Geosci. Model Dev. Discuss., 8, C2356–C2358, 2015
www.geosci-model-dev-discuss.net/8/C2356/2015/

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Importance of bitwise identical reproducibility in earth system modeling and status report" *by* L. Liu et al.

Anonymous Referee #1

Received and published: 1 October 2015

A brief response to challenge a couple of points made in response to my review.

The authors repeat the following truism: "Reproducibility is a fundamental principle of scientific research", which means little without defining what is meant by reproducibility. In most scientific fields (e.g. physics, biology, psychology) reproducibility does not mean that the exact data of a study can be replicated to 8 decimal points, but rather that the results paint a similar conclusion. I would expect that the same will be true in any computational science involving floating point calculations (e.g. cosmology, fluid mechanics). Is climate modelling unusual among the computation sciences in this respect?

In point 2 on page C2164, the authors make an argument as to why bit-level reproducibility is essential in climate modelling. The argument follows a chain of reasoning,

which rests crucially upon the following assertion: "For the simulation results that are sensitive to round-off errors, it is almost impossible to reproduce the results scientifically but not exactly." It is far from impossible to do this. There are many papers using coupled atmosphere-ocean for which the runs could look very different due to differences in forcing or states of variability, but for which this would make little differences to the conclusion of the paper (e.g. studies of the impact of a model change, or in the mechanisms behind a particular process). On the other hand, if the *conclusions of a study* (not just the simulation data itself) are sensitive to the precise initial state and computational platform on which the experiments are performed, e.g. because the conclusions are only valid for particular phases of a mode of internal variability, then the study should never have been published on the basis of a single experiment. If such a paper is published, and an attempt to repeat the experiment produces very different conclusions, then the reproduction attempt has been a useful exercise in that it has demonstrated that the conclusions of the original paper are unsafe and that more ensemble members or a longer integration is required, or that the result is only valid for an individual model. This is a perfectly reasonable way for our science to progress, and does not require bit-level reproducibility across multiple platforms.

As a final note, within my own institute, bit-level reproducibility *on a single platform* is an important requirement for model development. We take a great deal of care to ensure that repeated model runs produce bit-identical results, even when changing the number of cores on which the model is run. We find this a useful tool for identifying bugs in new code, and in filling gaps in recent experiments following an archive failure. However, even with this capability and with the "whole simulation setting" still available, we are unable to produce identical results when we upgrade to a new HPC platform. We would if we could because it would save significant effort in port validation, but it is simply not possible. So the chances of another centre bit-reproducing the results of a simulation are small, even with access to the "whole simulation setting". This is also demonstrated by the authors' own results, which show only a $\sim$30% success rate for the simulations for which they had the full information required.

Interactive comment on Geosci. Model Dev. Discuss., 8, 4375, 2015.