Geosci. Model Dev. Discuss., 7, C610–C614, 2014
www.geosci-model-dev-discuss.net/7/C610/2014/

Geoscientific Model Development Discussions

Open Access

GMDD

7, C610–C614, 2014

Interactive Comment

Interactive comment on "Probabilistic calibration of a Greenland Ice Sheet model using spatially-resolved synthetic observations: toward projections of ice mass loss with uncertainties" *by* W. Chang et al.

D.J. McNeall (Referee)

doug.mcneall@metoffice.gov.uk

Received and published: 22 May 2014

Introduction

Chang et al. (2014) set out to create a method to make probabilistic predictions of future ice sheet behaviour, given a comparison of output from an ice sheet simulator and observations of the real ice sheet. They use up-to-date statistical techniques, applied to an ensemble of a computationally efficient ice sheet simulator. They stop short of making probabilistic predictions of the real system, instead testing their methodology using synthetic data in a perfect model experiment.

The paper is a good and carefully worked example of a modern approach to probabilistic projection using complex simulators in the Earth sciences. While much of the groundwork is published elsewhere, this paper coherently develops, applies and tests these methods in a new and useful sphere. I have no hesitation in recommending it for publication in GMD, subject to minor corrections, and a single, conceptually simple extension of the testing regime, outlined below.

Main points

The goals of the research are stated by the paper thus:

"(1) to identify a method for quantifying the agreement between ice sheet model output and observations that incorporates spatial information,

(2) to characterize the interactions among input parameters, and

(3) to produce illustrative projections of sea level rise from the Greenland Ice Sheet based on synthetic data."

The paper succeeds in its own terms. Goal (1) is achieved, using a likelihood function and a well constructed principal-components emulator, for high dimensional observations - crucially, incorporating a sensible discrepancy function. Goal (2) is achieved through visual inspection of pairs plots of the 2 dimensional likelihood surface for parameters. Goal (3) is achieved by projecting likely parts of parameter space through the 21st Century, using a separately built emulator.

The paper stops short of making probabilistic predictions of the real ice sheet, but this is appropriate given the scope of GMD as a model and method development journal.

The paper is clearly written, and provides broadly appropriate mathematical detail of the emulator in the supplementary material. A slightly more comprehensive summary of the emulator technique in the main text would be welcome.

GMDD

7, C610–C614, 2014

Interactive Comment

The testing methodology of the emulator is mostly appropriate: testing the probabilistic methodology in both “forward” projection (of the ice sheet) and “backwards” (of the parameter space) is a nice touch. However, the demonstration of robustness of the major result of this this paper - that the probabilistic prediction methodology works well - would be greatly advanced if the testing of the probabilistic framework were more broadly applied across the ensemble. At present, the paper presents a “leave-one-out” perfect model test, where only a single ensemble member is “left out”! Such a test is more generally known as “leave-one-out cross validation”, where each member is left out in turn.

The supplementary material does use two more (extreme) ensemble members for testing the probabilistic methodology. However, there is a real risk that the authors “got lucky” with some well behaved ensemble members. Simply looping the leave-one-out process across the ensemble (as in e.g. McNeall et al., 2013, already cited in the text) would provide a more comprehensive test of the methods. Even better, would be to hold out a larger test set of ensemble members, as in e.g. Carslaw (2013), in order that there is not too much similarity between each training set.

This would allow a more quantitative assessment of the correctness of the probability distributions assigned to the parameter sets. Does the probabilistic method calibrate well? There are many examples of appropriate measures of this in the statistics or meteorology literature. Are the ensemble members nearer the edge of the ensemble subject to greater error, or more difficult to emulate? With such a small testing set, only a qualitative assessment of the methodology is possible. I would suggest acceptance of the paper, conditional on a test set of ensemble members of a sizeable fraction of the ensemble - perhaps at least a third, chosen at random. If computational effort is not a consideration, I would recommend a leave-one-out or leave-n-out test across the entire ensemble.

Minor points

There appears little justification of the use of 10 PCs in the emulator. What procedure was used to choose the 10 PCs, and why was 10 chosen as "good enough"?

The "future work" section describes the aims of the authors to extend the methodology to the full two-dimensional thickness map of the ice sheet, rather than the one-dimensional thickness profile. Given the apparent availability of model data (as compared to observations), why did this work use only thickness profiles?

Clearly, leaving a discrepancy term out when discrepancy was added to the synthetic data, will result in a mis-specified probability distribution for the input parameters (and subsequent predictions of the ice sheet). The authors have missed a trick here - It would be very useful to show, comprehensively across the ensemble, how much error a mis-specified discrepancy term adds to predictions. It might also be worth demonstrating how much uncertainty a well-specified-but-uncertain discrepancy term adds to the predictions, and to the identifiability of the input parameters.

Figure 1. could show the entire ensemble (perhaps greyed out), and highlight the subset of ensemble members.

The accuracy of the emulator as demonstrated in figure 2. is impressive. Again, it would be useful to show how this varies across the entire ensemble. There are ideas for doing this using similar PC emulation techniques for one dimensional data in Challenor et al (2010), and McNeall (2008).

If the authors are to extend the testing of the probabilistic methodology across the ensemble, a graphical representation of the strength of interactions between parameters - summarised across the entire ensemble - would be most welcome. The pairs plots as used show this nicely for a single ensemble member, but are not appropriate for large ensembles.

Carslaw KS; Lee LA; Reddington CL; Pringle KJ; Rap A; Forster PM; Mann GW; Spracklen DV; Woodhouse MT; Regayre LA; Pierce JR (2013) Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, pp.67-71.

Challenor, P., McNeall, D. and Gattiker, J. (2010) Assessing the probability of rare climate events. In, O'Hagan, Anthony and West, Mike (eds.) The Oxford Handbook of Applied Bayesian Analysis. Oxford, GB, Oxford University Press, 403-430. (Oxford Handbooks in Mathematics).

McNeall, D.J. (2008) Dimension Reduction in the Bayesian Analysis of a Numerical Climate Model. Phd Thesis, University of Southampton.

Interactive comment on Geosci. Model Dev. Discuss., 7, 1905, 2014.