihood? Are you using the full likelihood that accounts for spatiotemporal dependence, or the likelihood that would assume independence?

9. L130-134, please clarify what the Gaspari-Cohn covariance function is and explain why you chose that covariance function.

10. L155, I don't think it is a good idea to only look at the frequencies corresponding to the top 50 highest power. This will cause you to focus on low-frequency variability, but the total variability is the integral of the whole spectrum and there are more high frequencies than low frequencies (so it will typically be the case in my experience that the integral over the higher frequencies is comparable to or larger than the integral over lower frequencies, even though the power is lower at higher frequencies). Put another way, short timescale variability typically dominates, even though the spectrum is higher at lower frequencies. As such, if you get the high-frequency variability wrong, this can have very important consequences even if the power at individual high frequencies is small.

11. Section 3.3.3, I don't understand what is being done here. Are you averaging within the region and then taking the 5th, 50th, and 95th percentiles over time? Or are you calculating those quantiles across individual gridcells within the region?

12. L237-240, I don't understand why the spatial model should depend on the number of runs in the ensemble. You should be able to better estimate the spatial model with more runs, but the spatial model shouldn't change. Am I misunderstanding what you are saying here?

13. Figure 1, what is going on with the gridcells showing very high autocorrelation values (and, likewise, with those showing very negative autocorrelation values)? Are these locations where the mean model is not performing well?

14. Figure 3, It looks to me like the emulator runs are more variable than the ESM runs. Is that the case, or just an artifact of how the plot is displayed? Can you give a direct comparison of the variability from the two? Also, there are some visible problems with the mean model especially for MPI-ESM1-2-LR in WAF that are not discussed in the text.

15. Figure 4, It looks to me like the emulator is under-representing skewness particularly in WAF for regionally averaged temperatures. I wonder if this is the result of a deficiency in the model for spatial correlation, or if rather the power transformation you are using is not sufficient for transforming to Gaussianity.

16. L268 - 276 and Figure 6. I don't understand whether you are saying that you under-estimate the spectrum when the power is high, or that you under-estimate the spectrum for high frequencies. Can you please clarify? I also don't understand how to read the inset example in Figure 6, which doesn't seem to give information about frequency. Rather than plotting the power in one vs. the power in the other, I'd recommend that you plot the ratio of spectra vs. frequency.