Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-142-RC2, 2016


Geoscientific Model Development Discussions

# *Interactive comment on* "A new and inexpensive non-bit-for-bit solution reproducibility test based on time step convergence (TSC1.0)" *by* Hui Wan et al.

**Anonymous Referee #2**

Received and published: 4 August 2016

Apologies for being so late with my initial comments.

Agree with other reviewers that the paper is overall well-written and clear. I do have some questions and concerns, which are outlined below.

o In the test scenario given the drastically shortened simulation length (5 minutes) with much shorter time steps (1 or 2 seconds), how often are the physical parameterizations (radiation and non-radiation physics) executed? Is it only once for the entire run? If only once, is this a weakness in the overall test design?

o Are all of the outputs from the physical parameterizations that are used in the dynamics applied as tendencies rather than adjustments? Presumably yes, since the



effects of any parameterization that applies its effects as a hard adjustment will not be mitigated by a much shorter time step.

o Is it true that the very rapid growth of a perturbation is due entirely to the physical parameterizations rather than the dynamics? If so, it would be good to point this out specifically, meaning that more traditional means of code verification could still be applied for changes to the dynamical core, assuming the ability to run the model adiabatically.

o Page 2, #50: Regarding the PerGro test using CAM4, presumably the test always fails due to Condition 1 from Rosinski and Williamson (1997): "During the first few time steps, differences between the original and ported code solutions should be within one to two orders of magnitude of machine rounding". If this is correct, it would help to clarify as the primary reason for failure.

o Page 2, #55: It is stated that "Recent versions of the model have become so complicated that rounding level differences in the initial condition can result in very rapid divergence of the simulations". It is not obvious, and no evidence is presented, that code "complication" is a reason for the faster growth. Is it possible, for example, that the initial condition has points which lie on a code branch ("if" test)? Or more generally, perhaps the new physics is driving some quantity such as temperature toward a value which lies on a branch, such as the freezing point of water? If implemented via a tendency equation, the computed value may be one mantissa bit greater than, or one mantissa bit less than, the actual freezing point of water. If a subsequent "if" test applies substantially different algorithms across "true" and "false" branches of a test versus the freezing point, this can be a reason for rapid growth not necessarily related to code complication. This exact scenario was encountered many years ago when testing growth behavior with the relatively simple BATS land model in CAM.

o Page 3, #65: It is stated that "The very fast evolution of initial perturbation is caused by multiple factors". What are those factors? Similar to the previous point, a weakness

of the paper is that it does not describe any of the reasons for rapid growth. There is only speculation that code complication is to blame.

o Page 5, #125: Generally commutative operations are not answer-changing. Instead perhaps the authors mean "associative operations"?

o Page 6, #170: How is the convergence rate of 0.4 calculated?

o Page 9, #285: Definition of the two separate domains is presumably land and ocean. It would help readability to state this up front, and also the reasons for the choice.

o Page 15, #495: If passing the test doesn't guarantee that the model will produce the same climate characteristics, isn't this a weakness of the procedure? I thought the main point of the procedure was to provide a mechanism to enable non-experts to confidently commit roundoff-level code changes to the repository.

The "major revisions" requested involve a much more thorough analysis of the reasons for rapid perturbation growth in CAM4 and CAM5. Speculation about "code complexity" is not adequate. The example cited by this reviewer of rapid growth caused by a simple land scheme (BATS) was really a bug not a feature of the scheme. It would be nice to have some assurance that this possibility (ill-formed or buggy algorithms) has been explored to some extent with the current CAM model.

---

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-142, 2016.