R: Computational Benefits of an Ensemble-Based Approach to Climate Modeling and Testing at Scale - Matthew Norman, Abigail Gaddis, Valentine Anantharaj, Kate Evans, Salil Mahajan, Patrick Worley

Computational Motivations

Current Challenges of Climate Modeling at Scale

- Throughput needs force ACME to use many processors at high resolution
- CAM-SE on 28km grid: only 32 columns of elements per node
- MPI wait consumes 40% of CAM-SE runtime, and it still only runs 1 SYPD
- GPU speed-up of CAM-SE's tracer transport is 3.3x, 2.6x, 1.8x, and 1.05x at 384, 128, 64, and 32 columns of elements per node, respectively
- End-to-end time for running 100 years of baseline climate was five weeks

Dominant Needs for ACME Going Forward

- More elements per node are necessary for GPU performance and lower MPI
- Faster throughput is needed to get high-resolution answers in feasible time

Ensembles Increase Throughput and Elements Per Node

Experiment Setup & Computational Results

Experimental Setup

- Three atmosphere-only CAM5 one-degree experiments starting at year 1850: (E1) A single 100-year run; (E2) five 20-year runs; and (E3) 100 one-year runs
- Ensembles have a randomly perturbed initial temperature field (mag. 10^{-14})
- Results confirm distinct climates within 23 days → first month discarded for E3

Computational Results

- E3 used **5x** more elements per node and **25x** greater throughput than E1
- E3 used **25%** of Titan and finished merely **12 hours** from submission, whereas E1 and E2 took roughly five weeks a piece to complete.
- E3 used only **60%** of the core hours on Titan that E1 and E2 used
- Two of the 100 ensembles failed for reasons that are being investigated

Computational Implications

- The ensemble experiments consume a "capability portion" of Titan, thus receiving **queue priority** and a better chance at winning **large allocations**
- The GPU tracer transport routines will run **3x** faster in the ensemble runs because of having more elements per node
- Over **50x** reduction in user time spent maintaining the runs
- **40%** improvement in usage of available core hours

Current Scientific Results

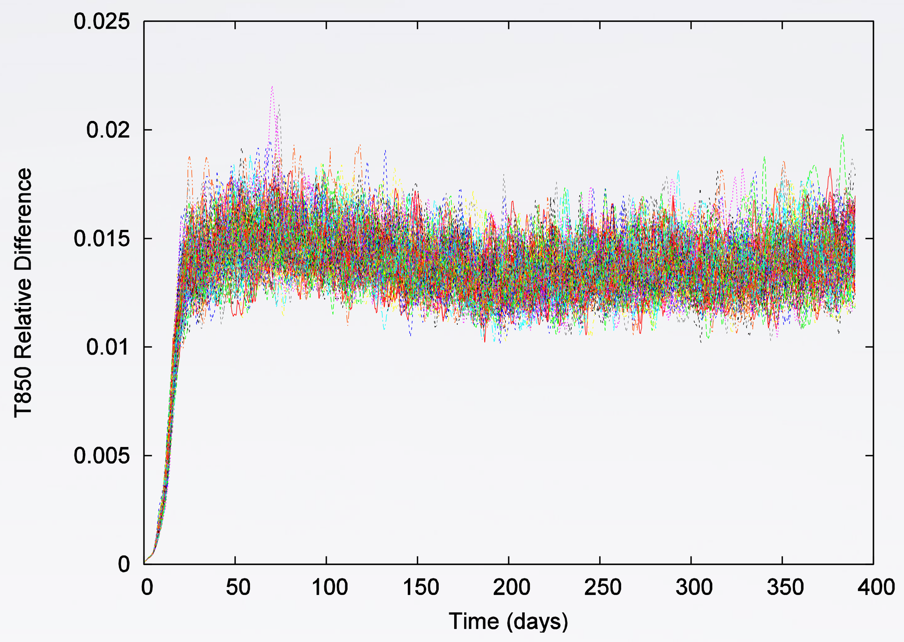

Figure 1: (above) L1 norm of relative difference in 850mb temperature between each of the 100 ensembles over time. Plot demonstrates ensembles are distinct after roughly 23 days.

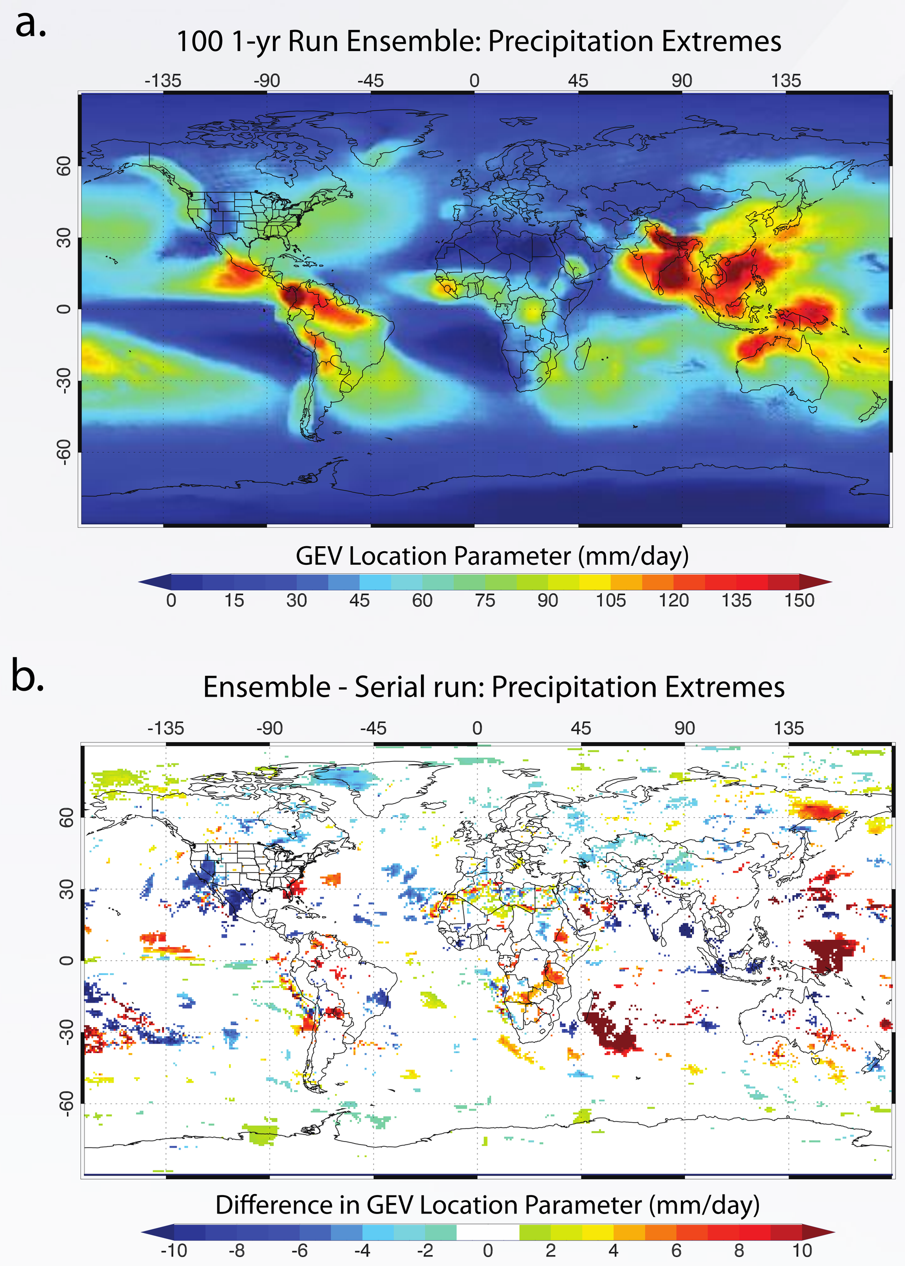

Figure 2: (above) (a) Precipitation extremes for the E3.
(b) Statistically significant difference in precipitation extremes between E1 and E3

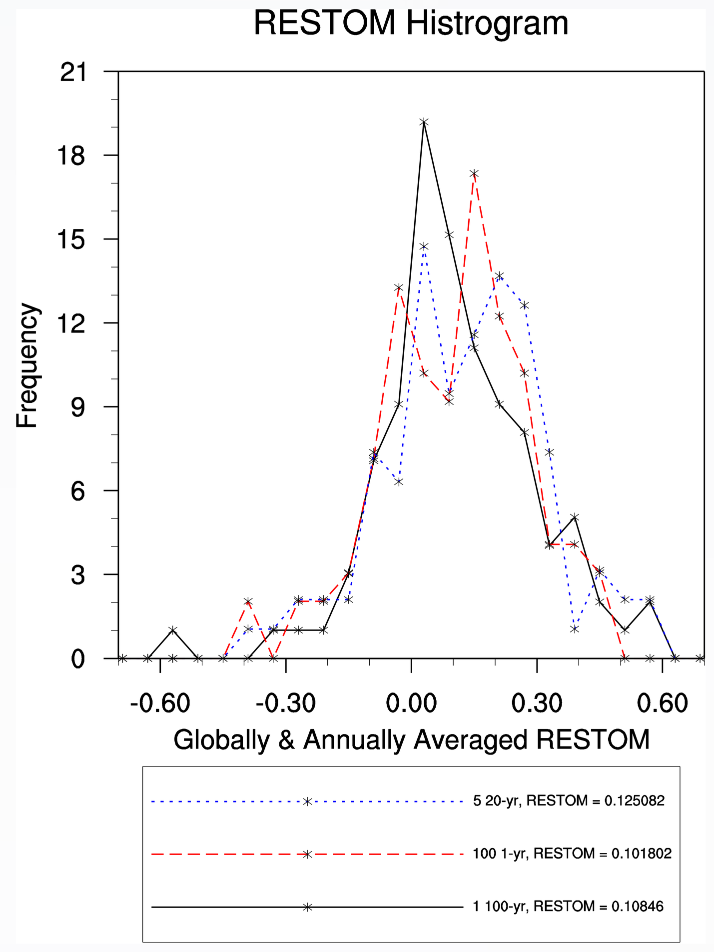

Figure 3: (left) (a) Histogram of globally and annually averaged radiation residual at the top of the model for E1, E2, and E3

Future Work

- More comprehensive similarity tests including externally developed tools
- Inclusion of the ocean and other processes in the ensemble framework
- Running with the GPU-ported portions of ACME as well as hi-res runs
- Developing cross-validation suite between compilers, flags, and machines

This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

ACME Accelerated Climate Modeling for Energy

For additional information, contact:
Matthew Norman
Computational Climate Scientist

National Center for Computational Sciences
Oak Ridge National Laboratory

normanmr@ornl.gov

OLCF Oak Ridge Leadership Computing Facility

U.S. DEPARTMENT OF ENERGY

climatemodeling.science.energy.gov/acme