

Interactive comment on “Computationally Efficient Emulators for Earth System Models” by Robert Link et al.

Robert Link et al.

robert.link@pnnl.gov

Received and published: 7 September 2018

1 General comments

- However, the language used is quite mathematical for a GMD paper. I think this could be addressed without loss of quality or conciseness.

When preparing the final draft of the manuscript, we will look for opportunities to reduce the density of the mathematics in the text. We do note, however, that it was important to us to provide enough detail for readers to be able to both recreate and evaluate the algorithm for themselves, if desired, and doing so requires a certain amount of mathematical specificity.

C1

- Also, as suggested by the first reviewer, this is not an emulator in the strict sense.

It would seem that there is some diversity in the way this terminology is used in different scientific communities. Amongst the researchers who develop these kinds of models the term “emulator” seems to be preferred; therefore, we have elected to keep to that convention.

- I also agree with the first reviewer in that, ESM outputs are not “observations”. “ESM outputs” would suffice.

We have adopted the first reviewer’s suggestion of “synthetic measurements” to refer to the data being used to train the model.

- A related point is that the model simulates global mean surface temperature from GCMs (general circulation models/global climate models - choose your favourite acronym) rather than ESMs. The CMIP5 definition of an ESM includes an interactive carbon cycle, going from emissions to concentrations to forcing to temperature. GCMs skip the emissions step, running from prescribed concentrations that have been calculated from a simple model, e.g. MAGICC, as they were in CMIP5.

There is nothing in the model that is specific to GCMs as contrasted with ESMs. The particular input data we chose to use as a demonstration were forced by concentration, but we could equally well have selected archival datasets that were produced with the carbon cycle turned on. Since the developers of CESM refer to their model as an “earth system model”, we chose to do the same, even when working with scenarios run in a mode more characteristic of a GCM.

C2

2 Specific comments

- In the introduction, the application of the model to extreme events is given as a justification for its creation. However, the model only produces annual mean temperature output in each grid cell. I am not aware of an extreme indicator that uses annual mean temperatures. Such indicators are usually calculated from daily climate model output (see Zhang et al 2011, 10.1002/wcc.147). This would be a natural extension to this model, but in its current form it is not capable of analysing "extremes" in the usual sense.

This is a good point. By "extreme events" we had in mind the tails of the distribution of annually averaged values. We will adjust the language in the final draft to clarify what we had in mind.

- I don't disagree with the authors about the notation convention: I understand the broadcasting concept used in their convention and agree it aids readability. I do find it hard to follow the equations though. If we have $|T_g\rangle = \mathbf{O}|\lambda\rangle$, then this suggests to me that $|T_g\rangle$ is a column vector of shape 855×1 formed by multiplication of \mathbf{O} (855×55296) by $|\lambda\rangle$ (55296×1). In eq(2) you have $T_g|w\rangle + |b\rangle$. Is T_g (not bracketed in eq(2)) times $|w\rangle$ a column vector times a column vector? ? How is this defined?

T_g (without brackets) is a scalar. On the other hand, $|T_g\rangle$ is a vector of global mean temperature values. When defined by the first equation in the quote, this vector is made up of the values of T_g for each year of each model in the input set. In other words, the name of a variable tells us what physical quantity the variable represents, and the decoration tells us how many we have and what kind of structure they are organized into. We will add some clarifying remarks on this point to the notation section.

C3

- And then in equation 3, there is $|T_g\rangle$ (a column) times $\langle w|$ (a row), which I think is 855×855 , then added to $|b\rangle$ (855×1)? and subtracted from \mathbf{O} (55296×855 - but how is this broadcasted?) If there are no typos in these equations, it would be helpful here to put in a diagram of the matrix dimensions in the equations 1 to 3.

The symbol $\langle w|$ is a row vector, with dimension 1×55296 (i.e., one value for each grid cell). The product $|T_g\rangle\langle w|$ is an outer product, the result of which is a matrix $(855 \times 1) \cdot (1 \times 55296) = (855 \times 55296)$. The vector $|b\rangle$ likewise has dimension (55296×1) (again, one value for each grid cell). Because this matches the number of columns in the matrix formed by the outer product, it can be broadcast in the usual way. The result is still a matrix (855×55296) , which is conformant with the matrix \mathbf{O} that it is being subtracted from.

We will clarify the dimensions of the vectors of pattern scaling coefficients, and we will add a figure that shows how these quantities fit together to produce the final matrix of residuals.

- σ values in table 1 and p5 line 9. I think these are the singular values of \mathbf{R} , but it is not really explained what these are or what they mean. This paragraph could do with some expansion of the key terms (rank deficient, discrete Fourier transform). Does dropping EOFs where $\sigma < \sigma_{threshold}$ guarantee full rank?

The σ are the singular values; we will clarify this in the final draft. We will also provide a brief explanation of what the singular values mean, and we will supply a reference to an approachable introduction to Fourier transforms and their applications.

Technically, having all $\sigma > 0$ is enough to guarantee full rank, so it would be more correct to say that the problem here is ill-conditioning, rather than rank deficiency. However, because we do not use the SVD to invert the matrix (only to find the

C4

principal components), it is not clear that the ill-conditioning causes any particular harm. Therefore, in the final draft we will regard the dropping of components with very small singular values as an implementation detail and omit the discussion of it in the text.

- Section 3: Can the four images in figure 1 be interpreted as ensemble members? If so, it would be good to state this.

Yes, they can. We will comment on this in the final draft.

- figures 4-6 and associated discussion in lines 24-28 on page 6: The periodic variability in EOFs 2, 3 and 5 - could these have a physical interpretation? For example there seems to be an El Nino style feature in EOFs 3 and 5. On the other hand, is there any evidence that the lower EOFs are not just noise?

We, too, had noticed the resemblance to El Nino in those components; however, it wasn't clear how to make a rigorous comparison between the patterns we see here and real-world El Nino events (since surface air temperature isn't really the right variable for computing a proper El Nino index). Developing a methodology for making such a comparison is outside the scope of this paper (though it would be interesting research in its own right), so we decided to characterize these components generically as periodic modes of variability, rather than to attribute a physical cause to them.

We feel very confident that the lower EOFs (we assume that by this you mean the ones with lower total power, not the ones earlier in the sequence) *are* mostly noise. In the time dimension their power spectra are almost completely flat, and the length scale of spatial correlations is just a few pixels. This is pretty much the definition of "noise" in this context. That said, there is still *some* structure, even in these noisy basis functions, and characterizing that structure with this

C5

model allows us to ensure that the noise in the output realizations has the *same* structure.

To put it another way, you could probably get an adequate representation of the noise in the system by just applying a random perturbation (i.e., without regard to space or time correlation) and then running a smoothing kernel over the result so as to reproduce the short-range correlations observed in the noisy components. But, what should be the width of that kernel, and how should that noise field be weighted relative to the structured components? Those things are an important part what we are trying to model with this technique.

- Section 4.2 got me thinking that as the model is trained on the RCP outputs, is there any difference in the results when taking just the set of realisations from RCP2.6 and RCP8.5? Certainly across ESMs, the variance across models increases with increasing global mean temperature. It would therefore not be correct to use a variability model that is trained on RCP8.5 for low forcing scenarios or those with a peak and decline. I note the authors address this in section 4.3, but I wonder if they have tested this.

We have worked with models trained on a single scenario, and for the most part the results are qualitatively similar to the multi-scenario results. We didn't try to run any statistical tests to detect differences, but with the limited amount of data available it seems unlikely that any such differences would be detectable. Therefore, although it's theoretically possible that by using variability from a model trained solely on a scenario of interest (supposing you know in advance what scenario that is) you might get more accurate results for that scenario. However, in practice the difference is likely to be small and perhaps offset by the effects of having less data to train on. Many of these topics would be worth revisiting in the future, particularly once improvements in the mean field response are in place.

C6

3 Technical Corrections

- page 5, line 3: allow → allows
Thanks. We will correct this.
- page 6, line 3: 143 seconds. What is the machine architecture here?
This was on a midrange workstation. We will mention this in the final draft.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2018-60>, 2018.