

Earth Syst. Dynam. Discuss., author comment AC1  
<https://doi.org/10.5194/esd-2021-59-AC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Reply on RC1

Shruti Nath et al.

---

Author 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-AC1>, 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:

We thank the Referee for their thorough review of the manuscript and the overall positive feedback. We found the insight provided - particularly on the technical parts as well as model verification (e.g. on the power spectra analysis) - extremely helpful and feel like the suggestions strengthened the paper in terms of future readers' understanding as well as robustness in methods. We tried our best to respond to the comments as below.

High-level Comments:

- 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.

Emulators have many avenues of application from exploration of model parameterisations to new scenario synthesis. In this study we indeed mainly focus on the applicability of MESMER-M to represent uncertainty in climate model projections resulting from internal variability, which is an irreducible model uncertainty and thus benefits from the significant reduction in computational costs provided by emulation. We chose to only use the

extreme climate scenario, SSP5-8.5, so as to demonstrate the applicability of the MESMER-M approach on the extreme end of climate response types. This however does not dismiss the applicability to more moderate climate response types displayed within other climate scenarios. From preliminary testing, we believe that the yearly to monthly temperature downscaling relationship will be to the most part scenario independent- although this is still to be fully investigated. The pattern scaling equation from Equation (1) has also been demonstrated to work across climate scenarios (see Beusch et al. (2021) (in review): <https://gmd.copernicus.org/preprints/gmd-2021-252/>) as long as a wider scenario space is given as training material.

To clarify this we will add our reasoning on focussing on SSP 5-8.5 in L59-60 as well as add some discussion points on the inter-scenario applicability of MESMER done according to Beusch et al. (2021) in the Conclusion and Outlook section highlighting the need for a representative training sample if inter-scenario exploration is pursued.

- 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?

We did look at inter-annual changes in skewness of monthly temperature distributions, which led us to include the yearly temperature dependent lambda parameter in the power transformer. Some discussion points within the text that arose from these inter-annual changes in temperature distributional properties can be seen in:

- L229-231: lambda values  $>1$  show a more positive shift in skewness of Northern-Hemispheric January temperature values (in line with other work e.g. Brodsky et al. 2020: <https://doi.org/10.1038/s41561-020-0576-3>, 2020)
- L232-234: The ice-free Arctic summer lead to a significant shift in skewness within Arctic July temperatures leading to a lambda parameter  $>1$

To furthermore investigate whether the emulator was able to capture such changes in distributional properties, we split the investigation of quantile deviations into the period 1870-2000 and 2000-2100. In this manner we show how well the emulator captures inter-annual changes in distributional properties for 2 periods experiencing different magnitudes of global warming. To be more specific, the quantile deviation calculations take into account the positioning of the actual ESM run within the emulated temperature distribution at each year, thus if the emulator was not able to represent the interannual changes in skewness we would see significant quantile deviations in the period 2000-2100 as compared to those in 1870-2000. Given that the quantile deviation values between the 2 periods correspond reasonably well, we believe that the emulator's representation of such interannual changes in distributional properties is sufficient. We can elaborate on the choice to separate the quantile deviations into the 2 periods in section 3.3.3 (i.e. it enables investigation of interannual changes) as well as mention the correspondence of the 2 period's deviations and what that entails in section 5.3.

Specific Comments:

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

The smoothed GMST is calculated using a locally weighted scatterplot smoothing (LOWESS), details of which can be found in the Beusch et al. (2020) paper.

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

We agree that this is confusing and will add  $m$  to the equation such that it is  $f_s(T_{\{s,y\}},m)$

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

We performed the BIC till  $n=8$ , however  $n$  greater than 6 was never chosen. We thus show an upper limit of 6 in the top panel's colorbar for Figure 2. For sake of simplicity we can however change  $n = 6$  in L101.

- 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)?

The mean response module focuses purely on the deterministic part of the monthly temperature response to evolving yearly temperatures. We therefore assume independence from the Gaussian noise when fitting for this part. Our method in using the BIC for choosing the order of the harmonic model follows a brute force approach where we calculate the BIC for the harmonic model fitted (using Ordinary Least Squares Regression) at each value of  $n$  and choose the value of  $n$  with the lowest BIC score.

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

We intentionally left out the noise term in Equation 4 as we are representing the time-correlation component (i.e.  $\eta_{\{m,s,y\}}^{\{time\}}$ ) of the AR(1) process that we use to represent overall variability ( $\eta_{\{m,s,y\}}$ ). The noise term is represented separately within the space component i.e.  $\eta_{\{m,s,y\}}^{\{space\}}$  (equation 6). The complete  $\eta_{\{m,s,y\}}$  is then the addition of the time and space components (combine equation 4 and equation 6) consisting of both lag-1 autocorrelations and noise terms. To make this clearer we could change L107 to read as

"Similar to previous MESMER developments, we represent the residual variability using an AR process. To produce spatio-temporally correlated noise terms, we delegate the noise component of the AR process as being space-dependent leaving the deterministic component as time-dependent, as shown in equation 3 (Beusch et al., 2020)."

And L116 to read as

"Following this, the noise component of the AR process, represented as  $\eta_{\{m,s,y\}}^{\{space\}}$ , derives its distribution based on spatial correlation structures and month-dependent skewness, expected to be non-stationary with regard to yearly temperatures..."

- 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."

We think this is a good suggestion and will change L118 accordingly.

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

We accept this suggestion and will add the equation for the Yeo-Johnson transformation before equation (5).

- L121, 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)?

We designed MESMER-M to follow a modular approach, such that each key element of the monthly temperature response to yearly temperature is accounted for in a separate, independent module. This allows for a simpler breaking down of such a complex downscaling problem, fulfilling the purpose of having lowered computational costs whilst allowing room for modular-wise modification. We thus delegated the representation of spatial correlations and month-dependent skewness to the noise term of the AR process used within the residual variability module ( $\eta_{m,s,y}^{\text{space}}$ ). In such, we sample spatially correlated terms from the transformed Gaussian space and then back-transform independently within ( $\eta_{m,s,y}^{\text{space}}$ ) before adding ( $\eta_{m,s,y}^{\text{time}}$ ). This design choice also seemed most appropriate as we train the month-specific, local Yeo Johnson Power Transformer to capture the covariation between monthly skewness and yearly temperatures. To this extent, first calculating the lag-1 autocorrelations between the raw residuals allows removal of any month-to-month processes remaining within the residuals, so that the power transformer can then derive the cleanest skewness to yearly temperature relationship.

- 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?

We are using a likelihood that assumes independence. Future developments could look into one that accounts for spatiotemporal dependence however this was not focused on in this study.

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

The Gaspari-Cohn function was used in previous MESMER developments (see Beusch et al. 2020, equation 8) and allows for exponentially vanishing correlations with distance such that anisotropy of spatial cross-correlations on regional scales is still retained. We chose it for consistency within MESMER as well as the aforementioned property. Since it was already elaborated on in the previous MESMER paper we do not go into its specifics. To make this clearer however we can add to L134

"... localized by point-wise multiplication with the smooth Gaspari Cohn correlation function (Gaspari and Cohn, 1999) which has exponentially vanishing correlations with distance  $r_m$ ."

- 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.

This is a good point and we will consider switching to looking at the representation of the top 50 highest frequencies instead.

- 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?

We take the regionally averaged temperature time series over which the percentiles are then calculated over time. We could elaborate on this by modifying L163 to read as

"The emulator is evaluated for its representation of regionally averaged monthly temperatures of all 26 SREX regions (Seneviratne et al., 2012) at each individual month (see Appendix B for details on SREX regions)."

- 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?

L237-240 point out that the localisation radii,  $r_m$ , will be larger with a larger number of ensemble runs. This follows from  $r_m$  being estimated using the leave-one-out approach, such that models with larger ensemble runs will have more years within their training set and thus more options to "leave-out" and calculate the likelihood over. Such encourages the selection of higher localisation radii as there is more evidence towards the spatial correlations existing up to higher distances. In contrast, models with fewer ensemble runs will prefer the simpler option and choose lower localisation radii due to limited options available to "leave-out" and calculate the likelihood over, while lower values are bounded by zero. We acknowledge that the determining factor for the localisation radius chosen is the model itself, however we mainly wanted to indicate that in coming closer to a localisation radius true to the model properties, the number of ensemble members could also be a bottleneck i.e. with less training material the model will settle for the simpler option.

- 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?

We interpret places where there are higher magnitudes of autocorrelations as areas where there are other month-to-month processes that are not fully covered by the mean response model. This is not necessarily a shortcoming on the side of the mean model but may just mean that there are additional month-to-month correlations arising from processes that do not directly follow from changes in yearly temperatures but for instance, are even amplified/dampened.

- 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.

The figures may be misleading as the number of emulations are much higher than number of ESM runs and we will correct this so as to show the emulations using a 2D histogram.

The overall variability representation can be deduced from the quantile deviation plots, where we have a typical problem of underdispersivity within the emulator due to localisation of spatial covariations (See Beusch et al. (2020) where this problem was

explored too).

Thanks for raising the mismatch between the MPI-ESM1-2-LR runs and their emulations in WAF. We had investigated this (not discussed in the paper), and strongly suspect that this is related to the dynamic vegetation scheme in this model. Tree fraction strongly decreases (by ~30%) in that model between 2030 and 2050, and we concluded that the resulting climate changes driven by biogeophysical changes in albedo and surface heat fluxes dominate the ESM runs during this period. We will add a short discussion on this at the end of this section.

- 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.

We acknowledge the problem in under-representing skewness and believe it arises from limited degrees of freedom (i.e. the power transformation solely relying on yearly temperature) to capture skewness. We discuss this in L305-310, later on in the paper,:

“To be exact, even though changes in skewness in local monthly temperature residuals with evolving local yearly temperatures is incorporated, such cannot completely represent delayed or compound effects, otherwise brought about by secondary/external feedbacks (Shinoda,2001; Potopová et al., 2016), which more largely affect extreme quantiles (Huybers et al., 2014; Guirguis et al., 2018; Loikith and Neelin, 2019).”

We can perhaps move this discussion up to after Figure 4 as it may be most relevant there given the strong visual evidence. Since the change in monthly temperatures over WAF may be dominated by other effects such as those arising from changes in land cover (see discussion in comment above), this is furthermore a strong example of the shortcomings of solely using yearly temperatures as input for inferring changes in distributional properties (this can also be added to the lines above).

Having noted these shortcomings, we have included ideas for further emulator developments within the Conclusion and Outlook section (e.g. develop a LCLM mean sub-module), but to directly link it to the problems seen in Figure 3 and 4 we can preface the suggestions with a reiteration of the shortcomings in the mean response and variability module due to yearly temperature being the sole input.

- L268 - 276 and Figure 6. I don't understand whether you are saying that you underestimate the spectrum when the power is high, or that you underestimate 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.

We mean that we underestimate the spectrum when the power is high. As a response to comment 10 however, we will in future look at the spectrum of high frequencies and will clarify this within the figure caption. The suggestion to plot ratio of spectra vs. frequency is also a good idea which we will implement.