1. Ensemble Data Assimilation

1.1 What is Data Assimilation?

Data Assimilation (DA) combines observations of a physical system with predictions from a numerical forecast model. DA can be used for many purposes, including:

- constructing initial conditions for forecasts,
- evaluating errors in the model and observations,
- finding appropriate values for model parameters,
- designing better observational systems.

The Data Assimilation Research Testbed (DART) is a community software facility that can be used for all the above purposes. DART provides a variety of ensemble filtering (EF) algorithms.

1.2 Sequential Ensemble Filtering

- Includes many flavors of ensemble filters:
  1. EAKF: Ensemble Adjustment Kalman Filter,
  2. EnPF: Ensemble Kalman Filter,
  3. KF model,
  4. particle filter,
  5. a fixed-lag ensemble Kalman smoother.

- Provides additional algorithms for improved performance:
  1. prior and posterior inflation,
  2. automatic adaptive inflation,
  3. horizontal, vertical, multivariate localization,
  4. hierarchical filter for adaptive localization,
  5. dynamic adjustment of localization cutoff radius,
  6. a priori sampling error correction.

- Output is in portable netCDF files and one custom-format observation file. Matlab(r) scripts are provided to investigate:
  1. rank histograms,
  2. bias and spread (by variable) as a function of height or time,
  3. ensemble trajectories, error, and spread,
  4. 400
dump of the ensemble forecasts,
  5. 5D plots of observation densities and rejection attributes.

1.3 Geophysical applications require extensions

The basic EF algorithm does not work well when applied to large geophysical problems. Model error, sampling error from using affordable ensemble sizes, and violation of linear and Gaussian assumptions all lead to oversampling in the ensemble prior. This can result in poor performance or filter divergence. DART has several self-tuning algorithms to address these problems that work for a wide variety of models and observations without the need for user expertise. Some of these are described in sections 3 and 4.

2. What is DART?

DART makes it easy to learn and apply EF data assimilation.

- Has an extensive tutorial and instruction set.
- Incorporating new models and new observation types requires only minimal coding of a small set of interface routines.
- Scales linearly to hundreds of processors. Parallel performance is independent of the forecast model. Even single-threaded models can be run in parallel.

2.1 Level 1

- CAM: Community Atmosphere Model (spectral, FV cores),
- WRF: Weather and Research Forecast Model,
- MIT; general circulation model; annulus,
- ROSE; Middle atmosphere dynamics and chemistry,
- GFDD: grid point GCM dynamical cores.
- Two-layer primitive equation model (NOAA/CCD),
- 7 Single column (WRF) model.

2.2 Level 2

- Observation types that have been used include:
  1. upper air radiosondes, ACARS, satellite drift winds,
  2. surface winds(10m), T and Q(D0m), Pz,
  3. scatterometer winds,
  4. Doppler radial velocity and reflectivity,
  5. GPS radio occultation, refractivity,
  6. ground-based GPS.

2.3 Level 3

- Inflates the adaptive spatial inflation and its evolution over time at a pair of locations. CAM T85 L winds at least 18 (+ 268 KPA) at the end of one month of assimilating observations every 6 hours. The field started with a uniform value of 1.0. (Figure 3)

2.4 Level 4

- Includes many flavors of ensemble filters:
  1. EAKF: Ensemble Adjustment Kalman Filter,
  2. EnPF: Ensemble Kalman Filter,
  3. KF model,
  4. particle filter,
  5. a fixed-lag ensemble Kalman smoother.

- Provides additional algorithms for improved performance:
  1. prior and posterior inflation,
  2. automatic adaptive inflation,
  3. horizontal, vertical, multivariate localization,
  4. hierarchical filter for adaptive localization,
  5. dynamic adjustment of localization cutoff radius,
  6. a priori sampling error correction.

- Output is in portable netCDF files and one custom-format observation file. Matlab(r) scripts are provided to investigate:
  1. rank histograms,
  2. bias and spread (by variable) as a function of height or time,
  3. ensemble trajectories, error, and spread,
  4. 400
dump of the ensemble forecasts,
  5. 5D plots of observation densities and rejection attributes.

1.4 Level 5

- Contains an extensive suite of tools to evaluate computational performance.
- Contains an extensive suite of tools to evaluate computational performance.
- Contains an extensive suite of tools to evaluate computational performance.
- Contains an extensive suite of tools to evaluate computational performance.

3. Hierarchical Filter for Adaptive Localization

Sampling error from using small ensembles leads to spurious large correlations among weakly related observations and state variables. This results in systematic underestimation of posterior variance and can lead to filter divergence. Localizing the impact of an observation to nearby state variables has been the traditional solution. This requires expert knowledge and trial-and-error to get appropriate localization. DART provides an algorithm to automatically compute localizations using a small group of ensembles.

4. Adaptive Inflation

Model error and violation of linear and Gaussian assumptions are additional sources of insufficient variance in the ensemble prior. This can be ameliorated by ‘inflation’, where the ensemble spread is increased while maintaining the mean and sample correlations among all prior variables. Traditionally, all variables at all locations have been inflated by a constant value, chosen by the user to optimize performance in a time window or timestep. This can be time and computer resources and can never be optimal for the entire ensemble. Often, a value of inflation too low in one region will lead to uncontrolled growth of variance in another. DART has an adaptive inflation algorithm that uses the set of observations affecting a state variable to determine the best value of inflation for that variable.

5. 3. Parallel Scaling

Scaling runs were done using a state-of-the-art global atmospheric climate model (CAM) at low and medium resolutions on a commodity-level Linux cluster from Aspen Systems, an Intel-based Linux cluster from IBM, and a Power S+ AIX system with a high-speed switch. The following timing results are from the Aspen Systems cluster, and are representative of the results obtained on the other systems.

6. Try this at home!

Our DART web site is http://www.image.ucar.edu/DARt. There you will find information about how to download the latest version of DART from our subversion server. Information on how to set up a full DART tutorial (included with the distribution), and contact information for the DART development group.

References