

## Computational Benefits of an Ensemble-Based Approach to Climate Modeling and Testing at Scale

Matthew Norman, **Abigail Gaddis**, Valentine Anantharaj, Katherine Evans, Salil Mahajan, Patrick Worley

The status quo for high-resolution climate simulation is to perform a very small ensemble (order five) of long simulations (roughly a century) for various scenarios arising from IPCC specifications. 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, CAM-SE is used and is scaled over many Processing Elements (PEs), and work per node is very small. At this scale, parallel data transfer overheads are 40% of the total runtime or more, and there are very few threadable indices to use on an accelerator (e.g. Graphics Processing Unit, GPU). Also, even at these scaling limits, ACME is barely achieving a “capability-scale” portion of Titan (i.e., > 25% of the machine), and throughput is still only around one SYPD for the 28km-mesh water cycle experiment targeted by ACME. This, in turn, means (1) a low benefit from using GPUs and (2) poor usage of computer allocations, and (3) less likelihood of receiving large computing allocations in the future.

This is 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 of the same configuration to discover and quantify the statistical differences between the two approaches and begin the process of understanding what science questions we may be able to answer in this manner. Other modeling centers have similar efforts underway, focused largely on developing tools to judge similarity between ensemble sets. This study is focused on the computational aspects. The benefit of using many separate ensembles is that they can be run in parallel. With one-degree mesh experiments, we used five times more columns of elements per node, used only 60% of the core hours, achieved an aggregate throughput 25 times faster, and realized Titan’s queue benefits for capability-scale jobs, which automatically receive priority boosts. The 100 ensembles 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. This is largely because small jobs on Titan cannot run more than two wallclock hours at a time, while large jobs can run up to 24 hours.

The key question is: did we achieve the same climate with the ensemble strategy? To address this, we generate histograms of globally and annually averaged variables of importance to climate and compare them against one another with RMSE and statistical tests. We will likely also leverage analysis tools developed elsewhere for judging similarity between ensemble sets. The relative scope of our immediate analysis is informed by our ACME task focus on computation and performance, and the data is available in storage for any further analysis.