Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-3-RC1, 2016

Geoscientific Model Development Discussions

Interactive comment on "Evaluating Statistical Consistency in the Ocean Model Component of the Community Earth System Model (pyCECT v2.0)" *by* A. H. Baker et al.

Anonymous Referee #1

Received and published: 22 March 2016

This paper tackles the important problem of developing statistical tests of "consistency" when a model is run to evaluate its climate. The fundamental problem is to develop tests on a chaotic system with sensitive dependence on initial conditions: the instantaneous state will obviously diverge under perturbation, but has it moved into a different basin of attraction? Current practice in most climate models is now brute force – run the model "long enough", a period determined by experience, custom, and institutional memory – and evaluate the results to see if the climate changed. This step is often subjective, and is in the eye (and metrics) of the beholder.

Baker et al (2015) attempted to solve this for the atmosphere. There an ensemble based on perturbing the initial conditions was used as the reference (151 1-year runs).

A PCA of 120 diagnostic fields is used (the 150-member ensemble is needed because N_ens > N_var is required for the PCA). That method showed reasonable success when applied to atmos-only runs of CAM, with some rate of false positives.

The ocean is considerably more difficult because of its different dynamical characteristics, and low-frequency modes of variability. This paper attempts to refashion the CAM-ECT method of Baker et al (2015) to CESM's ocean POP.

Unfortunately, this new paper is much less convincing than the old. There are several issues:

1. The setup is of an ocean-ice configuration driven by specified atmospheric BCs (what the authors call a "data model" p10L6). Those BCs are annually repeating, thus eliminating interesting modes of ocean variability, including basic coupled modes like ENSO. (An interesting and rather disturbing paper, Wittenberg 2014 in GRL, suggests that ENSO climate and other lower-frequency modes actually do not meet "climate consistency" as defined in this paper, even with the identical model!)

Furthermore it's subject to strong restoring of the salinity (actually its counterpart, freshwater) at the surface (p8L10). And at the resolution of the test, the model is rather dissipative (p7L35).

Thus the setup is a "low bar" compared to coupled models run for say, IPCC: dissipative and damped, with low-frequency modes filtered out. It's a necessary but not obviously sufficient condition of climate consistency of models as run in practice.

2. The test criteria (< 3\sigma over 90\% of the open ocean) is clearly a function the setup in (1). As is the ensemble size of 40 ... while it's well justified by the arguments of Sec 5 and Fig 13, it's not obvious how you'd extrapolate from the results of this paper to pick one for a different model, or even POP at a different resolution. All of the criteria of passing are very much dependent on the model and the setup, making the result rather weaker than one would hope. It might be strengthened if they could also

run a similar test with the hi-res POP configuration, and see how their criteria change. But perhaps that is too expensive, in which case one would question how this method would be used going forward.

3. The test cases include reproducing and strengthening the results of Hu et al for a different barotropic solver formulation; changing processor count; changing physics known to produce different climate. An additional test that might be interesting is to change the optimization level of the compiler, or to compare across different hardware as in Rosinski and Williamson. A false-positive there would be an interesting result.

4. These points above constitute very major revisions, in the opinion of this reviewer. Therefore I am not listing minor error or language improvements. However, I found Sec 2.2 very puzzling and it contains some fundamental errors. Eq 2 is not a solution of Eq 1, as elementary differentiation will show, nor does Eq 2 satisfy the boundary condition X(0) = X_0. Eq 1 is also an ODE not a PDE (so the \del should be a d). And I think they have confused the steady-state solution of Caya et al (from where Eq 1 is drawn) with what they are trying to establish.

The point of this section seems to be that perturbing initial conditions and perturbing other model parameters can be treated as equivalent. That is true under some conditions, and many papers from the data assimilation literature (including ensemble approaches to DA like Kalman filter) will show you how and under what conditions the two are equivalent. The authors should cite that literature but in this reviewer's opinion, the whole of Sec 2.2 as written, and the associated Fig 3, will have to go.

I hope these comments are constructive, and a stronger paper will emerge from the review process, as the problem Baker et al are attempting to solve is both fundamental and in need of a solution.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-3, 2016.