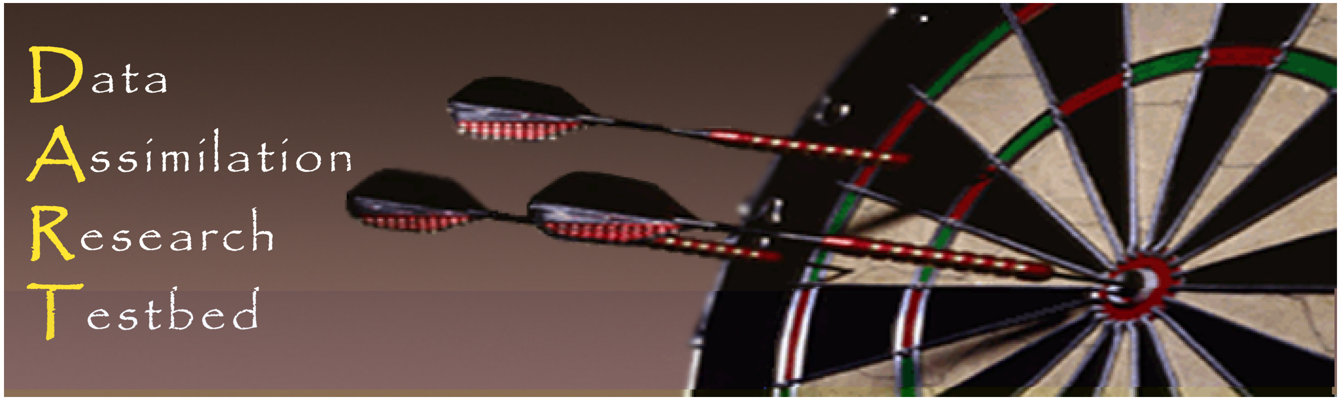
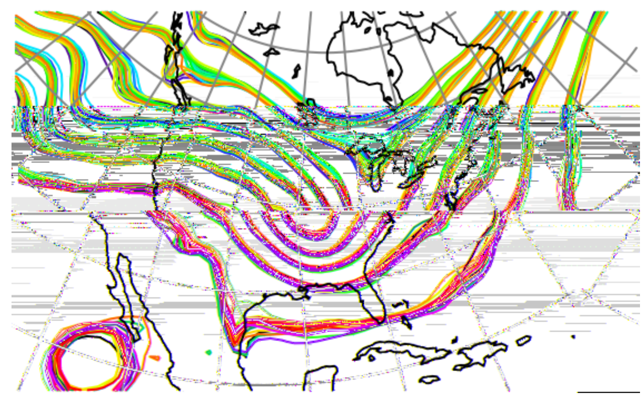


DART Tutorial Section 5: Comprehensive Filtering Theory: Non-Identity Observations and the Joint Phase Space



The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | National Center for
UCAR | Atmospheric Research

A More General Context for Filtering with Geophysical Models

Dynamical system governed by (stochastic) Difference Equation:

$$dx_t = f(x_t, t) + G(x_t, t)\, d\beta_t, \qquad t \geq 0 \tag{1}$$

Observations at discrete times:

$$y_k = h(x_k, t_k) + v_k; \qquad k = 1, 2, \ldots; \qquad t_{k+1} > t_k \geq t_0 \tag{2}$$

Observational error white in time and Gaussian (nice, not essential).

$$v_k \to N(0, R_k) \tag{3}$$

Complete history of observations is:

$$Y_\tau = \{ y_l; t_l \leq \tau \} \tag{4}$$

Goal: Find probability distribution for state at time t:

$$p(x, t \mid Y_t) \tag{5}$$

A More General Context for Filtering with Geophysical Models

State between observation times obtained from Difference Equation.
Need to update state given new observations:

$$p\left(x,t_k \mid Y_{t_k}\right) = p\left(x,t_k \mid y_k, Y_{t_{k-1}}\right) \tag{6}$$

Apply Bayes' rule:

$$p\left(x,t_k \mid Y_{t_k}\right) = \frac{p(y_k \mid x_k, Y_{t_{k-1}})\, p(x,t_k \mid Y_{t_{k-1}})}{p(y_k \mid Y_{t_{k-1}})} \tag{7}$$

Noise is white in time (3), so:

$$p\left(y_k \mid x_k, Y_{t_{k-1}}\right) = p\left(y_k \mid x_k\right) \tag{8}$$

Integrate numerator to get normalizing denominator:

$$p(y_k \mid Y_{t_{k-1}}) = \int p(y_k \mid x)\, p(x,t_k \mid Y_{t_{k-1}})\, dx \tag{9}$$

A More General Context for Filtering with Geophysical Models

Probability after new observation:

$$p\left(x,t_k \mid Y_{t_k}\right) = \frac{p\left(y_k \mid x\right)p\left(x,t_k \mid Y_{t_{k-1}}\right)}{\int p(y_k \mid \xi)p(\xi,t_k \mid Y_{t_{k-1}})d\xi} \qquad (10)$$

Exactly analogous to earlier derivation except that $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors.

EXCEPT, no guarantee we have prior sample for each observation.

SO, let's make sure we have priors by 'extending' state vector.

A More General Context for Filtering with Geophysical Models

Extending the state vector to joint state-observation vector.

$$y_k = h(x_k, t_k) + v_k; \qquad k = 1, 2, \ldots; \qquad t_{k+1} > t_k \geq t_0 \qquad (2)$$

Applying h to $\boldsymbol{x}$ at a given time gives expected values of observations.

Get prior sample of observations by applying h to each sample of state vector $\boldsymbol{x}$.

Let $z = [\boldsymbol{x}, \boldsymbol{y}]$ be the combined vector of state and observations.

A More General Context for Filtering with Geophysical Models

NOW, we have a prior for each observation:

$$p\left(z,t_k \mid Y_{t_k}\right) = \frac{p\left(y_k \mid z\right) p\left(z,t_k \mid Y_{t_{k-1}}\right)}{\int p(y_k \mid \xi) p(\xi,t_k \mid Y_{t_{k-1}}) d\xi} \qquad (10.\text{ext})$$

Dealing with Many Observations

One more issue: dealing with many observations in set y_k?

Let y_k be composed of s subsets of observations:

$$y_k = \left\{ y_k^1, y_k^2, \ldots, y_k^s \right\}$$

Observational errors for obs. in set i independent of those in set j.

Then:

$$p\left(y_k \mid z\right) = \prod_{i=1}^{s} p\left(y_k^i \mid z\right)$$

Can rewrite (10.ext) as series of products and normalizations.

Dealing with Many Observations

One more issue: dealing with many observations in set y_k?

Implication: can assimilate observation subsets sequentially.

If subsets are scalar (individual obs. have mutually independent error distributions), can assimilate each observation sequentially.

If not, have two options:

1. Repeat everything above with matrix algebra.
2. Do singular value decomposition; diagonalize obs. error covariance.
   Assimilate observations sequentially in rotated space.
   Rotate result back to original space.

Good news: *Most geophysical obs. have independent errors!*

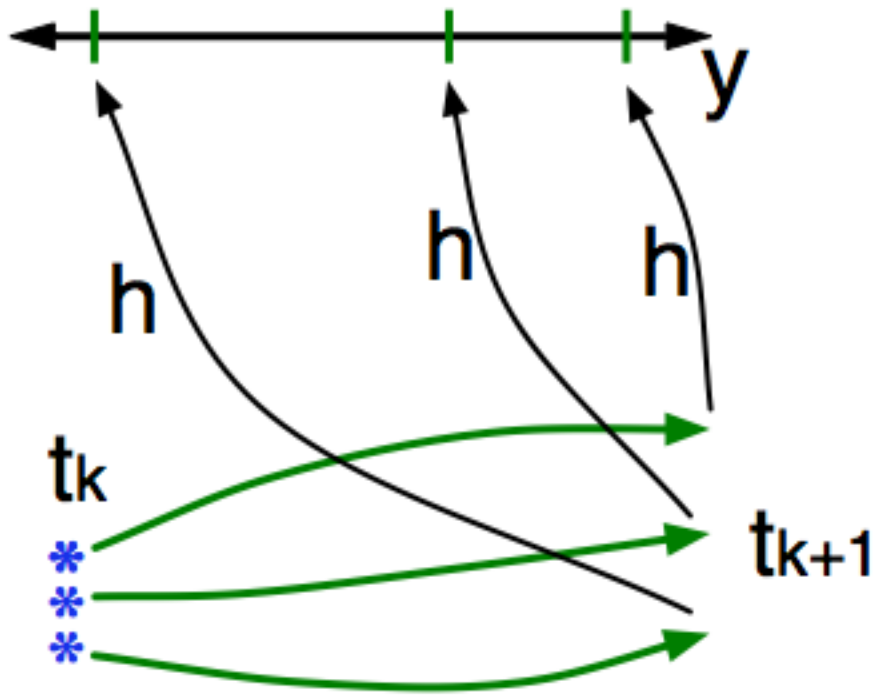
How an Ensemble Filter Works for Geophysical Data Assimilation

1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

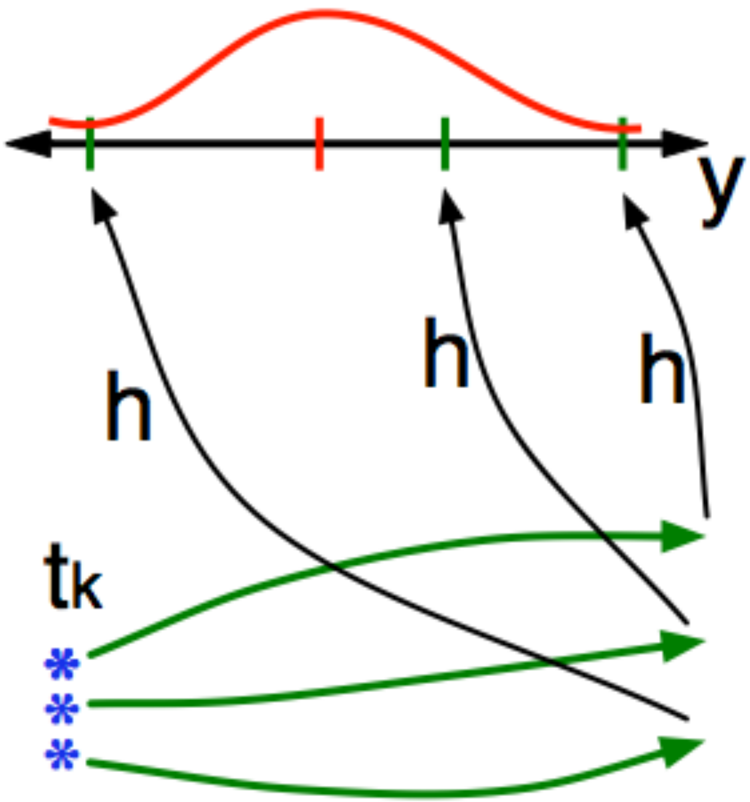
Ensemble state at time of next observation (prior)

2. Get prior ensemble sample of observation, $y = h(x)$, by applying forward operator **h** to each ensemble member.

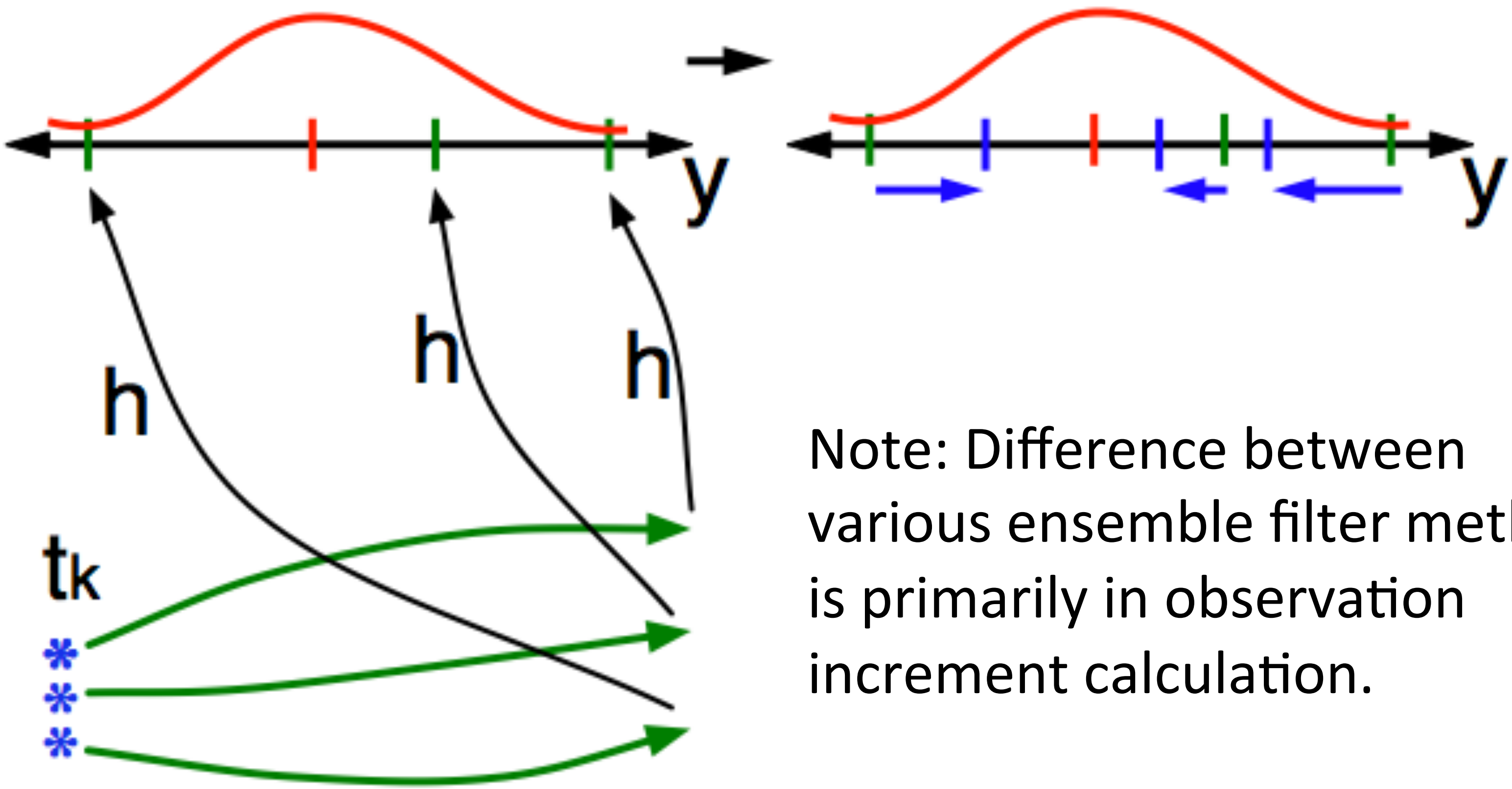


Theory: observations from instruments with uncorrelated errors can be done sequentially.

3. Get observed value and observational error distribution from observing system.

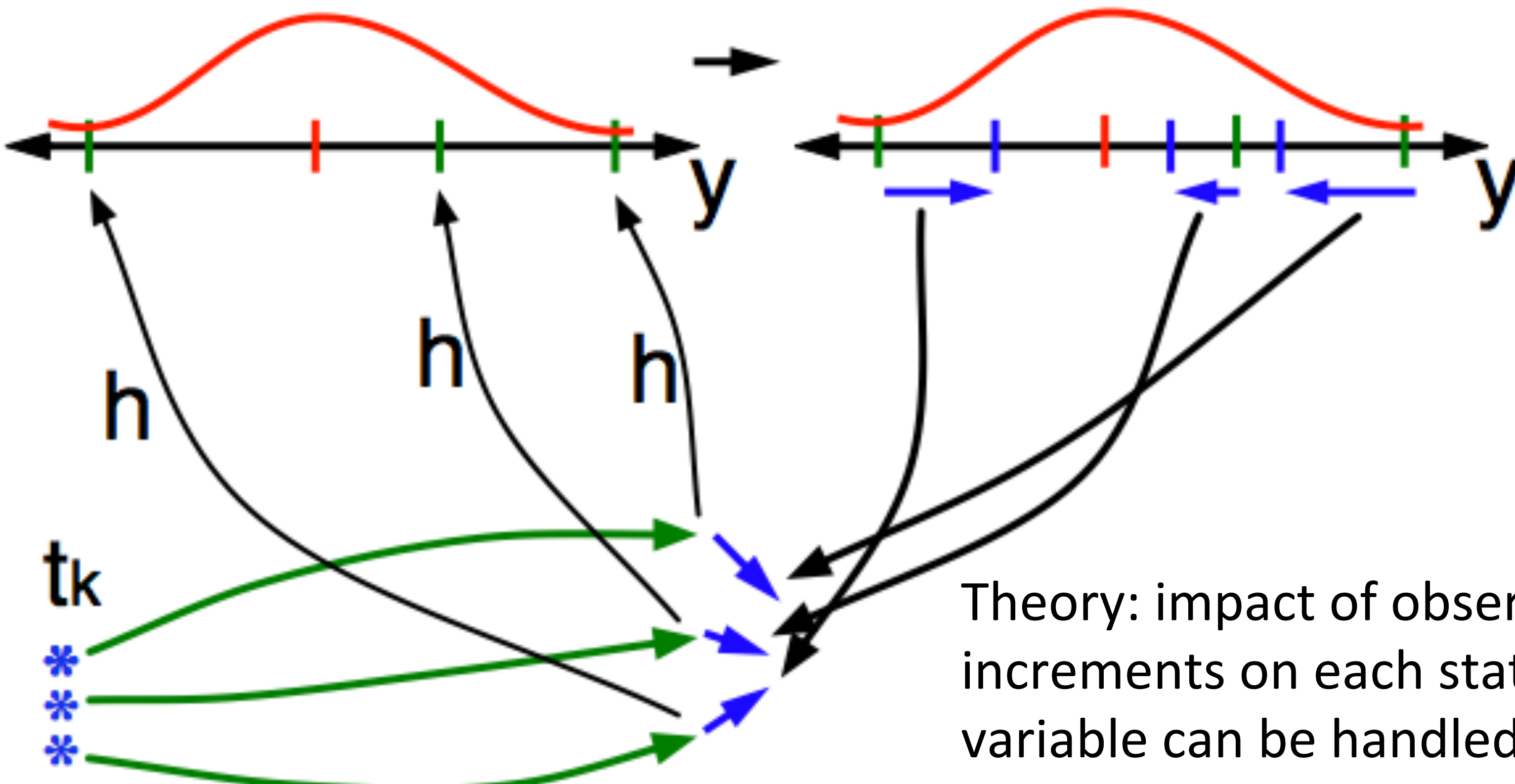


4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



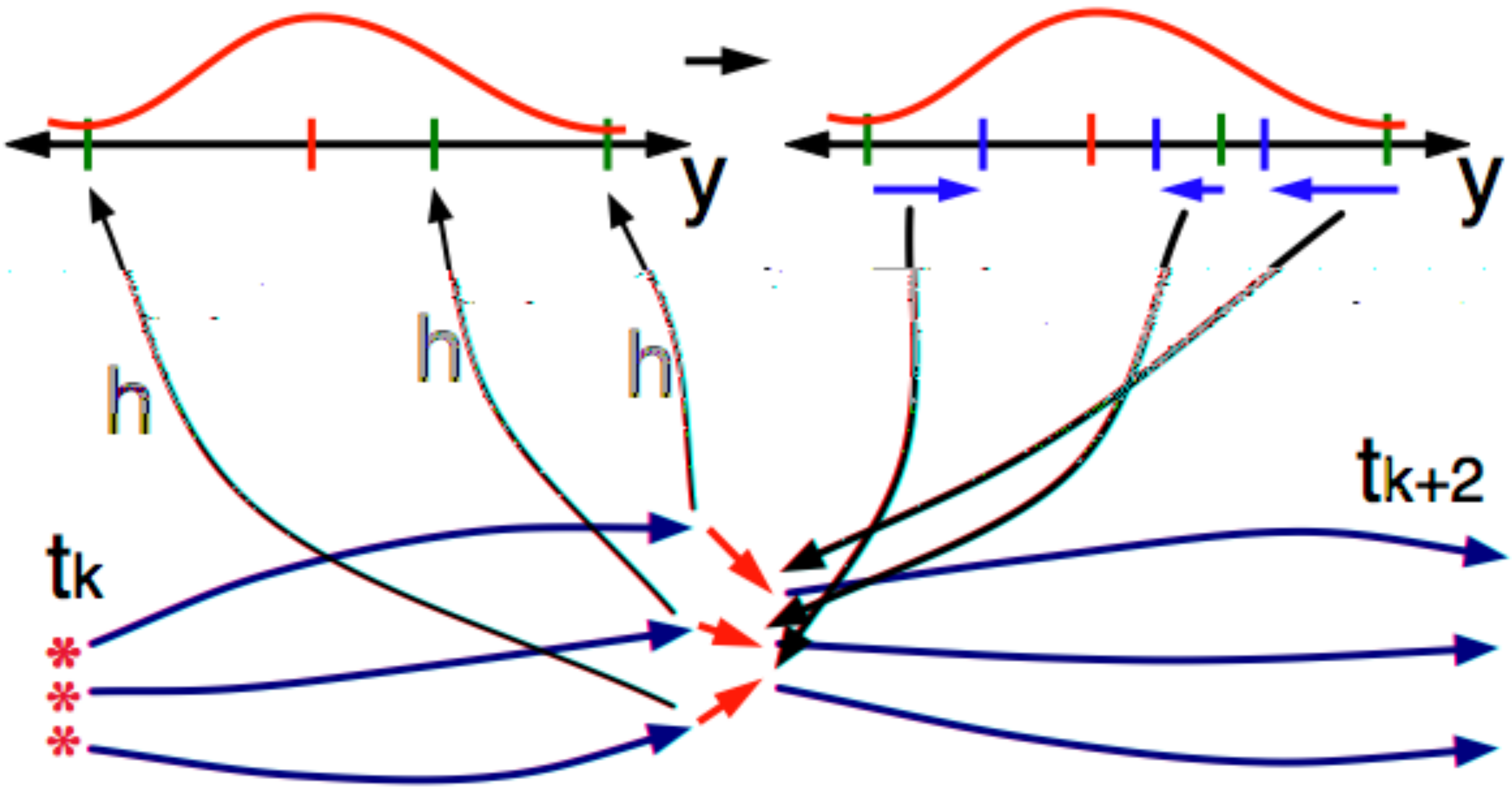
Note: Difference between various ensemble filter methods is primarily in observation increment calculation.

5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



Theory: impact of observation increments on each state variable can be handled independently!

6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …



Non-Identity Observation Operators in Lorenz_63:

Try observing mean(x, y), mean(y, z), mean(z, x) using `obs_seq.out.average` as input file.

Same error variance and frequency as previously.

In *models/lorenz_63/work* edit *input.nml*

```
&filter_nml
   …
   obs_sequence_in_name = "obs_seq.out.z"
```

Change to obs_seq.out.average

Execute *./filter* program to produce a new assimilation.
Look at the error statistics and time series with Matlab.
Record the error and spread values and compare to identity case.

Error is much larger!
Identity observations remove all regression error;
can be very misleading.

DART Tutorial Index to Sections