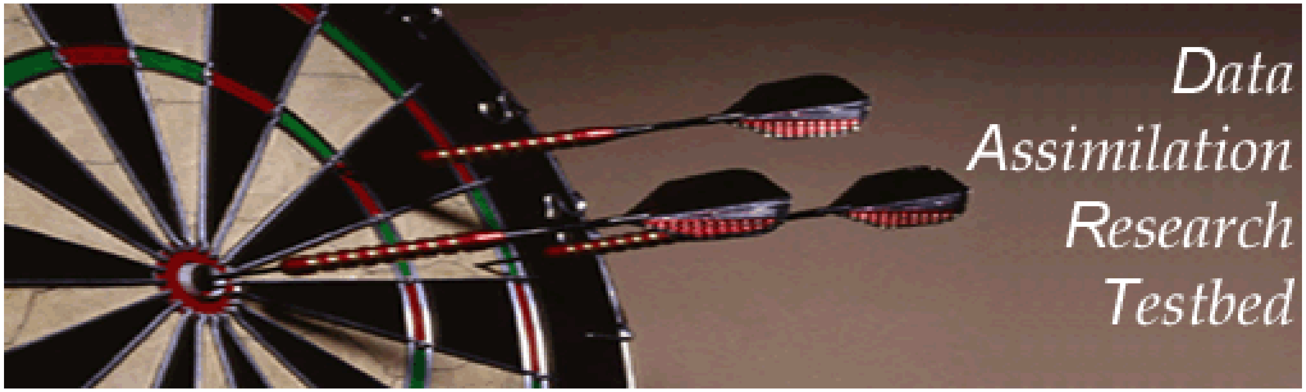
Capabilities and Limitations of Ensemble Data Assimilation

Jeffrey Anderson
National Center for Atmospheric Research
Data Assimilation Research Section



Ensemble filters work for many geophysical assimilation problems:

- Global and regional numerical weather prediction;
- Global and regional ocean prediction;
- Trace gas assimilation;
- Model parameter estimation.

However, there are some problems they cannot handle.

Talk will highlight things that can be difficult for filters.

Does your model / observing system have one of these?

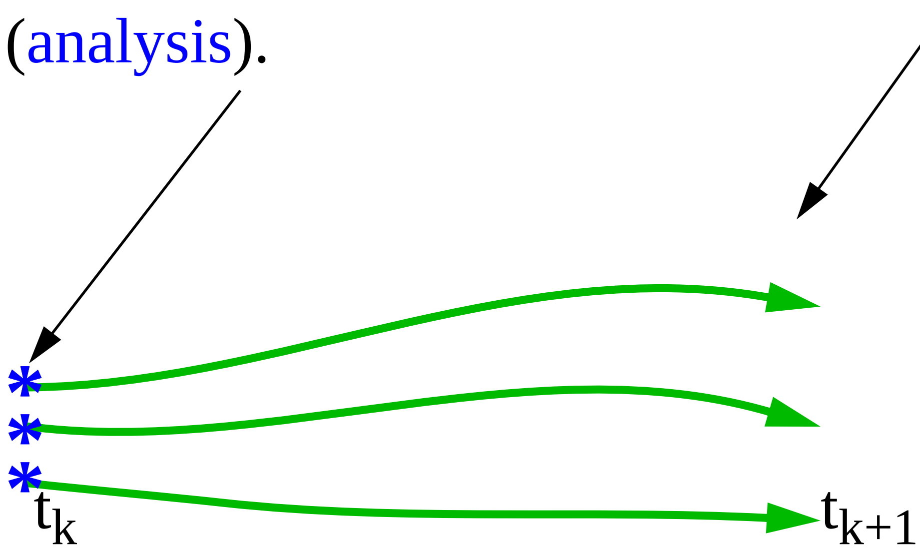
If not, ensemble filters are likely to be a powerful, flexible tool.

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.

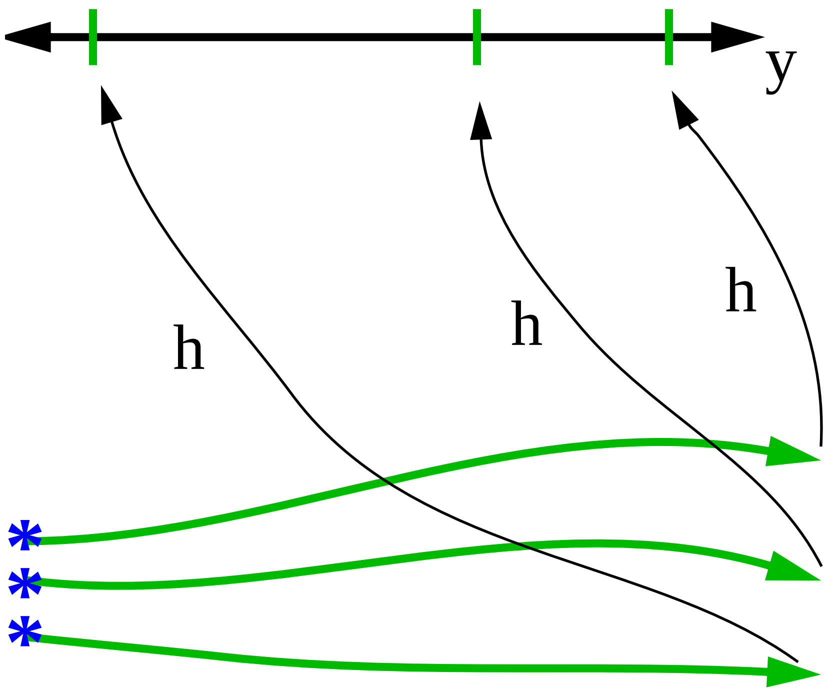
Ensemble state estimate after using previous observation (analysis).

Ensemble state at time of next observation (prior).



How an Ensemble Filter Works for Geophysical Data Assimilation

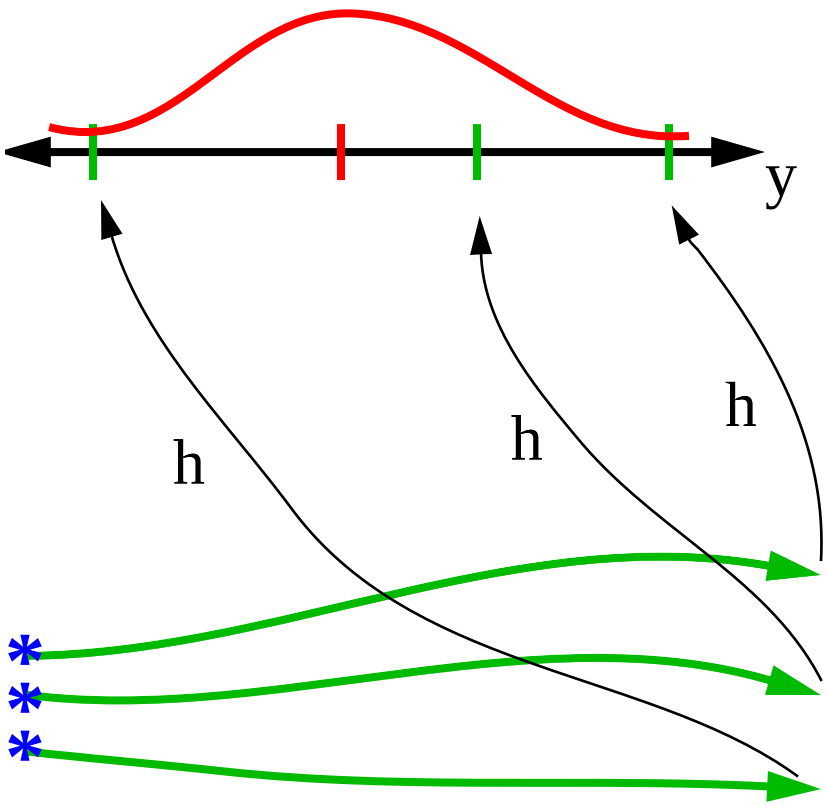
2. Get prior ensemble sample of observation, y=h(x), by applying forward operator h to each ensemble member.



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

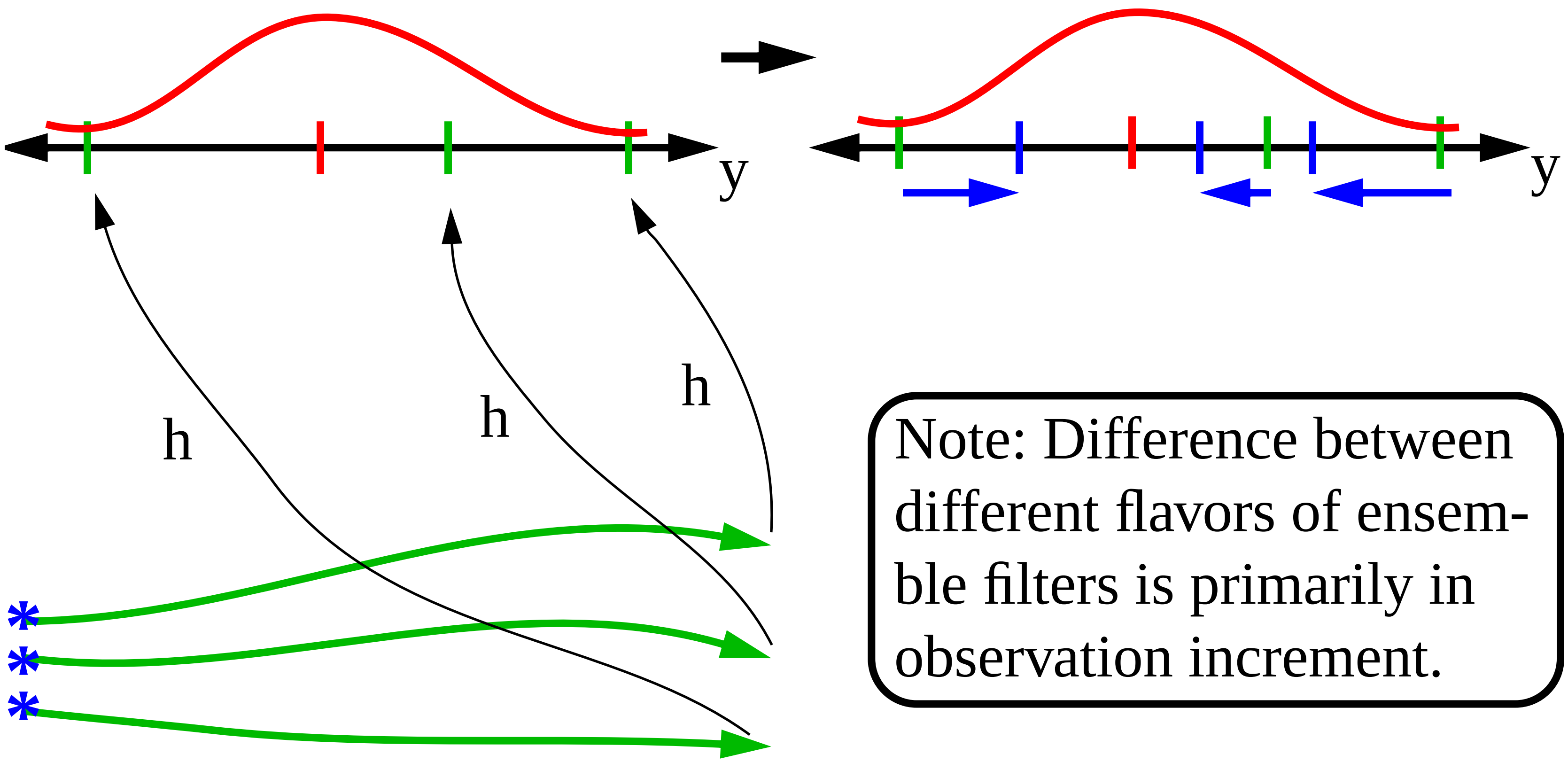
How an Ensemble Filter Works for Geophysical Data Assimilation

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



How an Ensemble Filter Works for Geophysical Data Assimilation

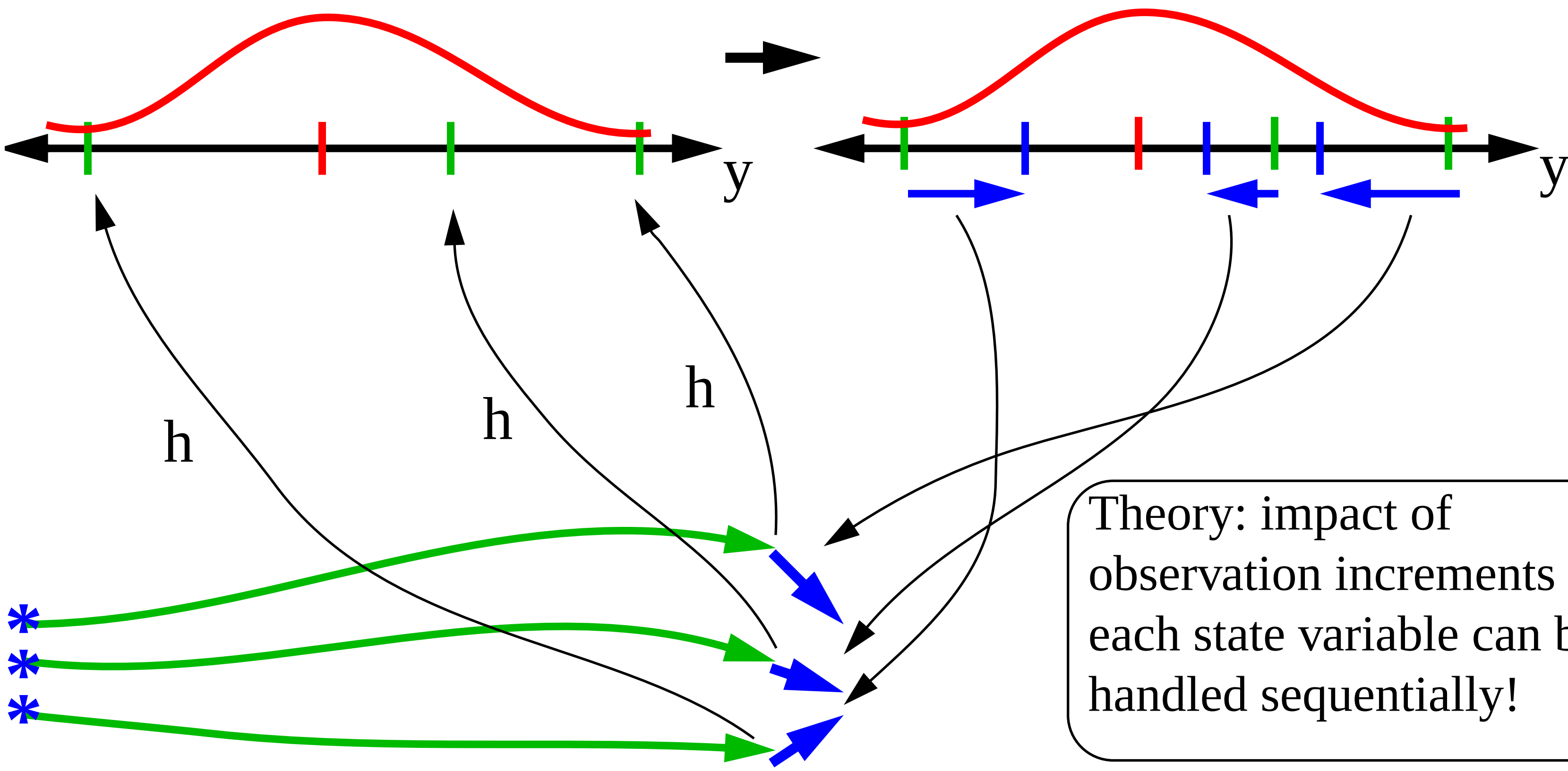
4. Find increment for each prior observation ensemble
(this is a scalar problem for uncorrelated observation errors).



Note: Difference between different flavors of ensemble filters is primarily in observation increment.

How an Ensemble Filter Works for Geophysical Data Assimilation

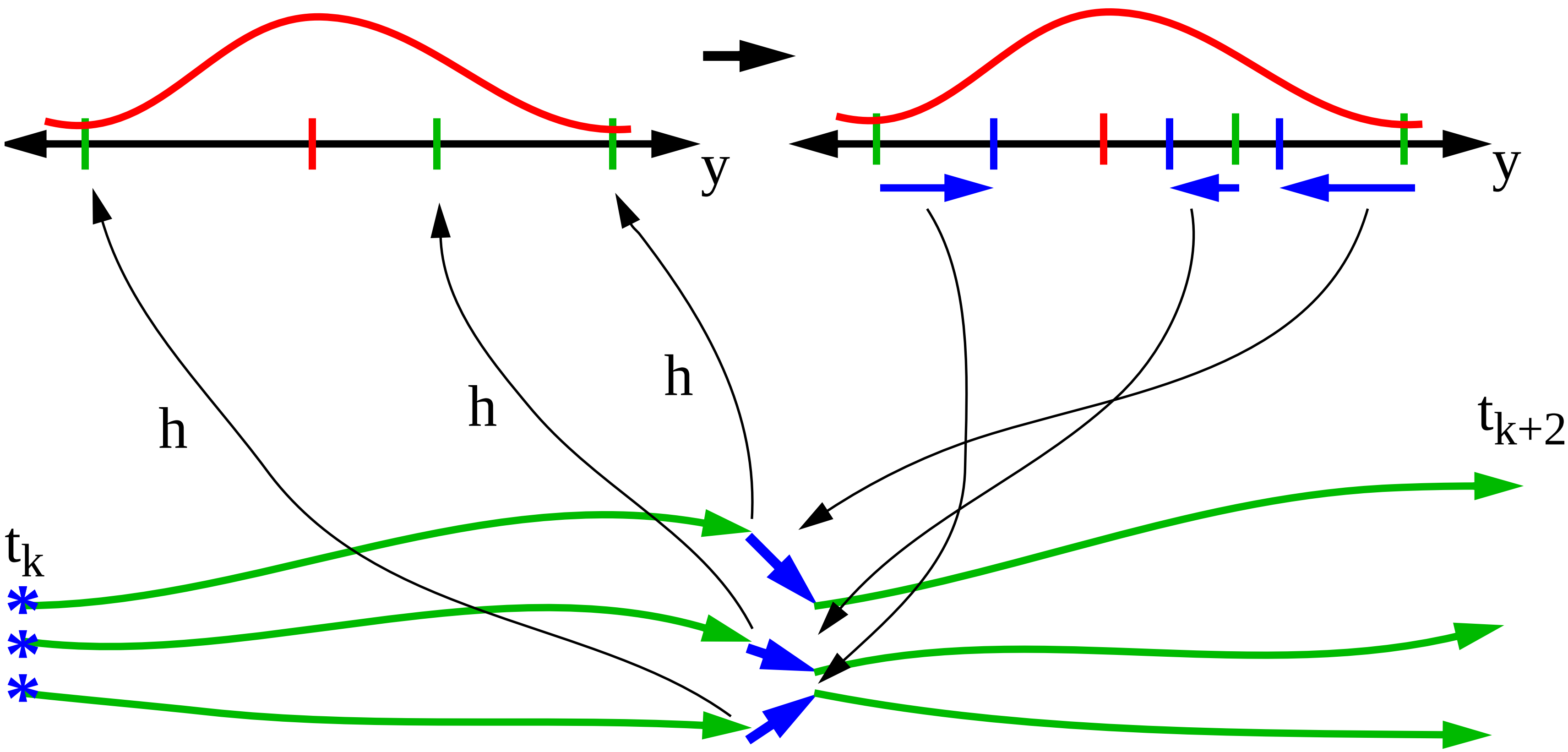
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



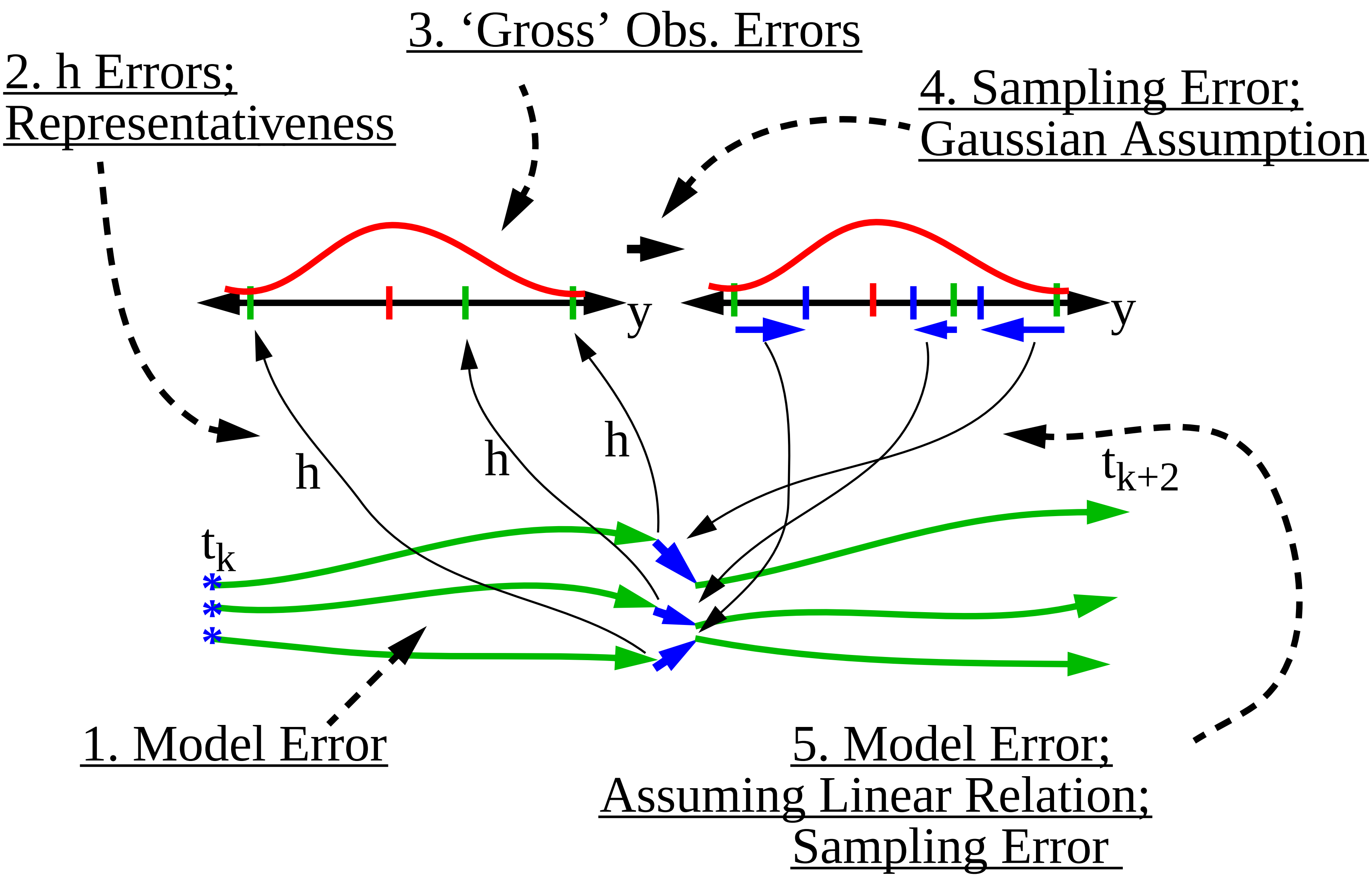
Theory: impact of observation increments on each state variable can be handled sequentially!

How an Ensemble Filter Works for Geophysical Data Assimilation

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

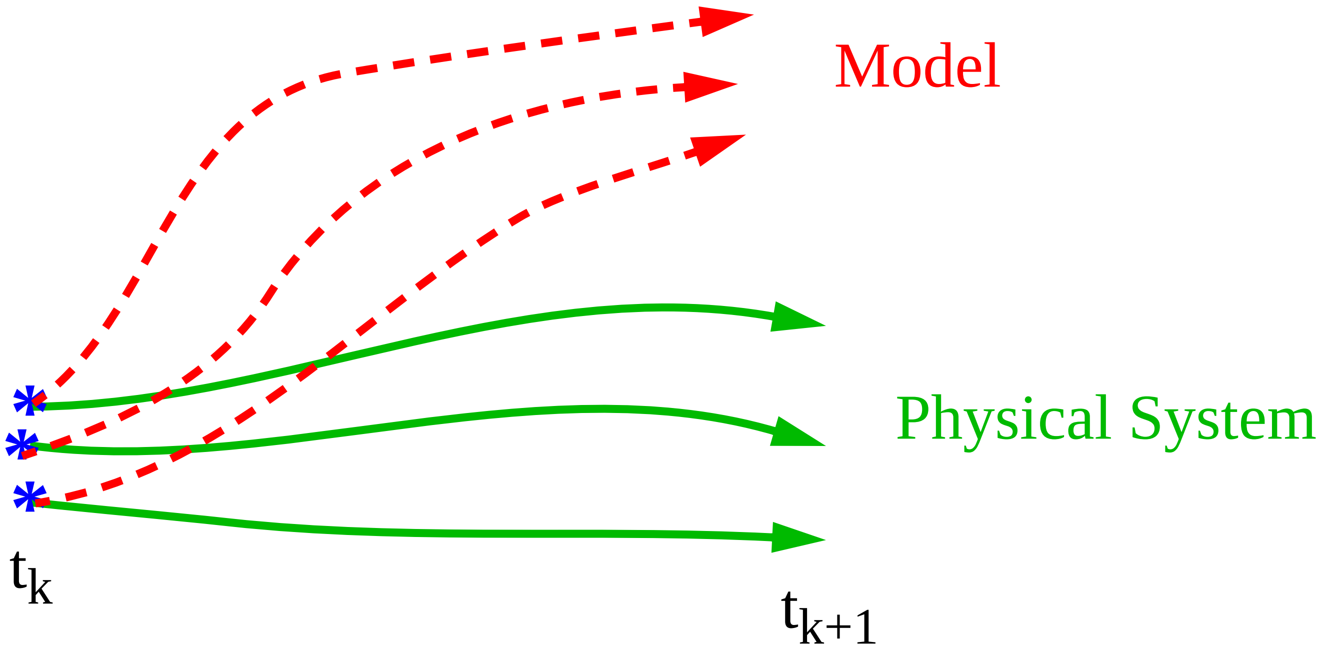


Some Error Sources in Ensemble Filters



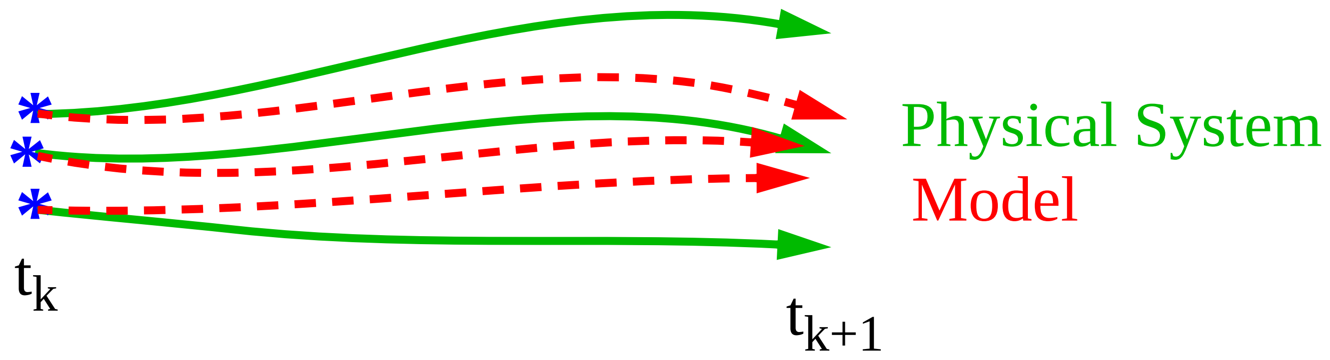
Challenges for Ensemble Filters: 1. Model Systematic Error

Assimilation can give observations too little weight.
Adaptive hierarchical Bayesian algorithms can correct this.
A REALLY bad model may contain no information.



Challenges: 2. Lack of model error variance growth

Assimilation can give observations too little weight.
Adaptive hierarchical Bayesian algorithms can correct this.
Far too little growth (or decay) can be cataclysmic.

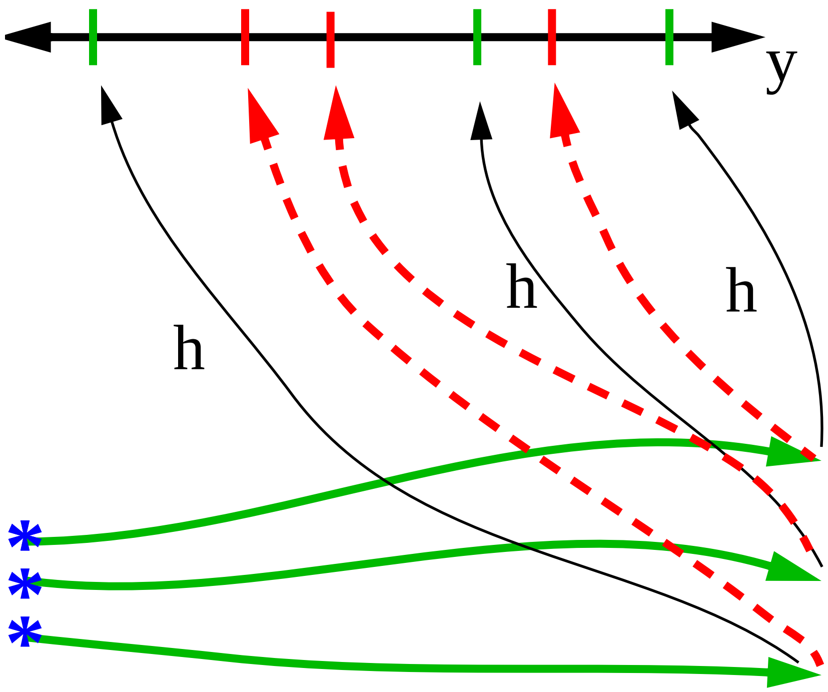


Challenges: 3. Model can't give good estimate of observed quantity

There is no h so that observation y=h(x).

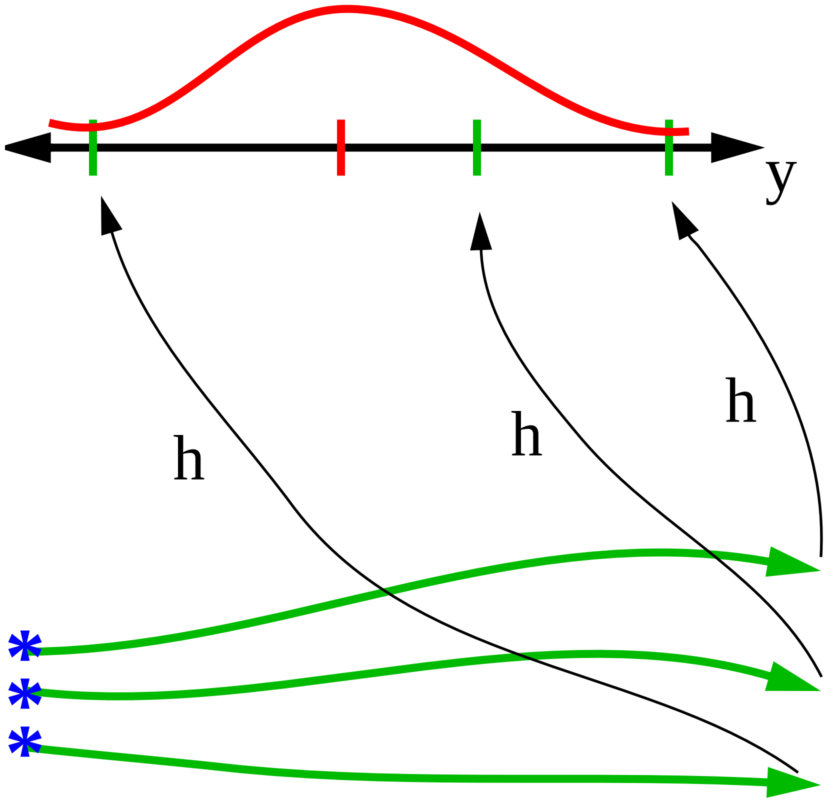
Physical/temporal scales or quantities needed to compute y are not in model.

Some error in h can be dealt with.



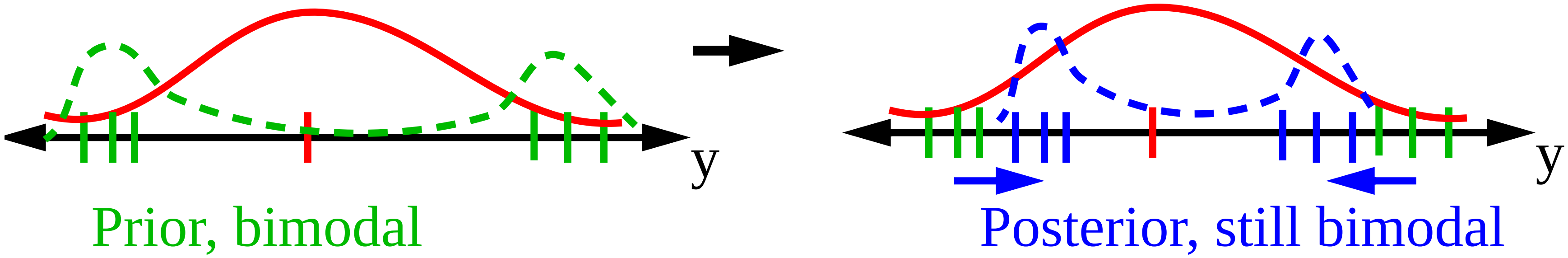
Challenges 4: Observational error distribution is poorly known

Bad estimates of variance (or shape) can bias filter.
Sporadic, extreme errors can be big problem.
Unknown instrument bias hard to detect, can be problem.



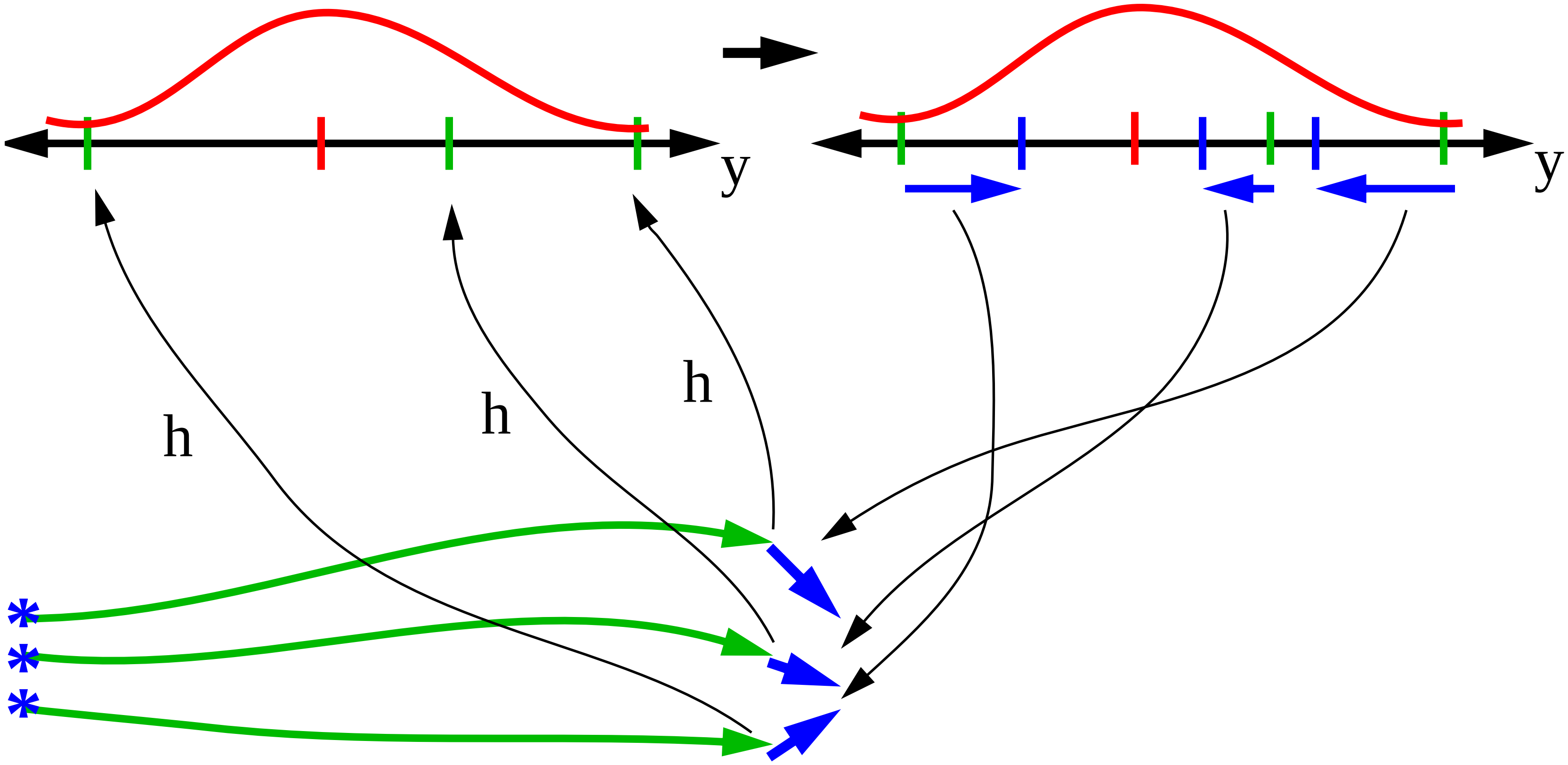
Challenges 5: Prior distribution is meaningfully non-Gaussian

If only in one dimension, filters can do okay.
If a handful of dimensions, kernel filters can work.
If more than a few, fundamentally out of luck.

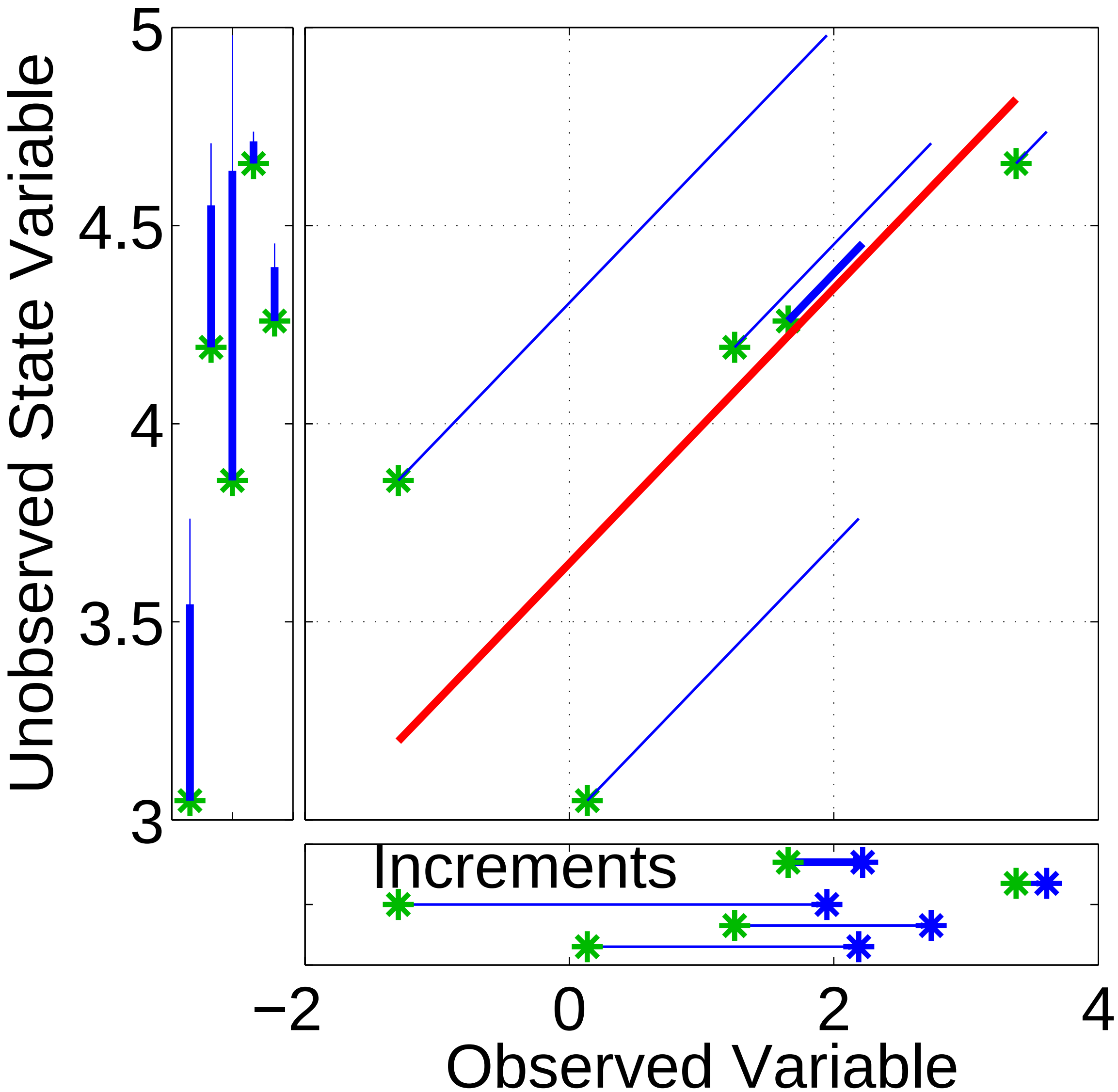


Challenge 6: Model sample covariance estimates are bad

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



Challenge 6: Model sample covariance estimates are bad



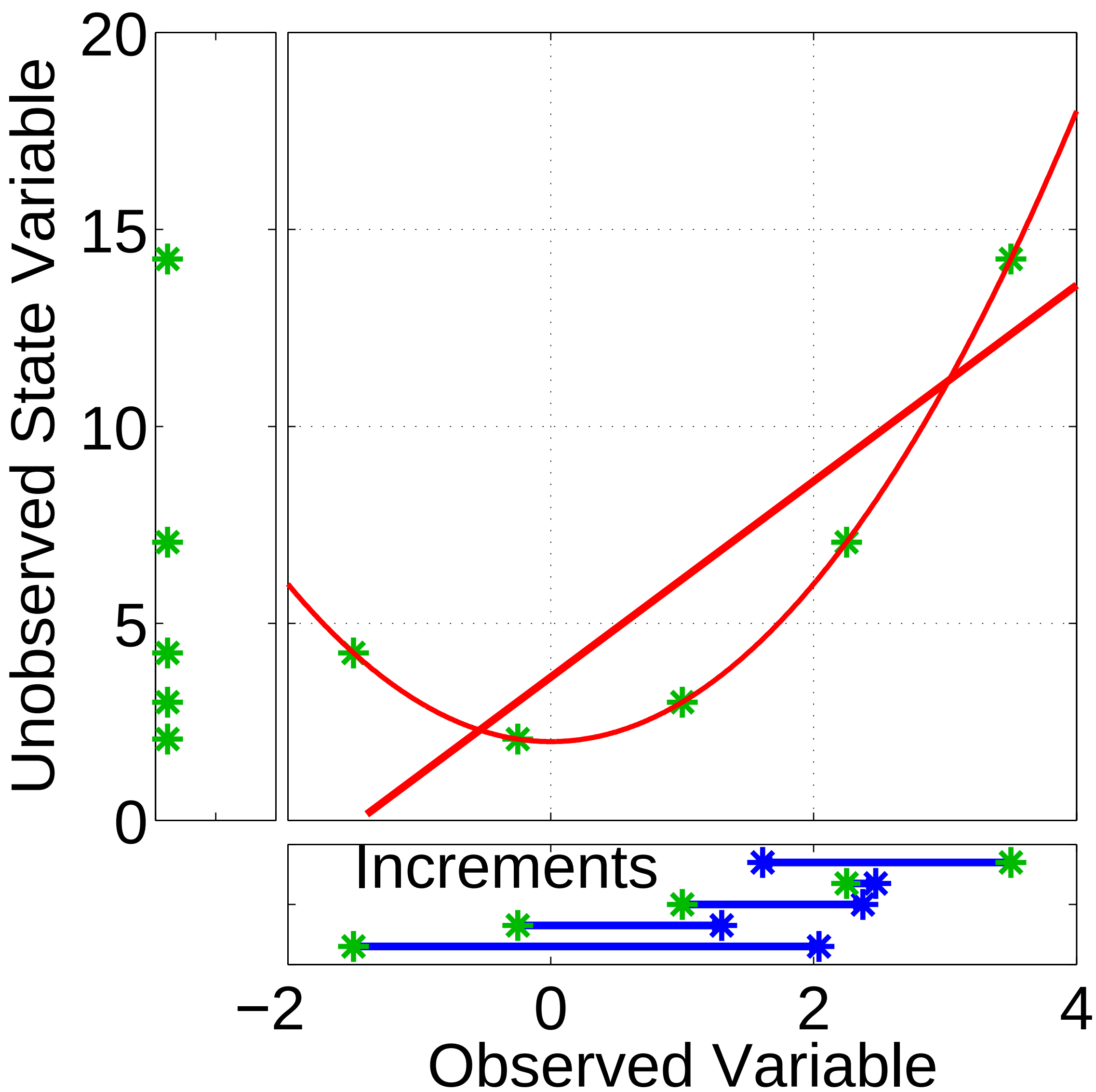
State variables updated by linear regression.

This is true but not apparent in Kalman context.

Joint prior distribution from model must be good.

Else, impact of observation is incorrect.

Challenge 7: Prior observation-state relations are non-linear



Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Regression error varies with value of observed variable.

This is nearly impossible to fix with ensemble filter.

Challenge 8: MANY state variables related to each observation

Ensemble filters normally use localization: observation only impacts subset of state variables.

One purpose is to avoid covariance matrix degeneracy, but this is NOT a problem with available hierarchical filter techniques.

BUT, filters become costly if each observation impacts many state variables.

Might reduce this cost by transforming state before assimilating.

For instance, if mean value of fields is related to many observations... convert to mean plus departure from mean for state.

Summary:

Modern ensemble filter techniques can deal with many error sources.

BUT, if your model

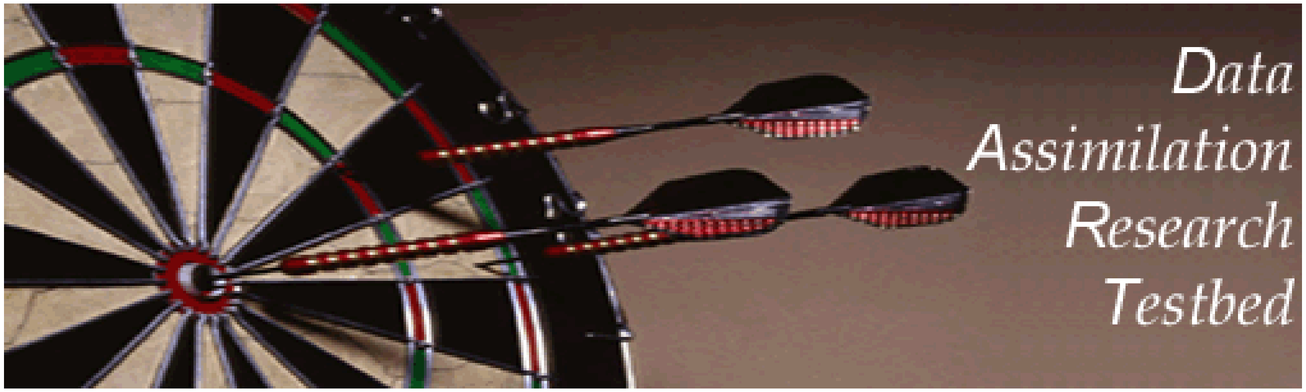
1. Can't produce a good estimate of the prior state,
2. Can't produce a reasonably growing error variance,
3. Doesn't produce reasonable prior covariances,

Then ensemble assimilation doesn't make sense.

If the prior distributions are non-Gaussian in more than ~3 dimensions, ensemble methods will not work with small sample sizes.

Could argue that systems that are like this are too complicated to understand and probably are not being studied or modeled...

Data Assimilation Research Testbed (DART)



Software to do everything here (and more) is in DART.

Requires F90 compiler, Matlab.

Available from www.image.ucar.edu/DAReS/.

DART compliant models (largest set ever with assim system?)

1. Many low-order models (Lorenz63, L84, L96, L2004,...).
2. Global 2-level PE model (from NOAA/CDC).
3. CGD's CAM 2.0, 3.0, 3.1 (global spectral model)

3a. CGD's CAM 3.1 FV (global finite volume model) with chemistry.

4. GFDL FMS B-grid GCM (global grid point model).
5. MIT GCM (from Jim Hansen; configured for annulus).
6. WRF model (regional prediction grid point).

6a. WRF column physics model.

7. NCEP GFS (operational global spectral; assisted by NOAA/CDC).
8. GFDL MOM3/4 (global grid point ocean model).
9. ACD's ROSE model (upper atmosphere with chemistry).
10. Cane-Zebiak 5 (tropical ocean/atmosphere model).

Also models from outside geophysics.
This allows for a hierarchical approach to filter development.

DART compliant Forward Operators and Datasets

Many linear and non-linear forward operators for low-order models.

U, V, T, Ps, Q, dewpoint for realistic models.

Radar reflectivity, doppler velocity, GPS refractivity for realistic models.

Mopitt CO retrievals.

Can ingest observations from reanalysis or operational BUFR files.

Can create synthetic (perfect model) observations for any of these.

Additional enhancements available for Earth System Analysis

1. Smoother: uses observations in past and future to estimate state.

2. High performance parallel implementations.
   Filter algorithms are naturally scalable.
   Run many copies of models.
   Impact of observations on model variables can be done in parallel.
   Algorithm has natural reformulations for different platforms.

3. Parameter estimation in large models.
   Use data to constrain parameters.
   Implementation is trivial.
   Interpretation is still VERY tricky.
   Did this for gravity wave drag efficiency in CAM.