Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2019-34-RC1, 2019

Earth System Dynamics Discussions

Interactive comment on "Emulating Earth System Model temperatures: from global mean temperature trajectories to grid-point level realizations on land" *by* Lea Beusch et al.

Anonymous Referee #1

Received and published: 6 August 2019

General comments

The authors propose a statistical model for emulating output from Earth System Models (ESMs). The model is composed of deterministic and stochastic components that are intended to capture the forced trend and variability, respectively.

There is clearly a lot of work to be done in developing cheap tools like emulators to get more information from our climate model archive, and I am glad to see another contribution to this field. However, I have a number of concerns about the model formulation and, echoing Comment 1 from Robert Link, the validation of the emulator output.

Specific comments

1. One of the challenges of fitting emulators to data or climate model output is separation of the forced and internal components (under the common assumption that they are linearly separable). The authors propose the use of a common approach of regressing onto a smoothed version of the global mean temperature (plus volcanic bursts), but do not provide evidence that this approach is successful. The method can and should be tested within one or multiple initial condition ensembles.

2. The spatial model for the innovations is presented with minimal justification. How was the exponential covariance model chosen versus one that is smoother in space? More importantly, given that the spatial structure of temperature variability depends on the prevailing wind direction, the geometry of the coasts, land surface type, etc., is an isotropic covariance model even appropriate?

3. Identifying parsimonious but sufficient metrics for validation of model ensembles is a challenging and unsolved problem. However, the authors are too qualitative in their evaluation of their emulator skill, which is composed primarily of visual inspection of emulated fields and plots like Figs. 9 and 10. Given the choice of spatial model discussed in (2), it would be helpful to see validation metrics on both the spatial and temporal correlation structure. In addition, the assumption of Gaussianity is built-in but never checked. Finally, validation metrics should be provided with respect to a reasonable null hypothesis, otherwise it is difficult to assess whether a certain error value is meaningful. For example, how large would a given error metric be if different realizations of an actual ESM were resampled, and then the metric of interest was calculated?

4. The writing could be improved to make the manuscript flow more smoothly. In particular, Section 2 could be reworked to more clearly identify what is missing in the current literature that the authors aim to ameliorate with this manuscript.

5. Lines 437-439 make strong statements about replacing single model ensembles with emulators such as the one proposed. Without further validation, I don't think the

authors can say "the latter can be readily mimicked by our emulator based on a single ESM run."

Technical corrections/minor points

1. There are minor grammatical and spelling errors throughout.

2. In the discussion of the forced component, the authors should additionally reference the various LIM-based methods (e.g. Frankignoul et al, 2017, Estimation of the SST Response to Anthropogenic and External Forcing and Its Impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation), signal to noise maximizing EOFs (e.g. Ting et al., 2009, Forced and Internal Twentieth-Century SST Trends in the North Atlantic), and low frequency component analysis (e.g. Wills et al., 2018, Disentangling Global Warming, Multidecadal Variability, and El Niño in Pacific Temperatures).

3. The citation of McKinnon and Deser (2018) is slightly misleading. The longer timescales related to coupled modes are explicitly modeled, such that the remaining variability has near-zero memory, and so can be block bootstrapped.

4. I was somewhat mystified by the comment on Line 378 that CMIP5 models do not reproduce the large-scale temperature response to atmospheric waves, which is incorrect. Any reasonable atmospheric model produces Rossby waves and is reasonably accurate at simulating the temperature response.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2019-34, 2019.