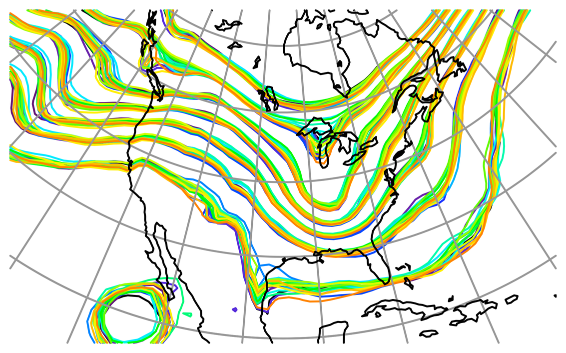

Ensemble Data Assimilation for Very Large Atmosphere, Ocean and Coupled Models with the Data Assimilation Research Testbed

K. Raeder, A. Karspeck, G. Romine, N. Collins, J. Hendricks, T. Hoar, J. Anderson, H. Kershaw, S. Karol, C. Schwartz, R. Sobash

National Center for Atmospheric Research
Boulder, Colorado, USA
dart@ucar.edu

1. DART Manhattan Highlights

- ★ Able to handle much larger model states. This is needed for higher-resolution and/or strongly-coupled DA with multiple components. Distributing the model state across all tasks during the entire filter run means no single task must store the entire state at any time.
- ★ One-sided MPI communication allows tasks to request remote data items from other tasks without interrupting their execution or arranging which data items will be needed in advance.
- ★ Computing the forward operators for all ensemble members at the same time leads to code that vectorizes better.
- ★ Native NetCDF support eliminates the conversion steps between NetCDF model files and a DART binary format file. This also reduces the high-water mark for disk requirements.
- ★ Ensemble data can be read and distributed across all tasks on a variable-by-variable basis, reducing the maximum memory requirements.
- ★ Diagnostic state space files are now written in parallel with state-space restart files, resulting in faster I/O and lower memory requirements.
- ★ Support for externally computed, forward observation operators.
- ★ Support for per-observation-type localization radii.

2. DART is ...

The Data Assimilation Research Testbed (DART) is an open source community software facility for ensemble data assimilation developed at the National Center for Atmospheric Research (NCAR). DART works with a wide variety of climate and weather models and observations and has been free and publically available for more than 10 years. Building an interface between DART and a new model does not require an adjoint and generally requires no modifications to the model code. DART works with dozens of models of varying complexity, including (but not limited to):

- weather models, e.g. WRF, COAMPS, COSMO, MPAS Atmosphere,
- components of climate models, e.g. CAM, POP, CLM, WACCM, MPAS Ocean, ROMS, MITgcm-Ocean, GCCOM, JULES, FESOM, CICE5,
- atmospheric chemistry models, e.g. CAM-CHEM, WRF-CHEM,
- ionosphere/thermosphere models, e.g. TIEGCM, GITM,
- low-order and simple research models

DART assimilates a wide variety of observation types including:

- temperature, winds, moisture from NCEP, MADIS, and SSEC,
- total precipitable water, radar observations, radio occultation observations from GPS satellites,
- temperature and salinity from the World Ocean Database,
- land observations such as snow cover fraction, ground water depth, tower fluxes, cosmic ray neutron intensity, and microwave brightness temperature observations.

DART provides both state-of-the-art ensemble data assimilation capabilities and an interactive educational platform to researchers and students.

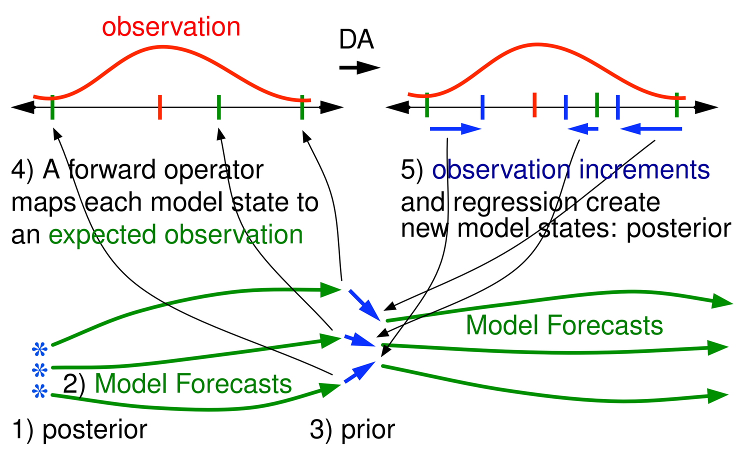

Figure 1: *Schematic for a toy ensemble size of 3.*

J. Anderson, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: *The Data Assimilation Research Testbed: A Community Data Assimilation Facility. BAMS* **90** No. 9 pp. 1283–1296

http://www.image.ucar.edu/DAReS/DART has information about how to download DART, the DART educational materials, and how to contact us.

3. High Resolution POP2

3.1 Background

The $1/10^\circ$ POP2 model is designed to resolve ocean eddies, which are considerably smaller than the typical resolution of 1.0°.

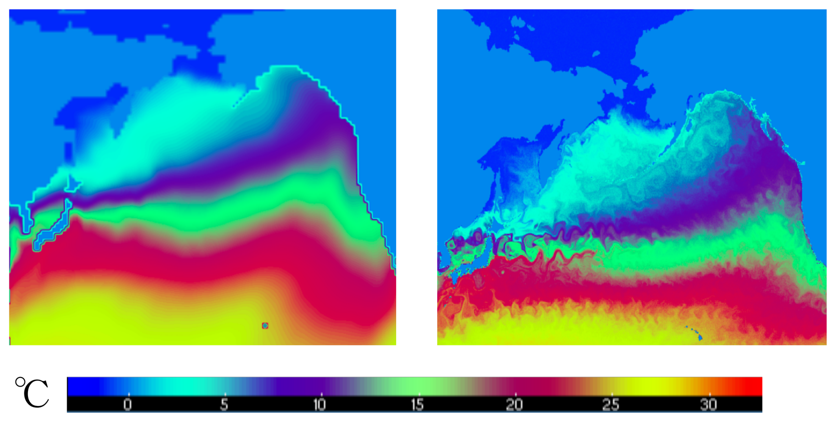

Figure 2: *Comparison of 1° (left) and $1/10^\circ$ (right) POP output reveals the necessity of high resolution for resolving ocean eddies. These pictures show just the North Pacific portion of the global ocean state.*

3.2 Memory and Input/Output

The higher resolution results in each state vector increasing to 17 Gb. Combined with static data and the rest of the program, this will not fit in the memory of a node on many computers, so each ensemble member must be distributed across multiple nodes. One-sided MPI2 communication makes this possible without creating excessive communication demands. For a $1/10^\circ$, POP2, 10 member ensemble, the memory scales well with increasing node count out to 128 nodes (using 16 tasks/node on yellowstone), dropping from 28 Gb/node on 56 nodes to 19 Gb/node on 128 nodes. Beyond 128 nodes the scaling flattens to 18 Gb/node, because the static data is not distributed, so there is a complete (redundant) copy on each task (total 9 Gb/node).

In the previous DART, the entire ensemble plus additional products, such as the ensemble mean and spread, were written to a single NetCDF file. Gathering that data to a single processor is impossible for a model like the $1/10^\circ$ POP, so DART Manhattan writes out the ensemble members and other DART diagnostic output in parallel to multiple NetCDF files. Even with this change, the I/O for the $1/10^\circ$ requires 70% of the assimilation time, while for the 1.0° it takes only 10%. These numbers vary with the number of processors and observations used.

4. Real Time Forecasting

4.1 WRF+DART Forecast System

NCAR has maintained a real-time, continuously cycled, ensemble data assimilation system since mid-March of 2015. This system currently includes an 80-member ensemble analysis that is updated every 6 hours with conventional observations on a mesoscale grid (Figure 3). The WRF model is used to advance the analysis state each cycle, and has been configured following testing of physics suites to find the combination with minimal systematic model bias as identified in prior analysis statistics. The assimilations are performed with DART configured as an ensemble adjustment Kalman filter. **This analysis system demonstrates the reliability and performance of the next-generation DART (Manhattan release) toolkit for generating initialized states for forecasts.**

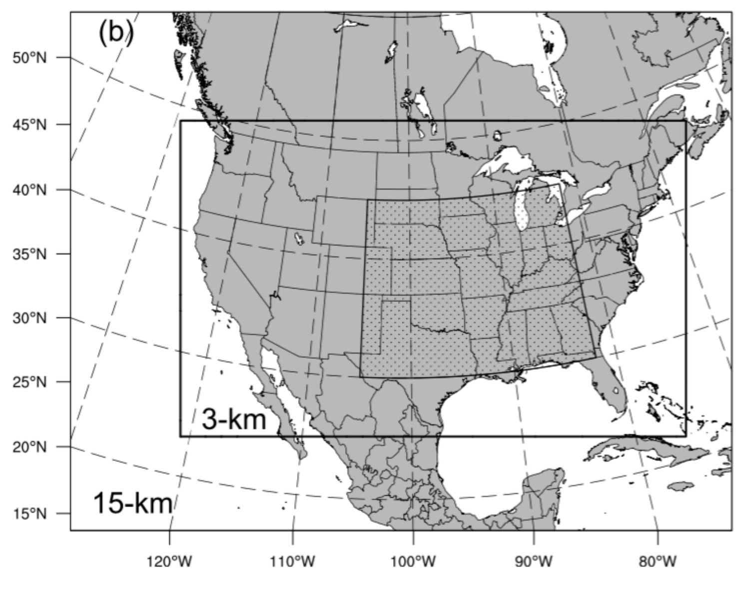

Figure 3: *The WRF+DART assimilation system uses the outer, 15-km grid, which covers areas beyond the conterminous US (CONUS). The ensemble forecasts, which start from the 15-km analyses, use the 3-km grid over the smaller, CONUS region.*

4.2 Ensemble Forecasts

The mesoscale analyses are used once daily (at 00 UTC) to initialize a 10-member convection-permitting (CP; 3-km horizontal grid spacing) ensemble forecast over CONUS. Forecasts are integrated for 48h with results posted to the web (http://ensemble.ucar.edu) where several novel approaches are demonstrated to convey probabilistic forecast information. For this project we also use DART for point based observation verification of the CP ensemble forecasts to investigate model error characteristics across a range of flow regimes. Collectively, this project demonstrates an ensemble analysis and forecast system design with a singular model core and physics suite that follows a consistent approach for forecast verification.

Evaluation of the NCAR ensemble system performance has motivated several additional studies including understanding model error characteristics and predictability of extreme weather events. The first example shows analysis increments averaged over a summer month to illustrate systematic model bias in surface temperature and moisture, particularly from the Plains and across the southern states (Fig. 4). Here, model surface conditions are too warm and dry, with assimilation of surface observations leading to cooling and moistening. Next, Fig. 5 shows a time series of threshold exceedance events for a severe storm surrogate (updraft helicity "UH"), relative to the observed number of severe storms. During the warm season, a surrogate threshold value of 75 gives about the right frequency of events, but a lower threshold is more appropriate during the cool season.

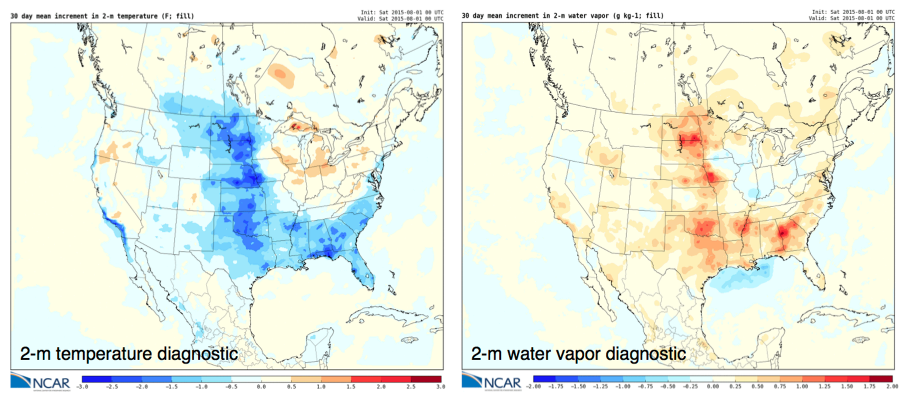

Figure 4: *30-day, mean increment of 2-m temperature (left) and 2-m water vapor produced by the DART assimilations.*

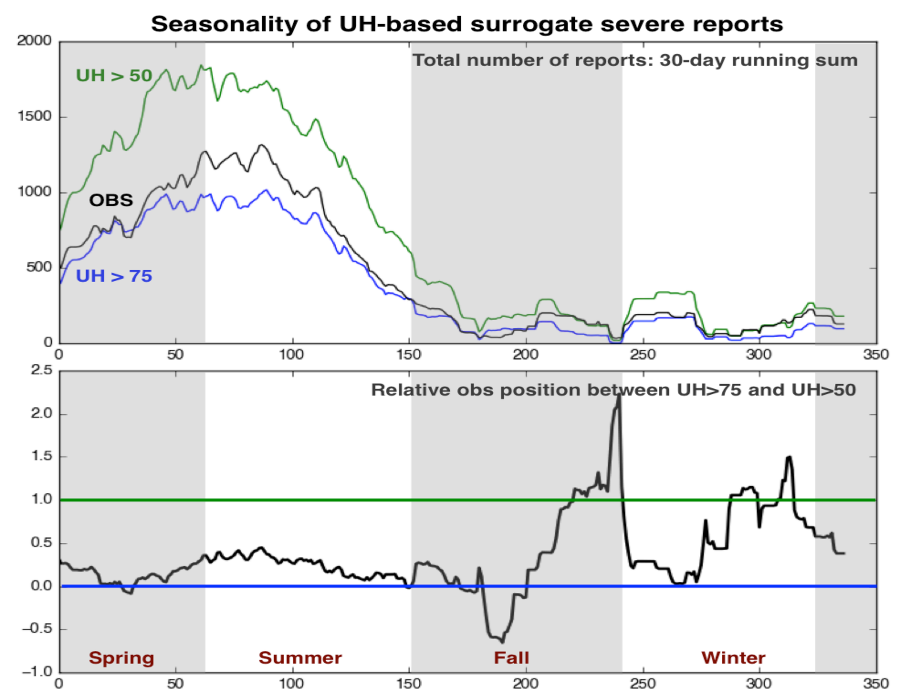

Figure 5: *Top: time series of the 30-day running sum of daily severe storm surrogate events for two thresholds of updraft helicity (50 m^2s^{-2}, green; 75 m^2s^{-2}, blue) and of the number of observed storm reports. Bottom: the relative event count between observed and surrogates at the two thresholds: (OBS - UH75)/(UH50 - UH75).*

5. Coupled Assimilation in CESM

5.1 "Weakly Coupled" Configuration

CESMDART is a prototype global coupled ensemble data assimilation system. In-situ ocean and atmosphere data from 1970-1981 (Figure 6) are assimilated in a "weakly coupled" framework using a 30 member ensemble adjustment Kalman filter (Figure 7). The model is run at nominal 1° resolution, in a standard CESM "workhorse" configuration.

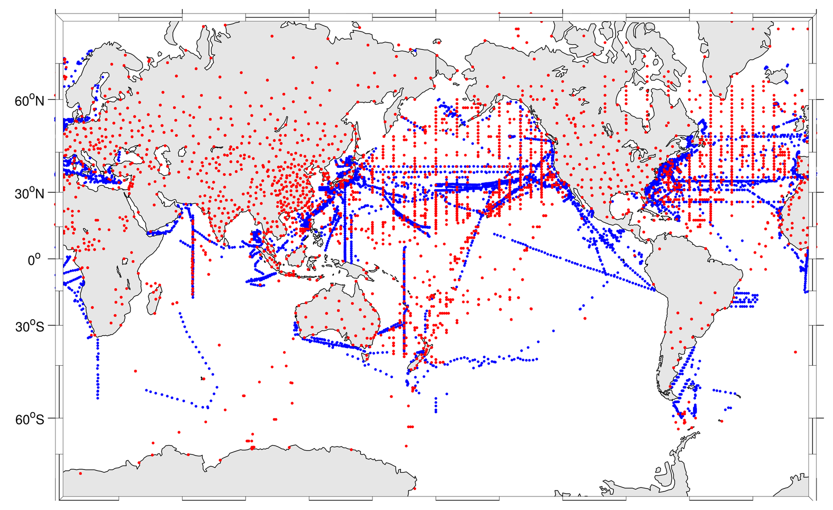

Figure 6: *Location of ocean (blue) and atmosphere (red) observations assimilated in the month of Jan 1975.*

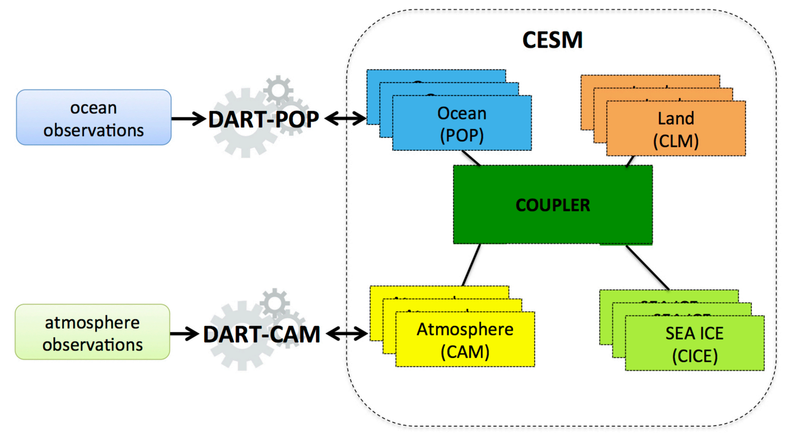

Figure 7: *The CESMDART weakly coupled framework assimilates ocean observations only into the ocean model, and similarly for the atmosphere. No land or sea ice observations are currently assimilated, but all 4 components are affected indirectly by all observations, through the interactions of the components through the CESM coupler during the forecasts.*

5.2 Evaluation of Analyses

Results are promising, indicating that the CESM can be constrained to a historical representation of the climate system.

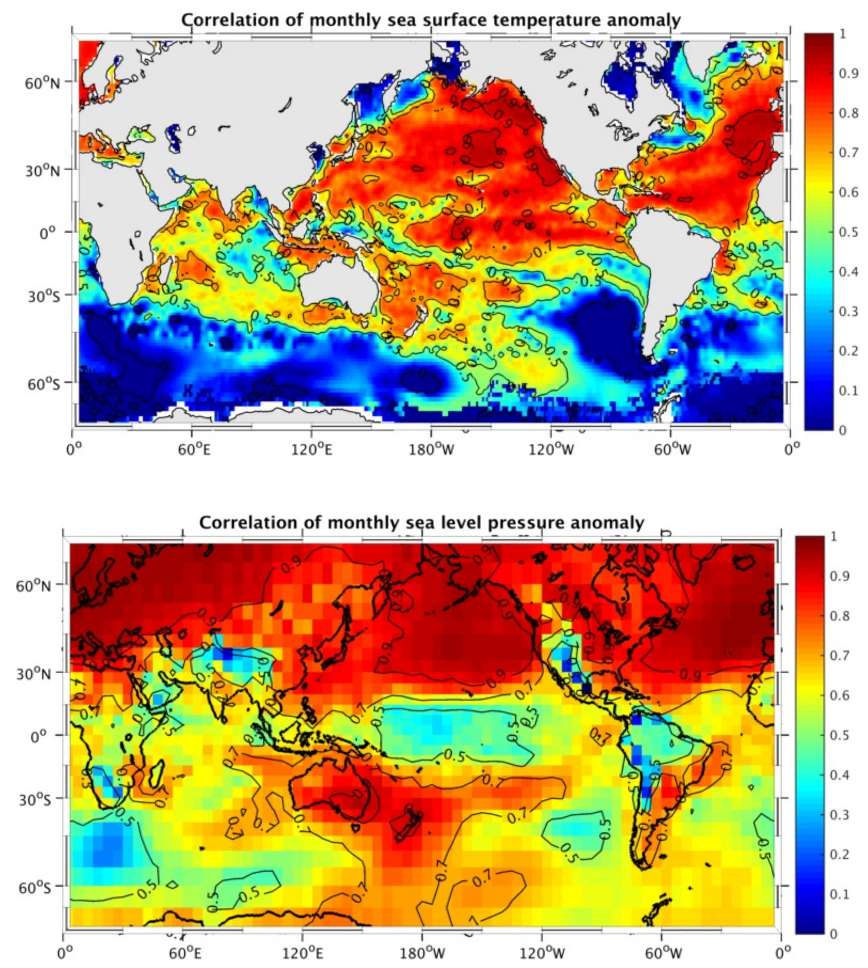

Figure 8: *Anomaly correlation between CESMDART and HADISST (top) and HADSLP (bottom). Agreement is high in regions where observations were available in the 1970s. SLP is completely independent data (no sea level pressure was assimilated). CESMDART does not use any SST products, but draws from the same raw in-situ sea surface temperature data sources sources as HADISST.*

6. Acknowledgements

The National Center for Atmospheric Research is sponsored by the National Science Foundation. Some computational resources were provided by the Computational and Information Systems Laboratory at NCAR.