Geosci. Model Dev. Discuss., 8, C2446–C2452, 2015
www.geosci-model-dev-discuss.net/8/C2446/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions
Open Access

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

L. Liu et al.

liuli-cess@tsinghua.edu.cn

Received and published: 9 October 2015

We thank the reviewer very much for the comments. We'd like to reply them one by one as follows.

1. I question the feasibility and, moreover, the necessity of trying to attain "world-wide 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.

Response: Bitwise identical is the strictest level of reproducibility. If reproducibility can

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

be achieved at bitwise identical level, the original results must be able to be repeated or scientifically reproduced. A package of the whole simulation setting can help achieve bitwise identical reproducibility and at least can guarantee repeatability. It has been well proved that simulation results of Earth system modeling can be sensitive to round-off errors. For the results that are sensitive to round-off errors, it is almost impossible to reproduce the results scientifically but not exactly. In other words, the original results are either irreproducible or exactly reproducible. Bitwise identical reproducibility, which guarantees the reproduction of exactly the same results, therefore is important to Earth system modeling. Worldwide bitwise identical reproducibility is defined as "original scientists of published results should ensure the whole simulation setting publicly available for bitwise identical reproduction, so that any fellow scientists can independently obtain the whole simulation setting and then independently repeat the original simulation or reproduce the original results (Liu et al, 2015)". It does not enforce every reproduction by fellow scientists at the bitwise identical level. We view this manuscript as the first step of bitwise identical reproducibility. It only wants to show bitwise identical reproducibility can be achieved in some cases, but cannot guarantee bitwise identical reproducibility can be achieved in any cases. Bitwise identical reproducibility requires further research efforts.

The manuscript is modified accordingly. Please refer to the context from page 3 line 7 to page 4 line 15.

2. 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).

Response: The approach of higher precision of summation can also help achieve bitwise identical results of multiple threads. Please refer to line 16 on page 18.

3. In the case of a large (and expensive) climate code, having to turn off performance optimizations does impact performance significantly, and this discrepancy will

GMDD
8, C2446–C2452, 2015

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.

Response: When using some compiler optimization restrictions (such as "-no-vec -mp1 -fp-model precise -fp-speculation=safe" for Intel compilers), –O3, -O2 and –O1 can achieve bitwise identical results with –O0. As a result, the performance reduction due to the compiler optimization restrictions will not be significant (for example, Figure 3). The manuscript is improved accordingly. Please refer to lines 25-32 on page 17.

4. 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.

Response: We define bitwise identical compiler version set and processor version set for this issue, in order to make computing environments bitwise compatible for a long time. Please refer to the last paragraph on page 18.

5. 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.

GMDD
8, C2446–C2452, 2015

Response: To the best of our knowledge, publicly open climate models and model platforms rarely provide systematic supports for reproducibility and repeatability. We may miss something and really hope referee #2 show us details. We believe that the concept about bitwise identical reproducibility and systematic solution for it is not old.

6. Page 2405, line 5: I agree that reproducibility is desirable on a single platform, but this is not even possible on all architectures.

Response: Yes. Bitwise identical reproducibility is almost the strictest kind of reproducibility. We find that it can be achieved on different versions of the same type of compilers or processors, while it is very difficult to achieve on all architectures. When bitwise identical reproduction cannot be achieved across different types of processors or compilers, this work can also provide as much as possible information to facilitate the repetition of the original simulation and the scientific reproduction of the original results. The manuscript is modified accordingly. Please refer to the last paragraph on page 23.

7. Page 2405, line 15: The fact that slight differences in computing environment result 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.

Response: As mentioned before, for the results that are sensitive to round-off errors, it is almost impossible to reproduce the results scientifically but not exactly. In other words, the original results are either irreproducible or exactly reproducible. It has been well proved that climate simulation results can be sensitive to round-off errors (Please refer to the reference Liu et al. (2015)). For such kind of model simulations, we currently

GMDD
8, C2446–C2452, 2015

cannot say there should be a problem with the hardware, software stack or model code.

8. 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.

Response: As mentioned before, the performance reduction for bitwise identical reproducibility is not significant, and worldwide bitwise identical reproducibility does not enforce every reproduction by fellow scientists at the bitwise identical level.

9. 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?

Response: as mentioned before, to the best of our knowledge, publicly open climate models and model platforms rarely provide systematic supports for reproducibility and repeatability. This work can improve not only reproducibility but also repeatability for Earth system modeling. Please refer to last paragraph on page 23.

10. Page 2409, line 1: I don't believe that you mean "infringement".

Response: please refer to lines 17-20 on page 6.

11. Page 2407, line 23: "Inheritance of reproducibility" is awkward and could be clarified. Response: please refer to lines 19-23 on page 5.

12. Page 2407: Challenges 2 and 3 are also relevant to just repeating the results.

Response: for this work, repetition is the first step of reproduction. At the same time targeting bitwise identical reproducibility, this work can also improve repeatability.

13. Page 2409: When the same computing environment is available, then obtaining the

GMDD

8, C2446–C2452, 2015

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.

Response: The challenge is that given the simulation results in published paper, the original computing environment is always unknown. Even when fellow scientists ask help from original scientists, original scientists may have forgotten the computing environment (Please refer the survey in the reference Liu et al. (2015). Many scientists involved in the survey have already forgotten the details of the simulations). Moreover, scientists may rarely care about the details of computing environment when they conduct simulations. Therefore the systematic implementation that automatically records detailed information of the simulation is necessary for Earth system modeling.

14. Section 3.1: "providing entrances for further downloads from simulation resource servers" - while convenient, is this a significant contribution?

Response: This implementation is necessary to make the simulation setting package as small as possible to be supplement of a published paper.

15. Section 3.2: This section contains rather straightforward information.

Response: We want to show the complete system design for bitwise identical reproducibility in this paper, in order to help the modeling groups who want to achieve worldwide bitwise identical reproducibility in the future. Therefore, we need to introduce the detailed implementations some of which may be straightforward.

16. 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?)

Response: lines 26-28 on page 14 has been modified.

17. Section 3.6, lines 6-10 : Couldn't any climate model in a version control system provide this information?

GMDD

8, C2446–C2452, 2015

Response: any climate model with a certain version control system can provide the version number and code patch. However, to the best of our knowledge, publicly open climate models and model platforms rarely automatically records the version number and code patch for a simulation. The C-Coupler platform can also record the code patch when the model code is not managed by any version control system.

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

Response: please refer to Fig. 3. We can still use higher optimization levels such as -O1, -O2 and -O3 to achieve bitwise identical reproducibility.

19. Section 4.2.5: Point number 3 has already been stated earlier.

Response: it has been removed.

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