Geosci. Model Dev. Discuss., 3, C857–C864, 2011
www.geosci-model-dev-discuss.net/3/C857/2011/

Geoscientific Model Development Discussions

Interactive comment on "Automated continuous verification and validation for numerical simulation" *by* P. E. Farrell et al.

P. E. Farrell et al.

david.ham@imperial.ac.uk

Received and published: 7 March 2011

We would like to thank the reviewers for their detailed comments. In particular, they have drawn our attention to what the scope of this paper is and indeed to what the scope of automated verification really is.

1 Scope

On reflection, we feel that what this paper demonstrates is a mechanism for the automated *verification* of a model. Automated systems do not contribute to the *validation* of a model. This is an important but subtle distinction which we freely admit we missed

the first time around.

Automated testing is essentially a guard against the introduction of bugs into a model. A programming bug is an incorrect implementation of the algorithm and is therefore a *verification* error. This is true even if the test which identifies the bug employs external data which might otherwise be used to *validate* the model.

Indeed, validity is a property of the algorithm which the model code implements rather than a property of the code itself. To the extent to which validation tests have established the validity of an algorithm, the repetition of those tests does not further establish validity. If a model which has been validated fails to perform as expected due to a change in the code, this is a *verification error*.

We will adjust the title of the revised paper to remove references to validation and include a description of why this is a verification but not a validation framework in the paper.

2 GCMs and other models

Both reviewers place emphasis on the validation and verification of Global Circulation Models and ask us to retarget the paper on the subject of GCMs. We would like to point out that this is not a paper primarily about GCMs. The current release of Fluidity-ICOM is, as the paper states, a flow solver which is used for the study of flow processes in a range of contexts including, but by no means exclusively, ocean processes. However it does not currently have, nor did we claim that it has, the large scale OGCM capabilities of the well known community ocean models.

Much more importantly, in the context of this paper, GCMs as a class are very atypical of geoscientific modelling. GCMs typically have large development teams and may have professional IT support. As a result, some of the GCMs in use have continuous verification mechanisms analogous to those presented here. Like the reviewers, we are unaware of publications by most of these efforts although we thank reviewer 2 for drawing our attention to Easterbrook and Johns (2009) which does provide some documentation of the process employed at the UK Met Office.

However, most computational science is conducted by individuals and small research groups. The common practice, with which we are sure that the reviewers are familiar, is that each member of the group has their own version of the code living on their own hard drive with or without some ad-hoc backup system. Code versions are handed on from one generation of PhD students to the next and verification and validation tests are, at best, performed intermittently. These models are nonetheless used in published scientific work. As we point out in the paper, we feel that it is very important that appropriate verification of models occurs. We therefore think that it is important that those who are following appropriate software engineering practice publish what they are doing. This both enables confidence in the users of those models with published verification methodology and encourages the adoption of appropriate methodology by the remainder of the community.

In this context, we think that publishing in GMD is particularly appropriate since the publication of development practice is explicitly one of the goals laid out in the original white paper. We encourage those other models also engaging in best practice to similarly document their efforts.

3 Point by point response to reviewer 1

1. As pointed out above, we feel that GCMs are atypical of the majority geoscientific model development and we do not feel that this is a paper primarily about GCM verification. However we take the reviewer's point that GCMs are an area in which automated verification is more common and we will include a brief survey.

However the lack of publications in this area makes an in-depth or comprehensive review a significant research project in its own right. We will include the following in the revised paper:

> We do not claim that this is the first invention of continuous code vertification. Several large, successful geoscientific models undergo similar efforts. The MIT general circulation model (Marshall et al., 1997) has an automated system that runs a verification suite on a variety of different machines. The summary page is available at http://mitgcm.org/public/testing.html. The Unified Model developed by the UK Met Office (Davies et al., 2005) also has an automated nightly verification system, as documented in Easterbrook and Johns (2009). Other research groups have developed suites of tests suitable for use in automated verification; these researchers may well have automated the process, although the authors were unable to find any information about any such automation on their websites. The Modular Ocean Model developed at Princeton (Griffies et al., 2004) documents an extensive collection of verification test cases in their manual (Griffies, 2009). The Regional Ocean Modeling System developed at Rutgers University (Song and Haidvogel, 1994) lists a collection of test cases at https://www.myroms.org/wiki/index.php/Test_Cases. Other large, successful projects do not appear to regard verification as an ongoing process. The contract setting up the consortium to develop the NEMO ocean model (Madec et al., 1998) states that the (quote) "testing and release of new versions" happens "typically once or twice a year" (NEMO Consortium, 2008). The SLIM Ice-Ocean model developed at the Université catholique de Louvain (White et al., 2008) states that (quote) "we ran as few simplistic, highly-idealised test cases as possible. Instead, whenever possible, we tested the model against realistic

flows, albeit often simple ones" (Deleersnijder et al., 2010). As realistic simulations are typically too expensive to run continuously, this suggests that no automated continuous verification system is in place.

2. The reviewer notes that we have not addressed the issue of parameterisation and tuning models for particular applications. We agree and would claim that this is a distinct process which is necessary for model validation (i.e., testing the relationship between the model and the real world system) but which interacts with the verification process only in that it may involve code changes. We do not feel that these code changes are different from any other code changes from a verification point of view. We absolutely concur with the reviewer that this is a very large problem, particularly in the ocean/atmosphere application area; however we concur with the reviewer that this is outside the scope of this paper. Like the reviewer, we are also interested in model tuning methodology but this paper is not on that subject.

3. The reviewer appears to have conflated two unrelated statements in different parts of the paper. We do comment in the introduction that simulation has increased in prominence as computing hardware has become cheaper. We think this is so obvious that it requires no citation. The 1982 reference two pages later is a comment on *software* engineering practice and makes the point that the methodology applied here is not a new idea. In that context, the fact that the paper was published a long time ago is exactly the point.

4. We agree completely with the reviewer that section 3 would be enhanced by the addition of a workflow diagram. We will include one along the lines of that attached to this comment in the revised paper.

5. The selection of test cases presented here is intended to be illustrative of the different test methodologies described in the paper. There are therefore tests against analytic solutions, MMS, laboratory data and DNS model results. This

selection of tests is not intended to be a recipe for validating GCMs because, once again, this is not a paper on this subject. It is also not a comprehensive list of the tests applied to Fluidity-ICOM. This is for the simple reason that there are over 400 tests in the automatic verification system and it would neither be feasible nor would it add to the core points of the paper to attempt to include them all.

   We will revise the introduction to the tests section to make the grounds for the choice of tests presented explicit.

6. We respectfully disagree with the reviewer's assertion that the continuous testing message has penetrated the numerical modelling community. We agree that it is not totally unknown and that some of the GCMs already employ the methodology extensively. We continue to maintain that this is atypical of the community as a whole and especially of the small-scale efforts which typify academic software development.

References

Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A., and Wood, N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere, Quarterly Journal of the Royal Meteorological Society, 131, 1759–1782, doi: 10.1256/qj.04.101, 2005.

Deleersnijder, E., Fichefet, T., Hanert, E., Legat, V., Remacle, J.-F., and Frazao, S. S.: Taking up the challenges of multi-scale marine modelling, http://sites-final.uclouvain.be/slim/assets/files/documents/ProposalWeb_2.pdf, excerpts from a proposal for a Concerted Research Actions (ARC) programme, 2010.

Easterbrook, S. M. and Johns, T. C.: Engineering the Software for Understanding Climate Change, Computing in Science Engineering, 11, 65–74, doi:10.1109/MCSE.2009.193, 2009.

Griffies, S. M.: Elements of MOM4p1, http://www.gfdl.noaa.gov/fms, 2009.

Griffies, S. M., Harrison, M. J., Pacanowski, R. C., and Rosati, A.: A Technical guide to MOM4, http://www.gfdl.noaa.gov/fms, 2004.

Madec, G., Delecluse, P., and Imbard, M.: OPA 8.1 ocean general circulation model reference manual, Paris VI, France, note 11, 1998.

Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C.: A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers, Journal of Geophysical Research, 102, 5753–5766, doi:10.1029/96JC02775, 1997.

NEMO Consortium: The Nemo Consortium Agreement, http://www.nemo-ocean.eu/, 2008.

Song, Y. and Haidvogel, D. B.: A Semi-implicit Ocean Circulation Model Using a Generalized Topography-Following Coordinate System, Journal of Computational Physics, 115, 228–244, doi:10.1006/jcph.1994.1189, 1994.

White, L., Deleersnijder, E., and Legat, V.: A three-dimensional unstructured mesh finite element shallow-water model, with application to the flows around an island and in a wind-driven, elongated basin, Ocean Modelling, 22, 26–47, doi:10.1016/j.ocemod.2008.01.001, 2008.

Interactive comment on Geosci. Model Dev. Discuss., 3, 1587, 2010.

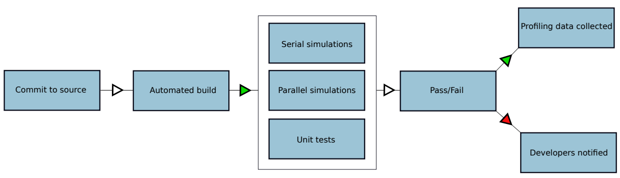

Fig. 1. Work flow diagram of the verification process.