Geosci. Model Dev. Discuss., 6, C1395–C1399, 2013
www.geosci-model-dev-discuss.net/6/C1395/2013/

Geoscientific Model Development Discussions
Open Access

Interactive comment on "The potential of an observational data set for calibration of a computationally expensive computer model" *by* D. J. McNeall et al.

D. J. McNeall et al.

doug.mcneall@metoffice.gov.uk

Received and published: 4 September 2013

All page and line references are to final discussion paper version.

General changes and updates

pg. 2371 l9. "This begs the question ..." changed to "This raises the question ...", due to incorrect use of the former.

pg. 2388 l11. We have rephrased this paragraph slightly, for clarity. The word "model" has been replaced with "simulator" where appropriate in the text, to distinguish computer model from statistical model.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

Affiliation of EJ Stone changed to School of Geographical Sciences (was “Department”).

There have been a small number of punctuation, spelling, and grammar corrections.

Review 1 M. Crucifix

Comment: First a remark on the form: the authors have chosen to make a use of the active form (e.g.: "we use this", "our metric") that is slightly more assertive than standard in the scientific literature. Whether this should be corrected is left as a decision of the editor.

Response: We believe that the form is within the range of scientific standards, appropriate to GMD, and have therefore only changed the text in a minor fashion.

Comment: p. 2371: There might me some semantic argument to be had about the definition of the words calibration and tuning. Here the authors define tuning as a form of point estimate (best matching) calibration. In other contexts, calibration and tuning refer to different processes, calibration implying the existence of a formal quantitative statistical framework while tuning being informal or qualitative. A reference to the definitions given by the authors might therefore be welcomed, especially if they are standard in the statistical literature.

Response: It appears that the standard definitions of tuning and calibration have settled somewhat in the literature (see e.g. Han et al. 2009). Tuning parameters don’t have counterparts in the real system, calibration parameters do. The paragraph has been changed to reflect this.

Han, G., T.J. Santner, and J.J. Rawlinson Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments Technometrics. 2009 November 1; 51(4): 464–474. 10.1198/TECH.2009.08126

Comment: p. 2377: Section 2.2.3 turns to be distracting and in fact not really helpful.

GMDD

6, C1395–C1399, 2013

Response: Given that the “other metrics” are not used or discussed further, this section has been removed, to aid the flow of the paper.

Comment: p. 2379, l. 21: It is read : “If the entire a priori input space is truly plausible (. . .) given an observation of the true system" (emphasis is mine). What is meant by this latter expression : “given an observation of the true system"?

Response: This paragraph has been rephrased to “If the entire ensemble input space is initially plausible, any one of the ensemble members might be a candidate for a future observation of the system.” Moved to the discussion (see next point).

Comment: The first part of section 2.4 seems to come at the wrong place. Considerations about the cost of observations or the interest and limit of the simulator in guiding observations are best left for the introductory material and the discussion, where they are already present. At this stage the text should be exempt of general considerations to help the reader to focus on the methodological and mathematical details. Furthermore, given that we are already p.11 of the manuscript (in web form), use of the future tense (“our metric will take") or the conditional (“we might use is making the reading impatient. Some editorial work is probably needed to present the results a bit earlier, and possibly use them to support some methodological choices.

Response: The initial part of this section containing some repeated material is either removed, or moved to the introduction, as appropriate. As the results depend heavily on the methodology for interpretation, we consider it difficult and perhaps counterproductive to move them forward in the paper.

Comment: p. 2380, l. 4 : extra “with".

Response: Removed

Comment: p. 2380, l. 19 : “Our metric will take into account not only the uncertainty in the emulator, but also the inherent problem of inverting the mapping for Y to X " : this is a point that might require clarification. It would be useful to be more explicit in stating

GMDD

6, C1395–C1399, 2013

which choices are critical in the quality of this mapping.

Response: On reflection, the point we were trying to make (that there are many plausible candidates for an input X, given an output Y) is better made elsewhere in the section. The quote is therefore removed.

Comment: p. 2382 Regarding the latter comment: little is said about the choice of the roughness lengths and emulator nuggets, known to be important. It is said that the BACCO package is used, which is fine, but the generic "These parameters are estimated empirically from the ensemble data, via an optimisation routine" is unsatisfactory: optimised on what? (probably leave one out criteria)? How ? Are their any visual diagnostics being involved? The leave one out approach implies the calibration of N − 1 emulators: do they all have the same roughness lengths and nuggets?

Response: We have added detail on the calculation of the roughness parameters (including a reference), and made it clear that the emulator does not use a nugget.

Comment: p. 2383, ll. 17 : What is meant here as the accuracy seems in fact to be the well calibrated character.

Response: We have changed the initial part of the sentence to “Once we have established that the emulator is accurate to an acceptable degree ...”, in order to highlight that the desired accuracy of the emulator is context dependent, and subjective.

Comment: p. 2388, ll. 36: Visualisations techniques are definitely important and the authors have delivered on this in the present article. This said I failed to make sense of this paragraph. Why speak of “projecting a set of lower dimensions in high dimensional space"? What is the point?

Response: We have changed the paragraph to: “Any method of summarising a set of volumes (e.g. “not implausible” regions) in high dimensional space, will be inadequate when projected onto a two dimensional surface for visualisation on the printed page. We welcome further developments in visualisation techniques.” ...in order to make the

GMDD
6, C1395–C1399, 2013

context for the statement clearer.

Comment: p. 2388, l. 14 and 18 : Shouldn't "at worst" read "at best" ?

Response: The original statement is correct. However, we now use "at least", to clarify the statement.

Comment: Finally, the Editorial board of Geophysical Model Development drawing attention on the importance of scientific reproducibility, the authors can only but be encouraged to provide code for their nice visual diagnostics.

Response: Code for the visual diagnostics is developed in the R statistical programming language, with the aim of creating an open source R package. This is as yet unfinished, but we are happy to provide the code on request.

Interactive comment on Geosci. Model Dev. Discuss., 6, 2369, 2013.