Geosci. Model Dev. Discuss., 7, C1513–C1523, 2014
www.geosci-model-dev-discuss.net/7/C1513/2014/

Geoscientific Model Development Discussions

Interactive comment on "Enhancing reproducibility of numerical simulation result on the C-Coupler platform" *by* L. Liu et al.

Anonymous Referee #2

Received and published: 28 August 2014

1 General remarks

With this paper the authors introduce a use-case of the C-Coupler platform as part of the C-Coupler software to assist Earth system model and component developers in the programming and testing phase. Parts of the software are already introduced in a companion paper by Li et al. (2014), and the paper at hand can only be understood in combination with this. In the present manuscript the authors claim that the C-Coupler platform is a useful tool to guarantee long-term maintenance and support of model configurations and whole experiments. In particular, the authors argue that running Earth system models with the C-Coupler platform framework help code developers and experimentors to accomplish bit-identical results when re-running a particular

model configuration during the development process or later in order to reproduce a scientific experiment. To achieve this, the C-Coupler platform collects system, model, experiment, and provenance metadata. These metadata are stored in a repository. Beyond that, the authors make suggestions on software development and the management of data and metadata. Finally, a specific use-case is discussed for which the authors make available the required software.

Some aspects addressed in this manuscript are very interesting and have the potential of being valuable assets to ongoing projects and discussions, while large parts of the manuscript are less relevant to be published in a peer-reviewed journal simply because they are already common practice. Thus, the overall topicality of this manuscript is low. To improve the scientific quality of the manuscript, the reader deserves some assessment of the state-of-the-art in the field covered here. Publications cited in this manuscript are mostly irrelevant for this topic, while various important publications in this field are ignored. This leaves the reader with the impression that the authors did not perform a sound bibliographic search. The paper has too many grammatical and spelling errors. My advice would be to consult a native speaker who carefully proofreads and corrects a revised version before resubmitting - even though GMD provides this service for the final publication in GMD.

I recommend to reject this version of the manuscript for a publication in GMD. Some aspects touched upon in this manuscript are of importance to the Earth system model community, but substantial rewriting is required to achieve a level of quality suited for a peer-reviewed journal with international reputation. Thus, I suggest to submit a completely revised manuscript and clearly identify the added value to ongoing international projects and discussions in this field. The manuscript can be improved by more clearly separating the four topics addressed in the current version, (1) the general metadata problem, (2) a description of the C-Coupler platform as a compile- and runtime framework, (3) reproducibility of numerical model results, and (4) strategies for code development. The authors may also think about shifting topic number 4 into a supplement

or appendix of the main paper, or a class book.

1.1 Metadata

Ignoring the bit-reproducibility aspect for a moment, having a complete list of provenance metadata that document how model output has been produced are highly desirable, following the good scientific practice of the past of keeping “laboratory journals”. In this sense, the whole idea of collecting – broadly speaking – complete provenance metadata is not new. (Here I use the term “provenance metadata” as a synonym to for all metadata to fully describe and document a numerical experiment.)

There is a general consensus and awareness in the climate community about providing the necessary metadata to allow for re-running and documentation of particular model experiments. In the last decade several international projects have been started to address this point, and papers have been published on this topic like Lawrence et al. (2012) and Guilyardi et al. (2013). These authors provide further references and links to past and ongoing projects. Li et al. have a valid point that metadata collected so far are still incomplete. Nevertheless, readers not directly involved in those discussions as authors or collaborators in related projects may miss an introduction into the theme and what has been achieved elsewhere so far. In particular it could be noteworthy to detail to which extent the work presented here can contribute to international efforts rather than leaving this as guesswork to the reader.

The authors could contribute to the discussion by identifying the missing pieces and come up with a concise but complete list of still missing metadata. On the other hand, to my knowledge such has already been thought about in the METAFOR project, and a subset went into the Common Information Model (CIM). I acknowledge the fact that the authors start a new effort to compile such a list and mention e.g. hardware, compiler version, compiler options, and some more. Nevertheless, for a full description their list is still incomplete. Library versions and compiler options used to generate the libraries

are missing. As an example, the authors mention the message passing interface (MPI) library. They discuss some problems and solutions in this context, but typically there are several more libraries involved.

When talking about data available in the CMIP5 archives the above is not sufficient. Typically those data are not directly produced by the Earth system model, but the archived data may have gone through various post-processing steps. Thus, provenance metadata describing the post-processing steps are required, plus information about the post-processing software (compiler, compiler options, possibly libraries (with compiler options) used to build the post-processing software.

1.2 Software framework

What makes the authors decide to build their own software to address the missing pieces in existing solutions rather than contributing to international efforts, e.g. the ones described by Lawrence et al. (2012) and Guilyardi et al. (2013)?

Ford et al. (2012) and references therein give an overview about existing tools for configuring, building and running models. Again, in the introduction the authors miss to provide a brief overview about the state of the art, and thus fail to derive from such an analysis those aspects that make their approach unique compared to existing solutions.

Using GIT and/or SVN in combination with log files for archiving does not sound very innovative. For model code this is state-of-the-art since many years, and I wonder whether this is very suited for the versioning of data. Clearly, a discussion is missing why the authors follow this approach rather than using some kind of data base which allows the query of information regarding a particular model run or experiment setup. It is tempting to defer a change in software solutions and strategies to forthcoming papers, but I would not accept this as an argument for the choice made here.

In two paragraphs the authors mention the use of checksum to identify data and discuss

problems associated with this approach. Again, this is not very innovative. The whole concept is well documented on the web and published elsewhere. References are missing.

1.3 Code development

Suggestions are given for code development. This can be achieved through the C-Coupler platform, or in one way or the other with many other “frameworks”. Suggestions made here are not new and already state-of-the-art. Clune and Rood (2011) already summarise software development practices and issues in their paper. They even go beyond the stage to simply keep producing bit-identical results and come up with concepts which have not yet found their way into the climate modelling community. Regarding the paper at hand I am missing a reference to this paper, followed by some hints where Li et al. go beyond what is already discussed in the literature. Concerning code development and code quality Easterbrook and Johns (2009) touch upon various aspects as well.

1.4 Bit-identical results

While the ability to get bit-identical results during the software development process is a very useful concept when working on technical code development I have the impression that this aspect is overemphasised in this paper. Climate scientist are usually interested in guaranteeing that a particular model is capable of producing an identical climate when run at a different site or in different hard- and software environments. In other words: the climate should be independent from the parallelisation, compiler option and hardware. As single realisations are only of little help here this reasoning brings us to ensemble members. Thus, it would be interesting for the reader to know to which extent the C-Coupler platform is capable of supporting such

ensemble runs and in particular to evaluate the ensemble member results. As Li at. explicitly talk about CCSM3, I wonder why they do not mention the CCSM web site http://www.cesm.ucar.edu/models/ccsm3.0/ccsm/doc/UsersGuide/UsersGuide/node9.html where many aspects of bit-reproducibility are explicitly mentioned. Furthermore, I found a presentation by Lapillone, Lardelli, and Fuhrer "Technical Test Suite for COSMO", held at the 14th COSMO General Meeting in 2012 At other sites in Europe this concept is used for almost two decades in the context of technical model development (e.g. paralellisation), and in my oppinion this is rather common practive than an innovative concept.

In the Li et al. paper it finally turns out that bit-reproducibility can only be achieved under certain very restrictive conditions. The C-Coupler platform does not perform any miracles to help achieving this beyond what we can already get with version control of our software and input data plus recording compiler versions and options. This is a disappointing result after reading in total 625 lines of text.

2 Specific remarks

2.1 Introduction

- The state-of-the-art is not captured.
- P4430 L18-24: The list of papers describing numerical models looks somewhat arbitrary. What is the reason for selecting those and ignoring other programming efforts?
- P4431 L1: I would say it is still the modelling groups and people that participate in MIPs with the models but not the models alone.

- P4431 L10-L12: Scientific reproducibility is not about hindcasting the climate of the past. I understand it as the ability to reproduce any climate obtained with a particular model setup (including input files and parameter settings) on any other platform independent of the compiler and hardware. In which sense does this deviate from technical reproducibility?
- P4432 L1 - P4433 L16: This whole discussion is not new. Those aspects have been discussed at several workshops and are not new to any group involved in the development of a numerical model.
- P4433 L4-L9: see also Easterbrook and Johns (2009) and Clune and Rood (2011)
- P4434 L1-L22: Mentioning the C-Coupler is confusing and the relation to the C-Coupler platform is not really clear.
- What is the main message you would like to get over? The whole paper should then be oriented along this main point.

2.2 Brief introduction to the C-Coupler1

The C-Coupler is already introduced in the companion paper. A reference should be sufficient.

2.3 Necessary information for archiving the technical reproducibility

- bullets 1 to 6: All this is not new and I claim that most professional modelling groups and companies (in particular benchmarkers from HPC vendors) already follow this strategy. It should go into a class book rather than into a publication for a peer-reviewed journal.

- Bullet 5: this is already described by Song et al. (2012), so why is it repeated here?
- A concise but complete list would be much more helpful.

2.4 Enhancement of the technical reproducibility on the C-Coupler platform

- P4440 L4-12: Wouldn't a directory structure be more appropriate?
- P4441 L11: not relevant for this topic
- P4444 bullets 1 to 8: this is not really feasible. Users who do not have the appropriate hard- and software are lost. Mechanisms to guarantee that the "correct" climate is reproduce would be much more feasible. For further development on a local system and to check certain development steps for bit-reproducibility users can generate their own reference data set.
- P4445 L5: It is a bit weird that suggestions are made for buying a particular product to achieve intercomparison about model results. Why do the authors suggest Intel products in first place and in favour of gcc (gfortran)?

2.5 Experiences and suggestions to the technical reproducibility

- P4448 L10-16: Is this discussion relevant for the understanding of the main topic?
- P4450 L5-L20: the topic covered here are already sufficiently discussed in previous sections.
- P4451 L5-L18: I do not get the point here. Usually, model development goes beyond pure technical improvement. Changes in the dynamic kernel or new physics options will change the numerical results.

- P4452 L1-L10: This topic has been discussed earlier.

2.6 Empirical evaluation

- P4453 L19: Is it the C-Coupler platform that acheives the technical reproducibility, or isn't it rather the numerical model components if not the humans, e.g. the programmers?

2.7 Discussion and conclusion

- P4454 L10-L11: What shall the motivation be for any scientist to use the C-Coupler platform for her own development? I claim that most model development groups have implemented their own strategies for clean model development and they do it for decades with quite some success (see Easterbrook and Johns, 2009 for an example). Of course there is always room for improvements, but it should be stated clearly where the added value is when using the C-Coupler platform.

2.8 Tables

What shall I do if I do not have any such system? I cannot ask my compute centre to buy one.

2.9 Figures

- Figure 1: What is the purpose of the colours? Simple b/w would be sufficient.
- Figure 2: What is the purpose of using a coloured figure here. Instead of using a figure a bullet list or simple table could cary the same information?

3 Citations used in this review

- Clune, T. and Rood, R., 2011: Software Testing and Verification in Climate Model Development, IEEE Software, vol. 28, no. 6, pp. 49-55, Nov.-Dec.
- Easterbrook, S.M. and T. Johns, 2009: Engineering the Software for Understanding Climate Change. IEEE Computing in Science and Engineering, Vol 11 (6), 65-74.
- Ford, R., Riley, G., Budich, R., Redler, R. (Eds.), 2012: Earth System Modelling - Volume 5 Tools for Configuring, Building and Running Models, Series: SpringerBriefs in Earth System Sciences, 97p. http://www.springer.com/earth+sciences+and+geography/book/978-3-642-23931-1
- Guilyardi, E., V. Balaji, Bryan Lawrence, Sarah Callaghan, Cecelia Deluca, Sébastien Denvil, Michael Lautenschlager, Mark Morgan, Sylvia Murphy, and Karl E. Taylor, 2013: Documenting Climate Models and Their Simulations. Bull. Amer. Meteor. Soc., 94, 623–627. doi: http://dx.doi.org/10.1175/BAMS-D-11-00035.1
- Lawrence, B.N., Balaji, V., Bentley, P., Callaghan, S., DeLuca, C., Denvil, S., Devine, G., Elkington, M., Ford, R.W., Guilyardi, E., Lautenschlager, M., Morgan, M., Moine, M.-P., Murphy, S., Pascoe, C., Ramthun, H., Slavin, P., Steenman-Clark, L., Toussaint, F., Treshansky, A., and Valcke, S., 2012: Describing Earth system simulations with the Metafor CIM, Geosci. Model Dev., 5, 1493-1500, doi:10.5194/gmd-5-1493-2012.
- Liu, L., Yang, G., Wang, B., Zhang, C., Li, R., Zhang, Z., Ji, Y., and Wang, L., 2014: C-Coupler1: a Chinese community coupler for Earth System Modelling, Geosci. Model Dev. Discuss., 7, 3889-3936, doi:10.5194/gmdd-7-3889-2014.

• Song, Z., Qiao, F., Lei, X., and Wang, C., 2012: Influence of parallel computational uncertainty on simulations of the Coupled General Climate Model, Geosci. Model Dev., 5, 313-319, doi:10.5194/gmd-5-313-2012.