Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2022-71-RC2, 2022

Comment on gmd-2022-71

Peter Düben (Referee)

Referee comment on "Characterizing uncertainties of Earth system modeling with heterogeneous many-core architecture computing" by Yangyang Yu et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-71-RC2, 2022

The paper is addressing a very important and underrepresented topic in scientific computing, namely a method to verify whether code which is ported to new hardware is still scientifically correct. The paper provides a new approach to tackle the problem using ensemble methods. The paper would still need to improve in clarity before being published in GMD following the comments below.

Major comments:

- The paper would need to be place better into the existing literature. For example: The problems that are discussed in the paper (and in fact also the solution) have been discussed in publications on the use of reduced numerical precision. Here, ensemble methods are used to diagnose the impact of a precision reduction. See for example:
  D. Dueben, A. Subramanian, A. Dawson and T. N. Palmer. A study of reduced precision to make superparametrisation more competitive using a hardware emulator in the OpenIFS model. Journal of Advances in Modeling Earth Systems, 9 (1), 566-584, 2017
  Tintó Prims, O., Acosta, M. C., Moore, A. M., Castrillo, M., Serradell, K., Cortés, A., and Doblas-Reyes, F. J.: How to use mixed precision in ocean models: exploring a potential reduction of numerical precision in NEMO 4.0 and ROMS 3.6, Geosci. Model Dev., 12, 3135–3148, https://doi.org/10.5194/gmd-12-3135-2019, 2019.
  The paper could also discuss the use of stochastic hardware, which would cause similar issues regarding checks of solution quality, see e.g.:
  https://royalsocietypublishing.org/doi/full/10.1098/rsta.2013.0276
  Finally, there is very interesting work by Oliver Fuhrer and team to port Earth system applications without changes of bit reproducibility, e.g.:
  https://ieeexplore.ieee.org/document/6877351
- There is one point that I do not understand in the discussion about many-core architecture: You can already trigger non-identical results when running a serial CPU code with a changed compile flag. This is well known. Why are many-core architecture any different? Do they produce results which are different between different realisations of the same code on the same machine in the same setting? If no, I do not understand

why many-core hardware is any different. If yes, you would need to provide more information how different different realisations actually are, for example in tables 1+2. Also, optimised MPI parallelisation when using many CPUs in parallel can also show differences in results when running the same code in the same setting with the same number of nodes/cores as messages can arrive in different orders. How is this the problem that you find with master and slave architecture different? Does the difference really justify having a fully figure 2 on the Sunway architecture?

- LL126: It should also be mentioned that the models are simulating chaotic dynamics resulting in differences between simulations to grow exponentially. Porting linear models would be simple.
- LL131: I understand the reason why you are applying the method to a small model in the paper. However, you should outline somewhere how you would apply your approach in a model as large as an Earth system model. Also, I do not understand what you mean by "building a software tool". You are presenting a method applied to a customised code. What is the tool you are talking about?
- Figure 6,8,10: The caption indicates that you are showing standard deviations, but the Figures seem to show ensemble members. I am confused as standard deviations should be a single number. Or is this a variable of the model?
- The English should be improved if possible.

Minor comments:

L19: Not really "potential differences". The model will most of the time not be bit reproducible when the hardware is changed and therefore different.

L24-25: I do not understand what is meant by "on-off switches" or why this is useful.

L32: The development does not require an increase in computing power. The increase in resolution does.

L116: "inevitable a perturbation" Is this reproducible between different runs or stochastic?

L145: I guess this should be a "\Delta x"?

L215: "of different modes" What is meant by this? Is the error calculated between the Intel mode and other modes when using the same random perturbation for each ensemble member?

L221: "the basic work" I do not understand.

“magnitude order perturbations” should be “perturbations of different order of magnitude” in several sentences throughout the paper

L249: What is QNLM?