

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

Anonymous Referee #1

Received and published: 30 April 2015

This paper provides a good summary of the motivation and issues associated with bitwise reproducibility of earth system models between different groups, machines and over time. It also provides a description of some solution strategies, a sample implementation, and some results of bitwise reproducibility tests.

There continue to be a number of challenges in practice, including several issues that were mentioned in the paper such as the ability to mimic the baseline hardware and software stack on a distinct system, the inability to predict future hardware, availability of source code and input datasets with respect to permissions/copyright/licensing issues, and performance tradeoffs for reduced compiler settings and bitwise reproducible global sums among other necessary deoptimizations.

But this paper is well formulated and a significant improvement over an earlier version.

C590

I recommend it be published subject to revision and would be willing to review it again before final acceptance.

General Comments:

I believe the authors should add a few sentences about the character of the climate systems, that it's a fairly stable but chaotic time dependent system and how a last bit roundoff error grows in the system.

I believe the authors should add a brief discussion of how bitwise reproducibility is tested, who makes that decision, and how that should be determined. In particular, a comment in the final section that suggests bitwise reproducibility can be determined from visual comparison of a plot should be removed. Also, it probably should be noted somewhere that if a run is not bitwise reproducible, it's fundamentally difficult to determine whether the solution is the same without a long run or a test that is proven to be statistically significant.

I think the authors assume that models are reproducible on the same machine, with the same executable, with the same processor counts, using the same software stack and same build, but that is not actually stated anywhere in the paper. It is an important assumption that NOT all models (or hardware) achieve, but it is almost certainly a requirement that should be noted in the paper somewhere. This is addressed partly in section 4.2.5 but recommend it be discussed earlier in the paper.

What if a simulation undergoes changes in input files, source code, or input values in the middle of the run. Are these changes recorded automatically with the C-Coupler tool?

Specific Comments:

Section 1: page 2408, line 28, point 3. infringement means that copyright is being violated. I think the authors want to say that code and input data can be downloaded from separate servers to protect intellectual property. That's not entirely true as intel-

C591

lectual property is often contained in source code or input datasets that are used for simulations even if that part of the model is not invoked.

Section 3.1: What if the results are not bitwise reproducible which will often be the case? Can a scientist perform new simulations and compare to the original simulations? How should the scientist proceed? Is there a test to verify the simulations are fundamentally "the same" without being bitwise reproducible?

Section 3.2: No need to define the filename precisely. Just state that the files used to store the setting information should be uniquely tied to the simulation and contain information about when it was created.

Section 3.3: Do the diagnostics have to be full precision with respect to the model? Is the ascii (base10) representation of the diagnostics adequate to ensure bitwise reproduciblty? Are global sums of fields enough to ensure that the entire field is bitwise identical everywhere? How many and which fields need to be diagnosed? Who makes those decisions? How long does a simulation need to be run to ensure the results truly are bitwise reproducible (ie. one timestep, one day, one year)?

Section 3.5: I think point 7 should be removed, but I understand why it's there. It's just not practical to make it part of the defined process.

Section 4.1.1: Some models can switch bitwise identical parallelization "on" for testing, but would choose to run in production with it "off" for performance. pg 2422 line 15: "There are always global summation" should be changed to "there are often global summation"

Section 4.1.2: How can a user know what the appropriate compiler settings should be so future hardware/software is bitwise reproducible? Is it adequate to test various compiler versions and hardware "at present"? If there is a performance degradation of the model at bitwise reproducible compiler settings, how does a scientist make the tradeoff between model performance and possible (not guaranteed) bitwise reproducibility in

C592

the future?

Section 4.2.5 (2.): I don't agree with this claim. If different compiler versions and/or different hardware processors produce different results, it is not reasonable to suggest there are bugs in the model code. There are many reasons this could happen including (as stated elsewhere in the paper) compiler optimization/flags, compatibility of different hardware processors with each other, etc. Maybe I'm missing the point and this needs clarification.

Section 4.2.5: I suggest this section be rearranged in the paper. Point 1 should be moved to a point earlier in the paper to introduce some basic starting requirements for bitwise reproducibility. Point 2 should be deleted unless it's further clarified. Point 3 is already made in the introduction and conclusions.

Section 5: what is piControl? page 2428, line 3, define "short-time" simulations more clearly (one timestep, one hour, one day, one year, one century?). What was your validation criteria in point 4 on page 2428? How many fields, what fields, how where they chosen? It would be nice to see what happens to the performance when the optimization is increased to scientifically acceptable levels but levels that are not bitwise reproducible in different processor versions or compiler versions. Can the authors demonstrate the claim made earlier that the performance difference is small. It would be nice to see a bit more results here, including efforts to reproduce simluations across different processor types and compilers and some suggestions about why this is difficult. For instance, is it possible at the lowest possible optimization with strict IEEE mathematical operations? But I don't expect the authors to address this at this time. This issue is briefly noted in section 4.2.3 point 2 and should be added to section 6.

Section 6: The idea of testing bit reproducibility from figures is introduced in section 6 without any prior discussion, and I believe this idea makes no sense at all. For many fields, you might not be able to distinguish differences in the first or second digit from figures, let alone bit reproduciblty. Figures can prove two simulations are not the same,

C593

but they will never prove that they are the same at the bitwise level.

Code Availability: I find it extremely ironic that the sample models used for this study are not available to the public to check bit reproducibility due to permission limitations.

Presentation Issues:

There are a number of challenging sections with respect to English language which need to be updated. I recommend the authors work with a native English speaker to help produce a final version. The list of suggested changes below should not be considered complete or correct, but merely point out some problems and potential solutions.

I find the terms "simulation setting", "compiler version set", "processor version set", and "parallel setting" to be generally difficult to understand. Please consider using other terms such as "simulation setup", "bit reproducible compilers", "hardware setup", "software setup", "bit reproducible hardware processors", "compiler version", "processor version", "process counts", "core counts", "processor counts", "mpi task counts", "processor layout", "task and thread layout", etc.

Then in particular,

change, "Februar" -> "February"

change, "researches" -> "research"

change, "worldwide bitwise identical reproducibility" -> "bitwise reproducibility" (in the context of the paper, bitwise reproducibility means identical results across different groups and machines)

change, "compiling flags" -> "compiler flags"

pg 2404 line 2, "i.e. bitwise identical result of numerical simulation can be reproduced."
-> "i.e. bitwise identical results of numerical simulations that can be reproduced."

C594

pg 2042 line 5 "intra" to "within"

pg 2404 line 24, remove "to participate in"

pg 2405 line 15, make ", and therefore" to a new sentence

pg 2405 line 18, "to referees" -> "on referees"

pg 2406 line 27, entire paragraph starting with "To achieve a bitwise" needs to be rewritten

pg 2407 line 19, "are generally origin" -> "generally originate"

pg 2407 line 20, "with protection of copyrights" -> "with copyright protection"

pg 2407 line 27, "requires to be" -> "must be"

pg 2408 line 2, "lack of the" -> "lack the"

pg 2408 line 21, "to find out" -> "to find"

pg 2408 line 25, "to be independent" -> "independent"

pg 2409 line 6, "succesfully bitwise identical reproduced" -> "bitwise reproducible"

pg 2410 line 8, "so as that fellow" -> "so fellow"

pg 2411 line 4, "to achieving the worldwide bitwise identical" -> "to achieve bitwise reproducibility"

pg 2411 line 7, "appropriately" -> "appropriate"

pg 2411 line 7, "of even" -> "or even"

pg 2411 line 11, "the entrances for further downloading them" -> "information to further download them"

pg 2411 line 17, "to prepare a new simulation resource server for downloading them"
-> "prepare a new simulation resource server"

C595

pg 2412 line 3, "to the reproduced" -> "with the reproduced"
pg 2413 line 6, "the worldwide bitwise identical reproducibility" -> "bitwise reproducibility"
pg 2414 line 17, "of the model code that is obtained from other modeling groups" -> "of other modeling groups"
pg 2414 line 25, entire sentence needs work
pg 2514 line 6, "When an input data file is not managed by Subversion, the whole data file will not be put into the simulation setting package for saving the storage space, while only the checksum is recorded" -> "When an input data file is not managed by Subversion, the data file is not put into the simulation setting package due to size concerns and only the checksum is recorded"
pg 2416 line 9, "to find" -> "find"
pg 2416 line 16, don't understand "Therefore, the deployment of processors for running a simulation is not concerned yet." please clarify.
pg 2416 line 23, the baseline results must be stored on the server. having them in a paper serves little use.
pg 2416 line 27, "able to be reproduced" -> "reproducible"
pg 2417 line 7, "to leave" -> "leave"
pg 2417 line 8, "so as that" -> "so"
pg 2417 line 9, "run the" -> "ran the"
pg 2417 line 19, "If a required input data file or code already exists under the specified local directories and is the same as the wanted (i.e., the checksum of the local data file is the same as the corresponding checksum recorded in the simulation setting package, or the code version under the local directory is the same as the correspond-

C596

ing code version recorded in the simulation setting package), it will not be downloaded again; otherwise, it will be downloaded from a simulation resource server, and scientists may be asked to input a username and password for the downloading." -> "If a required input data file or code already exist under the specified local directories and the checksums are identical, the data will not be downloaded again."

pg 2418 line 2, "to recover" -> "to update"
pg 2418 line 3, "procedure of reproducing" -> "procedure to reproduce"
pg 2419 line 1, "is hardly reproduced" -> "is difficult to reproduce"
pg 2419 line 17, sentence that starts with "Third" needs to be simplified. For instance, "new" appears far too many times.
pg 2419 line 24, "works" -> "work"
pg 2421 line 2, "result" -> "results"
pg 2421 line 3, "result" -> "results"
pg 2421 line 7, "Moreover, we give some suggestions for facilitating achievement of the bitwise identical reproducibility." -> "Next we provide some suggestions for dealing with these issues."
pg 2421 line 11, "does not require the models to achieve the" -> "does not require that models achieve"
pg 2422 line 7, "to find out the compiling flags" -> "to determine the compiler flags"
pg 2422 line 7, "the original calculation orders" -> "the original order of calculations".
pg 2422 line 8, "regarding the Intel compilers, the set of compiling" -> "for the Intel compilers, the set of compiler"
pg 2422 line 10, "at the optimization" -> "at an optimization"

C597

pg 2423 line 20, "in several aspects" -> "in several areas"
pg 2424 line 4, "used for the" -> "used for"
pg 2424 line 11, "scientists to" -> "scientists"
pg 2424 line 13, "of the input data" -> "about the input data"
pg 2424 line 17, "so as to save" -> "to save"
pg 2424 line 22, "should be concerned" -> "should be considered"
pg 2424 line 23, "highly depend" -> "highly depends"
pg 2426, line 4, "Bitwise identical tests have already been used for the model development (<http://www.cesm.ucar.edu/models/ccsm3.0/ccsm/doc/UsersGuide/UsersGuide/node9.html>, last access: 15 Februar 2015; Easterbrook and Johns, 2009), such as the bitwise identical parallelization, bitwise identical result in repeated runs, and bitwise identical result in initial run and restarted runs" -> Bitwise identical tests are already being used in model development (<http://www.cesm.ucar.edu/models/ccsm3.0/ccsm/doc/UsersGuide/UsersGuide/node9.html>, last access: 15 February 2015; Easterbrook and Johns, 2009) to check for bitwise identical parallelization, bitwise identical result in repeated runs, and bitwise identical result in restarted runs."
pg 2426 line 10, "result" -> "results"
pg 2426 line 11, "The bitwise identical reproducibility requires a simulation achieving the" -> "Bitwise reproducibility requires that a simulation achieve"
pg 2426 line 12, "result in the" -> "results in"
pg 2428 line 9, "using each other" -> "using other"
pg 2428 line 9, "version" -> "versions"
pg 2428 line 11, "each other computer platform" -> "other computer platforms"

C598

pg 2428 line 16, "totally 11" -> "a total of 11"
pg 2829 line 10, "for the model development" -> "for model development"
pg 2429 line 22, "result" -> "results"
pg 2429 line 27, "able to be" -> "is able to be"
pg 2430 line 7, "We therefore hope the experts in the field of computer will make the new versions of compilers and processors join in existing bitwise identical compiler version sets and processor version sets." -> "We therefore hope the experts in the field of computer hardware and systems software will make new versions of compilers and processors that are bitwise compatible in the future."
pg 2430 line 13, "works" -> "work"

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

C599