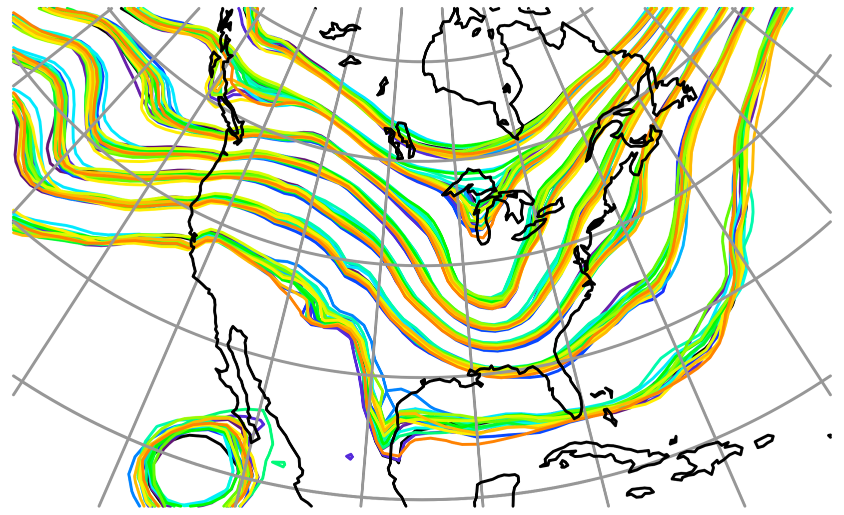


The Data Assimilation Research Testbed: Nonlinear Algorithms and Novel Applications for Community Ensemble Data Assimilation

J. Anderson[1], N. Collins[1], M. El Gharamti[1], T. Hoar[1], K. Raeder[1], F. Castruccio[1], J. Liang[2], J. Lin[3], J. McCreight[1], S. Noh[4], B. Raczka[3], and A. RafieeiNasab[1]

1 National Center for Atmospheric Research, Boulder, CO; 2 Insititute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; 3 University of Utah, Atmospheric Sciences, Salt Lake City, UT; 4 University of Texas at Arlington, Arlington, TX.

dart@ucar.edu

CISL Computational & Information Systems Lab

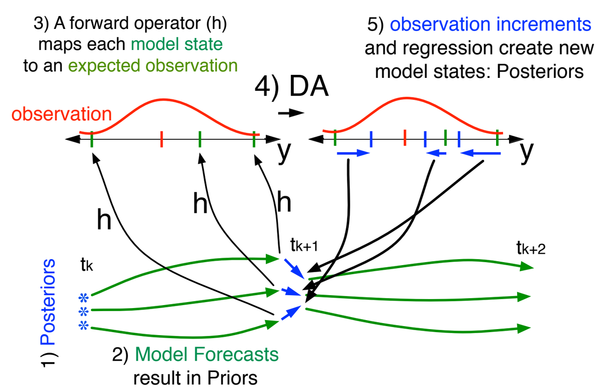
Data Assimilation Research Testbed

1. Overview

The Data Assimilation Research Testbed (DART) is a community facility for ensemble data assimilation developed and maintained at the National Center for Atmospheric Research (NCAR). DART provides data assimilation (DA) capabilities for nearly all NCAR community earth system models. The ensemble data assimilation tools provided by DART allow NCAR models to produce ensemble forecasts. Data assimilation involves combining short model forecasts with observations to produce ensemble analyses that can be used for subsequent forecasts of any length. This process of confronting the model with observations facilitates model evaluation and improvement. The ensemble analyses and forecasts from DART enable analysis and understanding of the earth system.

2. DART

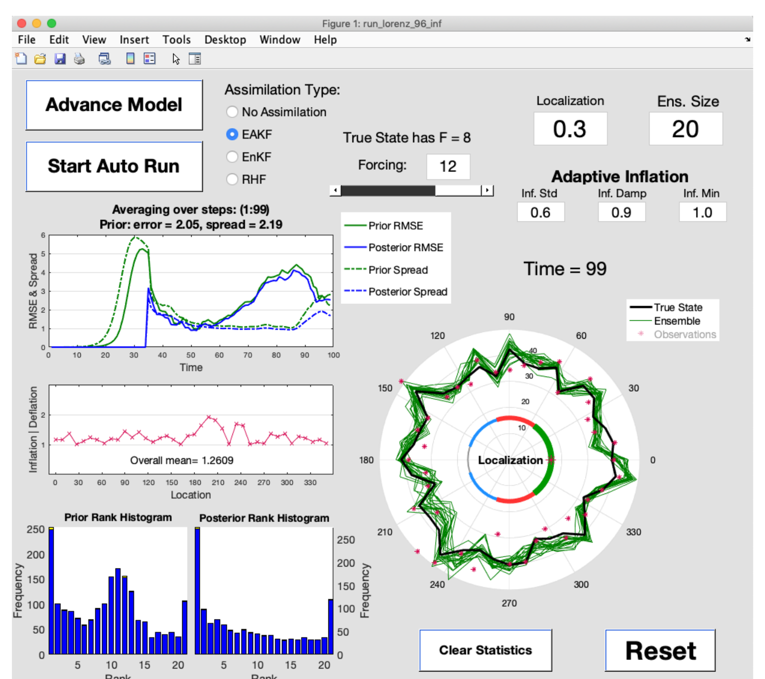
DART has been free and publically available for more than 10 years. Building an interface between DART and a new model does not require an adjoint and **generally requires no modifications to the model code**. DART works with dozens of models and a wide variety of observations.



dart.ucar.edu has information about how to download DART, the DART educational materials, and how to contact us.

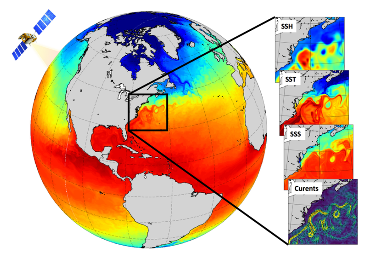
DART provides both state-of-the-art ensemble data assimilation capabilities and an interactive educational platform to researchers and students. DART contains a variety of instructional material to appeal to different types of learning:

- 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!



3. Eddy-Resolving POP *F. Castruccio*

Goal: Develop and implement proper high-resolution ocean initialization procedures and data assimilation infrastructure to support seasonal-to-decadal prediction with the high-resolution versions of CESM.



DART provides the infrastructure to support ensemble DA for the $1/10^\circ$ eddy-resolving POP ocean model. Running a 50+ member ensemble of global eddy-resolving ocean simulations is computationally prohibitive. We have implemented the far less expensive ensemble optimal interpolation (EnOI; Evensen, 2003) within DART. This eliminates the cost of running multiple simulations during the forecast step.

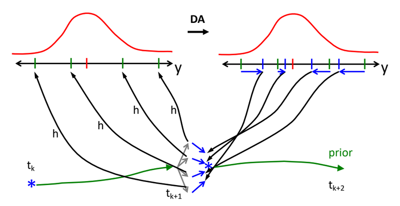


Figure 1: *As implemented within DART, the EnOI scheme uses a static (but seasonally varying) ensemble of precomputed perturbations to approximate samples from the forecast error covariance and uses a single model integration to approximate the forecast mean.*

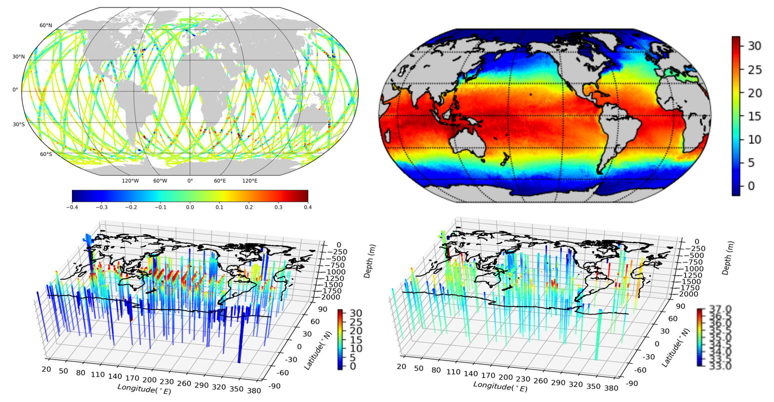


Figure 2: *A typical set of observations assimilated daily. 03/01/2005 shown. Clockwise from upper left: along-track SLA ≈100,000 obs, in-situ temperature from WOD13 ≈75,000 obs, in-situ salinity from WOD13 ≈25,000 obs, OISSTv2 ≈75,000 obs.*

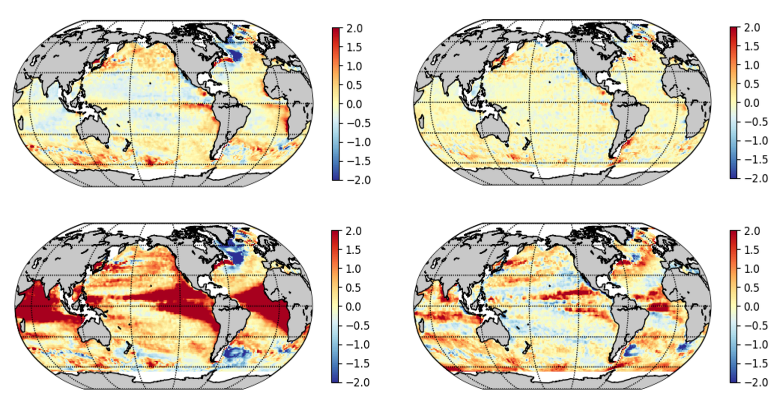


Figure 3: *The difference from the RG ARGO climatology (http://doi.org/10.17882/42182) and the mean Temperature from 2005-2009 (remapped to the RG ARGO climatology grid). Top row is 10m, bottom row is 100m.*

- Cost of integration is down to ≈600K core-hours from ≈10M per year of simulation with the EnKF.
- A highres hindcast forced with JRA (2000–2016) has been run.
- A highres reanalysis (2005–2016) with DART-EnOI is underway.

G. Evensen, 2003: *The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation.* *Ocean Dynamics* **53**:343 doi.org/10.1007/s10236-003-0036-9

This work is supported by NSF award *EaSM-3: The Role of Ocean Eddies in Decadal Prediction*

4. WRF-Hydro *J. McCreight, M. El Gharamti, S. Noh, A. RafieeiNasab*

The Weather Research and Forecasting Hydrologic model (WRF-Hydro) is a community modeling system and framework for hydrologic modeling and model coupling. Combined with DART, the facility is called HydroDART. Experiments assimilating streamflow every hour for a 3 month period that encompasses the landfall of hurricane Florence (2018) were performed to assess what is required for improved streamflow forecasts under dry, normal, and flood conditions.

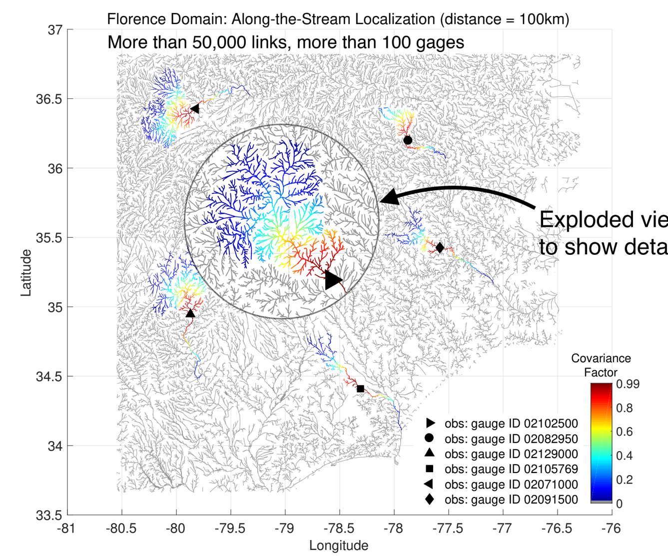


Figure 4: *The test domain 'Florence' for WRF-Hydro and DART. The domain is approximately 100,000 km^2 on the Carolina coast. The symbols are at stream gauge locations and the coloring depicts the localization of an observation. A localization distance of 100km is used for visualization only.*

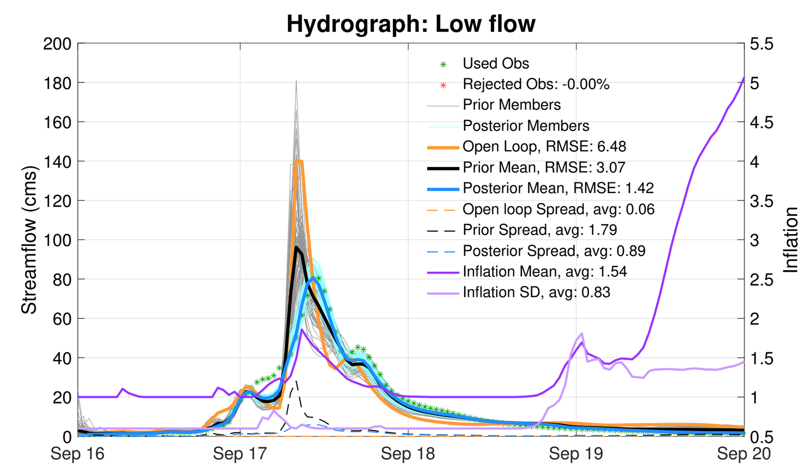


Figure 5: *A representative summary of the performance of the hourly assimilation with 80 ensemble members. The open loop run is clearly biased high and the timing of the streamflow does not match the gauge observations. The assimilation prior is generally much closer to the observed streamflow, and the posterior is better still.*

4.1 Gaussian Anamorphosis

The Kalman filter (KF) update assumes the state variables to have a Gaussian distribution. This is not true for many variables, e.g., streamflow and concentrations. Gaussian Anamorphosis (GA) in DA is the process of transforming the state variables and/or parameters into a Gaussian space prior to applying the KF correction. The transformations can be either constructed empirically following the marginal distribution or chosen from various analytic forms. After the DA update, the return to the physical space is obtained using the inverse of the anamorphosis function. Using GA:

- the KF update is applied to (near) Gaussian distributions,
- unphysical updates (e.g., negative thickness) are avoided,
- non-Gaussian features in the prior are retained in the posterior, and
- obtain more reliable distributions (no truncation is needed).

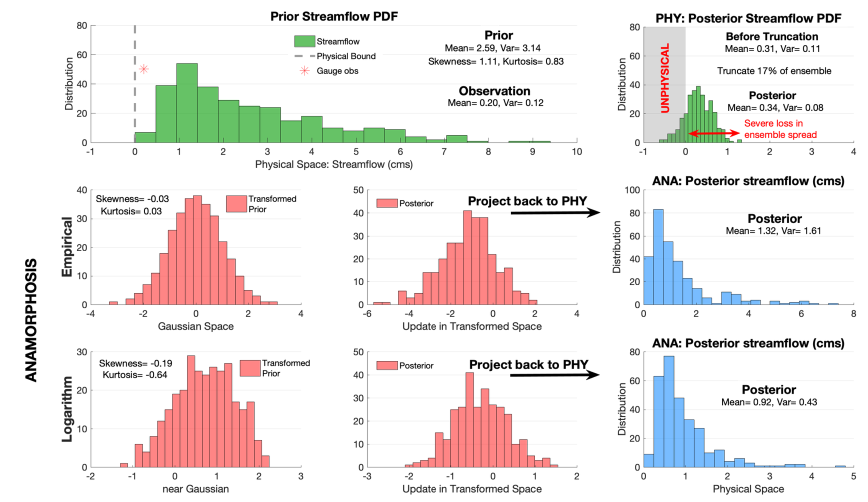


Figure 6: *Top: (Left) Prior streamflow distribution. The gauge observation is denoted by a red asterisk. The dashed gray line at 0 cms (cubic meters per second) depicts the lower bound for streamflow. (Right) Posterior streamflow distribution after the Kalman update. The shaded region shows the part of the ensemble that is unphysical. Middle: (Left) Transformed prior streamflow distribution using empirical Gaussian anamorphosis. (Center) Updated distribution in the Gaussian space. (Right) The posterior distribution in physical space. Bottom panels: (Left) Transformed prior streamflow distribution using an analytical Gaussian anamorphosis function; the logarithm. (Center) Updated distribution in the transformed space. (Right) The posterior distribution in the physical space.*

5. NOAH-MP MODIS & GRACE *J. Liang*

The Noah-multiparameterization land surface model (Noah-MP) and DART are being used to explore the assimilation of total water storage (TWS) observations from the Gravity Recovery and Climate Experiment (GRACE) satellites and MODIS (Moderate Resolution Imaging Spectroradiometer) snow cover fraction. The GRACE TWS observations are from a data product derived at the University of Texas at Austin. A single model state was replicated for 40 ensemble members and then advanced using an ensemble of CAM forcing for 3 years to generate spread in the initial ensemble. In the following captions OL refers to the Open Loop (no data assimilation), MOD refers to experiments assimilating only MODIS observations, and MOD GRA refers to experiments assimilating MODIS snow cover fraction and GRACE TWS.

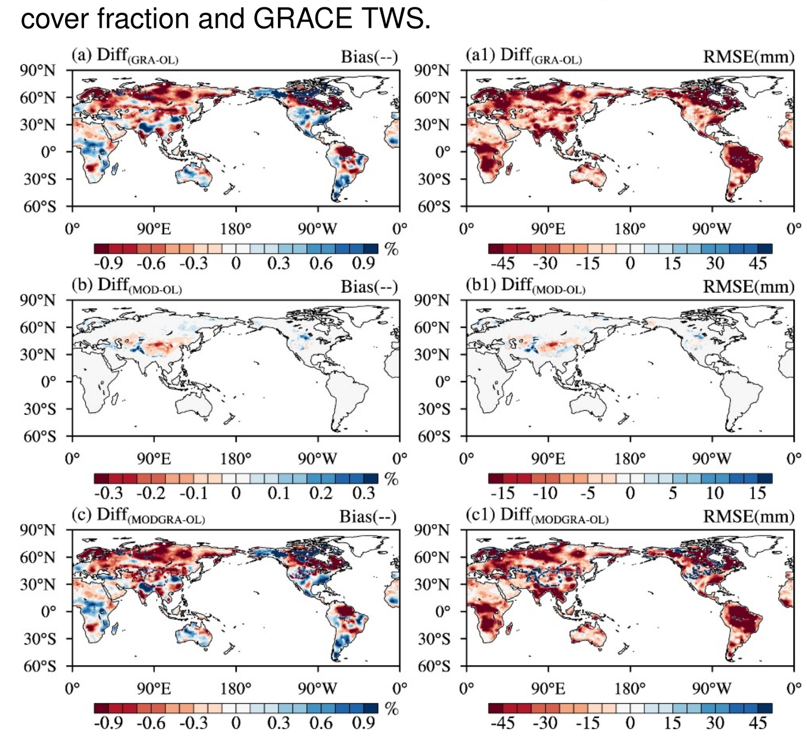


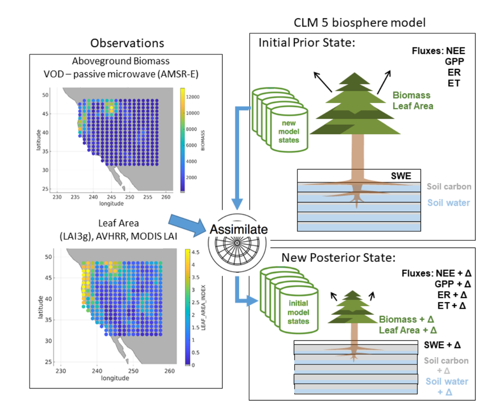
Figure 7: *Differences in the Bias and RMSE TWS between different DA runs and OL. From the top to bottom each row indicates GRA minus OL, MOD minus OL, and MOD GRA minus OL. For each gridcell, RMSE is calculated from the time series of each gridcell (2003.11-2004.05). The blank areas in the plots indicate regions where observation data are missing for more than half of the period.*

6. Carbon Monitoring *J. Lin, B. Raczka*

- How can satellite, atmospheric in-situ, and ecological observations be combined with atmospheric and biospheric models to inform carbon budgets in regions of complex terrain?
- How is satellite-retrieved SIF related to leaf-level physiology?
- What are the impacts of drought and disturbance on carbon cycling in mountainous regions?

We are developing and testing a new Carbon Monitoring System over mountains (CMS-Mountains) covering the Western U.S., where we can leverage numerous existing efforts in biospheric and atmospheric modeling. We are running the Community Land Model (CLM 5.0) and assimilating observations of above-ground biomass (Liu et al., 2015), and leaf area index (Zhu et al., 2013). The infrastructure for assimilating Solar-Induced Fluorescence (SIF) from GOME-2 and OCO-2, column amounts of CO2 (XCO2) from OCO-2 and snowcover is also being developed.

Observations of remotely-sensed leaf area and above-ground biomass are assimilated into CLM using DART, improving the accuracy of the model simulation across the Western US. Figure 8 represents the average ensemble behavior (80 members) across all grid cells within the spatial domain.



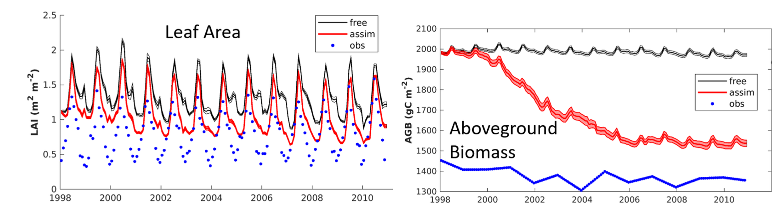


Figure 8: *The improvement in simulated carbon stocks comparing a simulation that uses no observations, "free", against a simulation that assimilates observations, "assim".*

The assimilation of remotely-sensed observations, leads to a redistribution of simulated biomass (Figure 9) and an overall reduction in land carbon sink to near neutral land-atmosphere carbon exchange (Figure 9) (Raczka et al., in prep). As the vast majority of biomass resides in the high elevation, complex terrain of the Rocky Mountains, this region is where the majority of the simulated biomass reduction occurs. Adjustments in leaf area were much more subtle with increases in leaf area within the central valley of California and southwest US (not shown).

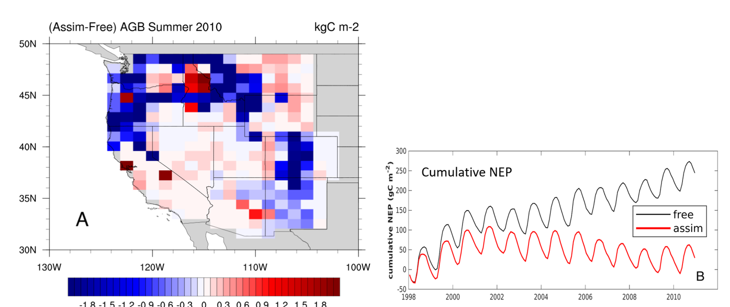


Figure 9: *Accounting for observations of aboveground biomass and leaf area leads to more accurate spatial distributions of biomass (panel A) and a decrease in carbon sink to the land of the Western US (panel B). Panel A describes the difference in aboveground biomass between the 'assimilated' and 'free' runs during summer (Jun, Jul, Aug) of the year 2010. Panel B shows the reduced carbon sink to land as quantified by the net ecosystem productivity (NEP). The assimilation was run from 1998-2011.*

7. CAM Reanalysis *K. Raeder*

An 80-member DART reanalysis using the latest version of the Community Atmospheric Model (CAM6-FV) at 1° resolution is in production. Every 6 hours we assimilate all conventional obs as well as AIRS and GPS. In addition to evaluating CAM's capabilities as a forecast model, several data products are being created and are available for public use. These include ensembles of dynamically consistent atmospheric forcing at 1° at hourly, 3-hourly, and daily resolutions (see Figure 11). These forcings represent the uncertainty in the knowledge of the atmosphere are all equally likely. These forcings can be used with other CESM components like CLM, POP, and CICE and are well suited for ensemble forecasts, sensitivity studies, and ensemble data assimilation. The full 80 member atmospheric state is available weekly. Results for 2011 thru 2015 are available now and 2017 will be available soon to facilitate the use of modern remote sensing platforms. 2000-present should be available within the next year.

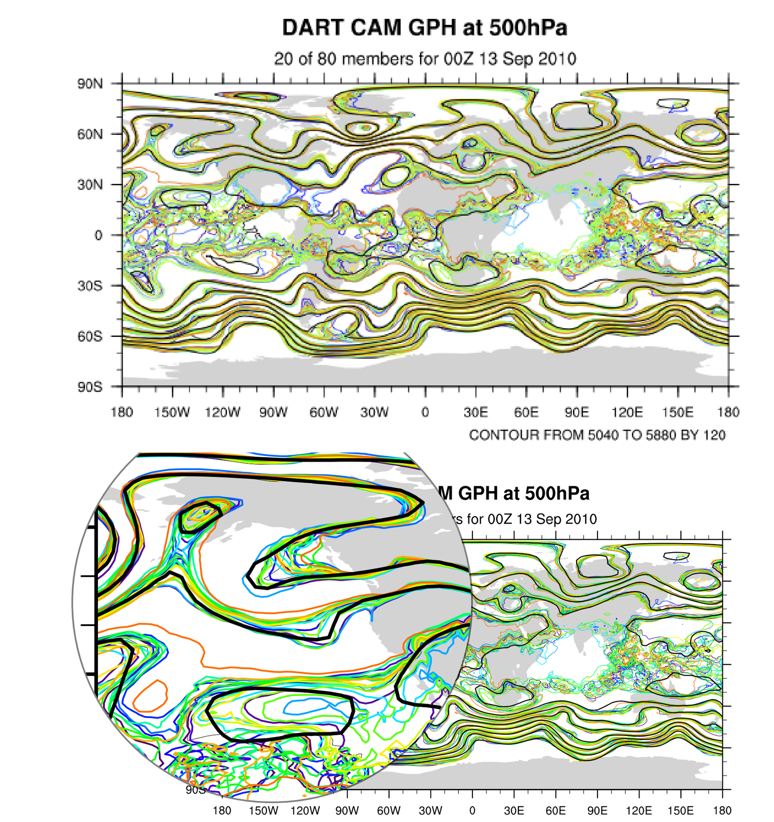


Figure 10: *Top: The variability of the height to the middle of the atmosphere is demonstrated by a subset of the 80 ensemble members. The colors represent individual ensemble members. The heavy black line is the mean of the 20 members shown here. The lower figure highlights this variability.*

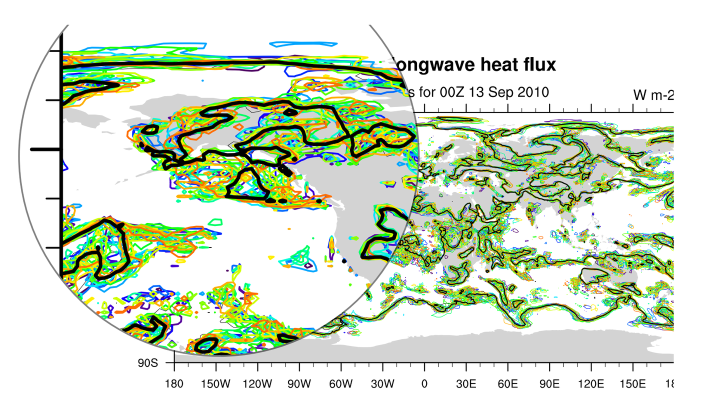


Figure 11: *The variability of the downward longwave heat flux at the surface. 20 of the 80 members are shown here. The colors represent individual ensemble members.*

The National Center for Atmospheric Research is sponsored by the National Science Foundation.

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.