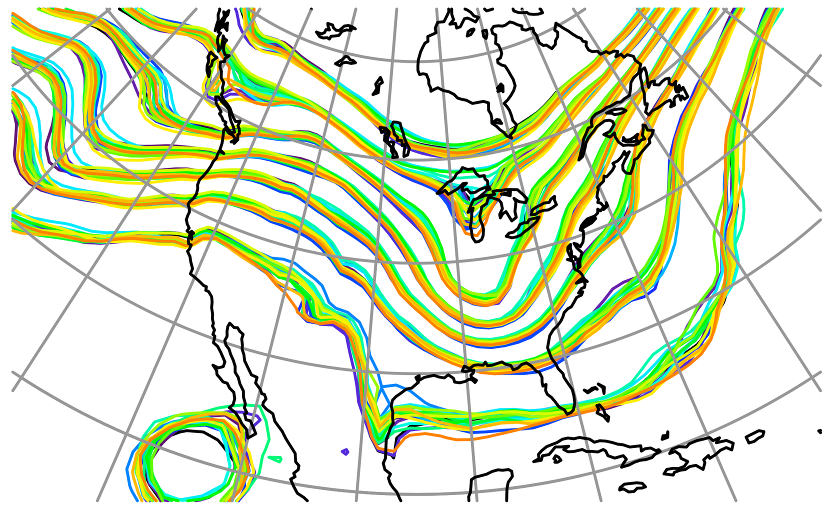


DART: New Research Using Ensemble Data Assimilation in Geophysical Models

K. Raeder, T. Hoar, S. Wuerth, B. Gaubert, L. Neef, S. Ha, M. Garcia, and X. Yang

National Center for Atmospheric Research
Boulder, Colorado, USA

dart@ucar.edu

Institute for Mathematics Applied to Geosciences

1. 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. 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, including:

- weather models, e.g. WRF, COAMPS, NOAH, MPAS Ocean and MPAS Atmosphere
- components of climate models, e.g. CAM, POP, CLM, WACCM, MITgcm-Ocean, GCCOM, ROMS, JULES
- ionosphere/thermosphere models, e.g. TIEGCM, GITM,
- low-order and simple models such as the Lorenz models for assimilation research and educational use.

DART assimilates dozens of observation types from a variety of sources, including:

- temperature, winds aloft, surface winds, moisture from NCEP, MADIS, and SSEC,
- total precipitable water, radar observations, radio occultation observations from GPS satellites,
- ocean 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.

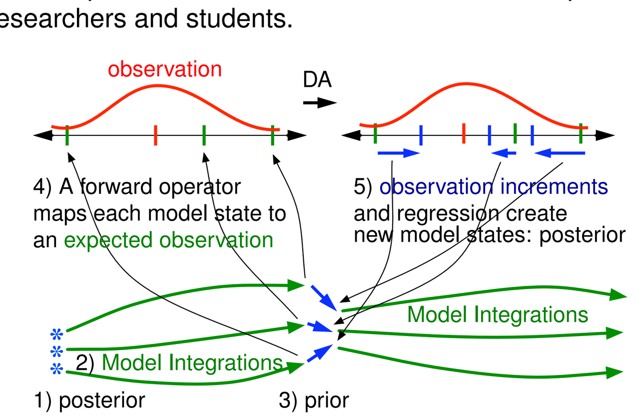


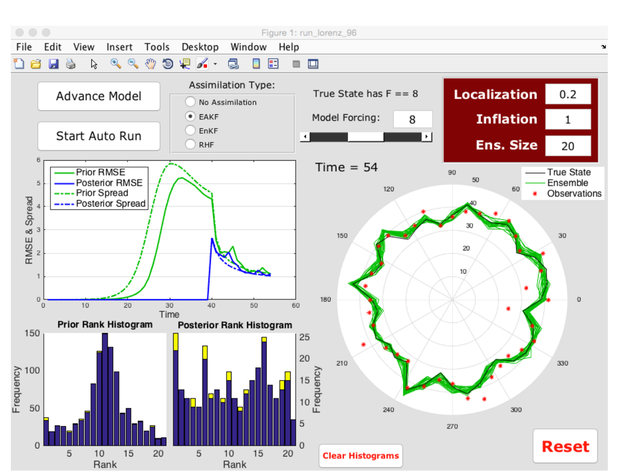
Figure 1: *Schematic for a toy ensemble size of 3.*

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

1.1 Education

DART contains a variety of instructional material to appeal to different types of learning:

- a tutorial directory with 23 self-paced modules,
- a MATLAB tutorial with point-and-click GUI examples,
- a user Application Program Interface (API),
- a web site dedicated to explaining how to use DART, and
- real live people to answer questions!



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.

2. OCO-2 Precipitable Water OSSE

Stephanie Wuerth (swuerth@berkeley.edu)
The Orbiting Carbon Observatory 2 (OCO-2) is NASA's first mission dedicated to observing the global carbon cycle. The measured radiances are used to calculate estimates of total column carbon dioxide, along with total precipitable water (TPW), winds, temperature, and surface pressure (1). These observations may be useful additions to the suite of observations ingested by numerical weather prediction centers. To assess the potential utility of OCO-2 TPW, we have performed a simple observing system simulation experiment (OSSE) using CAM-DART.

OSSE details: CAM5-FV, $1^\circ \times 1^\circ$ resolution, 30 ensemble members, run for 18 6-hour assimilation cycles. Perfect TPW observations were harvested from a free running CAM at the observation locations of OCO-2 corresponding to December 1-5, 2014. Observations were thinned to just one observation per model grid box (chosen randomly from the observations falling within these grid boxes).

Maps of innovation (posterior state - prior state) (Figure 2) demonstrate that the perfect TPW observations impacted the forecast for all state vector fields (Q, T, PS, V, and U (not shown)). Further, the ensemble forecasts generally bracket the truth (Figure 3) for all variables, indicating a healthy assimilation. These are promising preliminary results.

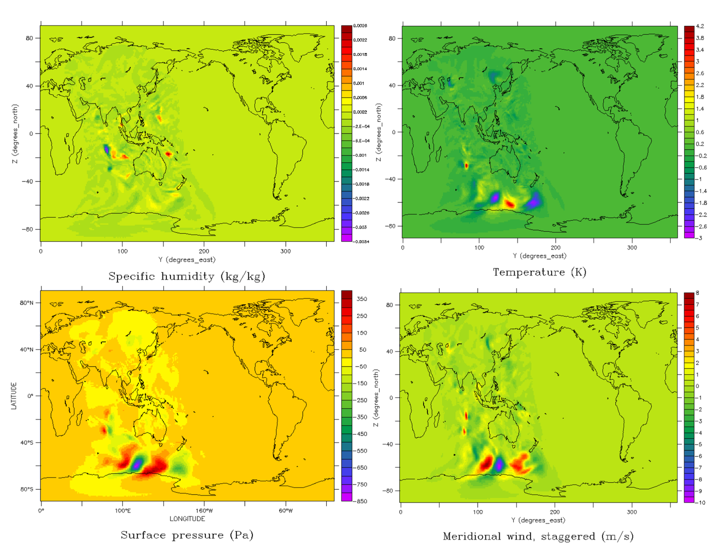


Figure 2: *Ensemble mean innovations generated by the assimilation in the 13th assimilation cycle near 525 hPa (except for surface pressure). The innovations are essentially 0 over large areas where OCO-2 took no observations in the cycle.*

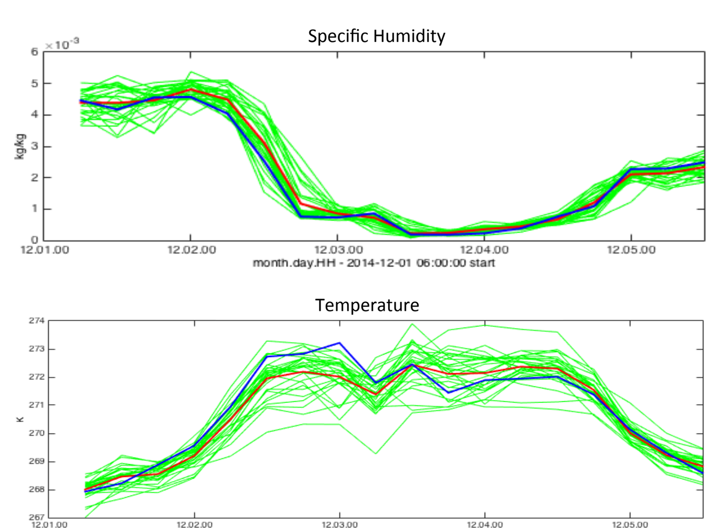


Figure 3: *Time series at a single point (70W, 29.7S, 525 hPa) of Q (top) and T (bottom) for all ensemble members, the ensemble mean, and the true values taken from the "truth" free run of CAM.*

The next test, also in the OSSE framework, will be to assimilate observations typically used for NWP (MET run); radiosonde, aircraft, and satellite drift wind observations of temperature (T), humidity (Q), and winds (U and V). Then assimilate MET plus the OCO-2 TPW observations. Differences between the MET run and the MET+TPW run can be attributed to the OCO-2 TPW's influence on the forecasts.

(1) Osterman et al., OCO-2 Data Product Users Guide, Operational L1 and L2 Data Versions 6 and 6R (2015)

3. CAM-Chem and MOPITT

Benjamin Gaubert (gaubert@ucar.edu)
In this study, a 30-member ensemble assimilation of meteorology plus chemistry has been performed from 2002 through 2007 using the coupled climate-chemistry CESM CAM-Chem model. The chemistry observations are carbon monoxide (CO) retrievals (V5J) measured by MOPITT (Measurement Of Pollution In The Troposphere).

MOPITT is a nadir sounder that measures global tropospheric CO. Our main focus is to understand the trends in CO concentrations. The meteorological observations are radiosondes, satellite drift winds, ACARS, and Aircraft as used in the NCEP/NCAR reanalysis. Figure 4 compares the time series of CO forecasted during the assimilation with the MOPITT observations, and several statistical measures of the quality of the assimilation.

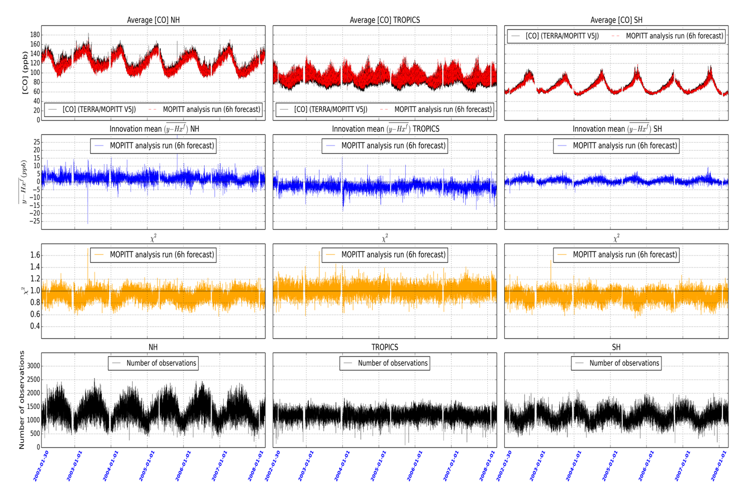


Figure 4: *CO average values and statistics for three regions: northern extra-tropics (20N-65N, left column), tropics (20S-20N, middle), southern extra-tropics (20S-65S, right). All regions consist vertically of the troposphere. The top row compares averaged MOPITT CO observations (black) with 6 hourly forecasted values (red, overlaying the black) from the assimilation. The second row shows the innovations resulting from the assimilation. The third row shows the χ^2 statistics of the forecasts from the MOPITT reanalysis run. The bottom row shows the number of successfully assimilated observations, for which statistics are calculated.*

The number of observations successfully assimilated remains roughly constant, disregarding the fluctuations in the number of observations available due to seasonal availability, occasional gaps in the data stream, and short term random fluctuations. This is a sign that the assimilation is healthy; the ensemble spread remains sufficient to encompass 96% of the observations. The average CO from the assimilation is quite close to the MOPITT values, and clearly shows the seasonal trends and several differences between the 3 regions: larger values in the northern extra-tropics, higher variability in the tropics, and lower values with small variability in the southern extra-tropics. The innovations show that there is a slight negative bias in the north, a slight positive bias in the tropics, and a small seasonal bias in the south. The innovations are small, supporting the conclusion that the assimilation is working well. The χ^2 plot shows that the errors are well balanced.

4. Teleconnections in WACCM

Lisa Neef (lneef@geomar.de)
Data assimilation can be used to investigate teleconnections in the earth system. This study investigates connections between the tropics and northern latitudes, where the Arctic and North Atlantic oscillations (AO/NAO), and sudden stratospheric warmings (SSW) occur.

The experiment consists of three assimilations during the 2009-10 winter, when a strong negative phase of the AO/NAO developed, and there was a SSW in January, 2010. All three assimilations were forced by actual sea surface temperatures (SST) which showed El Niño conditions in the equatorial Pacific. It used a 40 member ensemble of WACCM, which is a high-topped version of CAM and can simulate SSWs. It assimilated radiosonde, aircraft, satellite drift winds, and COSMIC radio occultation observations of temperature and winds. The free-running control ("No DA") assimilated no observations, a "Tropics" run assimilated observations only in the band from 30S to 30N, and the "Global" run assimilated observations over the whole globe. Figure 5 shows how much of the AO/NAO pattern can be generated from just El Niño SST forcing, El Niño + tropical atmospheric (Madden-Julian oscillation) forcing, and El Niño + all atmospheric forcing.

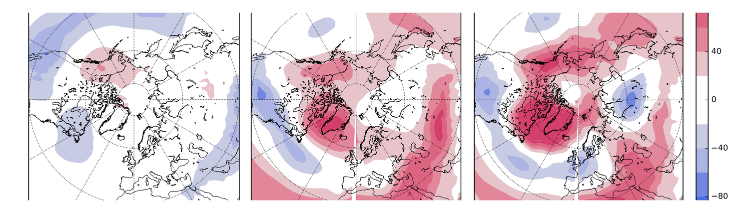


Figure 5: *500 hPa geopotential height anomaly (meters), averaged over Dec 2009 - Feb 2010, in three, 40-member, DART+WACCM ensembles (each plot shows the ensemble mean). Left; free-running ensemble forecast. Middle; only tropical observations are assimilated. Right; the global set of observations are assimilated. The Global assimilation is closest to what actually happened.*

A free-running ensemble shows only a weak anomaly, whereas an ensemble with full global assimilation (right) shows the extreme negative AO/NAO of 2009-10, with a high pressure anomaly over the Arctic and low pressure anomalies over midlatitudes, especially the North Atlantic. Assimilating data in only the tropical band (center), thereby constraining the subseasonal tropical weather, recovers large parts of the AO/NAO pattern. This suggests that tropical weather, including the Madden-Julian Oscillation, contributed to the extreme winter weather of 2009-10.

5. MPAS Variable Resolution Forecast

Soyoung Ha (syha@ucar.edu)
The Model for Prediction Across Scale (MPAS) is a global, nonhydrostatic, atmospheric (in this case) model based on unstructured centroidal Voronoi (hexagonal) meshes using C-grid staggering and selective grid refinement (Figure 6). The unstructured grid mesh facilitates a smooth transition from coarse to higher grid resolutions in the analysis. MPAS is jointly developed, primarily by NCAR and LANL/DOE.
http://mpas-dev.github.io

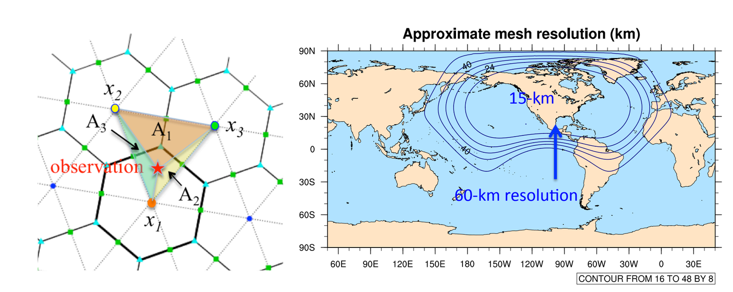


Figure 6: *Left; illustration of an MPAS grid cell and interpolation to an observation location. Right; global domain and variable resolution used in this study.*

MPAS has been interfaced with DART to enable data assimilation studies. MPAS+DART is used here in a retrospective case study comparing a 60-15 km variable mesh and 60-km uniform mesh by assimilating real observations (Figure 7) into a 96-member ensemble, using 6-hr cycling for May 25 - June 25, 2012.

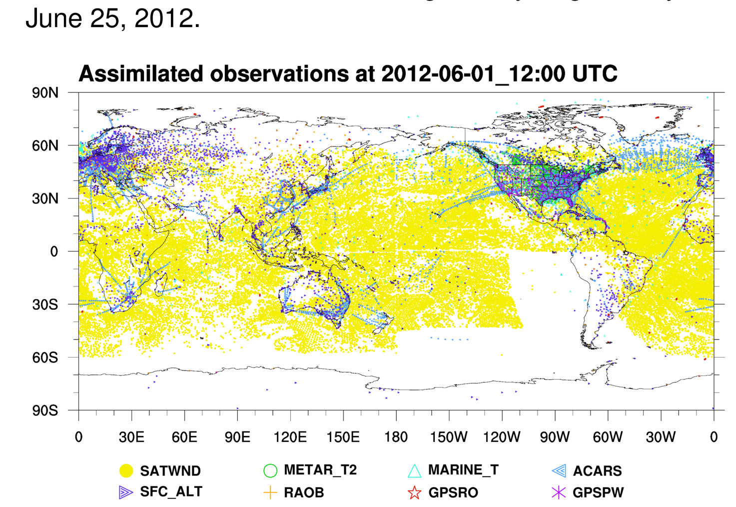


Figure 7: *Example observation set assimilated at a single time. A similar set is assimilated every 6 hours.*

Figures 8 and 9 are examples showing that the 60-15km variable mesh generally provides closer matches to the observations, both in the refined region and globally.

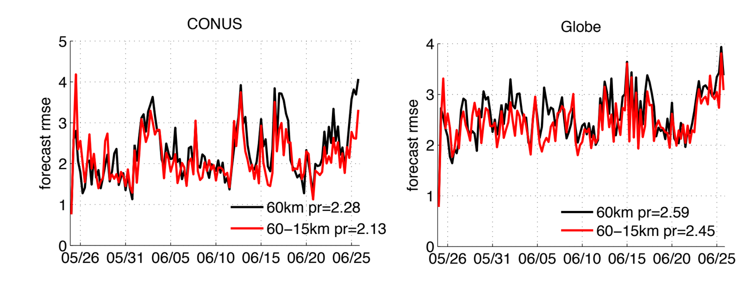


Figure 8: *The observation space, 6-hour forecast verification against surface altimeter observations for the CONUS region (left) and globally (right).*

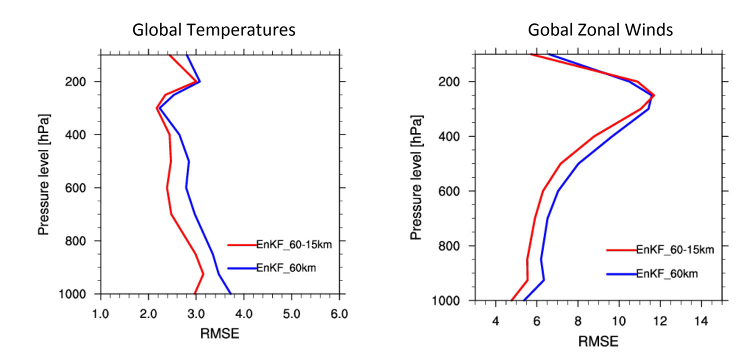


Figure 9: *The observation space, 5-day forecast verification against global FNL analyses for temperature (left) and zonal wind (right). The variable resolution forecasts generally have lower RMSE than the 60 km resolution forecasts.*

Figure 10 shows that the the 60-15km variable mesh successfully simulates mesoscale convective systems over the high-resolution (e.g., 15-km) CONUS area. Its maximum precipitation is much closer in magnitude to the observed and the squall line through South Dakota is resolved better.

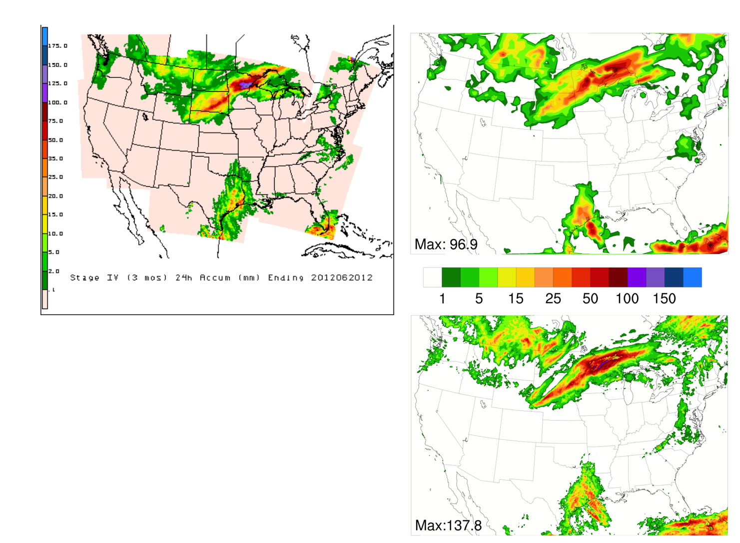


Figure 10: *24-hour accumulated rainfall(mm) as reported in NCEP Stage IV data (left) for 2012-06-20_12Z, and forecasted from the MPAS+DART 60-km (top right) and variable resolution (bottom right) analyses. The rainfall maxima are noted in the lower left corners of the 2 forecasts.*

6. High Resolution Ocean Studies

Mariangel Garcia (mgarcia@sciences.sdsu.edu)
The General Curvilinear Coastal Ocean Model (GCCOM) developed at the San Diego State University (SDSU) is used in conjunction with NCAR's Data Assimilation Research Testbed (DART). Our research included observation system simulation experiments (OSSEs) for a test case involving a very steep seamount and demonstrate that DART-GCCOM can successfully assimilate high-resolution observations (to tens of meters) with as few as 30 ensemble members.

GCCOM solves the three-dimensional primitive Navier-Stokes equation using the Boussinesq approximation in non-hydrostatic form under a fully three-dimensional, general curvilinear mesh. The domain is 3.6 km x 2.8 km and the 32 depths varied between 1 km at the deepest and 0.5 km at the top of the seamount. The grid for the domain is 97x32x32 with an nominal resolution of 30m, however the gridpoints are denser near the seamount. A series of experiments were performed assimilating U current velocities every 10 minutes for 6 hours with observation error variances between 0.1 and 1.0 m/s with localization half-widths of 250m, 500m, 1000m, and 2000m. The observations were located downstream of the seamount at 50 horizontal locations and 21 vertical locations for approximately 1000 observations every 10 minutes.

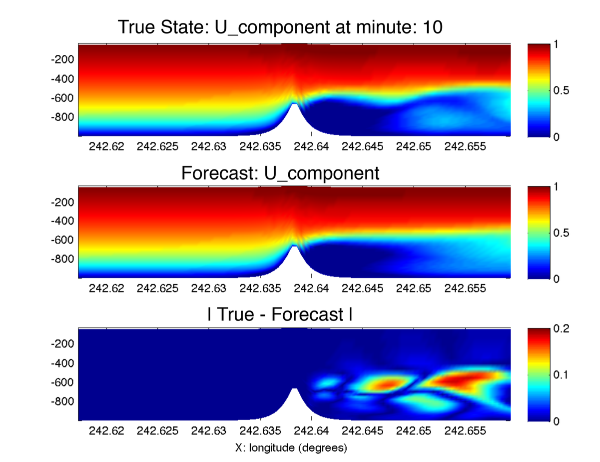


Figure 11: *Top: A vertical slice through the true state 10 minutes into the experiment. Middle: The mean of the 30 ensemble members at the same time before assimilation at that time (the prior). Bottom: The difference between the prior ensemble mean and the truth. Note the ensemble mean is still strongly influenced by the initial ensemble members.*

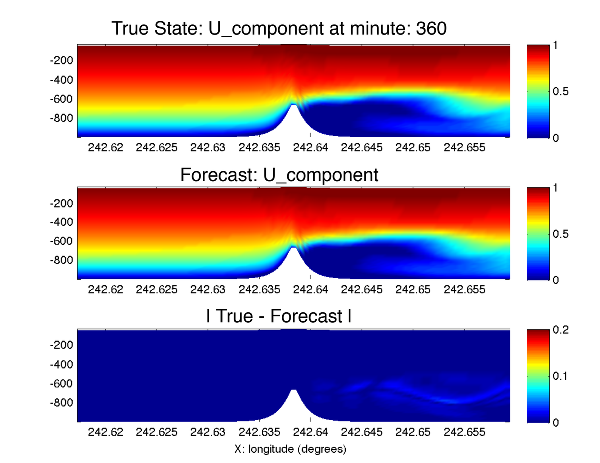


Figure 12: *Top: The true state 6 hours into the experiment. Middle: The mean of the 30 ensemble members at the same time. Bottom: The difference between the prior ensemble mean and the truth. The assimilation has effectively moved the ensemble to be consistent with the true state.*

7. Solar Induced Fluorescence in CLM

Xi Yang (xi_yang@brown.edu)
Figure 13 shows the impact of assimilating solar-induced fluorescence (SIF) on photosynthesis (FPSN) for an OSSE conducted for the Harvard Forest site using CLM45BGC. The run is from 2010-05 to 2010-08. Only SIF is assimilated – FPSN is evaluated. The CLM variables modified by the assimilation include the leaf and root carbon and nitrogen pools.

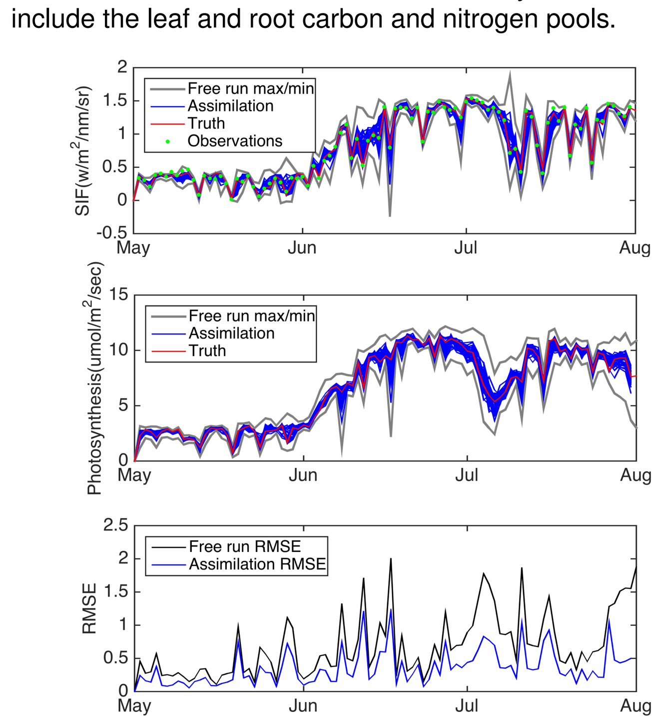


Figure 13: *Top: The SIF observations, true state, the unconstrained (free) run, and assimilation results. Middle: The FPSN for the free run and assimilation. Bottom: The RMSE for the free run and assimilations.*

8. Further Information

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