Geosci. Model Dev. Discuss., 8, C1779–C1782, 2015
www.geosci-model-dev-discuss.net/8/C1779/2015/

Geoscientific Model Development Discussions

Interactive comment on "Enhancement for bitwise identical reproducibility of Earth system modeling on the C-Coupler platform" *by* L. Liu et al.

Anonymous Referee #2

Received and published: 26 August 2015

Summary:

This paper focuses on enhancing bitwise reproducibility for the Community-Coupler platform and advocates the goal of "worldwide bitwise identical reproducibility".

General comments:

I question the feasibility and, moreover, the necessity of trying to attain "worldwide bitwise identical reproducibility". Indeed, even achieving reproducibility with same compiler (and options), hardware and software is not always feasible - and is becoming increasingly problematic on newer systems. For example, if one uses the FMA on AMD Cray machines, the result is not reproducible. Therefore, to achieve reproducible results on such machine, one must turn off this optimization. Running code on multi-core

platforms with many threads also presents many obstacles to reproducibility, such as a non-controllable summation order without targeted changes in the code (which are often costly and affect performance). In the case of a large (and expensive) climate code, having to turn off performance optimizations does impact performance significantly, and this discrepancy will only widen with architecture advances. Certainly one would not want to run a climate code with -O0 (as suggested on page 2422). One could argue that these bitwise-reproducible checks would only be run for short duration, but I would say that if the purpose for running them is to verify the code, one must use the compiler settings desired for a production run. Indeed, just because a code is correct with -O0, does not mean that it is with -O3. Overall, bitwise reproducibility places restrictions that limit optimizations - there has to be a better way to judge correctness that moves beyond requiring results to be bitwise identical.

Another issue is the availability of the desired hardware and software stack that was used for an initial simulations. In my experience, many HPC computing platforms are in constant flux as far as available compilers, and even hardware changes are frequent. Therefore, the span of time in which it is even possible to have the "same" environment is often quite short. By the time a research paper is published, a machine may have been completely upgraded or replaced.

In general, I also have reservations as to whether this manuscript "presents novel concepts, ideas, tools, or data" as well as a "sufficiently substantial advance in modeling science" as outlined by the GMD review guidelines. Many of the suggestions for repeatability and reproducibility seem either rather straightforward or already practiced in production climate models.

Specific comments:

(1) page 2405, line 5: I agree that reproducibility is desirable on a single platform, but this is not even possible on all architectures.

(2) page 2405, line 15: The fact that slight differences in computing environment re-

sult in differences in output (as in referred to in supplement) could also be viewed as a an argument against bitwise reproducibility and a motivation for better metrics for comparison. Though climate model systems do respond to subtle differences, in most cases one would hope that the scientific conclusion would be the same (that the mean climate is the same in some sense). If not, there could be a problem with the hardware or software stack. Or there could be poorly written code (unstable numerics) that is too sensitive.

(3) page 2406, line 3: "The bitwise identical reproducibility or Earth system modeling deserves to be a worldwide standard." I disagree with the statement in general and did not find the argument in the supplement sufficiently compelling. This requirement may force decisions (in code, hardware, etc.) that negatively impact performance and may not even be possible on many architectures.

(4) page 2406, lines 10-15: As stated here, I agree that being able to REPEAT an experiment is critical, and that scientists should be taking measures to ensure that. But is this aspect of the paper (if one were to abandon the bitwise reproducible requirement), a significant advance?

(5) page 2409, line 1: I don't believe that you mean "infringement".

(6) page 2407, line 23: "Inheritance of reproducibility" is awkward and could be clarified.

(7) page 2407: Challenges 2 and 3 are also relevant to just repeating the results.

(8) page 2409: When the same computing environment is available, then obtaining the settings and inputs for the original run is typically straightforward, so I am not seeing the intellectual contribution from this angle as scientists doing large climate runs already take such measures.

(9) section 3.1: "providing entrances for further downloads from simulation resource servers" - while convenient, is this a significant contribution?

(10) section 3.2: This section contains rather straightforward information.

(11) section 3.5:Does this section contain the sort of information better suited to the model documentation? (also does #7 go a bit too far?)

(12) section 3.5,lines 6-10 : Couldn't any climate model in a version control system provide this information?

(13) page 2422, line 12: I disagree with this statement about the performance impact being limited. This would be quite dependent on the system and the particular modification necessary to ensure bitwise reproducibility. The statement should be backed up with some data and specifics.

(14) page 2422, line 1: Using -O0 can be big performance hit, and this compiler setting is unlikely to be used by any production run.

(15) section 4.2.5: Point number 3 has already been stated earlier.

Interactive comment on Geosci. Model Dev. Discuss., 8, 2403, 2015.