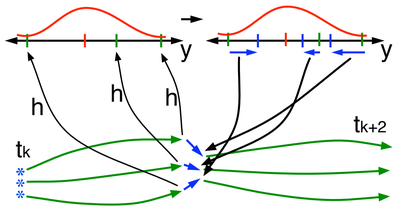
Ensemble Data Assimilation with the Community Land Model and the US National Water Model



T. Hoar, M. Gharamti, J. McCreight, A. RafieeiNasab, A. Fox, J. Anderson, N. Collins, J. Hendricks, K. Raeder

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

1. Ensemble Data Assimilation with DART

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 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 :

- weather models, e.g. WRF, COAMPS, COSMO, MPAS Atmosphere
- climate models components, e.g. CAM, POP, CLM, WACCM, MPAS Ocean, ROMS, FESOM, CICE5, WRF-Hydro, NOAH
- atmospheric chemistry models, e.g. CAM-CHEM, WRF-CHEM
- low-order and simple research models

DART assimilates a wide variety of observations :

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

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

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

2. DART 'Manhattan' Release Highlights

Much of the development effort has been to support larger model states using less memory and improve performance, particularly I/O.

- ★ Ability to handle much larger model states. Distributing the model state across all tasks 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 better vector code.
- ★ Native NetCDF support eliminates conversion between NetCDF model files and DART. 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 maximum memory requirements.
- ★ Diagnostic files are now written in parallel, resulting in faster I/O and lower memory requirements.
- ★ Supports externally-computed observation operators.
- ★ Supports per-observation-type localization radii.

3. Acknowledgments

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

4. Surface Fluxes with CLM4.5

Special Thanks to **Andy Fox** (University of Arizona)

CLM4.5 was used at the Niwot Ridge LTER site to explore the impact of latent heat, sensible heat, and net ecosystem production observations on relevant CLM variables.

- 9.7 km east of the Continental Divide, in a Subalpine Forest
- Spun up for 1500 years with site-specific information.
- 64 ensemble members using an ensemble reanalysis forcing
- Assimilating tower fluxes of latent heat (LE), sensible heat (H), and net ecosystem production (NEP).
- **unobserved** variables: LEAFC, LIVEROOTC, LIVESTEMC, DEADSTEMC, LITR1C, LITR2C, SOIL1C, SOIL2C, SOILLIQ

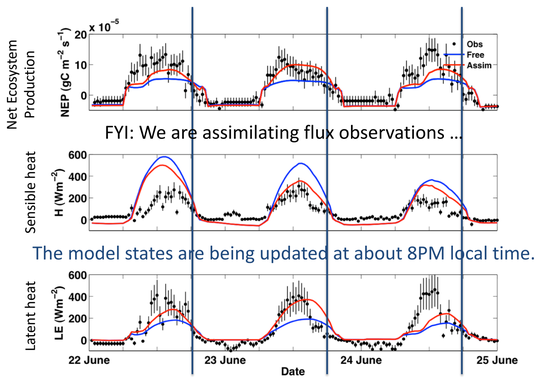


Figure 1: *The observations of the eddy covariance fluxes are available every 30 minutes. The free run of the model is shown in blue, the assimilation result is shown in red.*

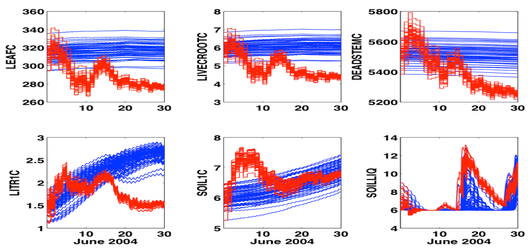


Figure 2: *The free run is shown in blue. The assimilation result is shown in red.*

Carbon Cycle

We have added a number of observations to the CLM-DART workflow, such as leaf area index (LAI) and aboveground biomass - key components of the terrestrial carbon cycle. DART's adaptive inflation algorithm allows the ensemble to become consistent with the observations and then maintain a useful ensemble spread. This effort will be moving to CLM5.0 to take advantage of the expanded treatment of the carbon and nitrogen cycles.

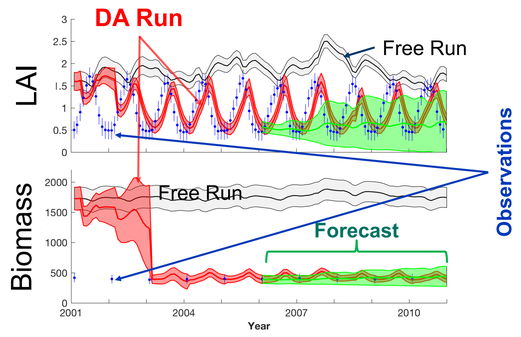


Figure 3: *This experiment highlights a study in a semi-arid region in New Mexico. Monthly LAI and annual biomass observations are able to provide a much better representation of the sesaonal cycle and accurate initial conditions for forecasting. The impact persists for years for different C pools.*

5. Soil Moisture with CLM5

The Community Land Model 5.0 was released in February 2018 and is the latest in a series of land models developed through the Community Earth System Model (CESM). CLM4.5 biogeophysics and biogeochemistry can be run from this release code. A new river model (MOSART) is also included. This release was a land-only release. The Functionally Assembled Terrestrial Ecosystem Simulator (FATES) is available as a research option.

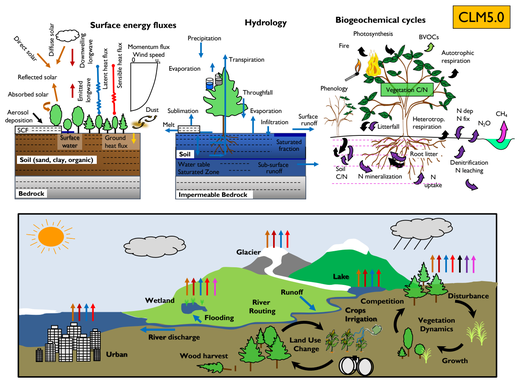


Figure 4: *The representation of the processes in CLM5.0 which was released this February.*

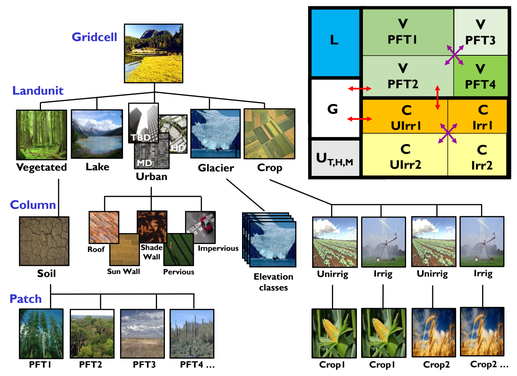


Figure 5: *The hierarchy of a single CLM5.0 gridcell. The only location information is at the gridcell level. We are working on an approach to more appropriately relate observation locations to specific Landunits, Columns or Patches.*

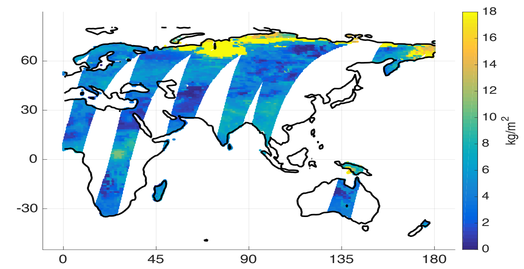


Figure 6: *SMAP locations from 23 July 2015 were the basis for a short proof-of-concept experiment. The soil moisture (liquid and ice) from CLM at the SMAP locations were used in an observing system simulation experiment.*

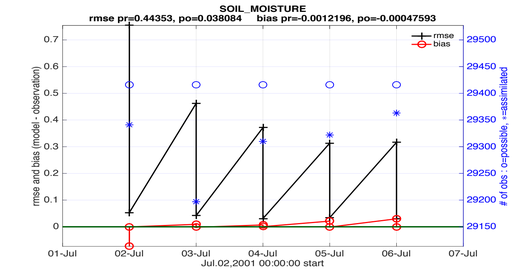


Figure 7: *The proof-of-concept result for soil moisture in the first layer of CLM. This is a 40 member ensemble.*

6. U.S. National Water Model

The NCAR-supported community Weather Research and Forecasting Hydrologic model (WRF-Hydro) is a modeling system and framework for hydrologic modeling and model coupling. In 2016 a configuration of WRF-Hydro was implemented as the National Water Model (NWM) for the continental United States.

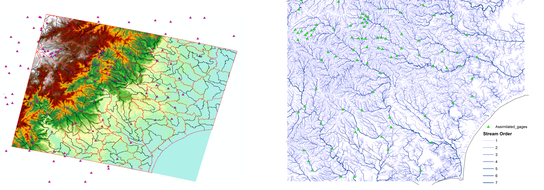


Figure 8: *Left: The test domain 'Matthew' for WRF-Hydro and DART. The domain is approximately 100,000 km². The triangles are at stream gage locations. The LSM runs at ≈ 1 km resolution. Right: A depiction of the Stream Order.*

WRF-Hydro is configured to use the Noah-MP Land Surface Model (LSM) to simulate land surface processes. Separate water routing modules perform diffusive wave surface routing and saturated subsurface flow routing on a 250m grid, and Muskingum-Cunge channel routing down National Hydrography Dataset (NHDPlusV2) stream reaches.

HydroDART can be run in multiple configurations, including a 'channel-only' configuration. A small case can be run in Docker.

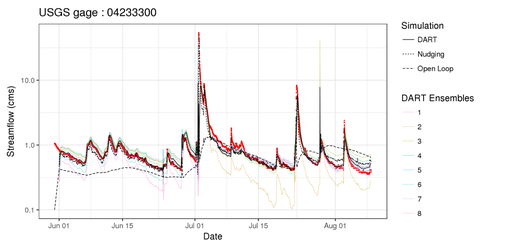


Figure 9: *The result from a test experiment on a 'channel-only' configuration. This is the verification gage.*

HydroDART is currently configured to assimilate streamflow from 115 gages in the Matthew domain and update the channel model state as well as estimate parameters regulating the efficiency of the link response to fluxes. There are 50225 links/reaches in the domain with 115 gages that are currently being assimilated. Research areas include localization, inflation, and parameter estimation.

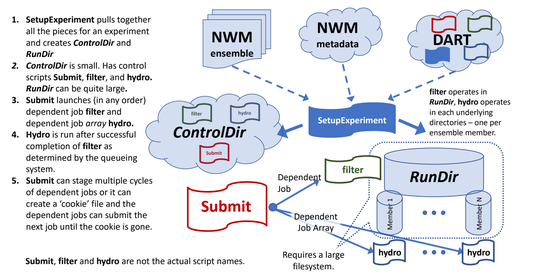


Figure 10: *Schematic of the workflow of WRF-Hydro and DART. Separate jobs perform the assimilation and a Job Array is used to run the ensemble of WRF-Hydro instances. Each job can be tuned for an optimal configuration. Since each job runs for a very short time, the jobs may backfill. The queueing system takes care of the job dependencies and automatically resubmits the next job in the series.*