

CGD SEMINAR



DATE: Tuesday, May 22, 2018
TIME: 11 a.m.
LOCATION: NCAR, 1850 Table Mesa Drive
Mesa Lab, Main Seminar Room
TITLE: Optimizing data compression for CESM
SPEAKER: Allison Baker, NCAR CISL

ABSTRACT:

Significant increases in computational power in recent years have enabled climate model simulations to run with higher resolutions and higher throughput, resulting in increasingly larger data volumes and straining computing center storage resources. Storage limitations are negatively impacting science objectives by forcing scientists to run fewer or shorter simulations and/or output data less frequently. Data compression offers a potential means of mitigating the data volume problem, and NCAR has been investigating applying compression to data from the CESM. Lossy data compression schemes are able to achieve substantial storage reductions, but by definition are unable to exactly reproduce original values. Therefore, striking a balance between meaningfully reducing data volume and preserving the integrity of the simulation data is a critical albeit non-trivial task, particularly given the large and diverse set of climate variables. In this talk, we first review data compression in general and describe the challenges and concerns when compressing climate data from CESM. We then discuss approaches and metrics for evaluating data loss due to compression, with examples from the CESM Large Ensemble Community Project. Finally, we discuss our current approach of customizing compression for each variable in order to compress optimally based on the characteristics of that variable. In practical terms, "optimal" compression may mean compressing some variables (or locations) quite aggressively while others are compressed quite carefully (or not at all).

Live webcast: <http://ucarconnect.ucar.edu/live>

For more information, contact Barbara Middlebrook, email bmiddleb@ucar.edu, phone: 303.497.1366