Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-142-SC1, 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.

W. Sacks

sacks@ucar.edu

Received and published: 30 June 2016

This is a clever idea, and the paper is very well written.

I'd like to be convinced that this technique truly has more power than seemingly simpler techniques. For example, can some of the same experiments be redone with this set of runs?:

(1) control: unmodified model with 1s time step

(2) baseline for comparison: unmodified model with 1s time step, with a roundoff-level perturbation in the temperature field

(3) test code: some change in the code with 1s time step

Basically, I'd like to be convinced that the "time step convergence" is truly needed here, and that it truly provides more power than just comparing two versions of the model with a short time step. Does the above, conceptually simpler test give false positives or false negatives in cases where the TSC test gives the correct answer?

I'd also like clarification on the following point: On a continuum from non-answer-changing to answer-changing, I see mention of the following types of changes: (1) bit-for-bit identical, (2) answer-changing only at the round-off level, (3) answer-changing only within the limits of numerical accuracy due to the discrete time step size, and (4) climate changing, according to criteria like SIEVE or CAM-ECT. The TSC test distinguishes changes at level 3 or lower from those at level 4. But is there actually a level in between (3) and (4): changes that affect the model evolution in an appreciable way, but are not large enough to cause statistically detectable changes in climate? It seems that many bugs might fall into this intermediate regime – e.g., accidentally flipping the sign on a minor term in an equation. Do the authors feel that there is a set of changes that falls between (3) and (4), and if so, how do they expect these changes to be categorized by the TSC test?

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