Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-276-RC1, 2018

Geoscientific Model Development Discussions

Interactive comment on "The Climate Generator: Stochastic climate representation for glacial cycle integration" *by* Mohammad Hizbul Bahar Arif et al.

M. Crucifix (Referee)

michel.crucifix@uclouvain.be

Received and published: 16 March 2018

1 Summary

The purpose of this article is to provide the rationale and documentation for a stochastic *climate generator* developed by the authors. The climate generator is stastical model based on a Bayesian neural network architecture, which provides gridpoint-wise climate outputs from a tuple of predictors (their Figure 1) which includes elevation, ice cover, greenhouse and orbital forcings, as well as the temperature predicted by an energy balance model. The model is trained on a series of experiments with two general circulation models: FAMOUS and CCSM. The purpose of the *climate generator* is to be coupled with an ice sheet model to simulate glacial-interglacial cycles. The

assessment of this *climate generator* is based on a series of metrics and benchmarks, which the authors present as a "Turing Test". The assessment relies mainly on the comparison of the model outputs (trained on FAMOUS) to simulations with FAMOUS for configurations in the training set.

2 Overall comments

2.1 Justification of the model

The 'Climate Generator' may actually be seen as a sophisticated machine-learning-based correction of energy-balance-model (EBM) outputs. Compared to other meta-modelling strategies proposed so far uses (as input) the the ice boundary conditions provided grid-point-wise. It does not need to rely an ice volume index aggregating the state of ice sheets at the global level.

The purpose of the introduction is to justify this effort, and it may fall a little bit short of this.
On the one hand, there is a dissonance between the claims of the introduction and what the paper actually offers. For example, the introduction explains that "Regression-based methods are relatively straightforward [. . .] but have an inadequate representation of observed variance and extreme events". Does the climate generator address these shortcomings (see below, remarks on variability)? The authors also explain that weather generators "replicate the statistical attributes of local climate variables [reviewer note: wouldn't it be *weather*?] rather than the observed sequence of events". In what sense would the "climate generator" *replicate* the statistical attributes of "climate" variables? Later in this review, I raise some questions about the lack of consistent representation for decadal or centennial modes of variability. Isn't it indeed what one would have expected from a "climate generator" ? Finally, more needs to be said

about how the climate generator will actually contribute to the grand challange of deciphering the mechanisms of glacial-interglacial cycles (if this is indeed what the authors are after).

2.2 The methodological implementation

2.2.1 BNN architecture

Crafting a BNN architecture is an art which, clearly, the authors master much better than this reviewer. The paper is thus an opportunity to introduce the reader to that art. We need to understand better the elements that makes the BNN designer opt for a certain architecture, when the objective is to emulate a climate model. As its stands the article features different architectures and compares their performance, but the main message is somehow diluted into the numerous tables and graphics, some of which are a little bit cryptic. In the end, I was left with what seems to be the key questions: why the authors originally opted such or such architecture, and how one could interpret the fact that some architectures seem to be working better than others? The authors should read this criticism as request for less rather than more content in the main article (authors are still free to put long tables in S.I.). At the end of the article, the reader should be satisfied that she understands the critical aspects in the development of BNN models for climate simulators, in order to be able to replicate the effort possibly with other climate simulators.

2.2.2 Modelling variability

The hypothesis (which the authors present as an assumption or approximation) that the predictive uncertainty of the BNN is in good part "due to the internal variability of the GCM climate" is certainly the most controversial technical aspect of the article.

There is in fact little to give substance to that hypothesis. As I understand it, the BNN (associated with the EBM) is only but an approximation of the GCM, and it is very unclear why the misfit would actually match the GCM internal variability. To mitigate this criticism, it must be observed that a regression model calibrated on GCM experiments may sometimes predict the model stationary mean so well, that the misfit between the predicted mean and a specific GCM experiment is mainly due to the finite sample length of the GCM experiment. Some have seen this before (Araya-Melo et al., 10.5194/cp-11-45-2015). However, turning this 'misfit' into a model for interannual or interdecadal variability requires some careful thoughts and discussion. The fact that, in the validation procedure, the difference between FAMOUS and CCSM is further used as an upper bound for the climate generator misfit (and thus, if I understood correctly, the acceptable level for self-inferred uncertainty) makes it even more confusing.

On the other hand, the modelled climate variability is assumed to be spatially uncorrelated, a point over which the authors do not seem to worry much about because ice sheets would integrate perturbations over long times. This defence is disputable. Spatial variability patterns such as those active in the ocean at the decadal and centennial scales may have specific and interesting consequences for the development of ice sheets. As suggested in the introduction of the present review, one could have argued that a *climate* generator should actually be a model which, compared to a weather generator, features carefully the structure of spatio-temportal decadal and centennial variability modes. For the same reason, the choice of a Gaussian noise is *a priori* arguable (as it may not represent the important consequences of extreme events).

2.2.3 Training

The authors describe the training procedure but say little about how the training can be made efficient. How many experiments are needed, how the choice of training simulations could be optimised, and to what extent the test cases used for the validation of the climate simulator are convincingly / sufficiently independent of the training experiments.

2.3 Turing test

With regret I must confess that I found the use of the “Turing test” phrase more obfuscating than enlightening. The Turing test was imagined in the context of artificial intelligence, where self-reference and induction are important concepts. In the present context the evaluation is essentially a benchmarking process, the output of which could best be summarised with a colorful table witch checkers, crosses, indicating clearly which criteria are met and which are unmet.

3 Other comments

- The tables and graphics are not always self-explanatory (especially the Taylor diagrams). Honestly, I found that this was a hard paper to read. Some editorial work is needed to streamline the paper.
- Section 2, which is currently oddly structured (one page of material at the section level, then section 2.1 subdivided into four subsections) could be more focused, with less emphasis on meta-digressions such as “reasoning behind the name of CG”. I suggest that both more impact and a better reading experience could be achieved by focusing on what is needed to simulate glacial-interglacial cycles, and how the stochastic climate generator will contribute to our understanding of glacial-interglacial cycles.
- The material under section 4.2 could best be illustrated with a flowchart. Make sure that all concepts are sufficiently explained to the non-experts (e.g. “hybrid

Monte Carlo sampling”). Among others, clarify concepts such as “sub-group of parameters” and “shape parameters”.
- p.8 : How do you “get an idea” of when the simulation has reached equilibrium. Can’t you use a formal criteria?
- p. 14 is quite descriptive. More effort could be given to extract the key message. The interest of the numerous tables should also be reconsidered.
- Insolation means “Incoming solar radiation”. Avoid “solar insolation”. In passing, why are the authors using 60°N insolation, rather than insolation predictors ($e \sin \varpi$, obliquity) which are not bound to a specific latitude?

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-276, 2018.