

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

Anonymous Referee #2

Received and published: 29 October 2015

This paper by Liu et al. discusses how bitwise reproducibility affects results obtained with earth system models, presents the results of a survey on the bitwise identical reproducibility of earth system models and proposes a standard to achieve such reproducibility world-wide.

As, in my opinion, the major points made are not supported sufficiently by evidence presented in this paper and in the literature cited in it, I recommend that revisions are required before this work can be published in GMD.

General comments

The paper consists basically of three parts. In the first part the importance of bitwise identical reproducibility is discussed. In software engineering bitwise identical repro-

C2720

ducibility has a merit as a verification tool when porting models to new hardware and software environments and to ensure that technical changes in the model's configuration, such as the number of CPUs used, do not impact the results. The paper states that bitwise identical reproducibility is relevant to earth system modelling as well because changes in roundoff errors affect the climate mean state, variability and trends simulated by the models. However, the paper does not support this claim with sufficient evidence. It shows that differences in roundoff errors will result in individual experiments to deviate. However, this is to be expected since both the earth system and the models are dynamical systems in a chaotic state. That such systems are very sensitive to small perturbations was first shown in Lorenz, J. *Atm. Sc.*, volume 20, 1963, pg. 130-141, and is now a well-known fact in climate science. The paper does not show that the mean climate state, variability and trends are affected by round off errors. To support this claim at least a comparison of two ensemble experiments, one of which is bitwise identical reproducible and one that is not, should be added. A discussion on how the uncertainties due to the lack of bitwise identical reproducibility in earth systems models compare to other uncertainties, such as in the physics incorporated in the models and in the data on the earth's climate, would also be a valuable addition to the paper.

In the second part the paper presents the results of a survey that is aimed at determining the current state of bitwise identical reproducibility in earth system models. The major conclusion drawn here is that bitwise identical reproducibility is not recognized as being important in earth system modelling. However, this conclusion depends on the assumption that the lack of bitwise identical reproducibility is a problem which has not been demonstrated by the paper. The survey itself only shows that the bitwise identical reproduction of results is difficult. In addition, the paper states that ensemble experiments are a viable alternative for bitwise identical reproducibility. Since the major model intercomparison projects are setup as a multi-model ensemble already this weakens the point that the paper is trying to make.

C2721

Finally, the paper presents a standard to achieve world-wide bitwise identical reproducibility and the actions that are required by scientists, journals, vendors of hardware and software and model intercomparison projects. The advantages of such a standard discussed here is limited to the software engineering aspects of model development. A discussion on the potential advantages for climate science itself would be a valuable addition. For example, would the proposed standard reduce the uncertainties in the CMIP5 scenarios for this century?

As final general comment, some of the literature cited in the paper, especially Song et al., 2012, does not seem to support the points that the authors are trying to make.

Specific comments

Page 4376, line 26: Easterbrook and Johns, 2009, in their section 4.4 on verification and validation, report indeed that bitwise identical reproducibility is used as a verification tool for code correctness at the Hadly Center. However, it also mentions the limitations of its applicability and discusses several other methods of verification and validation that climate scientists use. Therefore, a discussion of what this implies for the advantages, that world-wide bitwise identical reproducibility can offer, would be a valuable addition to this section.

Page 4377, line 6: Monniaux, 2008 discusses floating-point computations in the context of mission critical systems. This raises the question in what sense earth system models are or resemble mission critical systems. This is something the paper should discuss here.

Page 4377, line 15: Liu et al., 2015, state that "Facilitation of the bitwise identical reproducibility also requires the efforts beyond the field of Earth system modeling. Through enlarging the bitwise identical compiler version set and processor version set, the bitwise identical result of a simulation can be reproduced more easily and flexibly.". This is clearly something that cannot be achieved by climate scientists. Therefore, a discussion of the implication of this statement for actually achieving world-wide bitwise

C2722

identical reproducibility seems in place.

Page 4378, line 15 and following: The historical experiments of two models from CMIP5 are used to illustrate the impact of roundoff errors on model results. In order to strengthen the point the authors are trying to make a comparison of these experiments with those available from the CMIP5 data archives and the historical records is required. For example, to show that the climates of the models are significantly altered by roundoff errors, the authors could compare the mean state, variability and trends in their ensemble experiments with those in the experiments from the CMIP5 archive and in the historical data.

Page 4380, lines 19 and following: The first sentence here states that the results of the survey show that the importance of bitwise identical reproducibility is not widely recognized in climate modelling. It is not made clear why this is the only conclusion that can be drawn from the survey. This conclusion also seems to contradict the work of Eastbrook and Johns, 2009, that is cited earlier in the paper. Also the cited work by Song et al., 2012, implies that bitwise identical reproducibility is not as important for many climate studies as claimed here when ensembles of sufficient size are used.

Page 4394, caption of Figure 4: An explanation is required on the meaning of the confidence curves. It is unclear whether these confidence curves are computed for each individual member plotted or for the complete ten member ensemble.

Technical comments

pg. 4376, lines 3: ", i.e., bitwise ... be reproduced," can be removed.

pg. 4376, line 18: "An increasing ..." should be a "A large ..."

pg. 4377, line 2: "was extremely" should be "is extremely"

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

C2723