

Data
Assimilation
Research
Testbed



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



©UCAR 2014



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:

$$dx_t = f(x_t, t) + G(x_t, t) d\beta_t, \quad t \geq 0 \quad (1)$$

Observations at discrete times:

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

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

$$v_k \rightarrow N(0, R_k) \quad (3)$$

Complete history of observations is:

$$Y_\tau = \{y_l; t_l \leq \tau\} \quad (4)$$

Goal: Find probability distribution for state at time t:

$$p(x, t | Y_t) \quad (5)$$

A More General Context for Filtering with Geophysical Models

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

$$p(x, t_k | Y_{t_k}) = p(x, t_k | y_k, Y_{t_{k-1}}) \quad (6)$$

Apply Bayes' rule:

$$p(x, t_k | Y_{t_k}) = \frac{p(y_k | x_k, Y_{t_{k-1}}) p(x, t_k | Y_{t_{k-1}})}{p(y_k | Y_{t_{k-1}})} \quad (7)$$

Noise is white in time (3), so:

$$p(y_k | x_k, Y_{t_{k-1}}) = p(y_k | x_k) \quad (8)$$

Integrate numerator to get normalizing denominator:

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

A More General Context for Filtering with Geophysical Models

Probability after new observation:

$$p(x, t_k | Y_{t_k}) = \frac{p(y_k | x) p(x, t_k | Y_{t_{k-1}})}{\int p(y_k | \xi) p(\xi, t_k | Y_{t_{k-1}}) d\xi} \quad (10)$$

Exactly analogous to earlier derivation except that x and 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.

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

Applying h to x at a given time gives expected values of observations.

Get prior sample of observations by applying h to each sample of state vector x .

Let $z = [x, y]$ 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:

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

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:

$$y_k = \{y_k^1, y_k^2, \dots, y_k^s\}$$

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

Then:

$$p(y_k | z) = \prod_{i=1}^s p(y_k^i | z)$$

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**)

t_k



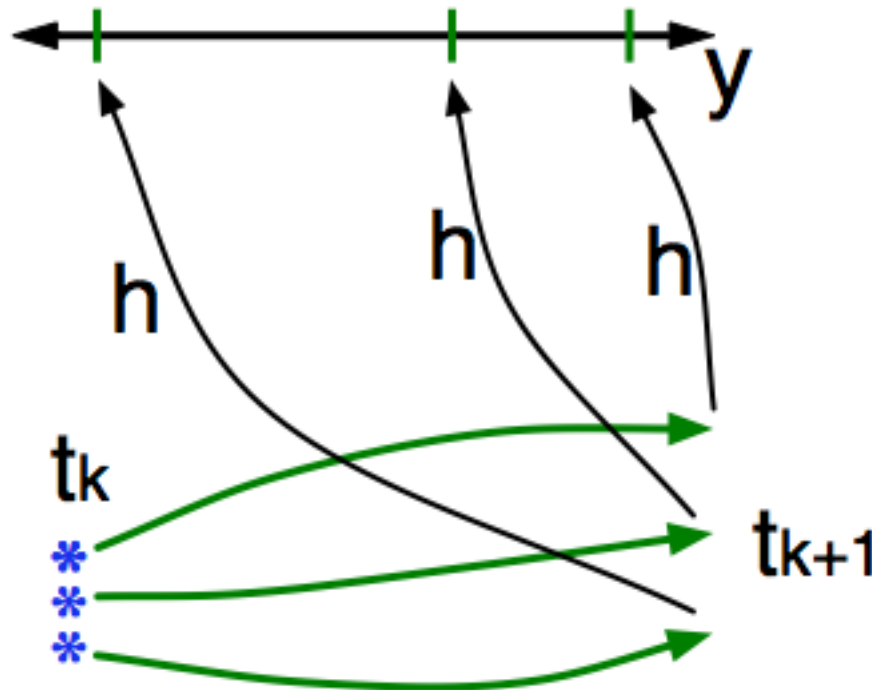
Ensemble state
at time of next
observation
(**prior**)

t_{k+1}



How an Ensemble Filter Works for Geophysical Data Assimilation

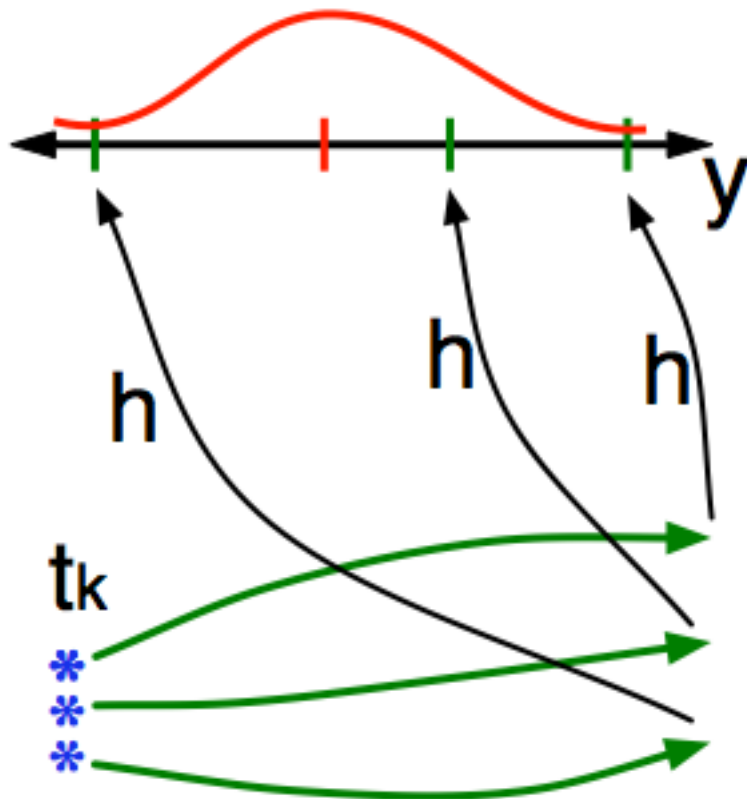
2. Get prior ensemble sample of observation, $y = h(x)$, by applying forward operator h to each ensemble member.



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

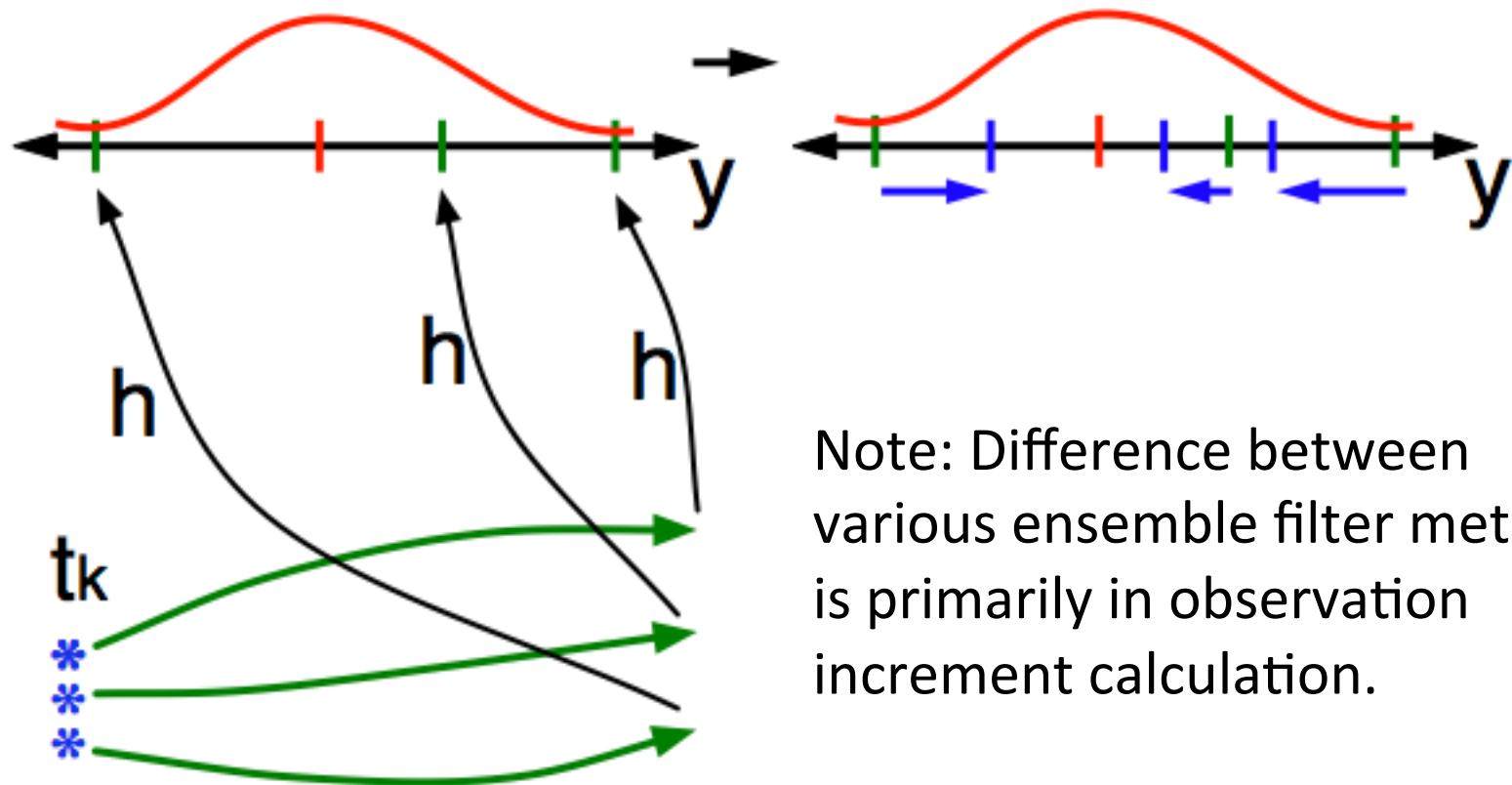
How an Ensemble Filter Works for Geophysical Data Assimilation

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



How an Ensemble Filter Works for Geophysical Data Assimilation

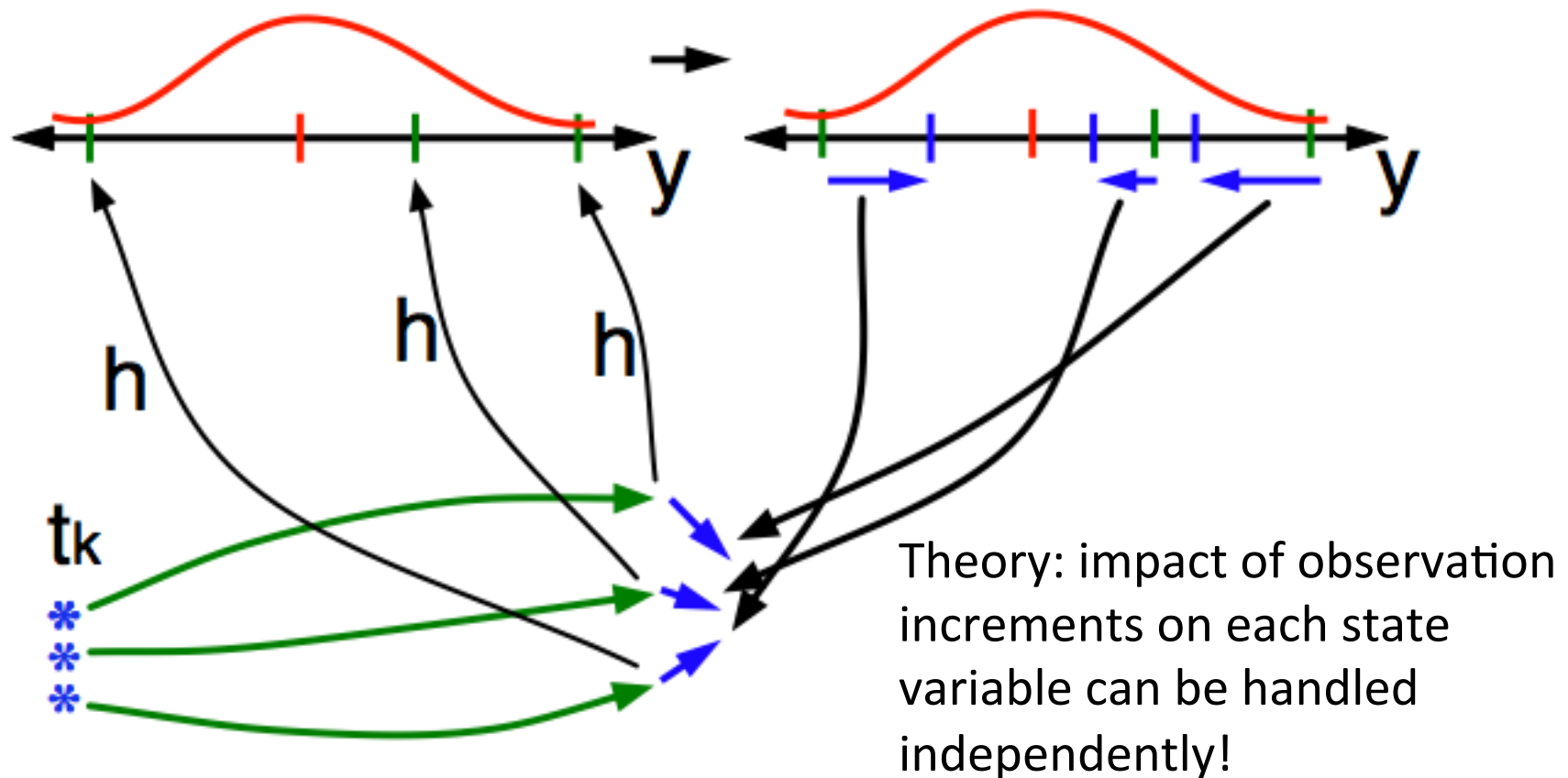
- Find the **increments** for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



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

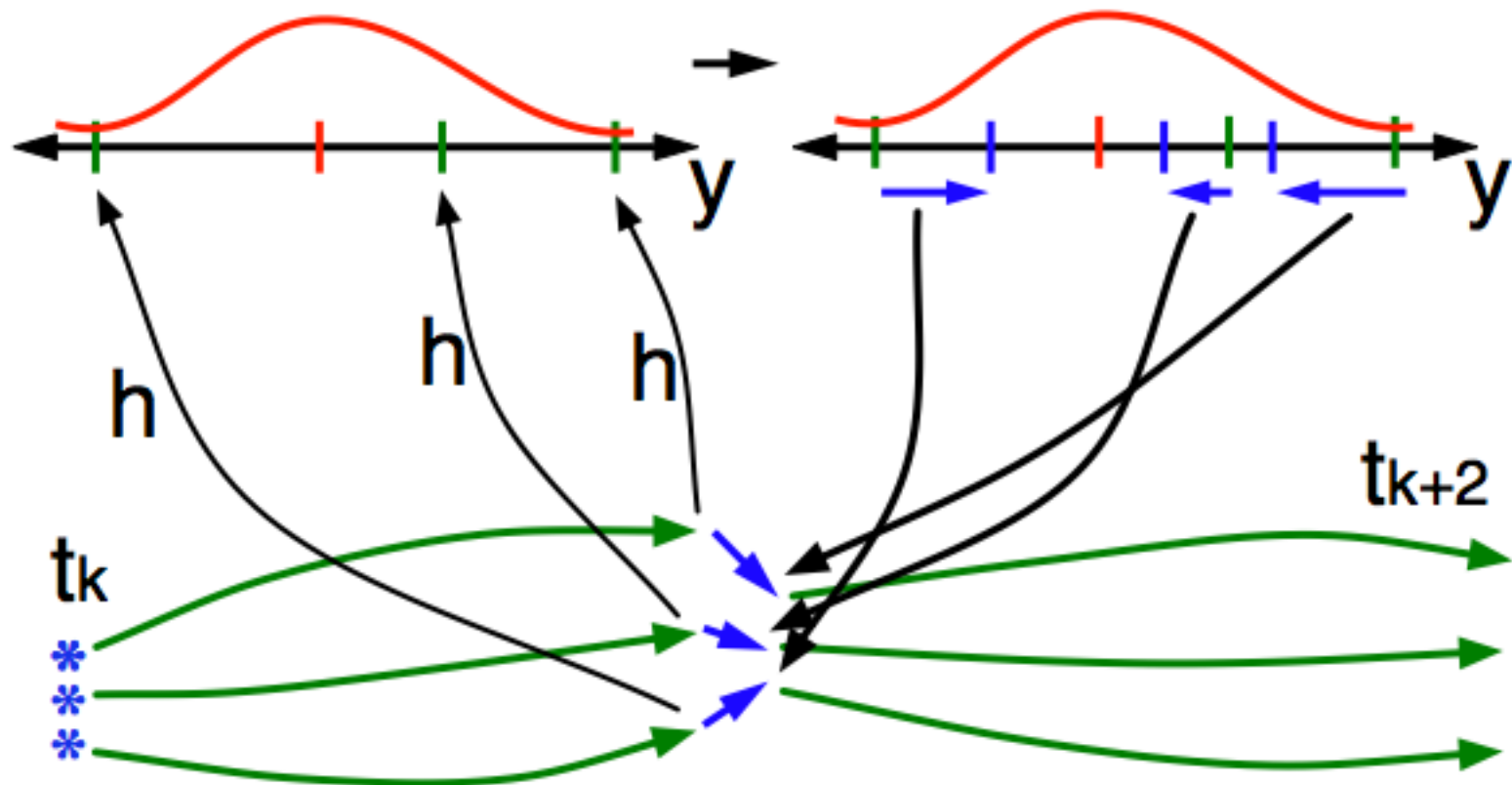
How an Ensemble Filter Works for Geophysical Data Assimilation

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



How an Ensemble Filter Works for Geophysical Data Assimilation

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 $\text{mean}(x, y)$, $\text{mean}(y, z)$, $\text{mean}(z, x)$ 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

1. Filtering For a One Variable System
2. The DART Directory Tree
3. DART Runtime Control and Documentation
4. How should observations of a state variable impact an unobserved state variable?
Multivariate assimilation.
5. Comprehensive Filtering Theory: Non-Identity Observations and the Joint Phase Space
6. Other Updates for An Observed Variable
7. Some Additional Low-Order Models
8. Dealing with Sampling Error
9. More on Dealing with Error; Inflation
10. Regression and Nonlinear Effects
11. Creating DART Executables
12. Adaptive Inflation
13. Hierarchical Group Filters and Localization
14. Quality control
15. DART Experiments: Control and Design
16. Diagnostic Output
17. Creating Observation Sequences
18. Lost in Phase Space: The Challenge of Not Knowing the Truth
19. DART-Compliant Models and Making Models Compliant
20. Model Parameter Estimation
21. Observation Types and Observing System Design
22. Parallel Algorithm Implementation
23. Location module design (not available)
24. Fixed lag smoother (not available)
25. A simple 1D advection model: Tracer Data Assimilation