

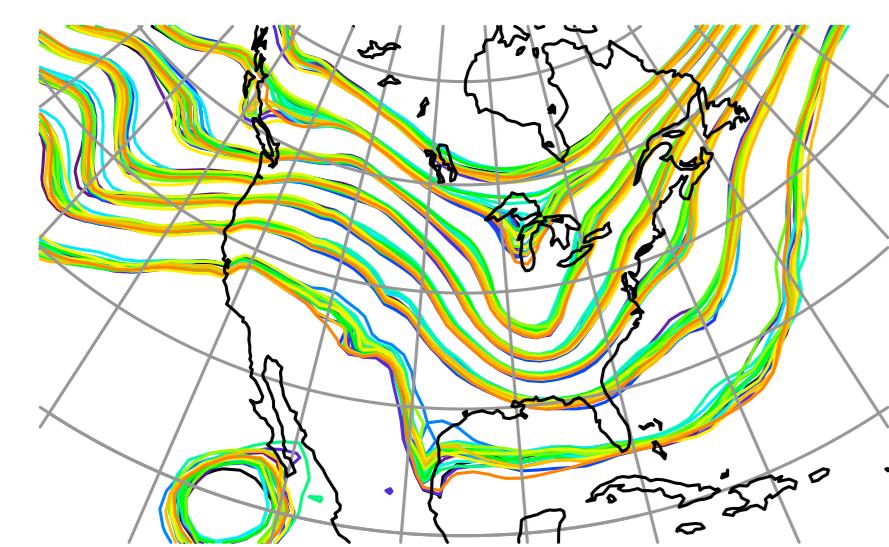
A CAM6+CLM5 Ensemble Reanalysis for Forcing

CESM Ice Models and Data Assimilation

K. Raeder¹, J. Anderson¹, M. El Gharamti¹, T. Hoar¹, B. Johnson¹, N. Collins¹, J. Steward²

1 National Center for Atmospheric Research; CISL/DARes, Boulder, CO;

2 Spire Global Inc., 1690 38th St, Boulder, CO, 80301, USA



1. Overview

There are new ensemble reanalysis datasets available which enable (ensemble) hindcasts in CESM's nonatmospheric components to be forced by actual atmospheric conditions. This forcing is in the form of CESM's "data atmosphere" (DATM) forcing files. This is also a crucial input for ensemble data assimilation using those components. The atmospheric ensemble provides variability, which helps to maintain the ensemble spread in the assimilation process of nonatmospheric components. The spread in the hindcast or assimilation results provides a justifiable estimate of the uncertainty. In the end, the reanalysis datasets enable direct comparison of CESM model states with related observations for dates of interest.

2. CAM Reanalysis

The Data Assimilation Research Testbed (DART) is a community facility for ensemble data assimilation developed and maintained at NCAR. Data assimilation combines short model hindcasts with observations to produce an ensemble description of the state of the Earth system. This description is a balance between the observations and the model formulations, which takes into account the uncertainties in each. This process of confronting the model with observations facilitates model evaluation and improvement. This 80-member DART reanalysis using CESM2.1's version of the Community Atmospheric Model (CAM6-FV) at 1° resolution spans 2011-2019. The compset = HIST_CAM60 CLM50%GBC-CROP_- CICE%PRES_DOCN%DOM_MOSART_- SGLC SWAV. The data ocean is daily, $1/4^{\circ}$ AVHRR. Every 6 hours we assimilate up to a million observations; most conventional atmospheric observations, temperature from AIRS, and atmospheric refractivity (\approx density) from GPS satellites (Figure 1).

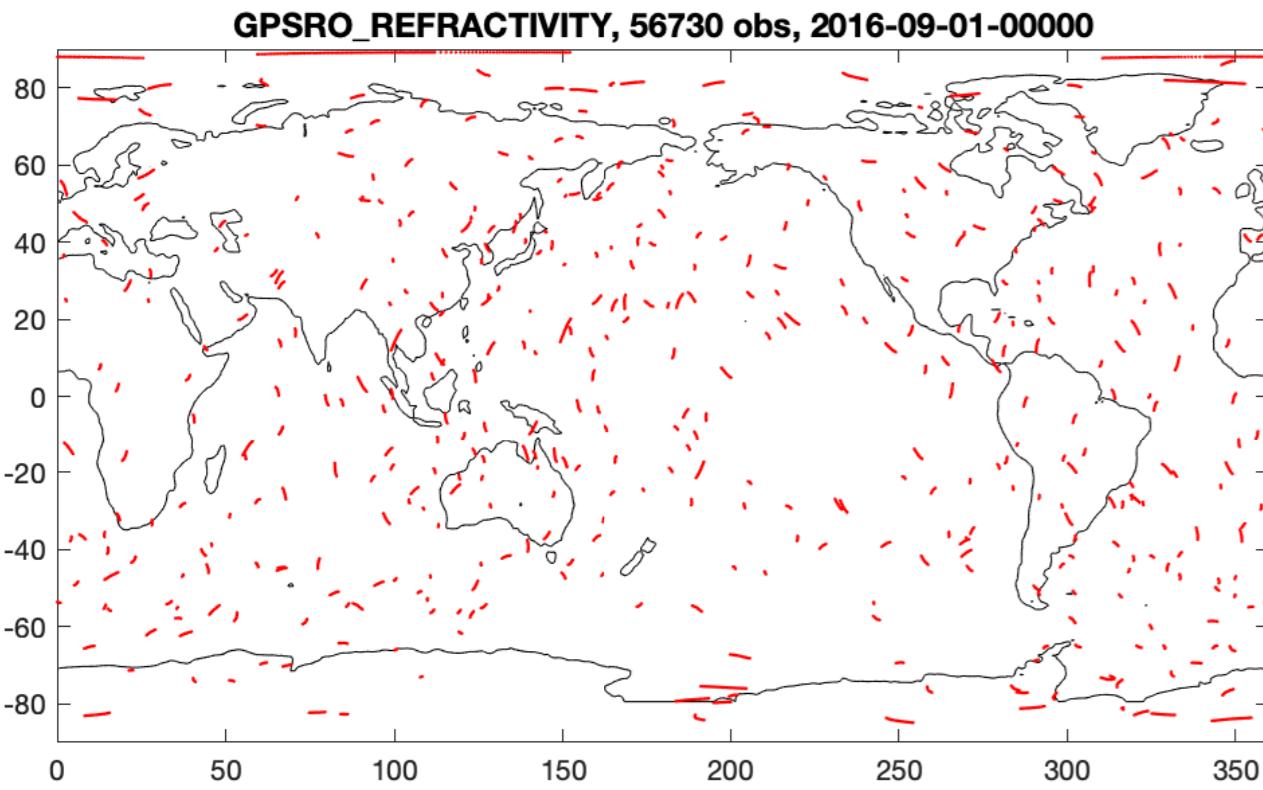


Figure 1: The horizontal locations of the refractivity observations from a variety of Global Positioning System receivers (e.g. COSMIC) in a 6 hour span centered on the date shown. This time has 56,730 of these observations.

The observations directly change the CAM+DART "state vector" {PS, T, US, VS, Q, CLDLIQ, CLDICE} and indirectly change all of CAM's other variables and the active component models; CLM, CICE, MOSART. CAM's "other variables" include the surface fluxes which are passed to the other components (e.g. Figure 3).

3. Evaluation of the Reanalysis

An essential part of the Reanalysis was an evaluation of the performance by comparing CAM's estimates of the observations with the actual observations after every month of assimilation. Examples are shown in Figure 2 for the polar regions. Our evaluation used different regions, included additional statistics such as bias, and included vertical profiles of all of the statistics. The "total spread" shown in the figure is a combination of the uncertainties of the model (ensemble spread) and the observations (observation error estimates). This is useful as a measuring stick to judge whether the difference between the model and observations is meaningful. For example, if the RMSE is smaller than the total spread, then we judge that the model and observations are consistent with each other and the assimilation is performing as hoped. The monthly diagnostic pictures are included in the datasets. They can be used to judge the suitability of the reanalysis for a chosen application. DART has software for making similar evaluations, based on users' needs.

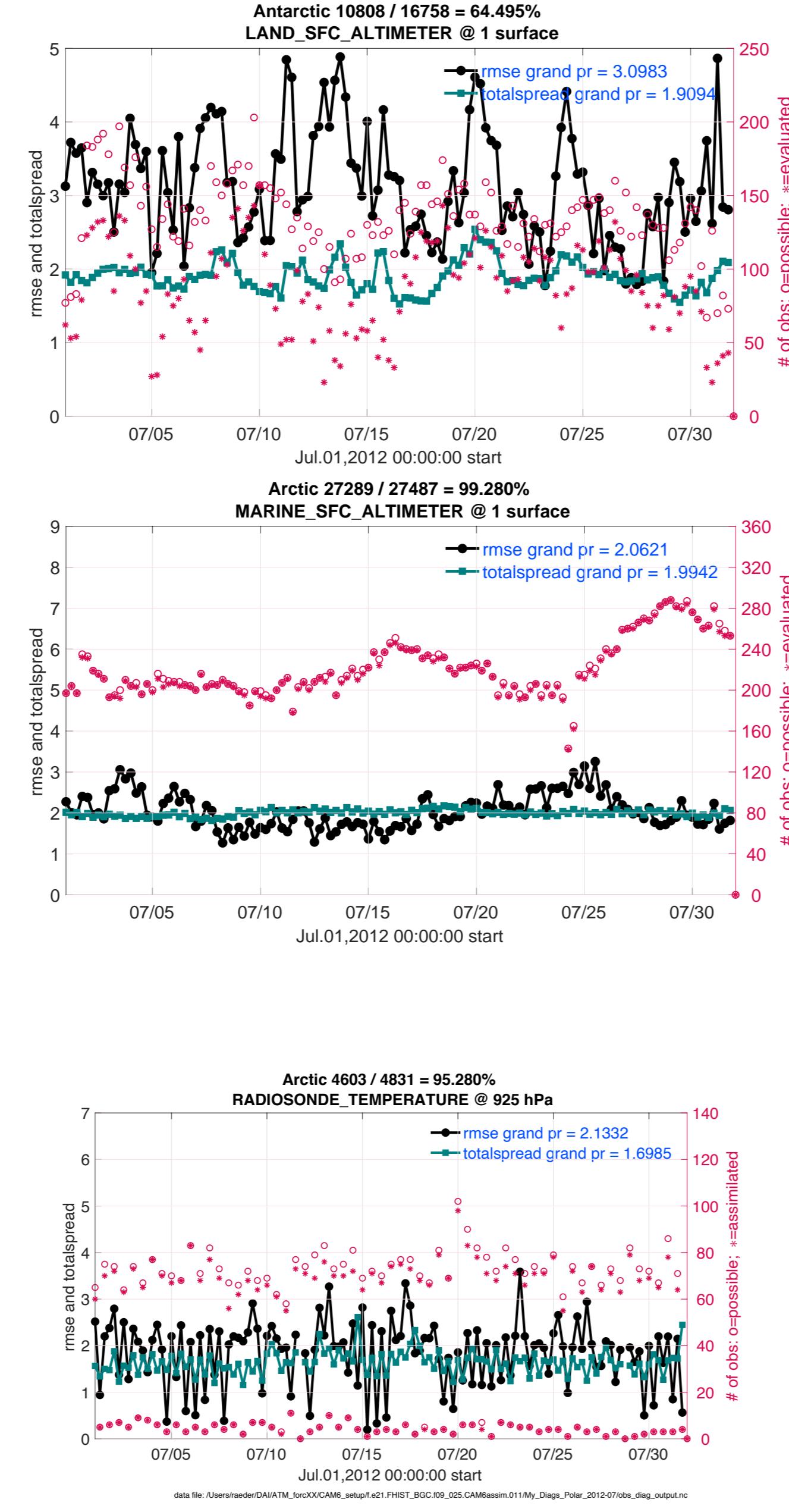


Figure 2: Root mean square error (black) and total spread (green) (left axis) of the reanalysis observation estimates relative to the actual observations in the Arctic ($lat > 66 N$) or Antarctic ($lat > 66 S$), for July 2012. The red circles mark the number of available observations (right axis) and the red * mark the number used or evaluated. The title also lists the total observations in the month ("available/used = %"), the observation type, and atmospheric layer.

4. Data Products Useful for Surface Component Models

The observation space diagnostics and the datasets described below are freely available in NCAR's Research Data Archive ds345.0 (compressed) and in /glade/collections/rda/data/ds345.0/cpl/unzipped (uncompressed).

The most relevant dataset for surface model hindcasts is the ensembles of atmospheric forcing. They have CAM's 1° horizontal resolution and hourly, 3-hourly, and daily frequencies. The ensemble represents the uncertainty in the knowledge of the atmosphere; each member is equally valid (See Figure 3).

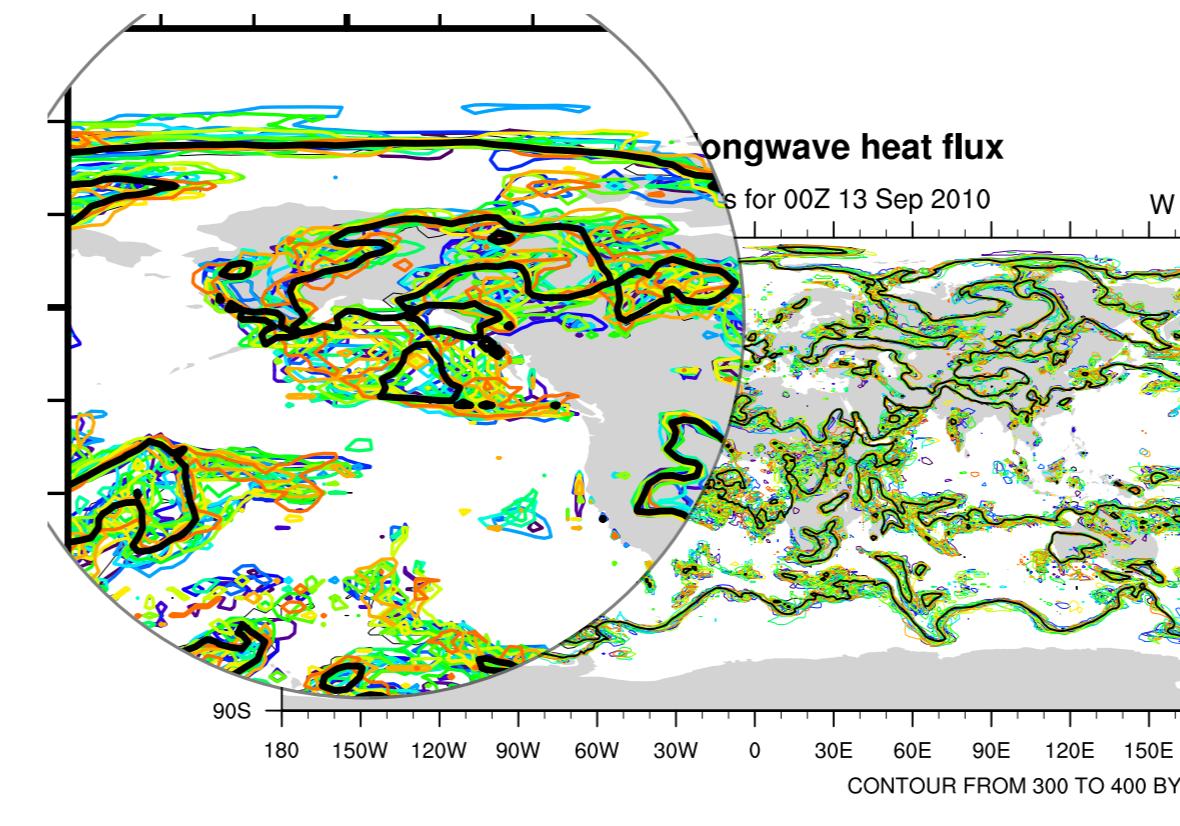


Figure 3: The variability of the downward longwave heat flux at the surface. 20 of the 80 members are shown here. The colors represent individual ensemble members.

These forcings are in the form of DATM coupler history files, which can be easily used with all of the other CESM components; CICE, CISM, CLM(CTSM), POP(MOM), They are well suited for ensemble forecasts, sensitivity studies, ensemble data assimilation, and model parameter tuning efforts. They are packaged into year-long files, one member per file, for convenient use. The forcing files contain the variables which are written when histaux {a2x3hr, a2x24hr, a2x1hr, a2x1hr, r2x} = .true., which were the settings used in the J1850G CISM spinup case, plus river forcing of other components. J compsets have all components active, except for the atmosphere,

A third dataset is an 80 member ensemble of restart files for CICE, CAM, CLM, and MOSART, which is available weekly. They are consistent with CAM's representation of the actual atmospheric states. These can be used as initial conditions for hindcasts of historical environmental events.

5. Using the DATM forcing Files

Figure 4 shows some of the data motion in an assimilation which uses DATM files to force POP. Currently CLM and CICE can be substituted for POP in that figure because there are interfaces to DART for those components. If the DART branch of the figure is left off, then it represents an ensemble hindcast with no assimilation. Then any component (except CAM) can replace POP. Finally, the ensemble size can be 1, which is just a traditional hindcast forced by a single data atmosphere.

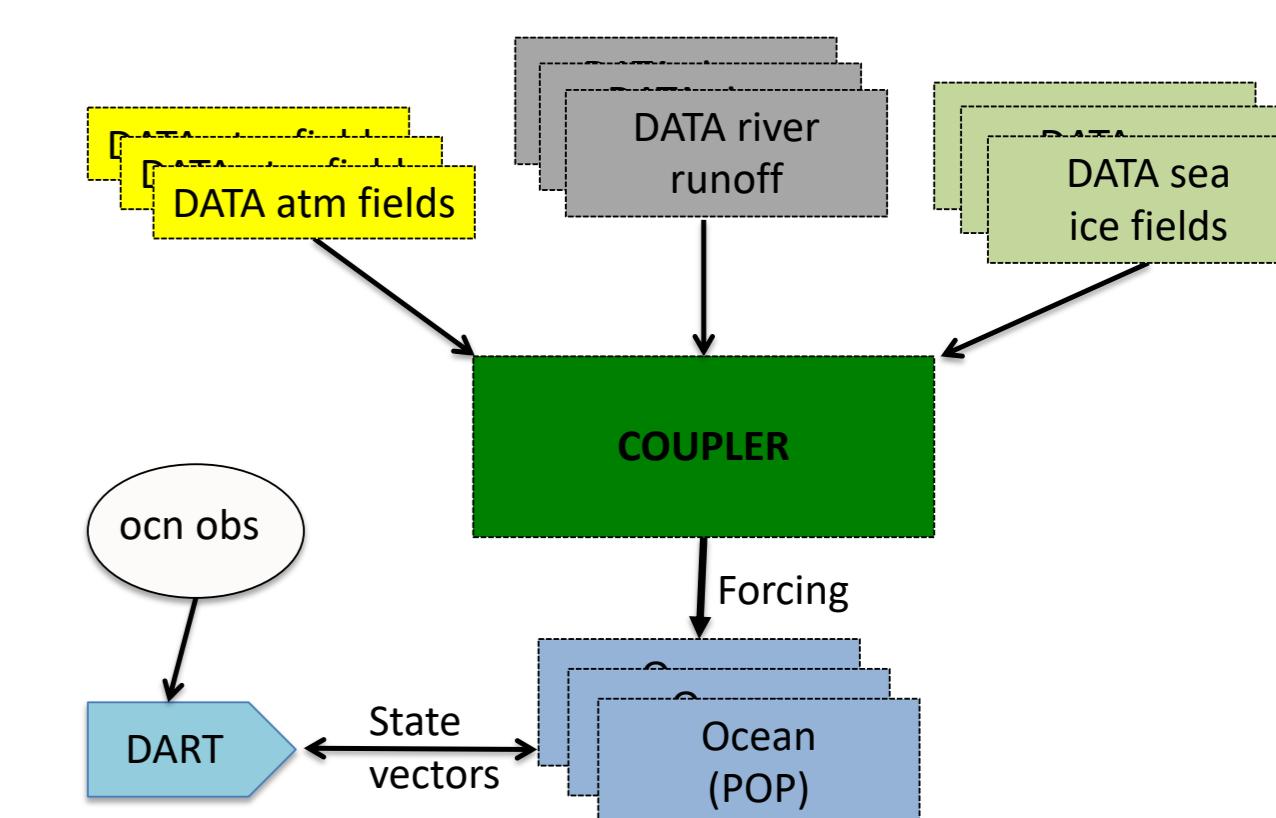


Figure 4: CESM's coupler, run in multidriver mode, ingests an ensemble of DATM forcing files and passes the data to the active components which need it.

6. Plans

- Extend the Reanalysis through 2020
- Collaborate with people in the polar community to use the data for:
 - recreating recent historical polar conditions in CESM,
 - model improvement through identification of biases,
 - investigating model response to variability in the forcing.
- Other ideas from the polar community.
- Evaluate the feasibility and usefulness of developing an interface between CISM and DART.

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. We would like to acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

