Data Management Plan

Overview:
All model simulations and data products created, processed, or analyzed under this proposed work will be managed in compliance with NSF Grant Proposal Guidelines. This proposal data management plan will ensure efficient and safe data management during analysis and research activity, and will outline our guidelines to providing well-documented access to model outputs and data products used in this proposed work in accordance with OCE policy.

Data Standards and Types:
This proposal will mainly lead to the production and processing of model outputs from the CESM and TPOSE experiments described in this description document. CESM and TPOSE model ouputs are provided as gridded outputs at 5 day-mean frequencies, and are created and made available for dissemination in NetCDF Format.

Software and Derived Models
Both model code for the CESM and MITgcm models used to run the model simulations are based on Fortran and are freely available online. Processing and analysis of model ouputs will primarily be conducted using Python and the NetCDF Operator (NCO), both of which are also freely available online. Lagrangian analysis will be conducted using a PARCELS, a Python-based Lagrangian simulator package that is available on the developer's web platform.

Computing Environment and Collaborative Platforms
CESM model outputs will be stored and analyzed on NCAR Cheyenne, a 5.34-petaflops, high-performance computer. The eddy resolving and coarse resolution of CESM will be processed and stored in the Globally Accessible Data Environment (GLADE), a high-performance and secure GPFS shared file system technology built for analysis, visualization resources, and secure data sharing across collaborators. To ensure data security and integrity, model outputs will be backed up and archived in the High Performance Storage System (HPSS), a robotic tape libraries with high storage capacity in the NCAR-Wyoming Supercomputing Center (NWSC).

The Cheyenne server will allow direct and secure data, code, and model outpout sharing and transfer by authorized collaborators. Several of the PIs, including the lead PI, have existing accounts under GLADE/Cheyenne, which will make collaborative analysis efficient.

Ocean state estimates simulations will be conducted on the Stampede computing system at the Texas Advanced Computing Center. A local secure server-class machine at SIO will be used for analysis, visualization, storage, and archival of the B-TPOSE output.

Data Archiving and Distribution Methodology:
Model outputs from the CESM simulations will be archived in HPSS, and made available online on the NCAR CESM Climate Data Guide website for access by the scientific community and the general public. B-TPOSE outputs from the Ocean State Estimate model simulations will be made publicly on the UCSD ECCO website available via a server hosted at Scripps Institution of Oceanography. Metadata, and guidelines for accessing data will be posted along with the model outputs.

Contact information through email address will also be provided for questions and comments. Model output files from both CESM and TPOSE will be provided in the common and widely used NetCDF (Network Common Data Form) format. NetCDF is a flexible, selfdescribing format that allows for efficient bundling of data along with related metadata.

Publications and Reproducibility:
The PIs expect several publications to result from the proposed work, organized around the science

questions outlined in the project description. Results from this proposal will be disseminated in peer-reviewed journal publications and presented at national and international conferences. All data and model simulations arising from or used in these publications will be published on an online repository (zenodo), with direct links from the publications webpage and documents.

Code, analytical methods, datasets and model outputs, and detailed guidelines within code to ensure reproducibility of results will be made publicly available and archived on an online repository (zenodo). Access to full model outputs from CESM and TPOSE used in this proposal will be provided in the model groups' respective websites, shared widely to the scientific community and the public, and will contain appropriate file transfer links and guidelines.

Curriculum Materials:

Lesson plans, hands on experiments, and visualization software developed for and with the NGSS early implementer teachers will be made available on the California Science Teacher Association (CSTA) website, similarly to previous efforts and shared with teachers throughout California at various workshops held by the CSTA.