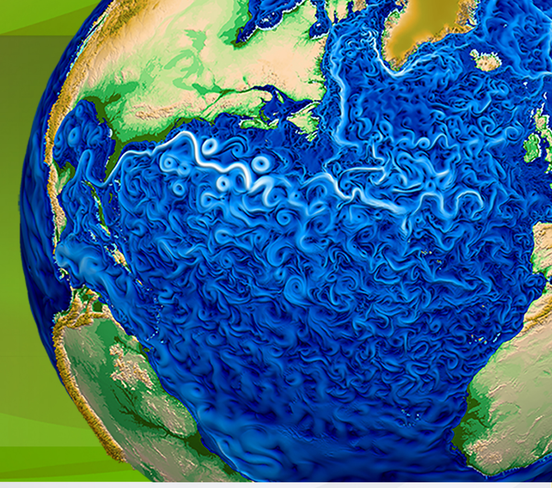
P: Parallel Ensemble Simulations for ACME Performance and Verification

Abigail Gaddis, Matthew Norman, Katherine Evans, Salil Mahajan, Mark Taylor



Issue

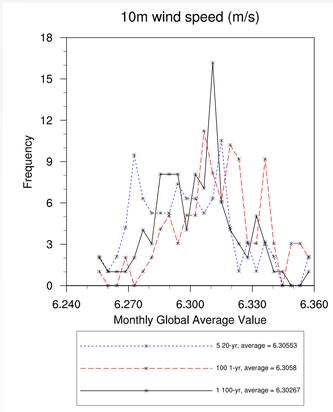
Summary:
Metrics were developed to assess whether ACME code modifications are climate-changing. As a baseline, a statistical test was applied to three sets of similar simulations including a large ensemble of short simulations run in parallel. Unexpected statistically significant differences between variables in the sets were found, leading to identification of a bug in the in-line interpolation routine.

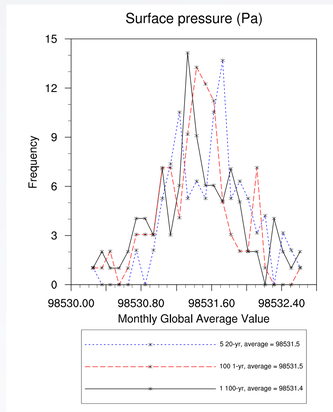
Background:
A large ensemble of simulations run in parallel for a short period of time is more computationally efficient and allows for quick model verification, if the climate generated is similar to that of a longer simulation. A framework for automatic, efficient model testing requires a set of Pass/Fail type tests interpretable without consulting climate scientists. The status quo for climate simulation is to perform a small ensemble (order five) of long simulations (roughly a century). To succeed in feasible time, a throughput constraint of five Simulated model Years Per wallclock Day (SYPD) is generally accepted as necessary. To achieve this using CAM-SE, the model is scaled over many processing elements, and work per node is small. Improvements in model performance from more efficient computer architecture use are important for throughput, scalability, and receiving large computing allocations in the future. Additionally, testing non bit-for-bit changes to the model becomes possible and efficient with higher parallelism.

Variable	K-S AB	K-S BC	K-S AC
PSL	0.95	1	0.96
PRECC	0.602	0.483	0.634
PRECL	0.624	0.692	0.66
U10	0.39	0.772	0.862
QREFHT	0.886	0.985	0.979
CLDTOT	0.176	0.141	0.922
FLUT	0.827	0.902	0.974
TS	0.926	0.963	0.838
BURDENSO4	0.0137	0.00729	0.937
Z3 (995mb)	0	1.46E-06	0.291

A: 100 year, B: 1 year, C: 20 year
Selected KS test p-values for monthly average of different model variables. Data are not globally averaged. If $p < 0.01$, the null hypothesis that the data is from the same distribution is rejected at 99% confidence.

Probability distribution functions of monthly average wind speed at 10mb (Left) and surface pressure (Right). The data have been interpolated using the in-line interpolation scheme and a global weighted average is calculated from the interpolated latitude-longitude values. The PDFs are similar for all three sets of simulations





Solution Attempts and Methods

Simulations and Performance:
This is part of a pilot study investigating the merits of an ensemble-based approach to climate science and model evaluation rather than the traditional single, long simulation approach. Along with a single 100-year atmospheric simulation with annually cycled ocean conditions, we ran two additional experiments: five 20-year runs and 100 one-year runs (98 of which completed successfully) of the same configuration. The ensembles can be run in parallel; they completed in merely 12 hours from job submission, whereas the single 100-year and five 20-year simulations took roughly five weeks a piece end-to-end due to queue wait times, and job / node failures that inhibited automatic resubmission.

Statistical measures of similarity:
- The Kolmogorov–Smirnov test was chosen because it is non-parametric and compares the largest difference between the entire probability distribution functions of each variable, rather than only comparing averages or standard deviations.
- Of 98 variables tested, only 8 reject the null hypothesis that the 100 1-year runs are the same as either the 100 year or the 5 20 year runs
- Only 2 reject the null hypothesis that the 5 20 year and 1 100 year simulations are the same
- However, we expected full agreement. Also, all p-values are consistently lower for the 100-1 year comparisons

Ideas and Lessons Learned

Conclusions:
- Running ensembles in parallel increases performance greatly
- A test suite is needed to clarify what is meant by “climate-changing” code modifications
- KS tests and other statistical measures provide ways to quickly check code validity across model variables
- There is likely a bug in the in-line interpolation routine in CAM-SE
- Differences between simulations run in parallel and serially are few, but similarity should be verified with further simulations

Ideas:
- In-line interpolation routine needs further investigation
    - Probably related to variable initialization to -999
    - Spurious numbers appear more frequently in parallel simulation output
    - Find and fix the bug
    - Until the issue is fixed, do not use (no interp in user_cam_nl)
- Baseline and tuning cases should be as simple as possible

ACME Accelerated Climate Modeling for Energy

For additional information, contact:
Staff Member
Title
Institution or Organization
(555) 555-1234
staff.member@email.gov

climatemodeling.science.energy.gov/acme

U.S. DEPARTMENT OF ENERGY