Geosci. Model Dev. Discuss., 8, C1300–C1304, 2015
www.geosci-model-dev-discuss.net/8/C1300/2015/

Geoscientific Model Development Discussions
Open Access

Interactive comment on "A new ensemble-based consistency test for the Community Earth System Model" *by* A. H. Baker et al.

A. H. Baker et al.

abaker@ucar.edu

Received and published: 3 July 2015

Thank you for your thorough review and suggestions for improvement. We address all comments below.

**Major comments:

(1) Assumption of non-skewed distributions

Indeed, we are looking at 120 variables, and the distribution of all are not exactly symmetric, though most are close to normal. However, the RMSZ scores are not used to make the consistency determination (global means are used), but are provided to scientists as they provide insight in many cases (please also see response to Minor Comment #5).

(2) T-scores vs. Z-scores

We are relying on the global means for our consistency test, but one could additionally calculate T-scores if that was of interest. We note, however, that we do need to have a large ensemble to be able to calculate the PCs in a stable way (that does not depend heavily on the initial perturbations used for the ensemble), so it is likely the use of a T-score (instead of a Z-score) would not provide much additional insight.

(3) Is PCA necessary?

The PCA is necessary in order to determine straightforward and objective pass/fail criteria. For example, assume we have 26 variables labeled A, B, C, ...Z. If variables A - F are linearly dependent, then an issue in the code could lead to 6 variables (A-F) failing. A different code issue could lead to only one variable (e.g., Z) failing. So to design a test cutoff based on how many variables fail would be arbitrary if the dependencies are not considered. Using PCA removes this issue by yielding a linearly independent set, and we can more easily compare the number of PCs that fail between different cases. More importantly, with PCs, we are able to determine (and satisfy) false positive rates, which is important to the CESM scientists. The cutoff of 3 PCs corresponds to the specified 0.5% false positive rate (section 3.4). One can change the parameters to achieve different false positive rates. We do not assume that the leading PCs are more important, as we have found subtle errors in the code that do not affect the large PCs, but are relevant (i.e., we want to detect them) nevertheless.

We agree that looking at 50 components (instead of 120) does not provide much computational savings, but it is not clear that looking at the remainder of the PCs would be beneficial as the first 50 represent nearly all of the variance. However, we note that the 50 is only the default for the CESM-ECT python code, and a user can chose to look at more or less if desired.

**Minor comments:

GMDD
8, C1300–C1304, 2015

(1) What variables?

The two variables with large distributions are TAUY and VU (meridional surface stress and meridional flux of zonal momentum), and we added these to the revised text.

Our emphasis is intended to be on the procedure itself and not on specific variables, as different CESM configurations output different variables by default (and an individual user may wish to alter the output variable list). Note also that the CESM-ECT python code allows the user to specify variables that should be excluded from the PCA analysis, so there is a lot of flexibility around numbers of variables. However, we will add an appendix to the revised manuscript so that the user can replicate our list. Note that it is quite similar to the example output (monthly) variable list that can be found in the CAM documentation (e.g., http://www.cesm.ucar.edu/models/cesm1.2/cam/docs/ug5_3/hist_flds_fv_cam5.html).

(2) Averaging time

For this study we are looking at annual temporal averages, which is indicated in the last paragraph in Section 3.2. Please let us know if further clarification is needed.

(3) Perturbations

We are perturbing the entire temperature field and have modified the text to clarify this point.

(4) Incorrect citation formats

Thank you for pointing these out. We have corrected them in the revision.

(5) Better roadmap for the methods

We agree that Section 3 needs to clarify that Z-scores are not used to evaluate consistency and emphasize that only the PCs of the global means are used by CESM-ECT. We discuss the Z-scores in the manuscript as they are included as optional output from CESM-ECT for legacy reasons as many scientists are accustomed to viewing variable

Z-scores. The revised version clarifies this point, and we rearrange the discussion so that the Z-scores are mentioned *after* the discussion of PCs.

(6) Validation vs. Evaluation

There is indeed much disagreement in the community over the use of the terms "verification" and "validation". We were not aware of the two references that you cite, but have included them and the "evaluation" option in the revised manuscript.

(7) "The result is that the tolerance for rounding accumulation..."

While we have not performed a thorough analysis of the rapid accumulation of rounding errors in the entire CAM5 package, we have analyzed individual parameterizations that produce large growth. For example, the radiation code (RRTMG) uses a Monte Carlo method to evaluate the impact of clouds on the radiative fluxes. The seed for the random number generator is based on low order bits from the model pressure field (the high order bits are lopped off). Thus a roundoff level perturbation applied to the temperature field produces roundoff level changes to the pressure field, which in turn produces a completely different sequence of random numbers. The resulting differences in computed radiative fluxes are not at the roundoff level, but rather arise from being based on two very different states arising from different random number sequences. An initial RMS temperature difference of 1.e-12 becomes a temperature difference of 1.e-3 after a single time step.

(8) Should say “computationally” ...

We have corrected this typo in the revision.

(9) What is “ne= 30”?.

"ne" indicates the number of elements on the edge of a cube in the spectral-element version of CAM, and we will add this information to the revision. Note that "ne=30" is approximately a 1 degree grid (already mentioned in 4.1).

GMDD
8, C1300–C1304, 2015

Interactive Comment

Interactive comment on Geosci. Model Dev. Discuss., 8, 3823, 2015.

GMDD

8, C1300–C1304, 2015