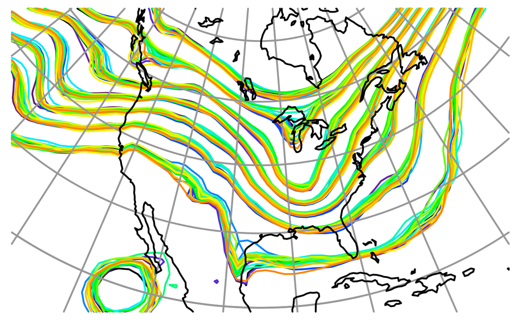
The Latest from the Data Assimilation Research Testbed: Powerful New Assimilation Algorithms, Advances in Efficiency and Capabilities, New Model and Observation Interfaces, and Novel Results.



K. Raeder[1], M. Gharamti[1], B.K. Johnson[1] H. Kershaw[1], X. Huo[2], L. Kugler[3], T. Kodikara[4], B. K. Johnson[1], B. Raczka[1], M. Smith[1], J. Liu[5], J. D. Labriola[6], F. Ishraque[7], D. Hagan[8], J. Anderson[1]

1 National Center for Atmospheric Research; CISL/DAReS, Boulder, CO; 2 U. AZ School of Natural Resources and the Environment;
3 U. Vienna Department of Meteorology and Geophysics; 4 German Aerospace Center (DLR) Institute for Solar-Terrestrial Physics;
5 Drexel Univ. Civil, Architectural & Environmental Engineering; 6 National Severe Storms Laboratory;
7 Princeton Univ. Department of Geosciences; 8 Nanjing Univ. of Information Science & Technology

dart@ucar.edu

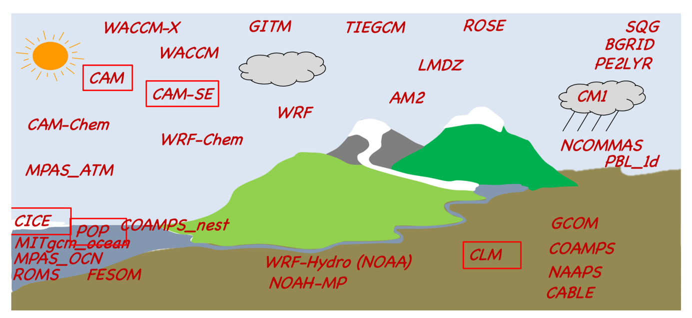
1. DART Is ...

A flexible suite of ensemble data assimilation tools for accelerating Earth sytem research. Data assimilation (DA) is the combination of prior information, usually from numerical models, with information from observations. DART includes:

Tutorials and Documentation
From fundamental DA concepts to DART architecture to extensive web-based information.

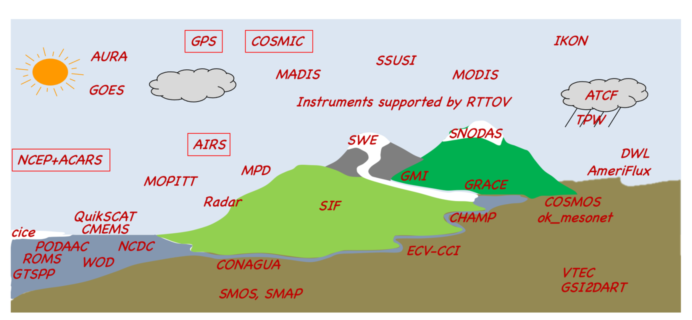
Model Interfaces

+ Low-order models used for rapid DA research and clear tutorials: Lorenz '63, Lorenz '96, simple advection (NEW; See 4.6), ...
+ Many widely-used geophysical models (next picture)
+ New interfaces require no model changes.



Observation Interfaces

+ These "forward operators" calculate the model estimates of observations. (See next picture)
+ Or pre-computed observations can be imported from a model.
+ Synthetic observations can be generated for OSSE experiments.



Numerous Core Assimilation Algorithms

+ NEW Quantile Conserving Filter (Anderson's AMS2023 poster)
+ NEW Randomized Dormant EnKF (El Gharamti's AMS2023 talk)
+ traditional "square-root" filters (EAKF, EnKF, ...)
+ rank histogram filter
+ particle filters

Support algorithms for efficient DA in Earth system models:

+ localization
+ ensemble inflation
+ sampling error correction
+ highly parallel computation
+ efficient interprocess communication

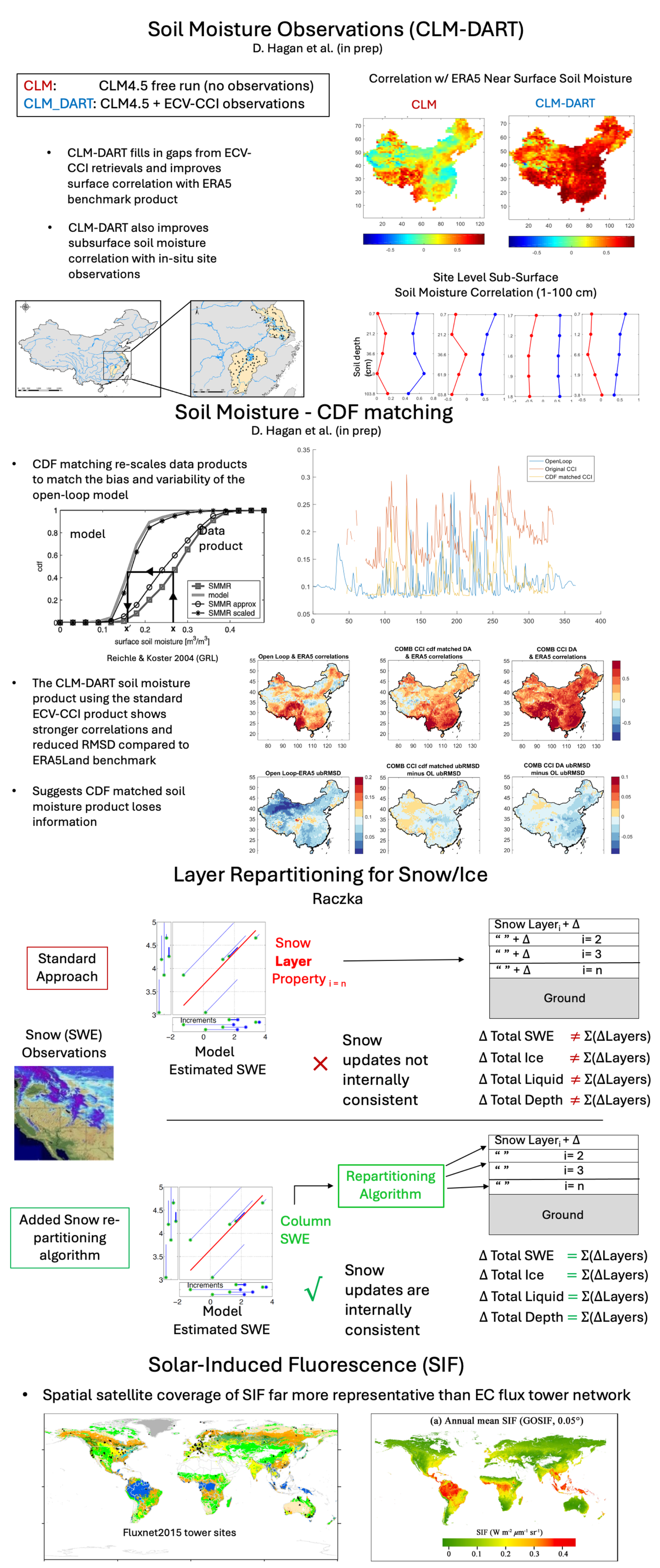
Postprocessing tools
effectively analyse assimilation output

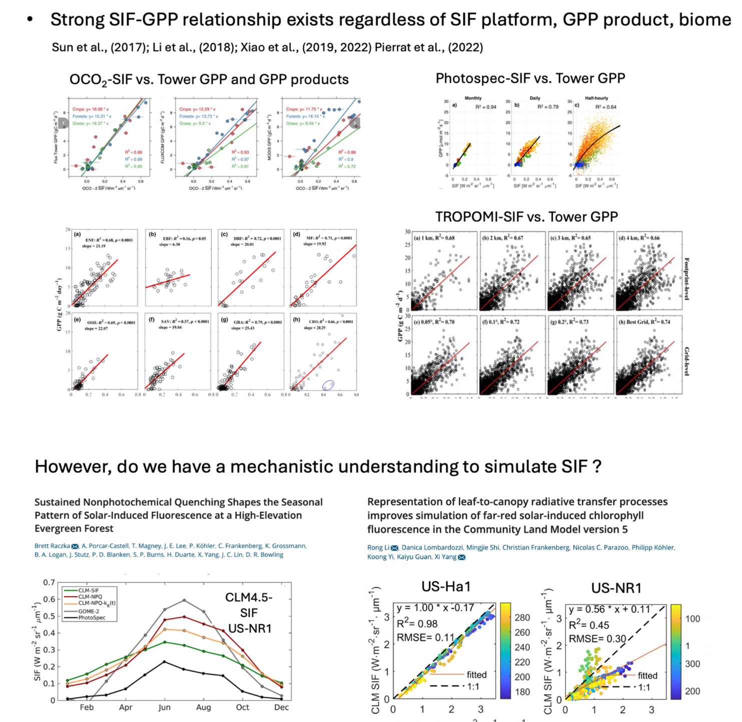
- in state space (on the model grid),
- in observation space (at the observation locations)
- from an ensemble statistical perspective.

Bonus!
Sensitivity analysis and analysis uncertainties can be derived from the ensembles.

2. Surface Observations and Models, and the CAM6+DART Reanalysis

Observations of surface and subsurface quantities (e.g. snow cover and soil moisture) can be assimilated into "surface" models with the help of an ensemble of atmospheric surface forcing. An example is the CAM6+DART Reanalysis, which covers years 2011-2020 (K. Raeder AMS2023), which slashes the cost of surface assimilations. These are most conveniently run in surface model configurations of the Community Earth System Model (CESM).





2.1 Interface to RTTOV v13; L. Kugler

Forward operators for the RTTOV model for assimilation of satellite radiances from NOAA-15...18; both RTTOV-direct (visible, infrared, and microwave) as well as RTTOV-scatt (microwave) computations. RTTOV = Radiative Transfer for Advanced TIROS Operational Vertical Sounder; TIROS = Television and Infrared Operational Satellite.

3. DART EXCLUSIVE: New Assimilation Algorithms

3.1 Anderson: Non-Gaussian and Nonlinear Ensemble DA Algorithms

DART now implements a novel efficient algorithm that allows use of *arbitrary* continuous observation priors and likelihoods for the generation of observation increments. The key innovation is selecting posterior ensemble members with the same quantiles with respect to the continuous posterior distribution as the prior ensemble had with respect to the prior continuous distribution. This is a generalization of both previously-documented square-root ensemble Kalman filters for normal distributions and non-parametric ensemble filters such as the rank histogram filter. Examples of new continuous priors that can be implemented include gamma, inverse gamma, beta, a sum of normal kernels, and a bounded rank histogram, which is a general non-parametric technique that works well for any application.

Then, doing the regression of observation quantile increments in a probit-transformed bivariate quantile space guarantees that the posterior ensembles for state variables also have all the advantages of the observation space quantile conserving posteriors.

For details see J. Anderson's AMS2023 oral and poster presentations.

3.2 Gharamti: Hybrid Ensemble-Variational Data Assimilation for Streamflow and Flood Prediction

The updated WRF-Hydro has been coupled to DART and named HydroDART (El Gharamti et al., 2021). Stream flow data from 194 gauges in Florida were assimilated from Sep 15th - Oct 15th, 2022, which includes the flooding due to Hurricane Ian. The boxplots (next column) show the prior root-mean-squared-errors (RMSE) resulting from 12 experiments:

OL open loop (no DA)

E80 a typical DA run using 80 members

E20-H0.# a hybrid prior sample covariance; 20 members linearly combined with a static background covariance matrix given weight 0.#, held constant.

E20-Ha Like E20-H0.# but # can evolve in time.

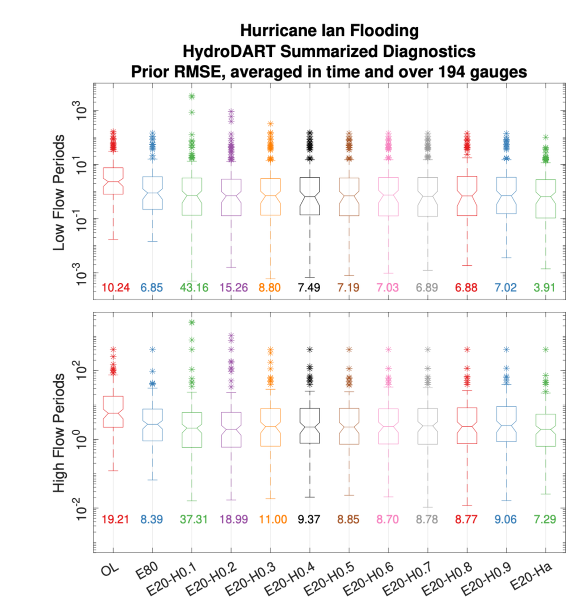


Figure 9: *Assimilation performance from 12 experiments (see text) performed for each of 2 stream flow regimes; low (top) and high (bottom). Each box summarizes the time-averaged RMSE of ensemble-mean stream flow relative to the observations (m^3s^{-1}), at the 194 gauges. * are the outliers. The overall RMSEs (averaged over all gauges) for each experiment are reported underneath the box plots.*

E20-Ha clearly outperforms E80 especially for low-flow periods where the standard EnKF suffers from low ensemble variability. On average, the estimates suggested by the E20-Ha scheme are 43% and 13% more accurate than those obtained using E80 for the low flow and high flow periods, respectively.

3.3 Gharamti: A Randomized Dormant Ensemble Kalman Filter

This new variant of the Ensemble Kalman Filter aims to improve the estimate of the background ensemble perturbations and mitigate variance underestimation. It uses prior ensembles constructed from active and dormant state realizations. At each assimilation cycle, a subset of the ensemble is randomly selected to go through the analysis scheme of the EnKF ("active members"). The remaining dormant members do not take part in the analysis. After the update, both active and dormant members are used to perform a forecast to get to the next data assimilation cycle.

This has several advantages over the fully-active EnKF:

+ The background ensemble spread given by the RD-EnKF is often larger, producing better consistency between the prior RMSE and ensemble spread.
+ sample covariances have better statistical properties (e.g., rank), which makes the algorithm computationally more stable.
+ Less tuning of localization.
+ Robust to changing observation networks.

For details see M. Gharamti's AMS2023 oral presentation.

4. Performance and Usability

4.1 Liu; State Compaction

Model grid points with no active variables are excluded from the state passed to filter. This enables DA with very large, but sparse, state vectors; This Red Sea model domain (resolution=0.01 degrees) is 90% (inactive) land. Only implemented for MIT gcm at the moment.

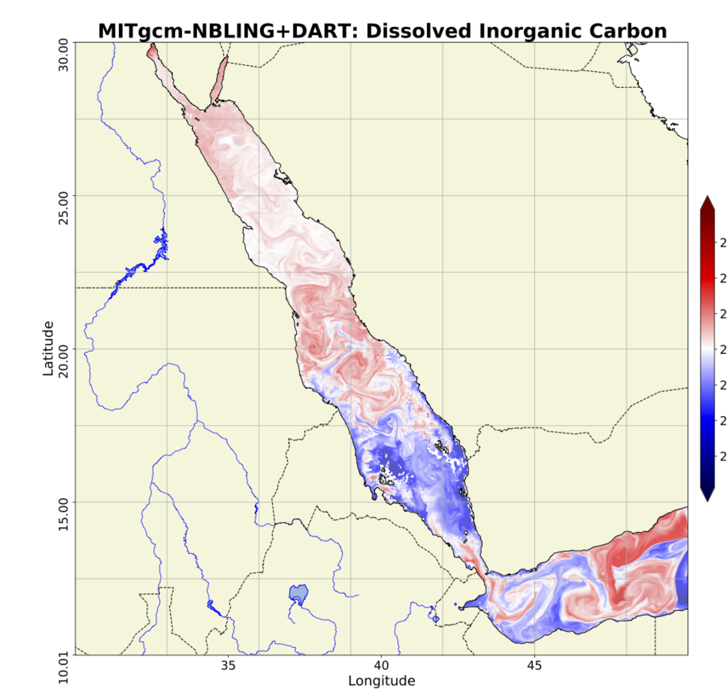


Figure 10: *This dissolved organic carbon distribution was generated by an atmosphere+ocean+biogeochemistry model and DART.*

4.2 Liu, Smith; Improved Caching

- Ed Liu (summer student) discovered and profiled, Marlee Smith fixed.
- Redundant caching in the get close obs cached and get close state cached subroutines has been removed.
- Savings of up to 20% run time.

4.3 Kershaw; Simplified Building Executables

- Reduced the number of files in DART by 30%
- Simplified the building of a single executable (default = build all)
- Enabled modifying a local ("work" directory) copy of a source file instead of the copy in its usual home.

4.4 Collins, Kershaw; Flexibility

New mechanism for defining obs quantities, localization, and model interface building:

- Removes the need to have a hard coded list of integers for DART quantities.
- Users can add new quantities by defining a QTY NAME

4.5 Labriola; CM1 and localization upgrades

- The DART interface to Cloud Model v1 (CM1) can now handle mixed periodic boundary conditions; the x andor y dimension can be periodic or not. It also now handles interpolation of 3D fields such as reflectivity.
- The threed cartesian location module can use multiple localization radii, so that each observation type can have a localization radius appropriate for its correlation characteristics.

4.6 Ishraque; Tracer advection model

This new model interface was implemented by a SIPaRCS summer student to solve a long-standing need for a low order model (Lorenz 96) which uses a semi-Langrangian tracer advection scheme.

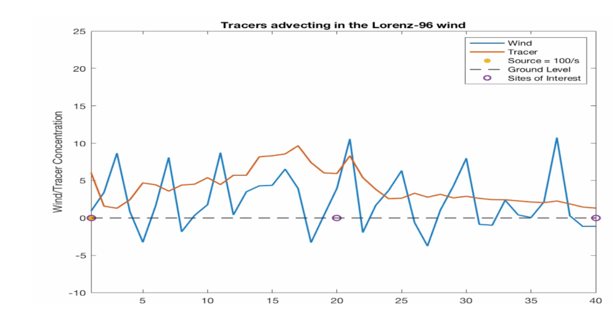


Figure 11: *Model state {wind,tracer} at 40 sites on a circle. Tracer source: site 1, $100s^{-1}$*

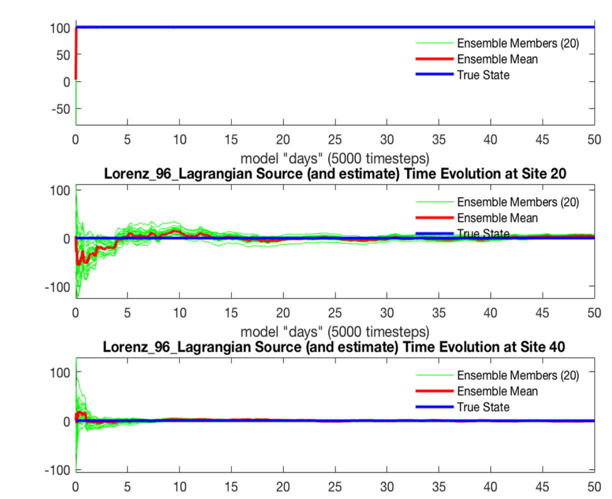


Figure 12: *Time evolution of the assimilation estimate of the tracer source strength at 3 sites.*

The assimilation not only constrains the wind and tracer to be close to the truth (not shown), but can identify the location and strength of the tracer source. Another example can be seen in J. Anderson's AMS2023 poster.

5. Acknowledgements

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. Any opinions, findings and conclusions or recommendations expressed in this material do not necessarily reflect the views of NSF.

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.