Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-60-AC2, 2018

Geoscientific Model Development Discussions

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 Major comments

1. Please don't use 'emulator' at all in your manuscript. What you have implemented is a 'surrogate model' or 'metamodel'. An emulator is a type of surrogate model / metamodel which is an interpolator and gives a probability distribution for outputs corresponding to inputs it is not trained at. See [1] for details. I understand that others in the earth system modelling community have used 'emulator' in the same way as you, but whoever started using it first and passed down this definition is incorrect for doing so.

The reviewer's point is well taken; however, as the reviewer points out, the term "emulator" is firmly entrenched in the earth system modeling community. Moreover, it is the term used for this kind of model in other papers in this journal. Therefore, we have elected to stick with the terminology that is customary among our target audience.

2. Your literature review in the introduction is very limited. This would be fine if the journal was very narrowly focused, but GMD has a very broad appeal. Statisticians from the UQ (Uncertainty Quantification) community have, for example, done a lot research on surrogate modelling methods for very expensive models like an ESM. Names that come to mind are Nathan Urban (Los Alamos), Jonty Rougier (Univ. of Bristol), Michael Goldstein (Durham univ.). A series of workshops were held in Cambridge earlier this year (http://www.newton.ac.uk/event/unq/workshops) which will help you find statisticians working in this field. Non-statisticians like David Sexton at the Met office in the UK are also working on quantifying uncertainty of ESMs. There are probably lots of other research that you can also mention, but above is just a start.

When preparing the final draft of our manuscript, we will broaden the discussion of related literature.

3. Please use more standard mathematical notation for defining vectors and matrices. In particular, please don't use $|x\rangle$ and $\langle x|$ for column and row vectors. I have never seen this notation being used before. It is much more standard to use x for a column vector and xT (i.e. the transpose of x) for a row vector.

In the final draft we will replace the Dirac notation currently used in the manuscript with $\vec{x}$ and $\vec{x}^T$.

4. Please also stick to normal conventions for matrix and vector algebra, e.g. page 2 line 28 you state "Occasionally we will add a matrix and a vector e.g. $B = A + |x\rangle$. Please do not do this! I know that you explain what you mean when you do this, but mathematically it is not the correct way of doing things because you're effectively defining $|x\rangle$ to be a vector sometimes and a matrix and other times. If you want to add a matrix with a vector in the way you describe then define a new matrix which has as its columns or rows the vector you want. This is a much clearer way of defining things.

In all cases $\vec{x}$ (formerly $|x\rangle$) is a vector. Writing $\mathbf{B} = \mathbf{A} + \vec{x}$ is merely a shorthand way of saying $\mathbf{B} = \mathbf{A} + \mathbf{M}(\vec{x})$, where $\mathbf{M}(\vec{x})$ is the outer product of $\vec{x}$ with a suitably-sized vector of ones (i.e., $\mathbf{M}(\vec{x}) = \vec{1}\vec{x}^T$). Making this operation explicit doesn't make the discussion any clearer; quite the contrary, it forces the reader to stop and figure out what the new operator actually does. Likewise, defining a new matrix each time we need to perform this operation proliferates symbols unnecessarily and obscures the relationship between the vector and matrix versions of the same quantity. For these reasons, we have elected to keep the broadcast notation.

5. Page 3, line 8. You state "We refer to the ESM data as "observations" . . . ". NO, NO, NO! Please do not do this. I spent about an hour reading your paper thinking you were using real observations and then I realized I had not properly read this very important line in your methods. If you want to use your ESM output as measurements then call them "synthetic measurements". Everyone knows what this is. But if you just state measurements or observations, we think you're using the real thing.

In the final draft we will refer to the ESM output as "synthetic measurements".

6. There are lots of subject specific words / jargon used here which you assume that the reader knows the meaning of because you don't define them. Examples include: 'linear pattern scaling', 'discrete Fourier Transform', 'randomizing the phases of G', 'spatial coherence', etc . . .

In the final draft we will include explanations of any terminology that might be unfamiliar to a broader scientific audience.

7. For things that you define you don't give enough detail. For example, with EOFs (statisticians use principal components but it means the same thing) you need to say that we normally only choose the first n EOFs where n is determined such that most of the variance (or power as you mentioned) is explained. Only when we get to the results section do you talk about the number of EOFs that you're using.

We thank the reviewer for pointing out this important omission. Although principal components are often used for dimensionality reduction, in this case we are using them solely to diagonalize the covariance matrix. The final draft will include an explanation of this difference.

8. I was also completely lost with section 2.4. You seem This seems to be key part of the methods, so really needs to be explained better. You make statements that make no sense to the non-specialist – e.g. in the lines prior to equation 7 you state that you're making some minor modifications to the standard procedure, but why? And what is a zeroth basis vector and why are you defining it in this way. These things I read and go "okay", but I have no idea why you're doing this. This is just one example of many scattered throughout the manuscript.

The purpose of including it is explained in detail in the paragraph following equation 7. ("This property guarantees that . . . . This property is useful because . . . .") We will move this paragraph up before equation 7 so that the differences between this basis vector and the others and the motivation for including it at all are clear before we begin describing how it is calculated.

9. There's no motivation or justification for doing what you're doing. At the end of the introduction, you state that there's a need to quantify the uncertainty in ESM output. Fine. But then you jump straight into your approach of quantifying this uncertainty by generating it based on an error covariance matrix that is derived from defining the temporal and spatial correlation of the ESM output. Why is this a good way of defining the ESM uncertainty?

Our paper is not about quantifying uncertainty in ESM output at all. Instead, it is about providing a source of data, beyond what is available in public archives, for models that are *consumers* of ESM output. Uncertainty studies (in these models, not in the ESMs themselves) are just one reason why we might want to generate these datasets (we give two others in the introduction). In the final draft the opening paragraphs of the introduction will be rewritten to clarify these points.

10. Following on from the previous point, this seems to be a major flaw in this paper. Normally when we do uncertainty analysis, we propagate the uncertainty from the inputs (e.g. uncertainty in model parameters) through to the uncertainty in the ESM outputs. Instead, you seem to be using the spatial and temporal correlations of the ESM output as a means to generate the uncertainty in the ESM output with the metamodel (or "emulator" as you call it) as the vehicle for carrying out the extra ESM runs. Perhaps I'm mistaken and this isn't what you're doing. If I am mistaken then the fact that I have misunderstood this is a major problem. If you want people to be interested in your research, you first need to communicate it clearly and (sometimes) simply to them.

As we noted above, the purpose of this paper is not to study uncertainty in ESM

outputs. It is important to realize here that the variability in ESM outputs is an important contributor to the uncertainty in downstream models *and would still be so even if there were no parameter uncertainty at all in ESMs*. Indeed, for many downstream models the uncertainty contribution from variability is *more* important than that from parameter uncertainty. As we discuss in the introduction, many ESM emulators do not produce this variability, or produce it with significant limitations. The purpose of this work is to produce an emulator that produces the full range of variability seen in the ESM data and does so in a realistic way.

The revised manuscript will explain these matters in more detail than the current manuscript.

11. I am confused with how you train your metamodel (or "emulator" as you call it). When I train metamodel, this often involves multiple runs of the computationally expensive model (i.e. the ESM in your case). In your paper, it seems that you just need one run of the ESM to train the metamodel. Is this correct? Again, this isn't how we normally train metamodels so you need to be really clear about this. In fact there should really be a whole section in the methods explaining everything about how you constructed the metamodel. Maybe you feel what you've written is enough, but I'm just generally confused so you need to lay things out much more logically and clearly at the very least.

In section 2.2 (p.3, l. 5–7) we wrote:

> We used surface temperature data from all available 21st century runs for all four Representative Concentration Pathway (RCP) emissions scenarios (RCP2.6, RCP4.5, RCP6.0, and RCP8.5), for a total of 9 runs, each 95 years in length.

12. Table 1 makes little sense to me. If I really concentrate I can probably understand what's going on, but you could help the reader by using

less complicated words or phrasing it in a simpler way.

Table 1 is a summary of the steps in the algorithm described in the rest of the section. Although the reviewer appears not to have found it useful, other people we have shown it to have said that the summary helped them visualize how the pieces of the algorithm fit together. Therefore, we are inclined to keep it. In the final draft of the manuscript we will expand the table caption to clarify that this is a summary of the material in the rest of the section.

13. The results section seems to be too short (less than one page). Most papers I read have at least 3 or 4 subsections within the results section. These subsections have their own titles and help navigate the reader through the different aspects of the results. At the moment, the results just seems like a list of things. Your results show [sic] flow more like a story. You also don't really give a lot of detail, e.g. just one sentence for figure 5? What's the point of having it in there?

In dividing the paper into sections, we construed "Results" narrowly to mean "the artifacts produced by running the model". Conversely, we categorized analysis of the properties of the model output as "Discussion", which is why that section comprises 10 of the paper's 17 pages, with three subsections and two subsubsections. In the final draft we will merge the two sections into a single "Results and Discussion" section.

In addition, the final draft will describe some of the salient features in the map figures.

14. I didn't really read through the discussion in depth, but in section 4.1.1 (and may other subsections) you describe further results that were carried out. Thse should be in the results section. The purpose of the discussion section is: (a) to give an explanation for why your results look the way they do; (b) to put your results in context of other

comparable studies. I see a bit of (a) in the discussion but no evidence of (b).

As explained above, we appear to have a different convention than the reviewer regarding the distinction between "results" and "discussion". In the final draft the two will be merged into a single section.

2 Minor comments

- Panels of figures: label them with letters. E.g. Figure 1a would refer to panel at the top left of figure 1.

These will be added in the final draft.

- Figures: captions lack enough details

The captions will be expanded in the final draft.

- When submitting for review, it's more helpful to put all the figures and tables at the end of the manuscript. This makes it easier for the reader to refer to a particular figure when reading a particular part of the results.

Opinions differ on this; many scientists prefer in-text figures and tables. The GMD author guidelines leave it up to the author.