An Introduction to Ensemble Kalman Filtering

Jeffrey Anderson
NCAR Data Assimilation Research Section (DAReS)

AMS NWP/WAF, 26 June, 2007

Overview

1. The Data Assimilation Problem
2. A Bayesian View of Ensemble Kalman Filtering
3. Challenges to Ensemble Filters
4. Adaptive Ensemble Filter Algorithms
5. Model and Observing System Development with Ensembles

The Data Assimilation Problem:

The Data Assimilation Problem:

A (Physical) system:

Atmosphere, coupled climate system...

The Data Assimilation Problem:

A (Physical) system:

Atmosphere, coupled climate system...

A model of the physical system:

Represents system with discrete vector: the model ‘state vector’.

Approximates time evolution of system (poorly).

The Data Assimilation Problem:

A (Physical) system:

Atmosphere, coupled climate system...

A model of the physical system:

Represents system with discrete vector: the model ‘state vector’.

Approximates time evolution of system (poorly).

Observations of the system:

Have a (sometimes poor) estimate of observation error.

Could be sparse and irregular in time (and space).

Relation to model ‘state vector’ may be complicated.

May have very low information content.

Data Assimilation increases information about all three pieces:

Data Assimilation increases information about all three pieces:

(Physical) system:

Estimates of state (analyses, posteriors...).

Initial conditions for forecasts.

Enhanced (physical) understanding.

Data Assimilation increases information about all three pieces:

(Physical) system:

Estimates of state (analyses, posteriors...).

Initial conditions for forecasts.

Enhanced (physical) understanding.

Model:

Estimates of model errors.

Relative characteristics of different models.

Improved model (find good values for model ‘parameters’).

Data Assimilation increases information about all three pieces:

(Physical) system:

Estimates of state (analyses, posteriors...).

Initial conditions for forecasts.

Enhanced (physical) understanding.

Model:

Estimates of model errors.

Relative characteristics of different models.

Improved model (find good values for model ‘parameters’).

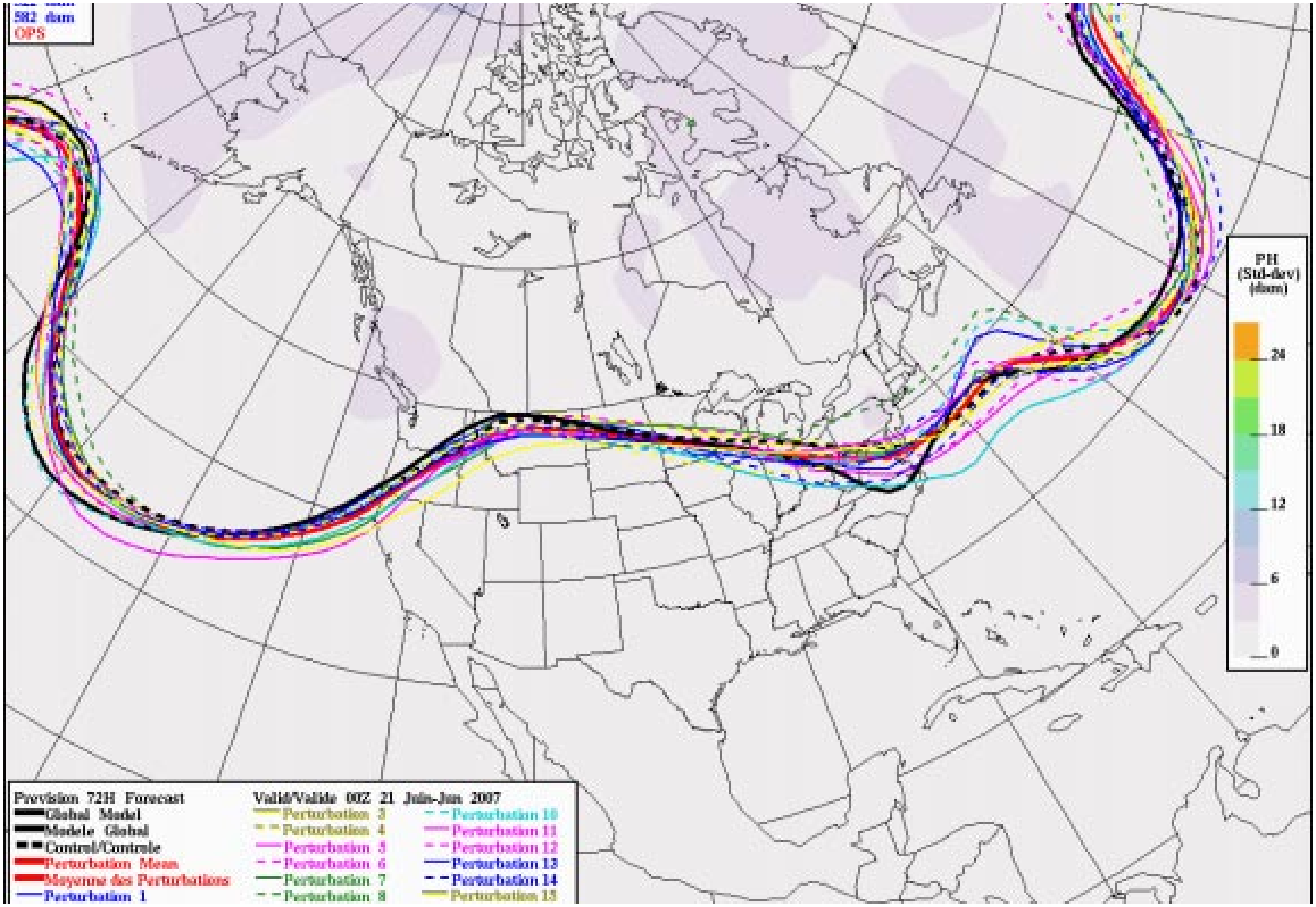
Observations:

Estimates of observation errors.

Information content of existing or planned observations.

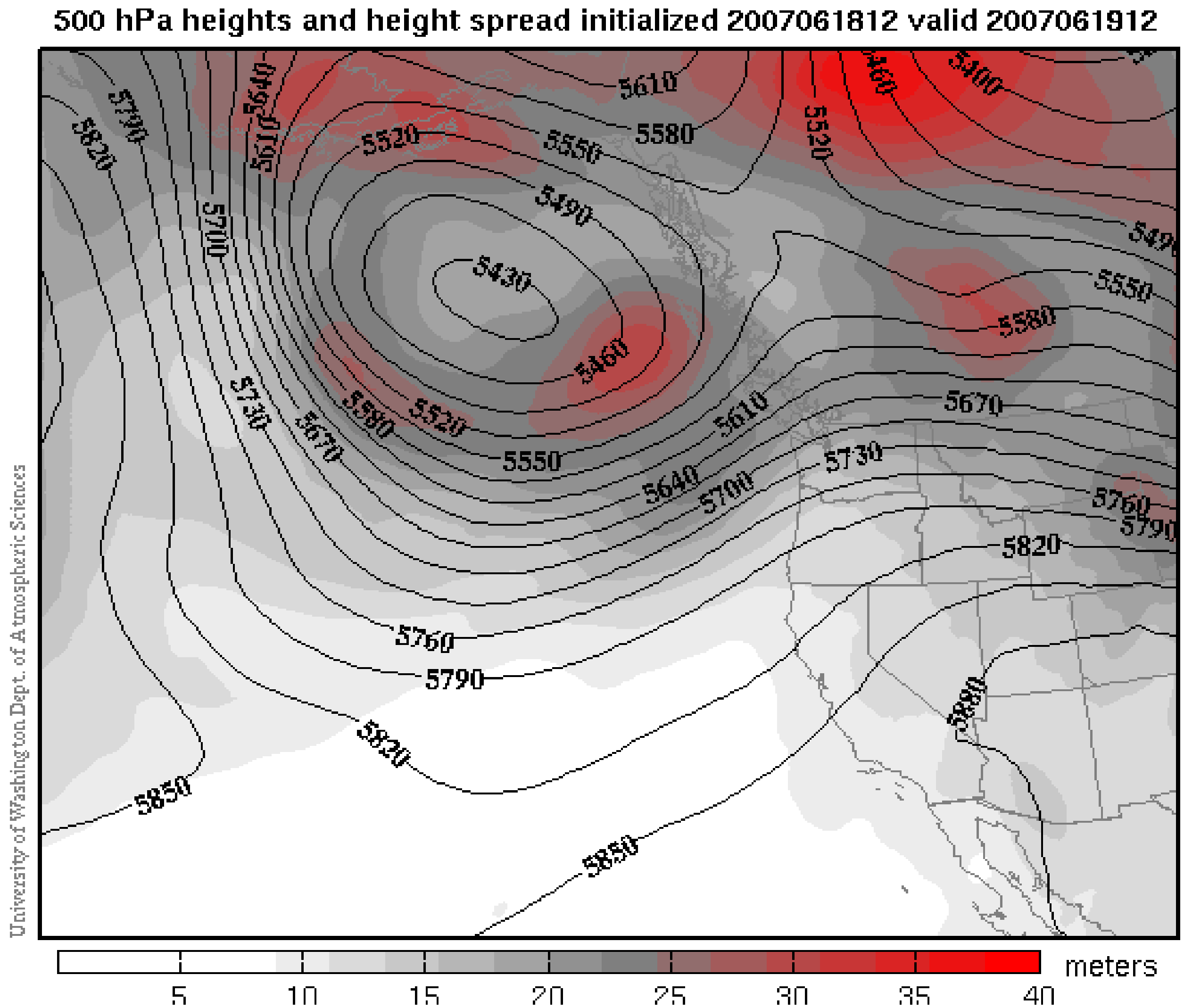
Observing system designs that provide increased information.

Ensemble Filter Products now Available to Forecasters



Environment Canada Operational GEM 16 member ENKF

Ensemble Filter Products now Available to Forecasters



University of Washington WRF EnKF

Ensemble Filter Products now Available to Forecasters

1. Many other ensemble forecast products are available.

2. Most suffer from the same challenges outlined below.

3. Other ensemble generation methods may have additional issues.

An introduction to Ensemble Filtering.

1: Single variable and observation of that variable.

Let’s think of it as temperature at SLC.

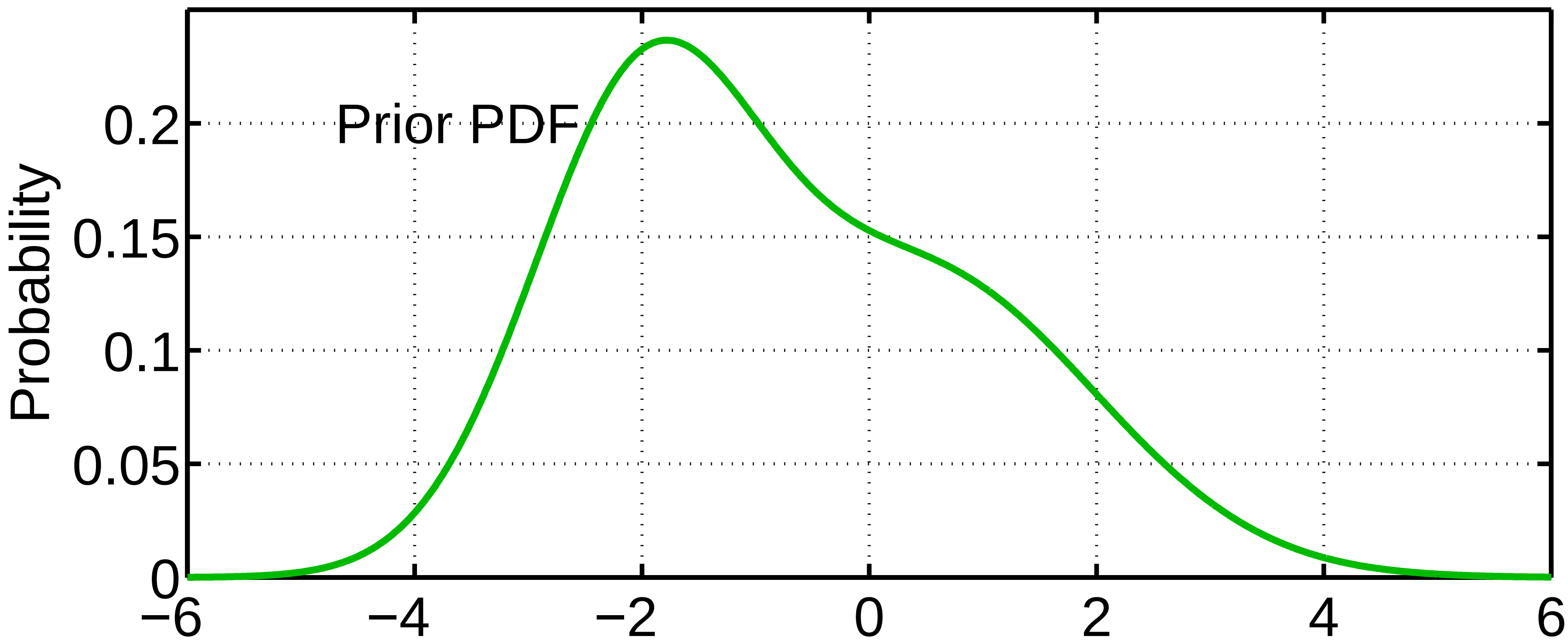
(Slides are for a mid-winter ski day...inversion in the valley)

2: Single observed variable, single unobserved variable.

SLC temperature, Park City temperature.

That’s all there is... (without loss of generality).

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)p(A|C)}{p(B|C)} = \dfrac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

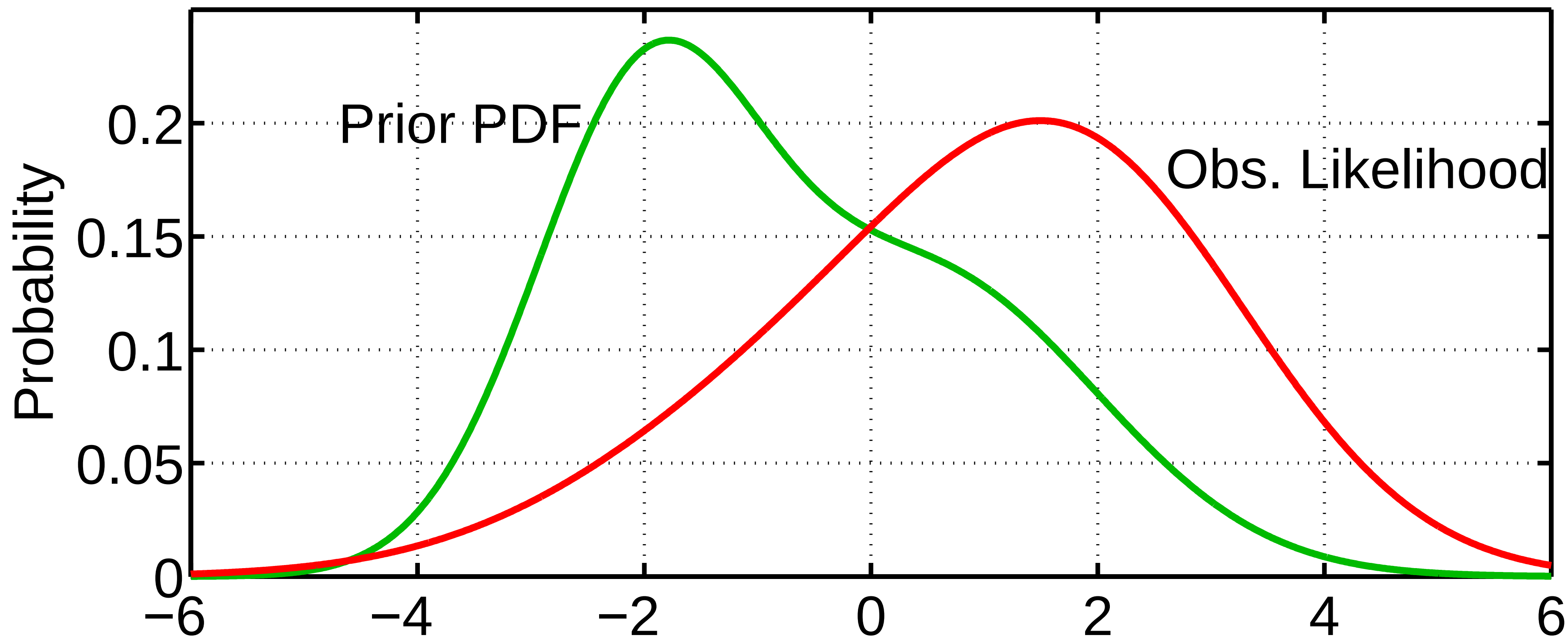


A: Prior estimate based on all previous information, C.

B: An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)p(A|C)}{p(B|C)} = \dfrac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

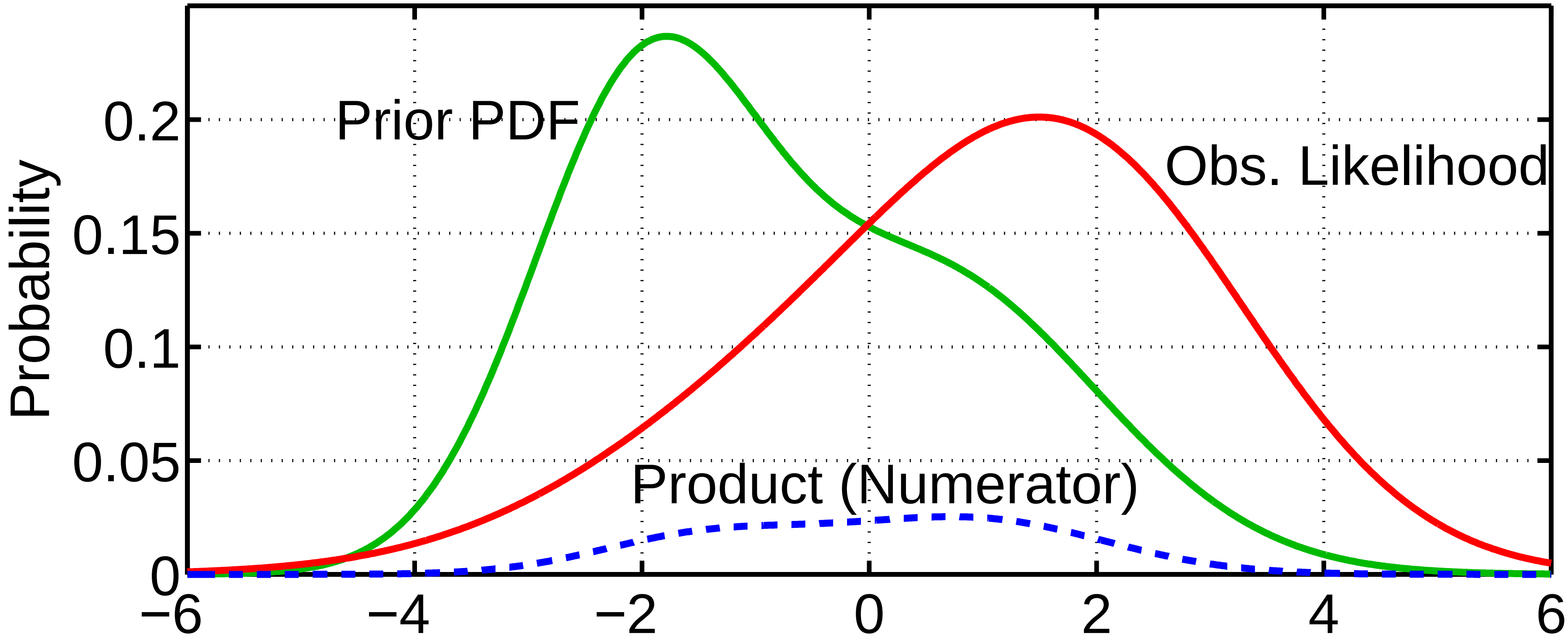


A: Prior estimate based on all previous information, C.

B: An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)\,p(A|C)}{p(B|C)} = \dfrac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$

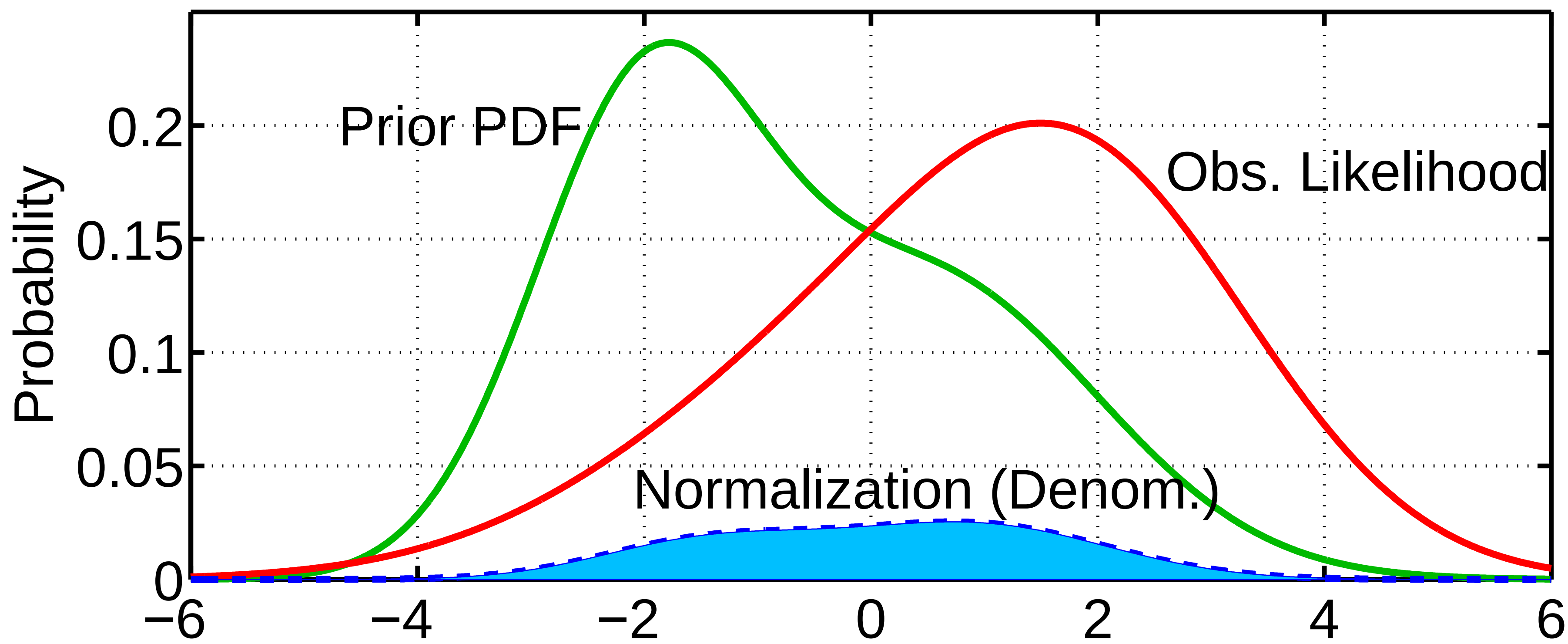


A: Prior estimate based on all previous information, C.

B: An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)p(A|C)}{p(B|C)} = \dfrac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

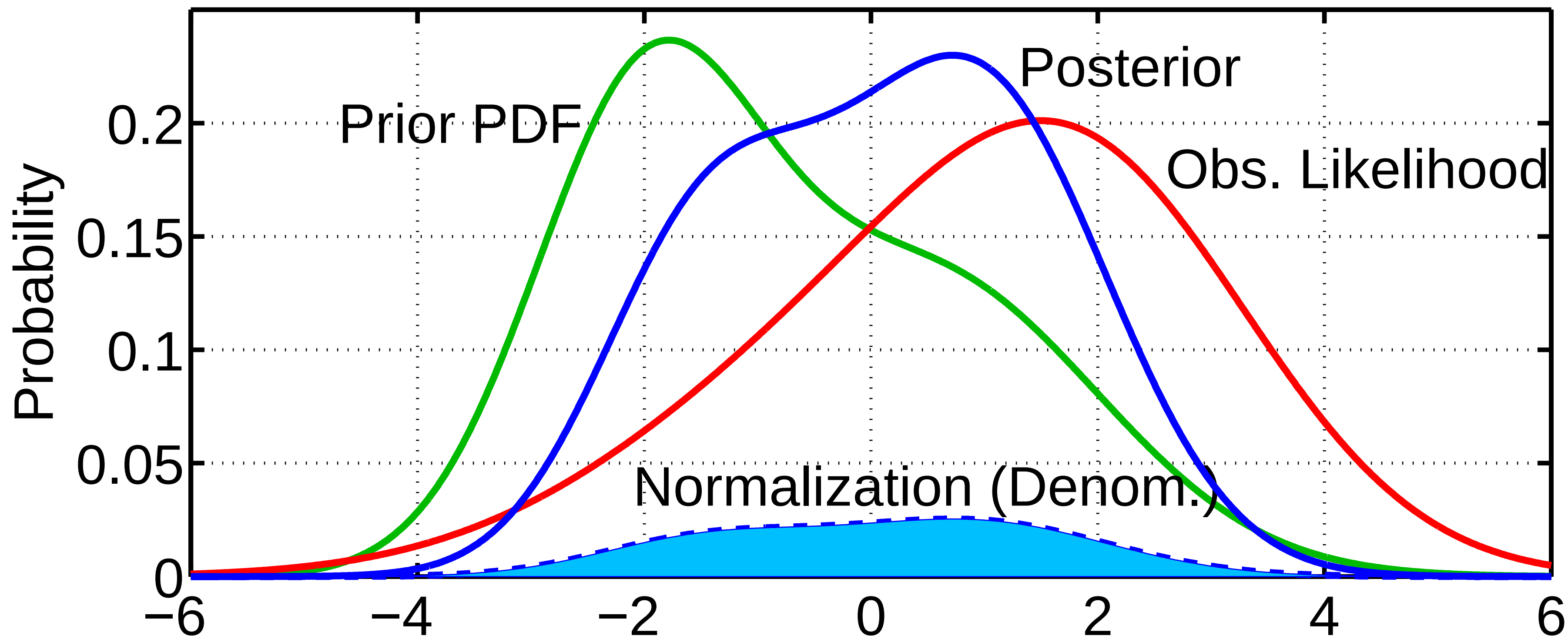


A: Prior estimate based on all previous information, C.

B: An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)\,p(A|C)}{p(B|C)} = \dfrac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$



A: Prior estimate based on all previous information, C.

B: An additional observation.

$p(A|BC)$: Posterior (updated estimate) based on C and B.

Consistent Color Scheme Throughout

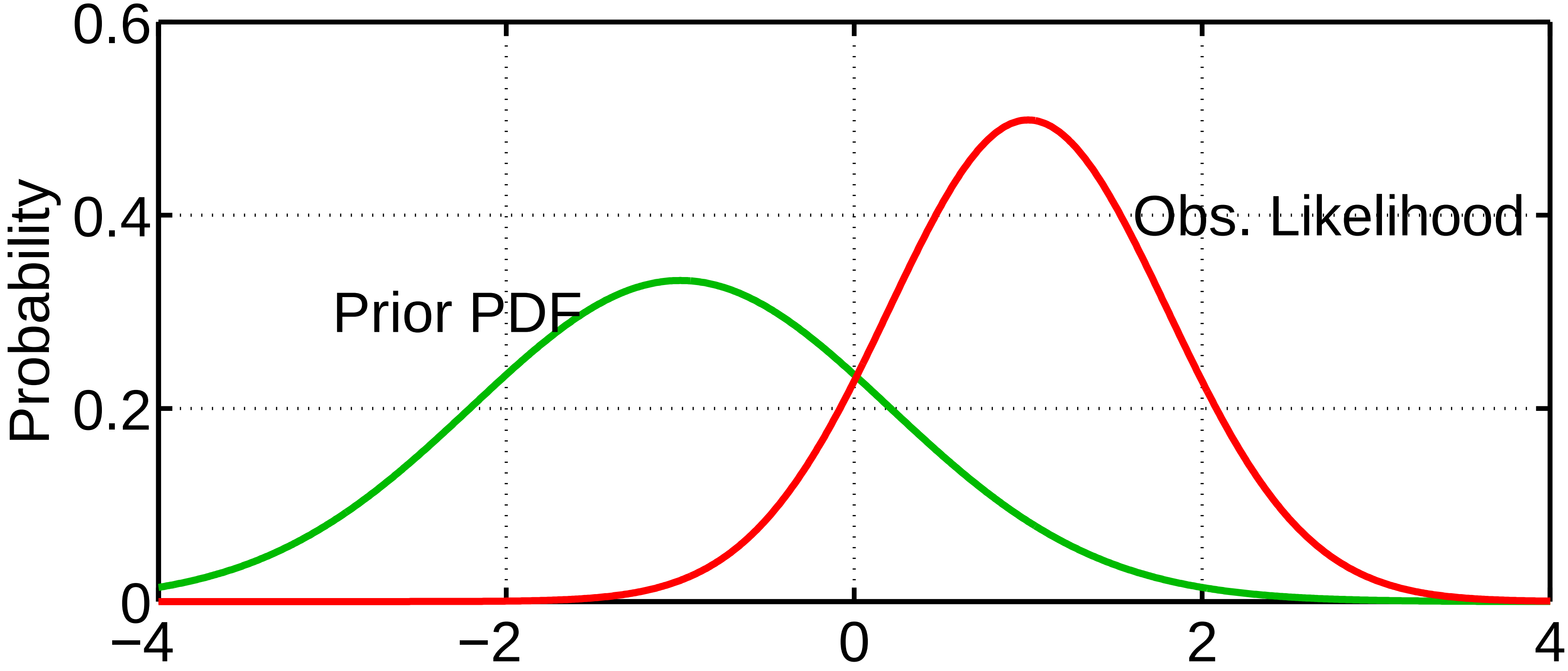
Green = Prior

Red = Observation

Blue = Posterior

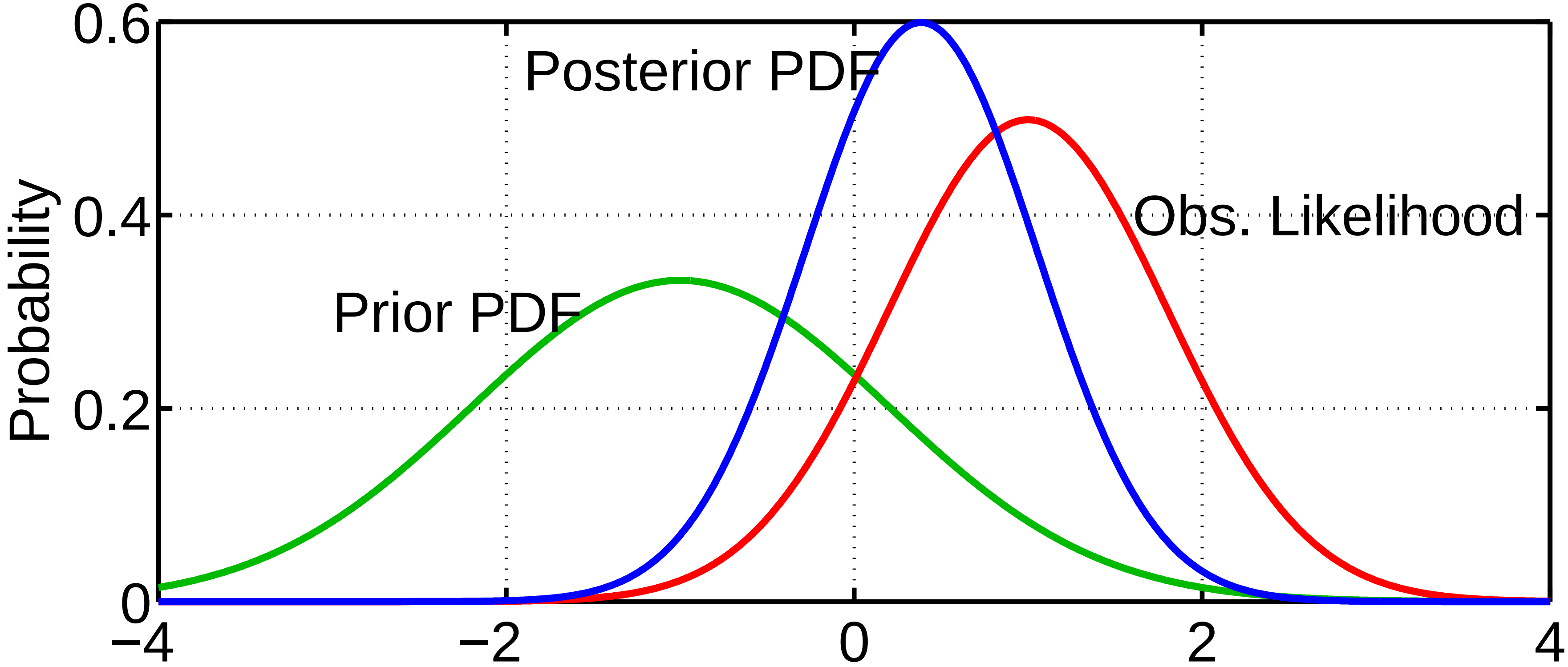
Bayes rule: $p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

This product is closed for Gaussian distributions.



Bayes rule: $p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

This product is closed for Gaussian distributions.



Product of two Gaussians:

Product of d-dimensional normals with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is normal.

$$\mathbf{N}(\mu_1, \Sigma_1)\boldsymbol{N}(\mu_2, \Sigma_2) = c\mathbf{N}(\mu, \Sigma)$$

Product of two Gaussians:

Product of d-dimensional normals with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is normal.

$$\mathbf{N}(\mu_1, \Sigma_1)\boldsymbol{N}(\mu_2, \Sigma_2) = c\mathbf{N}(\mu, \Sigma)$$

Covariance: $\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$

Mean: $\mu = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$

Product of two Gaussians:

Product of d-dimensional normals with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is normal.

$$\mathbf{N}(\mu_1, \Sigma_1)N(\mu_2, \Sigma_2) = c\mathbf{N}(\mu, \Sigma)$$

Covariance: $\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$

Mean: $\mu = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$

Weight: $c = \dfrac{1}{(2\Pi)^{d/2}\left|\Sigma_1 + \Sigma_2\right|^{1/2}} \exp\left\{-\dfrac{1}{2}[(\mu_2 - \mu_1)^T(\Sigma_1 + \Sigma_2)^{-1}(\mu_2 - \mu_1)]\right\}$

We'll ignore the weight since we immediately normalize products to be PDFs.

Product of two Gaussians:

Product of d-dimensional normals with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 is normal.

$$\mathbf{N}(\mu_1, \Sigma_1)\boldsymbol{N}(\mu_2, \Sigma_2) = c\mathbf{N}(\mu, \Sigma)$$

Covariance: $\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$

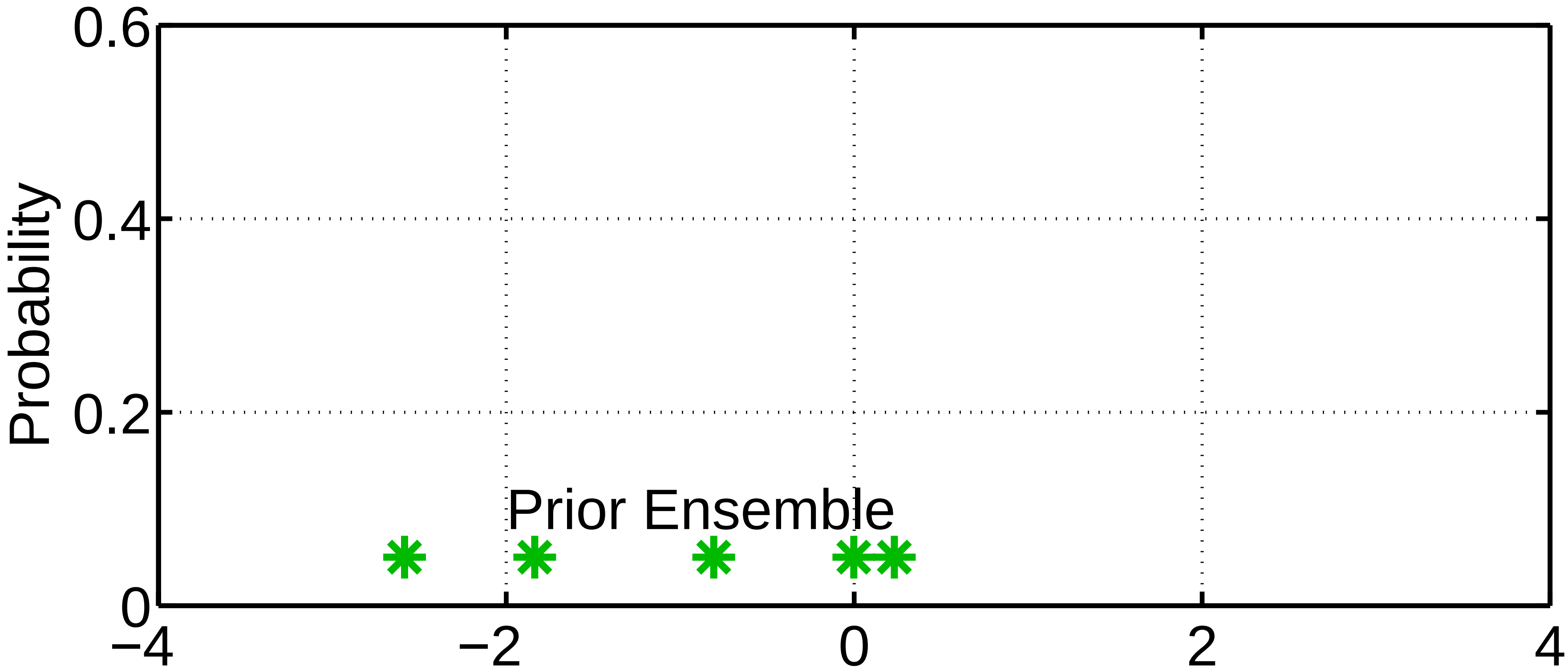
Mean: $\mu = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$

Weight: $c = \dfrac{1}{(2\Pi)^{d/2}|\Sigma_1 + \Sigma_2|^{1/2}} \exp\left\{-\frac{1}{2}[(\mu_2 - \mu_1)^T(\Sigma_1 + \Sigma_2)^{-1}(\mu_2 - \mu_1)]\right\}$

Easy to derive for 1D (d=1); just do products of exponentials.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)p(A|C)}{p(B|C)} = \dfrac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

Ensemble filters: Prior is available as finite sample.

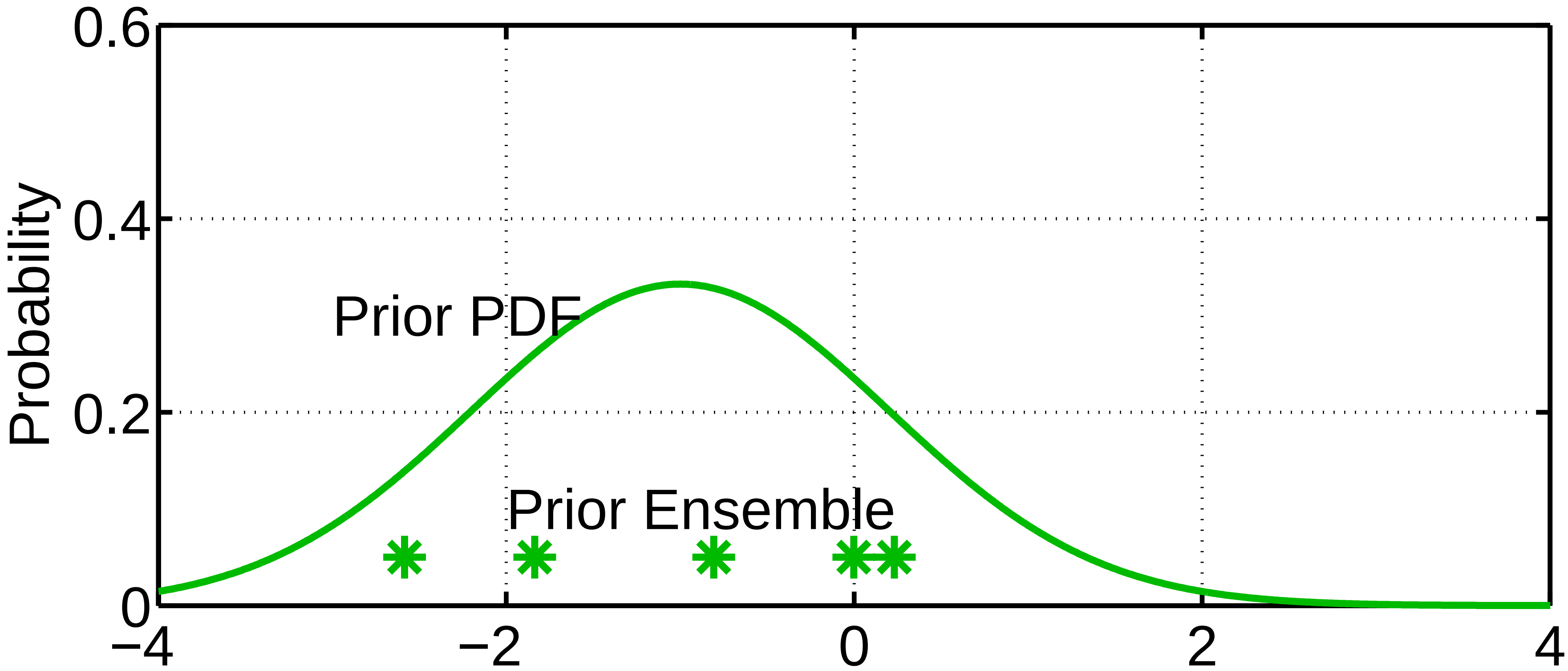


Don’t know much about properties of this sample.

May naively assume it is random draw from ‘truth’.

Bayes rule: $p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

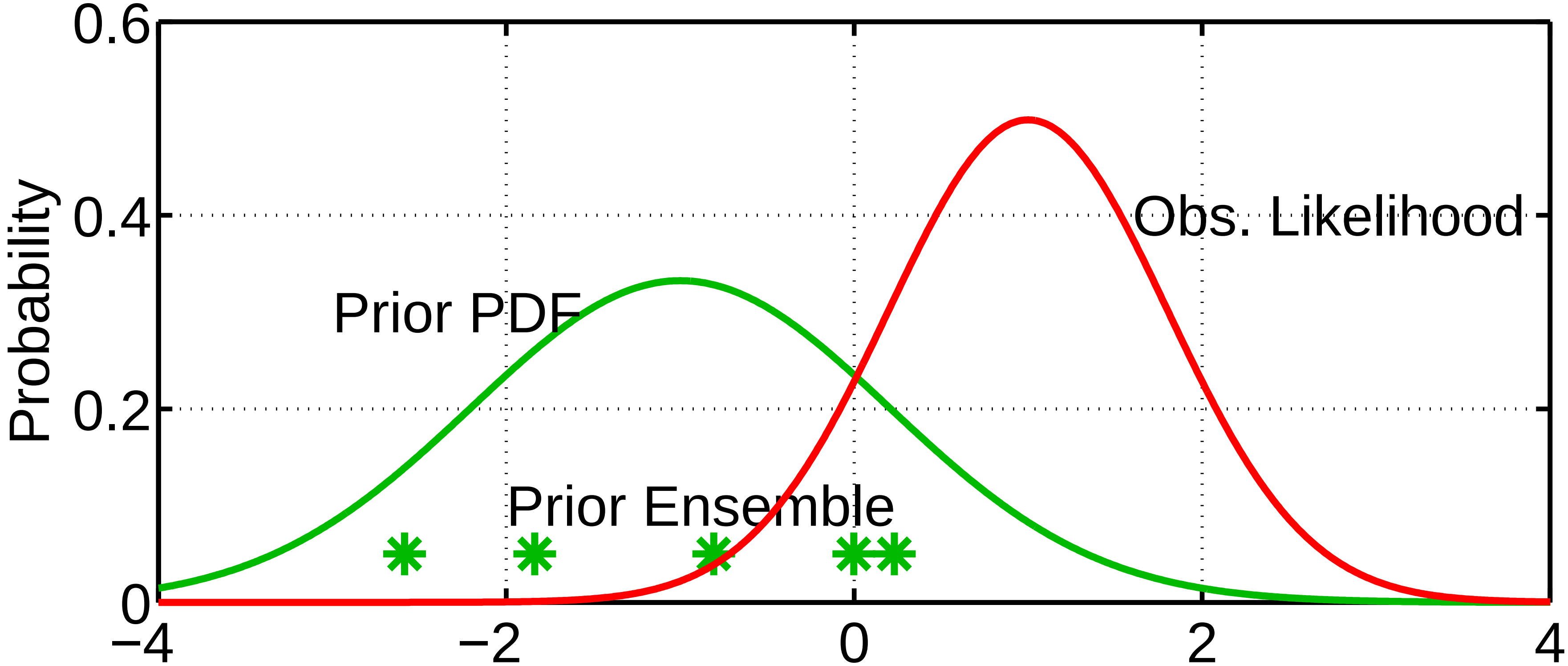
How can we take product of sample with continuous likelihood?



Fit a continuous (Gaussian for now) distribution to sample.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)p(A|C)}{p(B|C)} = \dfrac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$

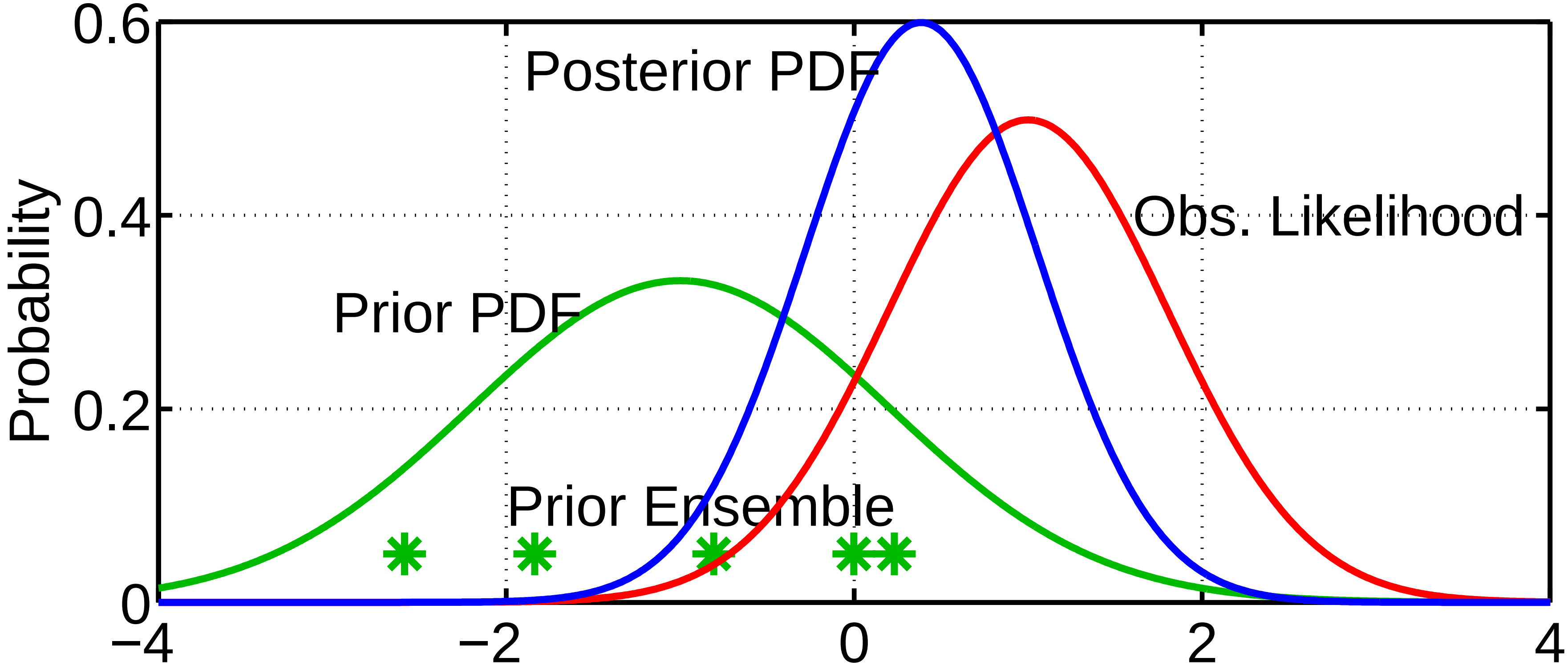
Observation likelihood usually continuous (nearly always Gaussian).



If Obs. Likelihood isn't Gaussian, can generalize methods below.
For instance, can fit set of Gaussian kernels to obs. likelihood.

Bayes rule: $p(A|BC) = \dfrac{p(B|AC)\,p(A|C)}{p(B|C)} = \dfrac{p(B|AC)\,p(A|C)}{\int p(B|x)\,p(x|C)\,dx}$

Product of prior Gaussian fit and Obs. likelihood is Gaussian.

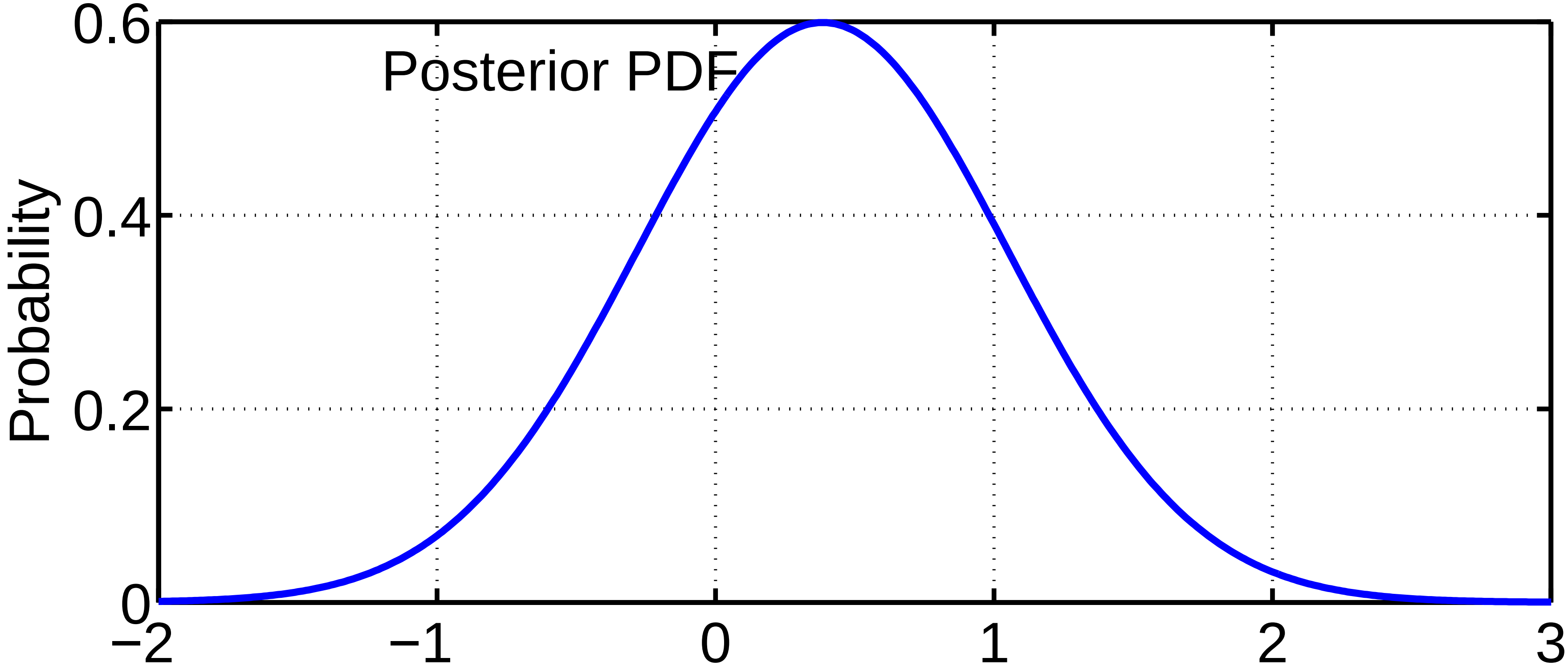


Computing continuous posterior is simple.

BUT, need to have a SAMPLE of this PDF.

Sampling Posterior PDF:

There are many ways to do this.



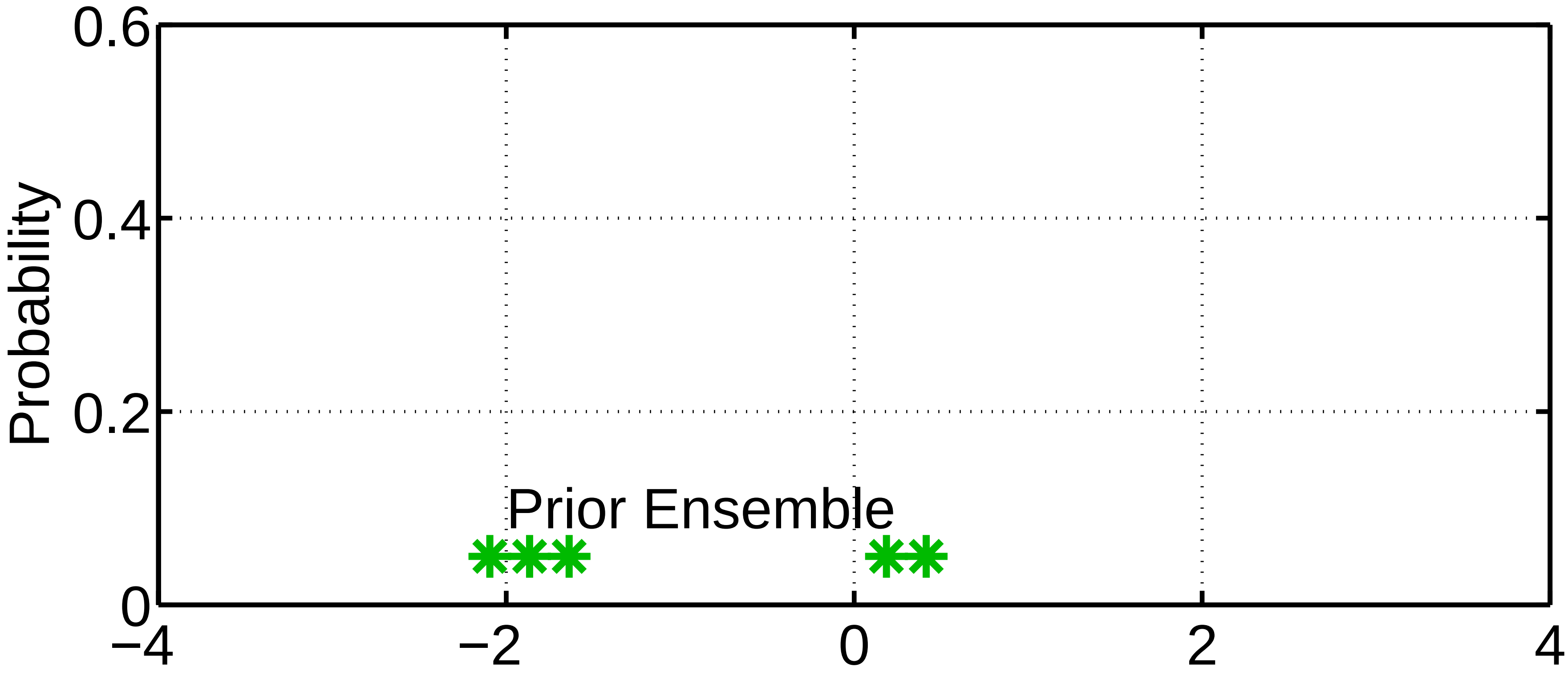
Exact properties of different methods may be unclear.

Trial and error still best way to see how they perform.

Will interact with properties of prediction models, etc.

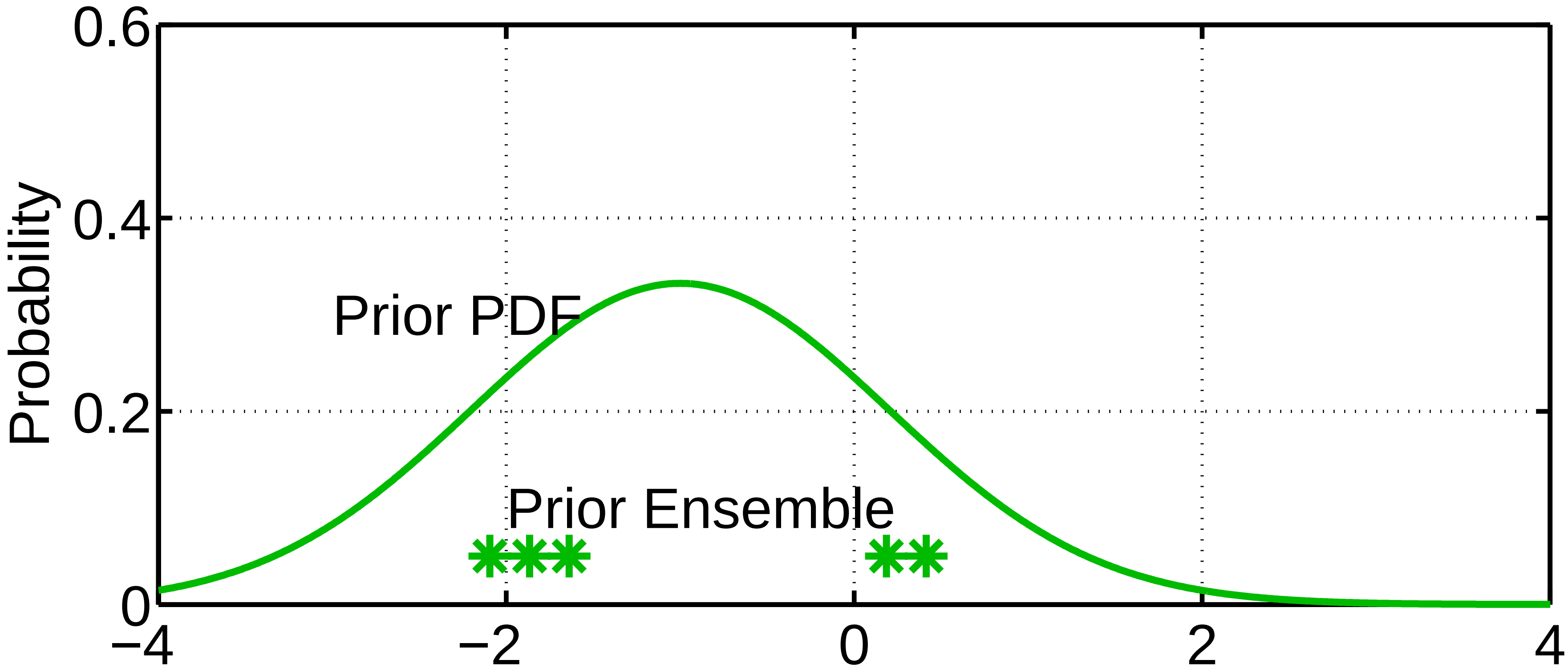
Ensemble Filter Algorithms:

Ensemble Adjustment (Kalman) Filter.



Ensemble Filter Algorithms:

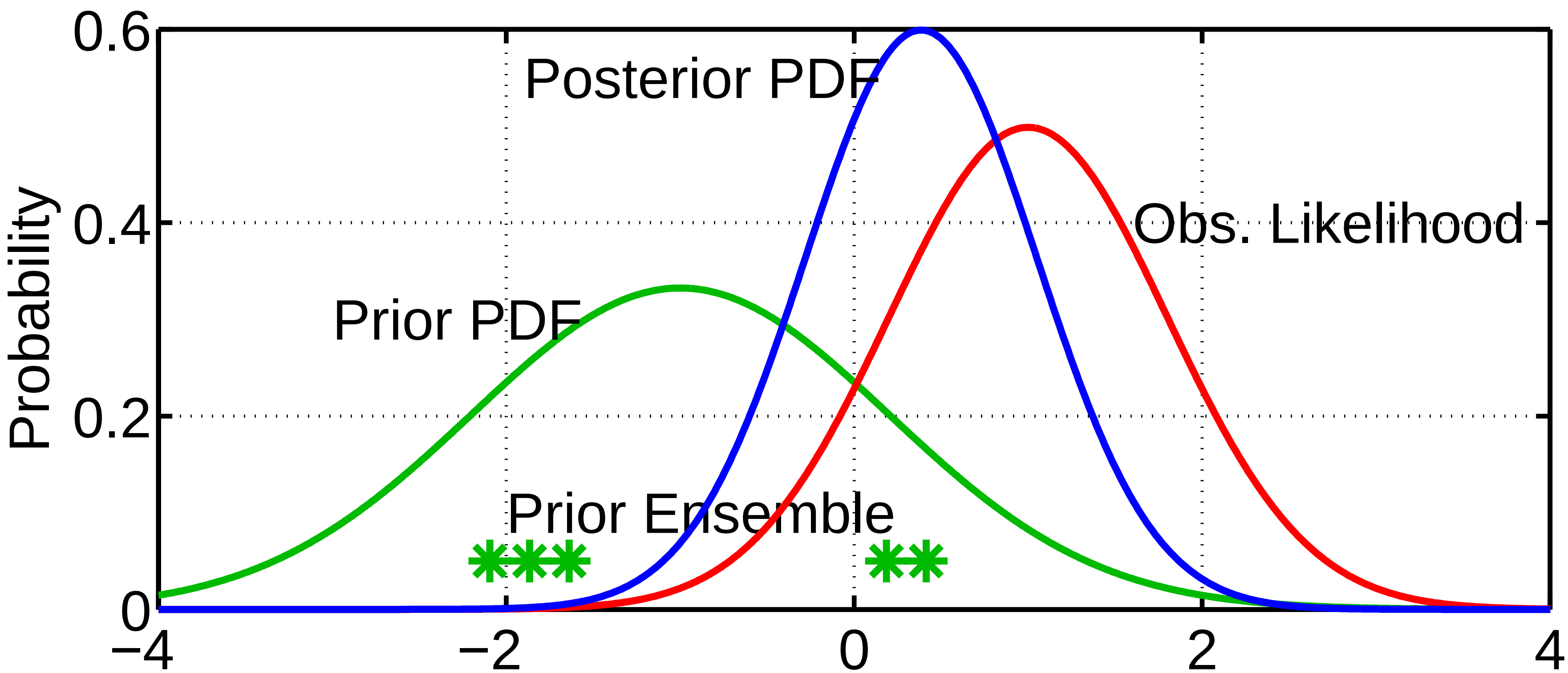
Ensemble Adjustment (Kalman) Filter.



Again, fit a Gaussian to sample.

Ensemble Filter Algorithms:

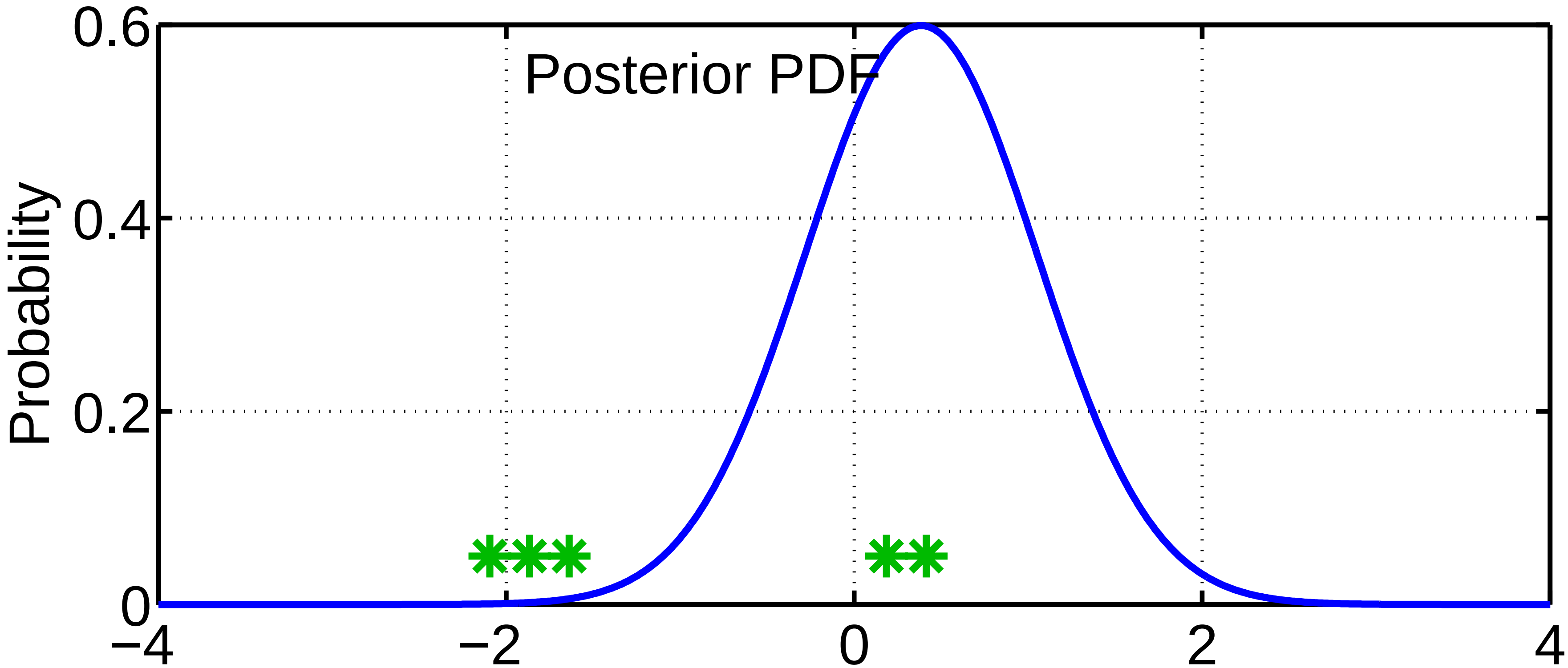
Ensemble Adjustment (Kalman) Filter.



Compute posterior PDF.

Ensemble Filter Algorithms:

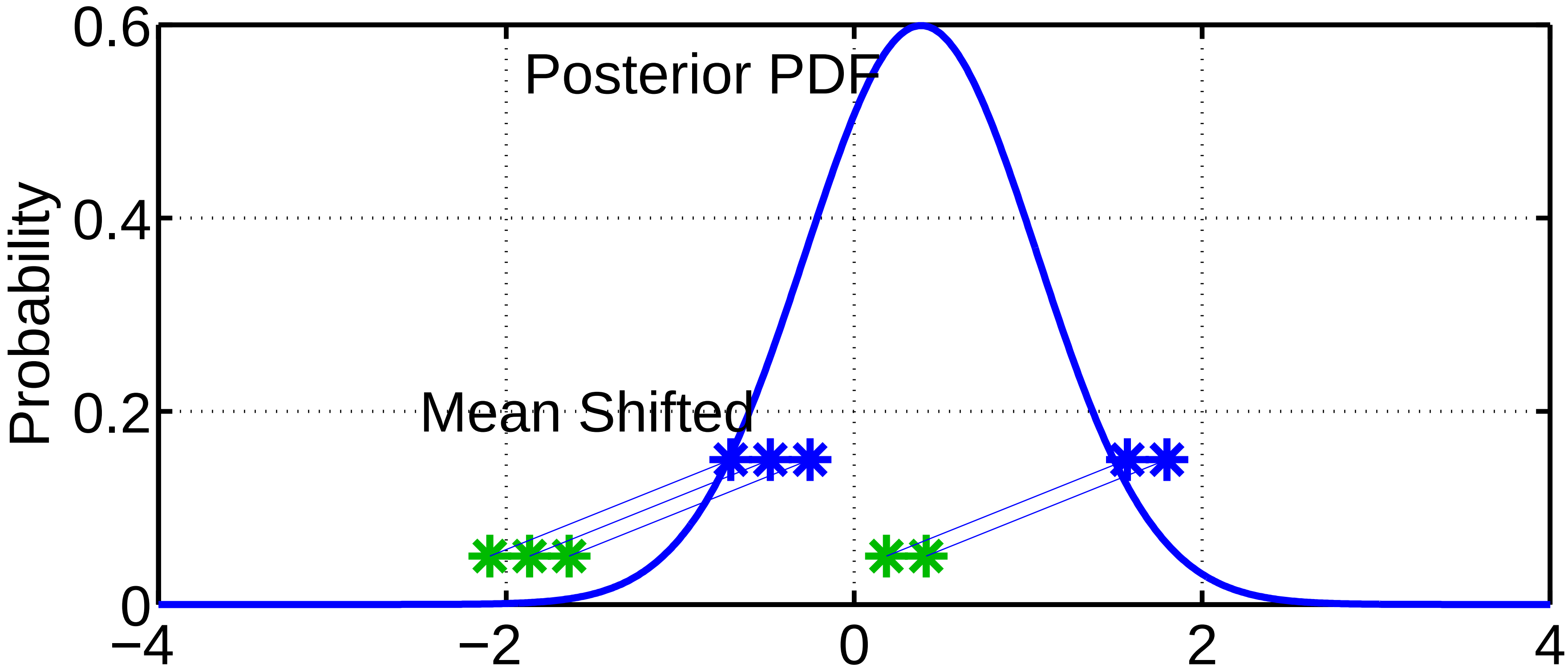
Ensemble Adjustment (Kalman) Filter.



Use deterministic algorithm to 'adjust' ensemble.

Ensemble Filter Algorithms:

Ensemble Adjustment (Kalman) Filter.

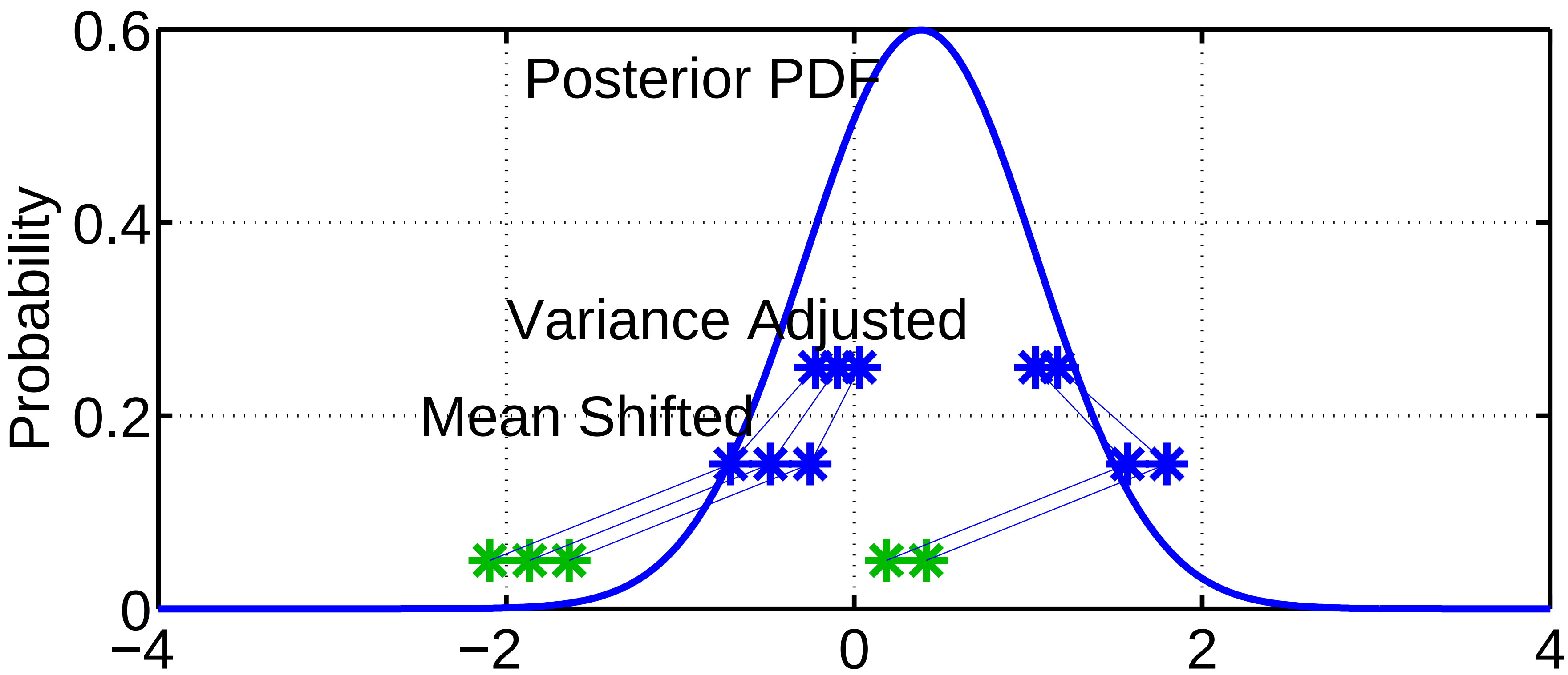


Use deterministic algorithm to 'adjust' ensemble.

First, 'shift' ensemble to have exact mean of posterior.

Ensemble Filter Algorithms:

Ensemble Adjustment (Kalman) Filter.

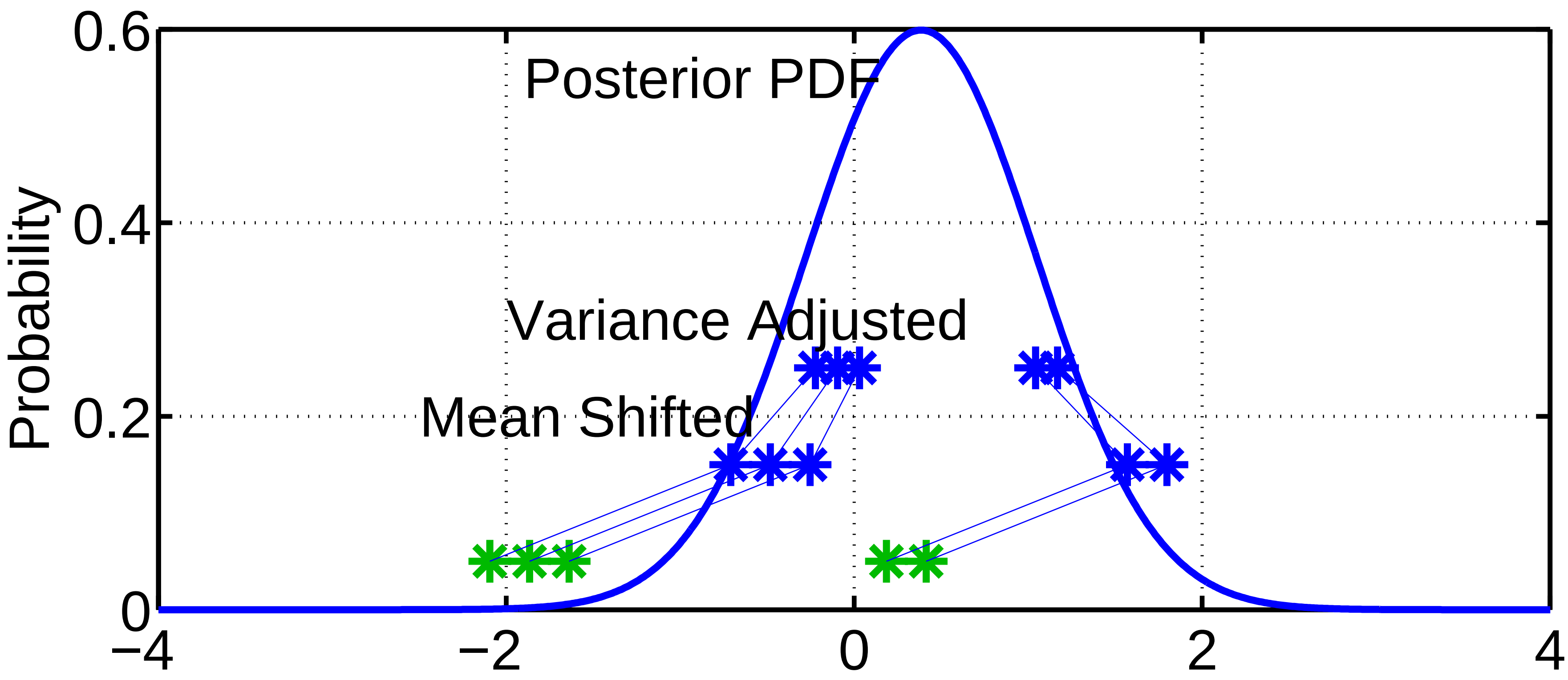


Use deterministic algorithm to 'adjust' ensemble.

- First, 'shift' ensemble to have exact mean of posterior.
- Second, use linear contraction to have exact variance of posterior.

Ensemble Filter Algorithms:

Ensemble Adjustment (Kalman) Filter.

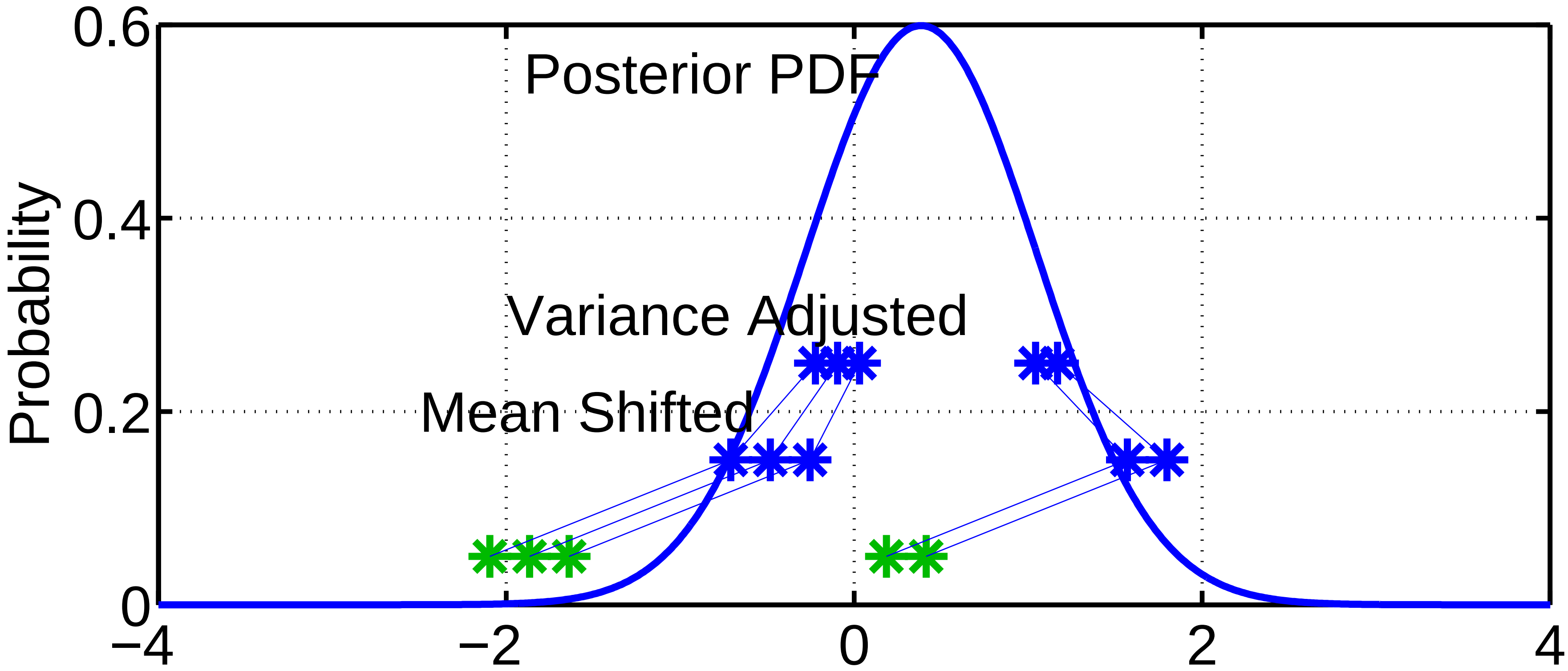


$$x_i^u = (x_i^p - \overline{x^p}) \cdot (\sigma^u / \sigma^p) + \overline{x^u} \quad i = 1,\ldots, \text{ensemble size.}$$

p is prior, u is update (posterior), overbar is ensemble mean, σ is standard deviation.

Ensemble Filter Algorithms:

Ensemble Adjustment (Kalman) Filter.



Bimodality maintained, but not appropriately positioned or weighted.
No problem with random outliers.

2: Single observed variable, single unobserved variable

SLC temperature, temperature at Park City.

So far, have known observation likelihood for single variable.

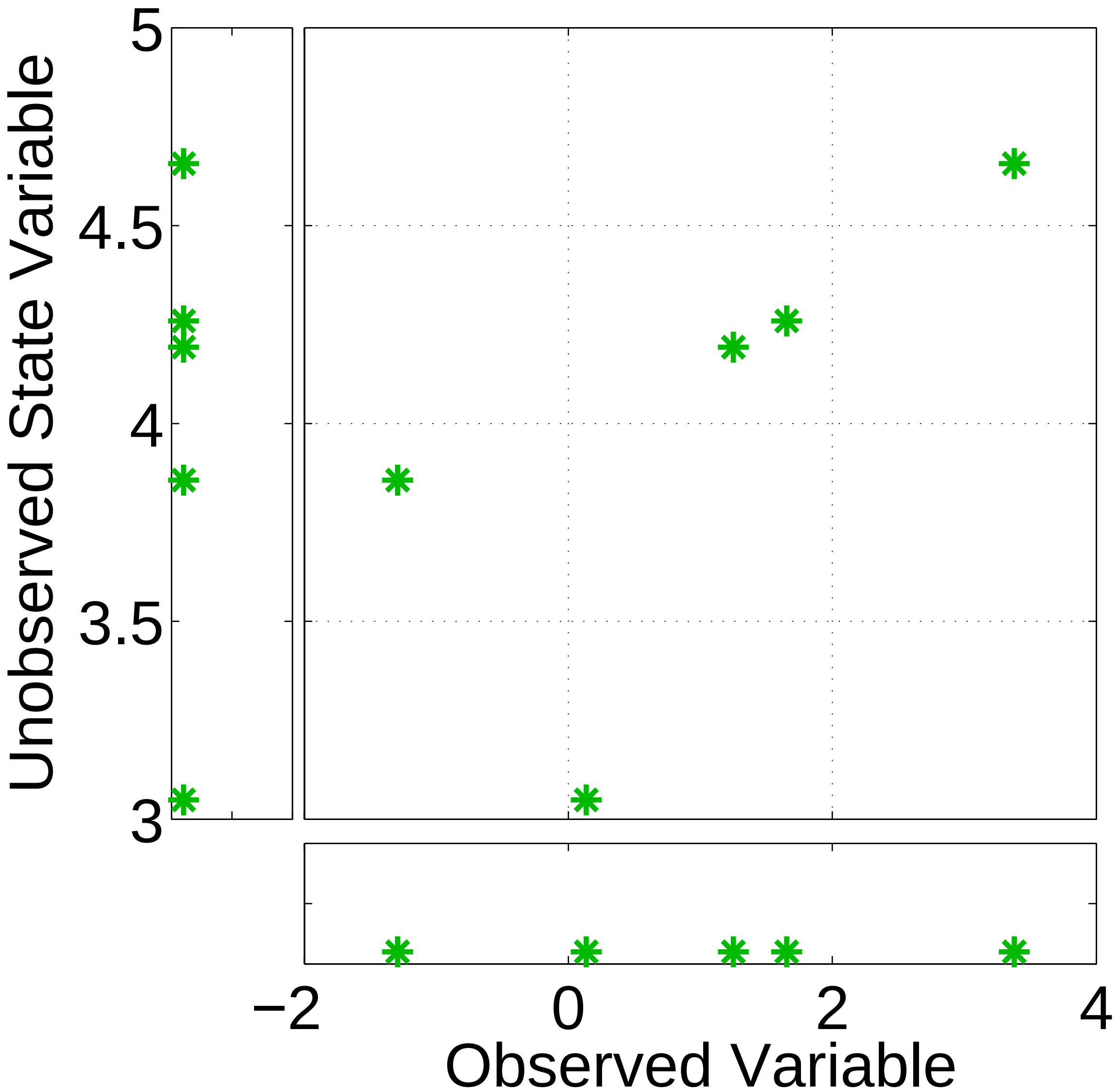
Now, suppose model state vector has an additional variable.

Will examine how ensemble methods update additional variable.

Basic method generalizes to any number of additional variables.

Related to Kalman filter in subtle ways.

Ensemble filters: Updating additional prior state variables

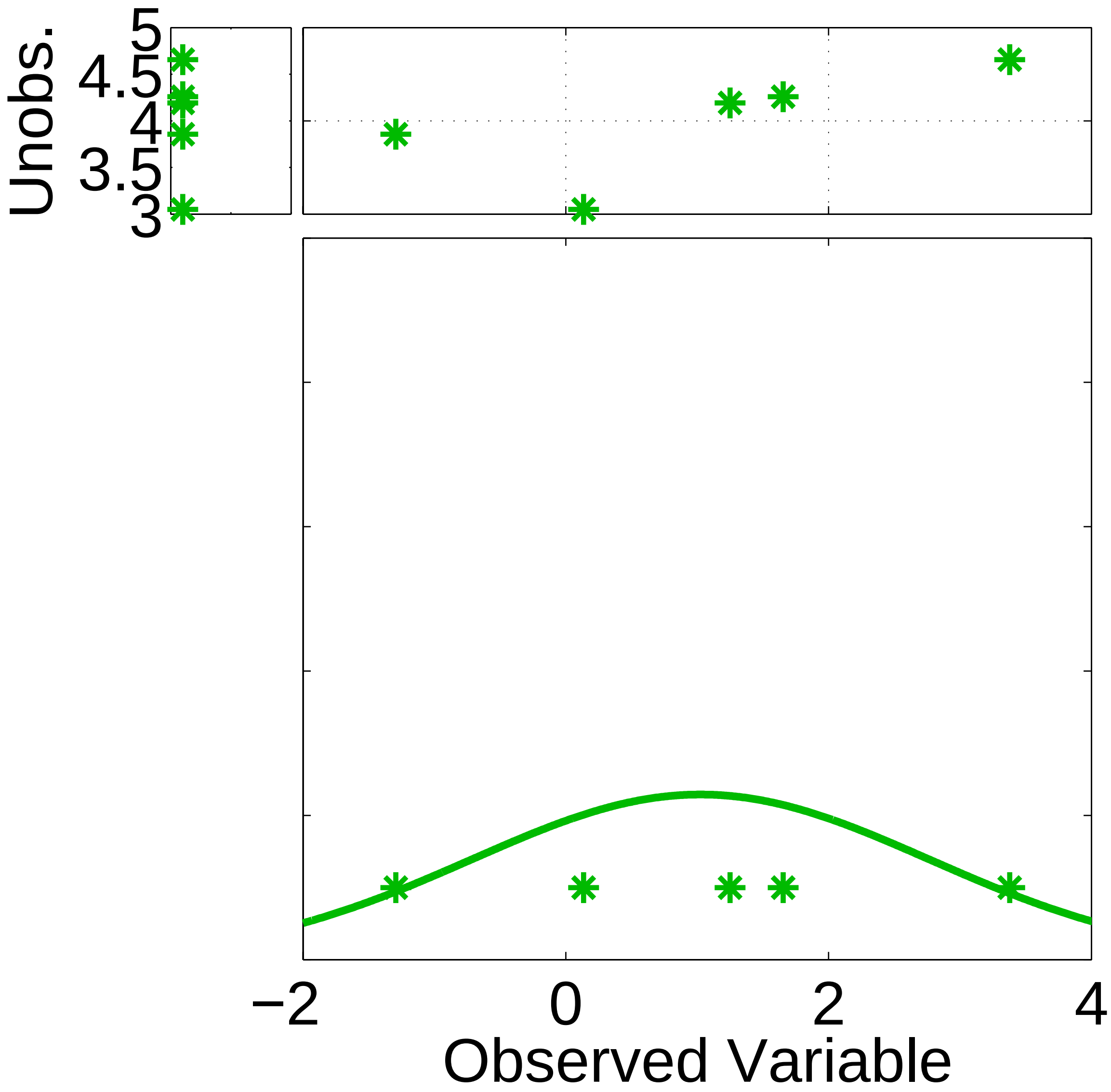


Assume that all we know is prior joint distribution.

One variable is observed.
(SLC temperature)
What should happen to unobserved variable?
(Park City temp.)

Looks like a nasty inversion...

Ensemble filters: Updating additional prior state variables

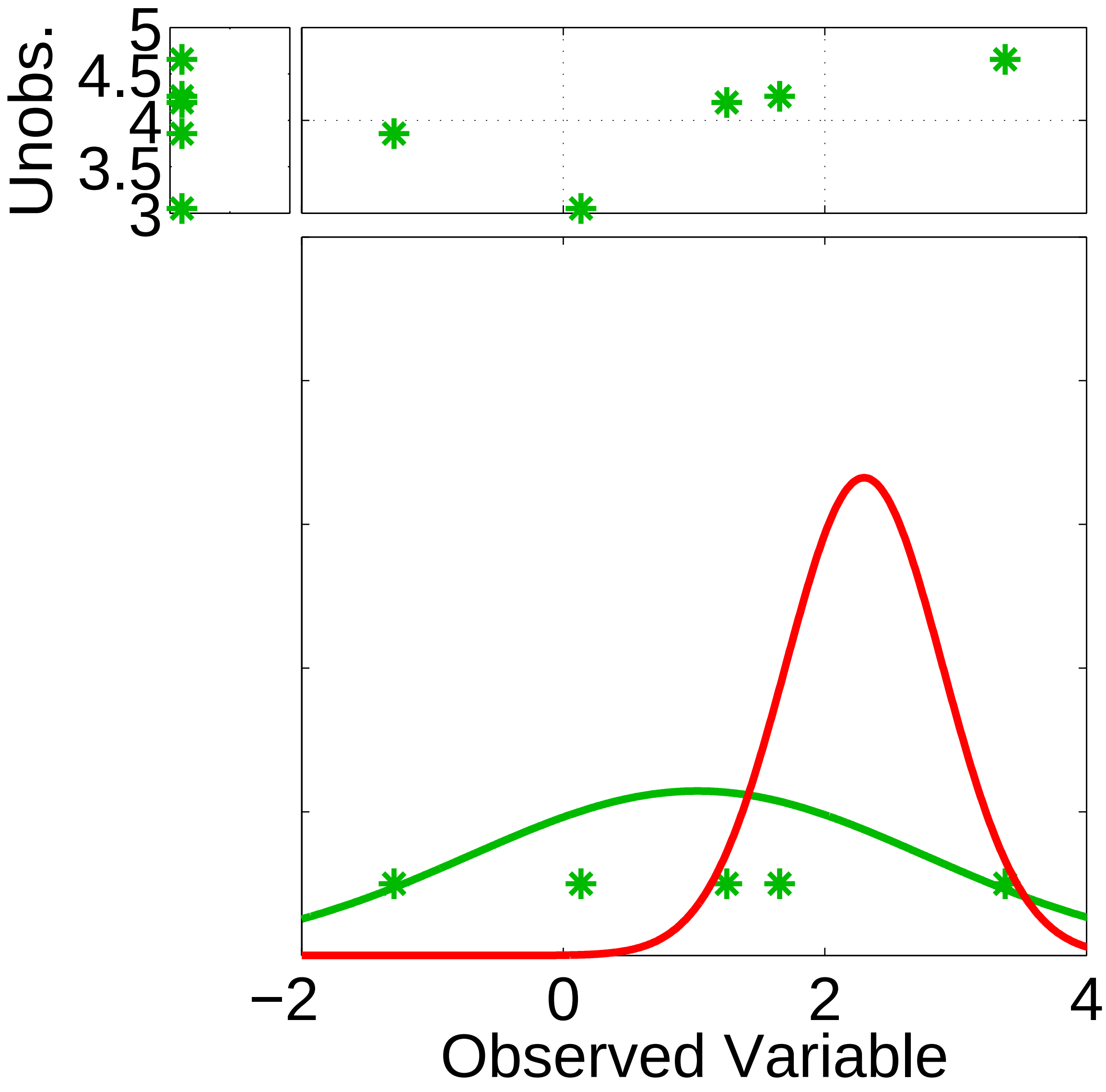


Assume that all we know is prior joint distribution.

One variable is observed.

Update observed variable with one of previous methods.

Ensemble filters: Updating additional prior state variables

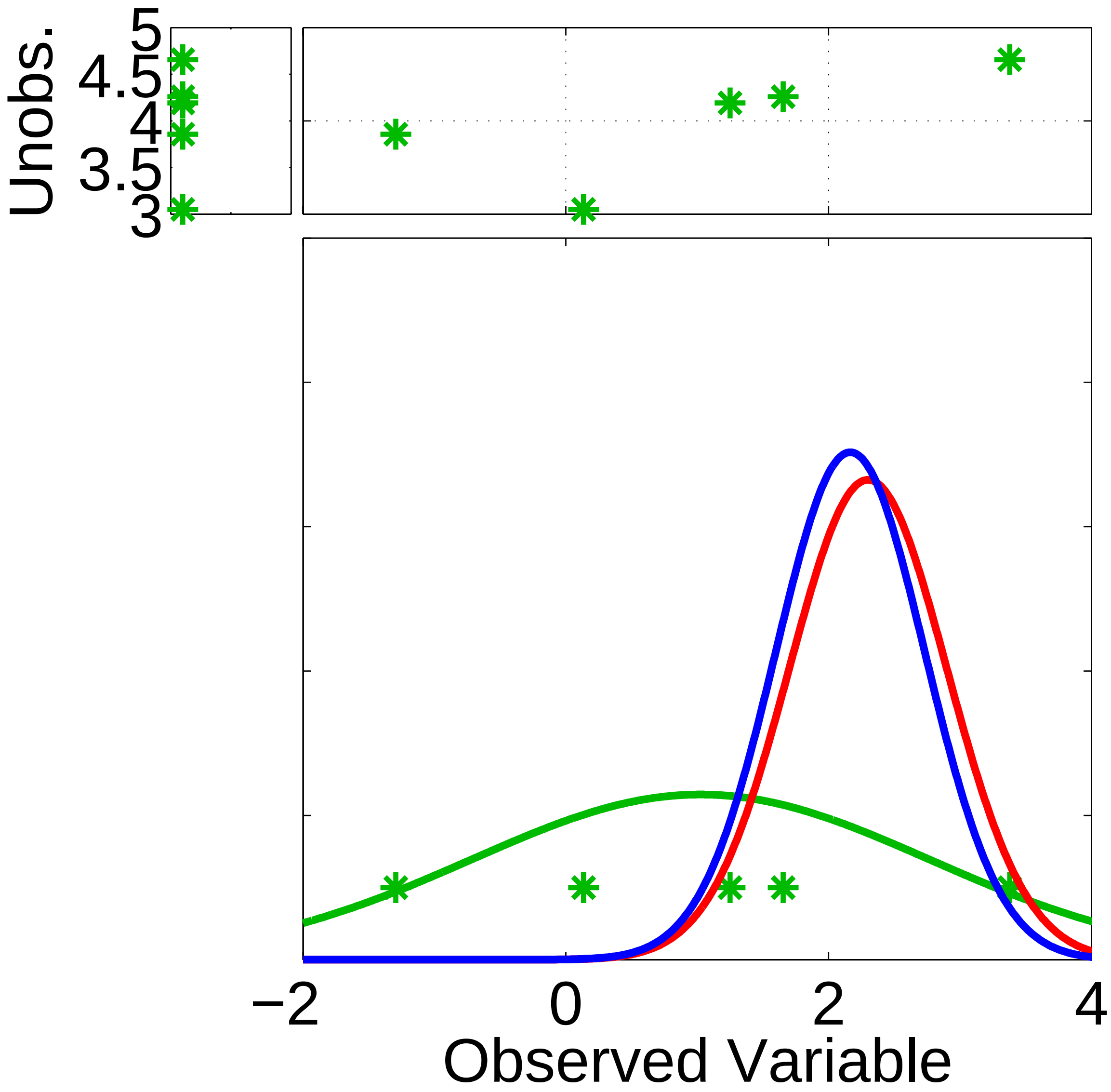


Assume that all we know is prior joint distribution.

One variable is observed.

Update observed variable with one of previous methods.

Ensemble filters: Updating additional prior state variables

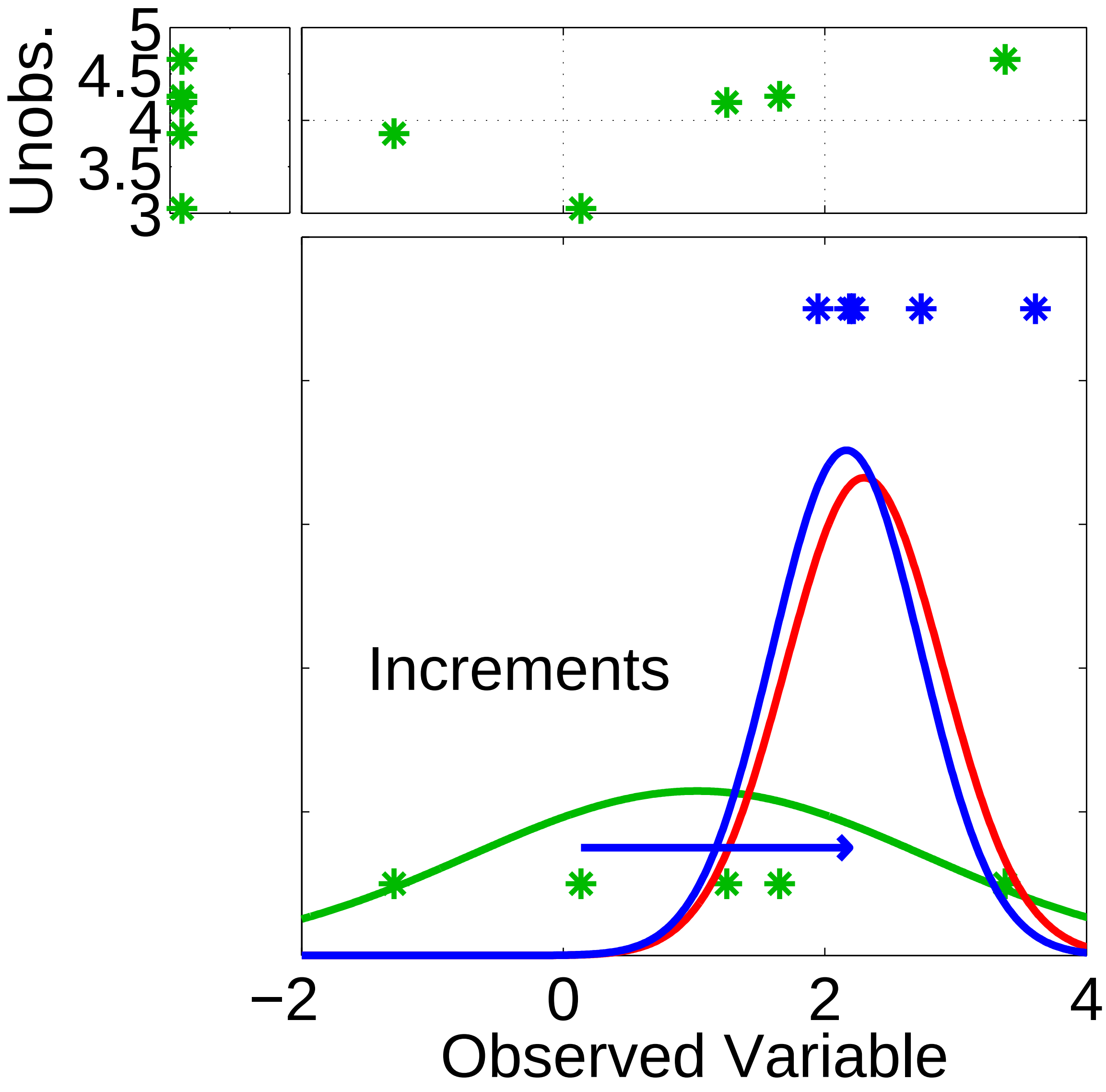


Assume that all we know is prior joint distribution.

One variable is observed.

Update observed variable with one of previous methods.

Ensemble filters: Updating additional prior state variables

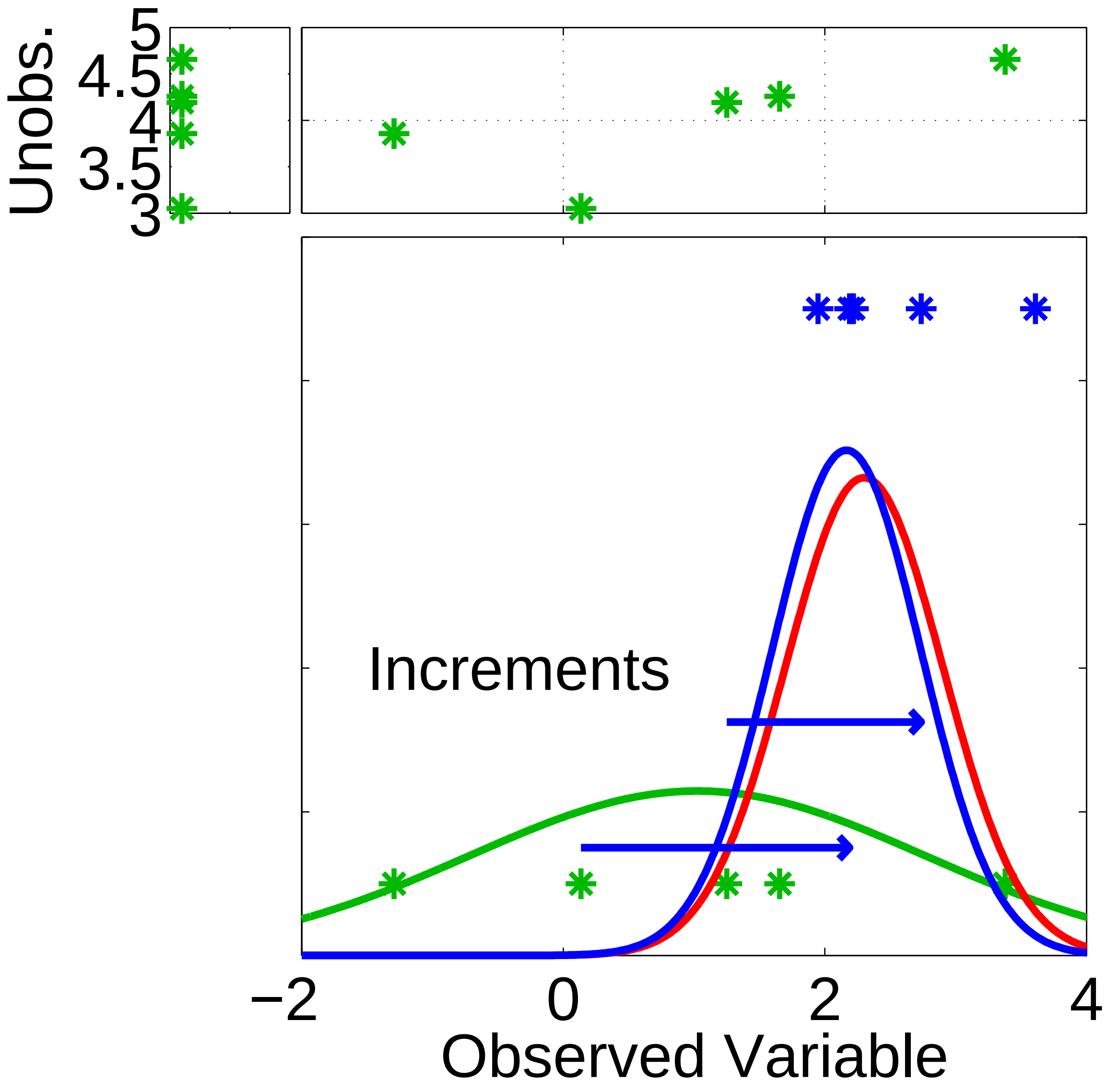


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

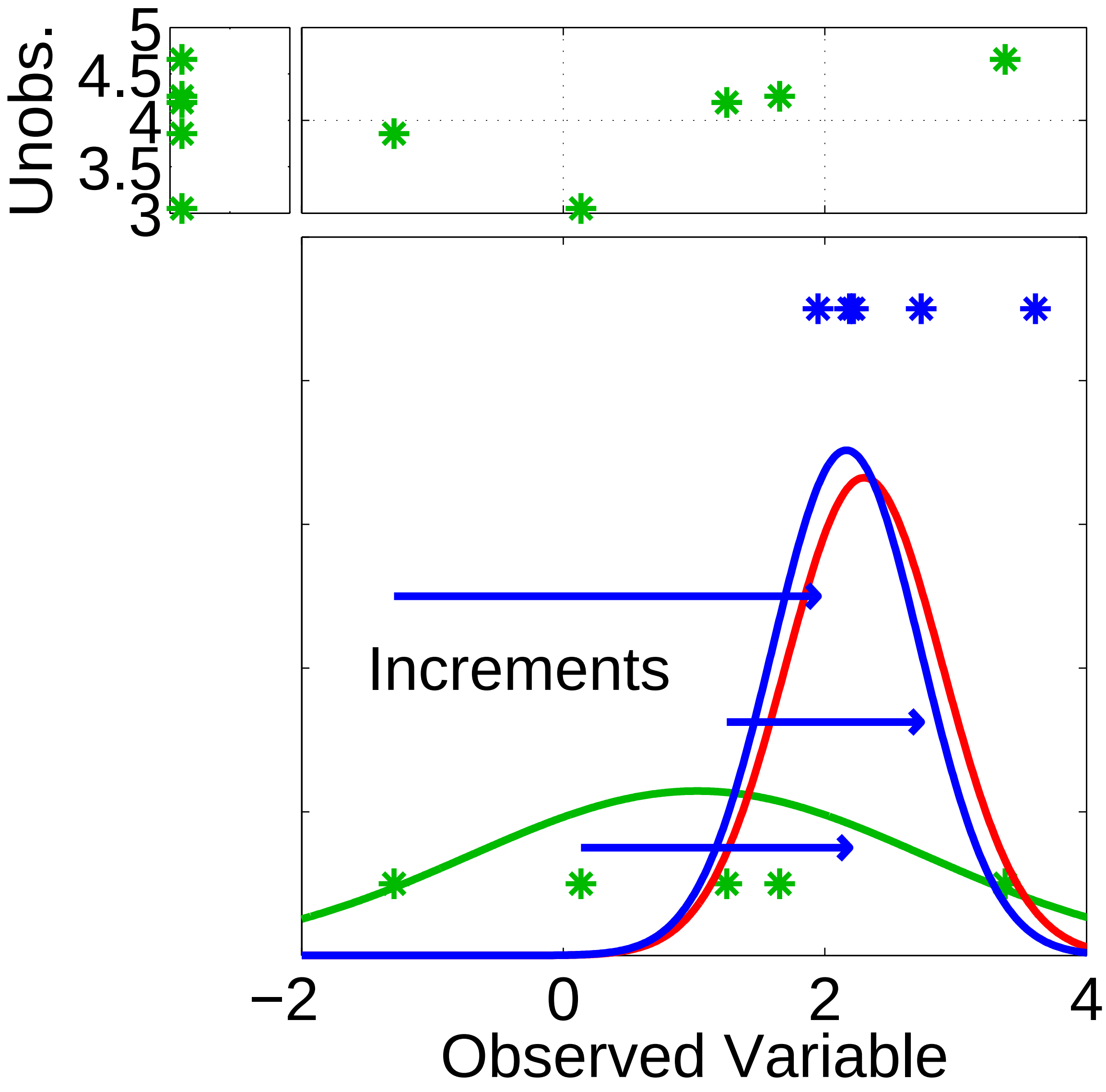


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

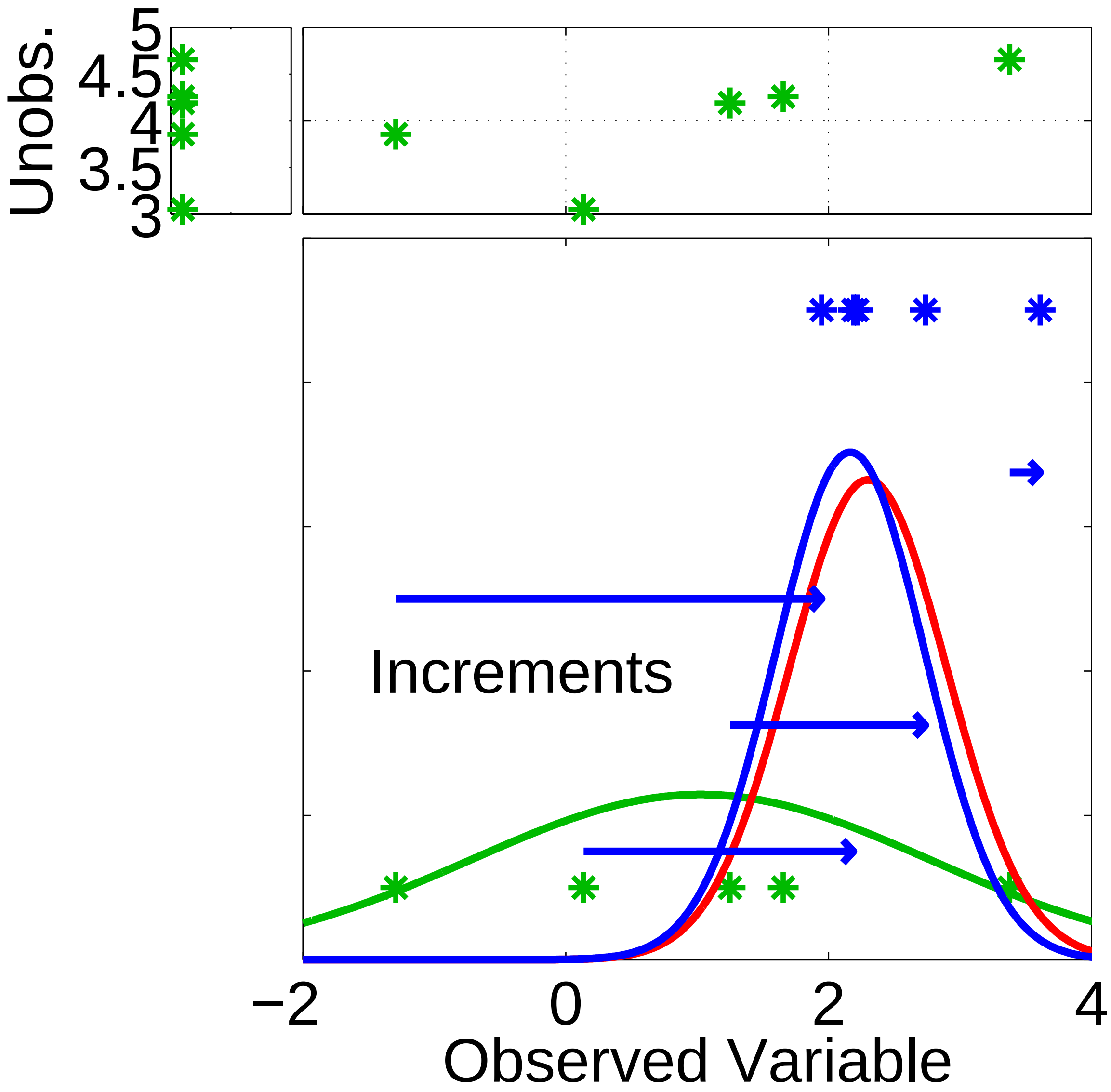


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

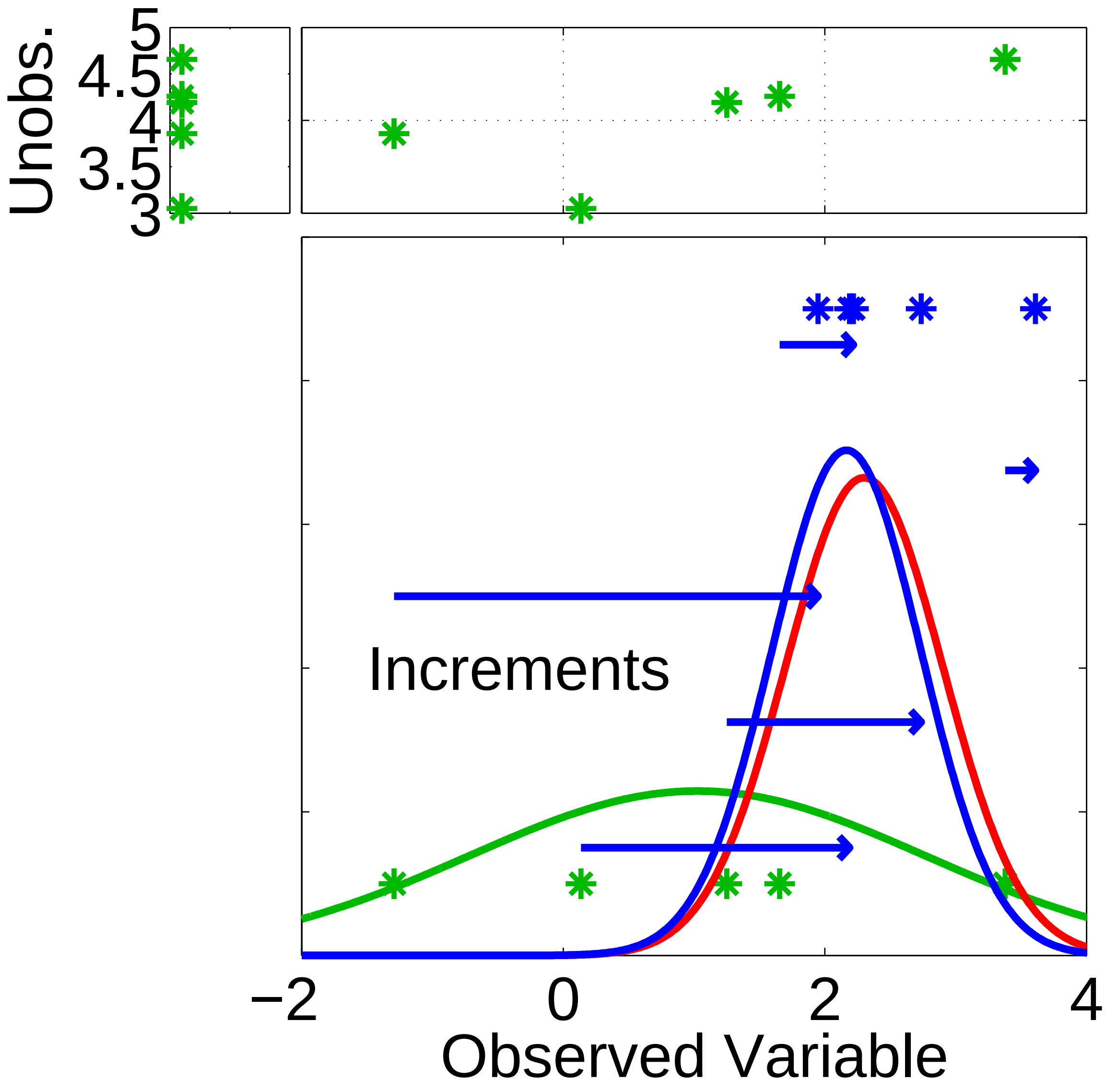


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

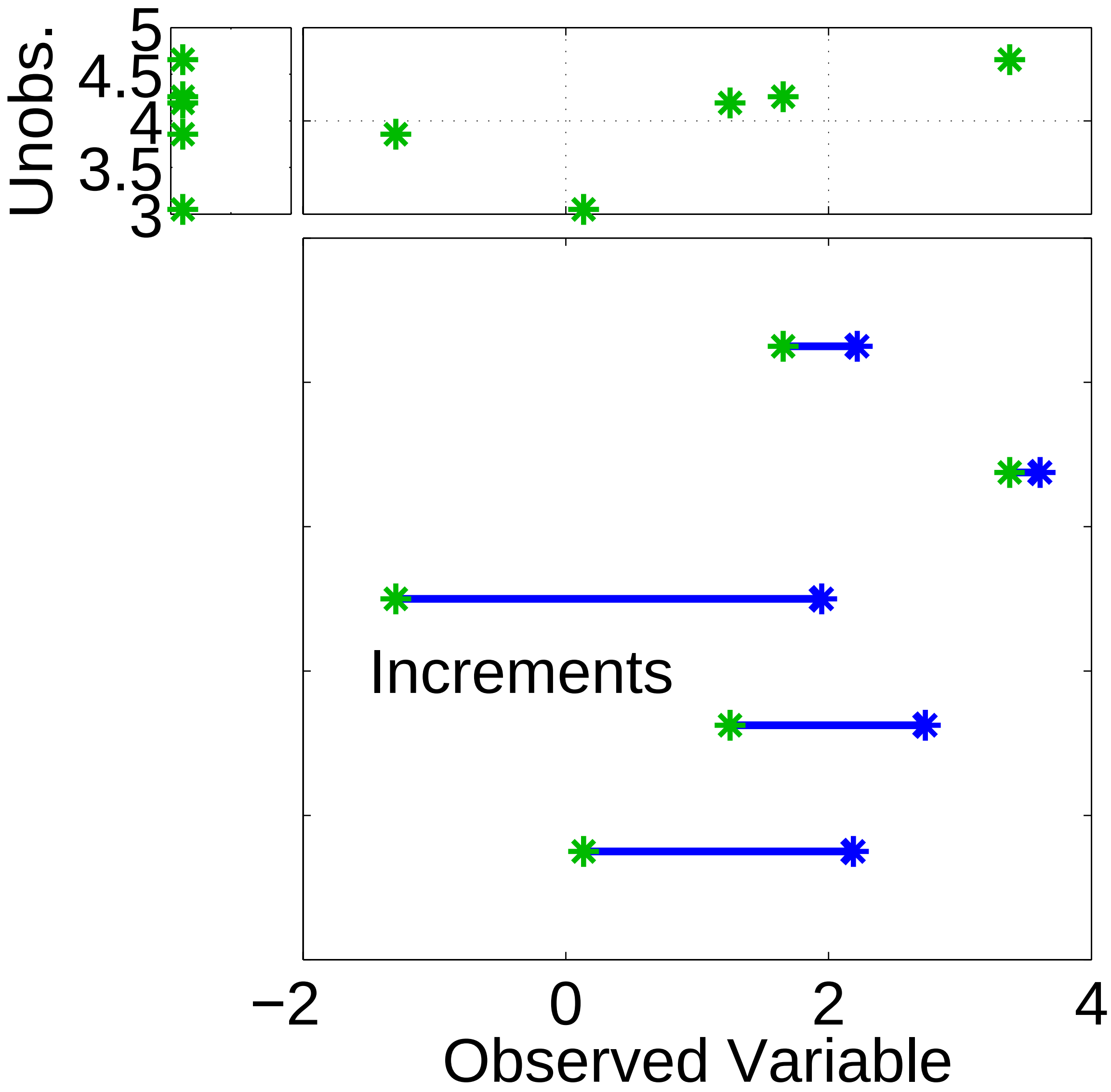


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

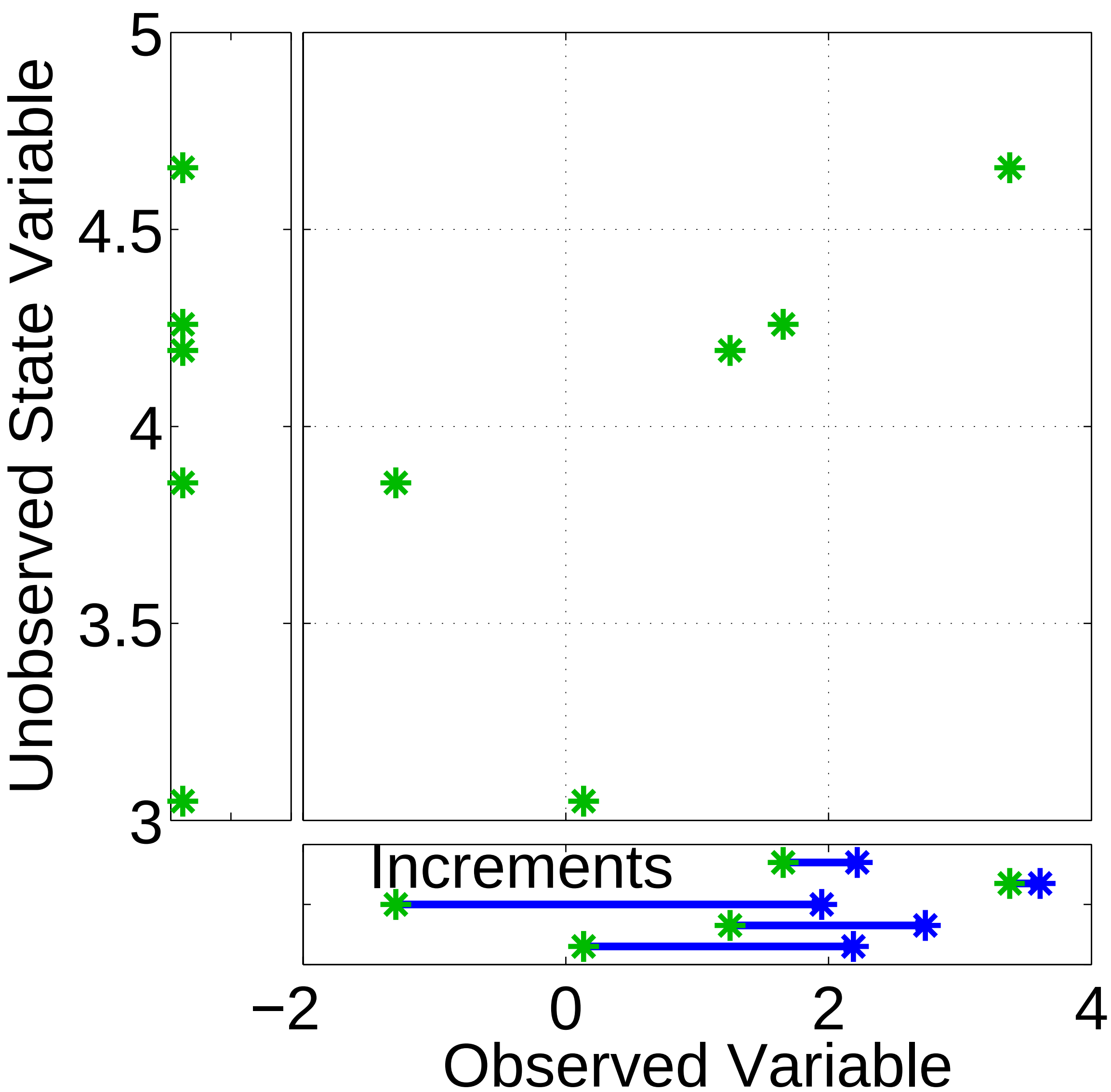


Assume that all we know is prior joint distribution.

One variable is observed.

Using only increments guarantees that if observation had no impact on observed variable, unobserved variable is unchanged (highly desirable).

Ensemble filters: Updating additional prior state variables



Assume that all we know is prior joint distribution.

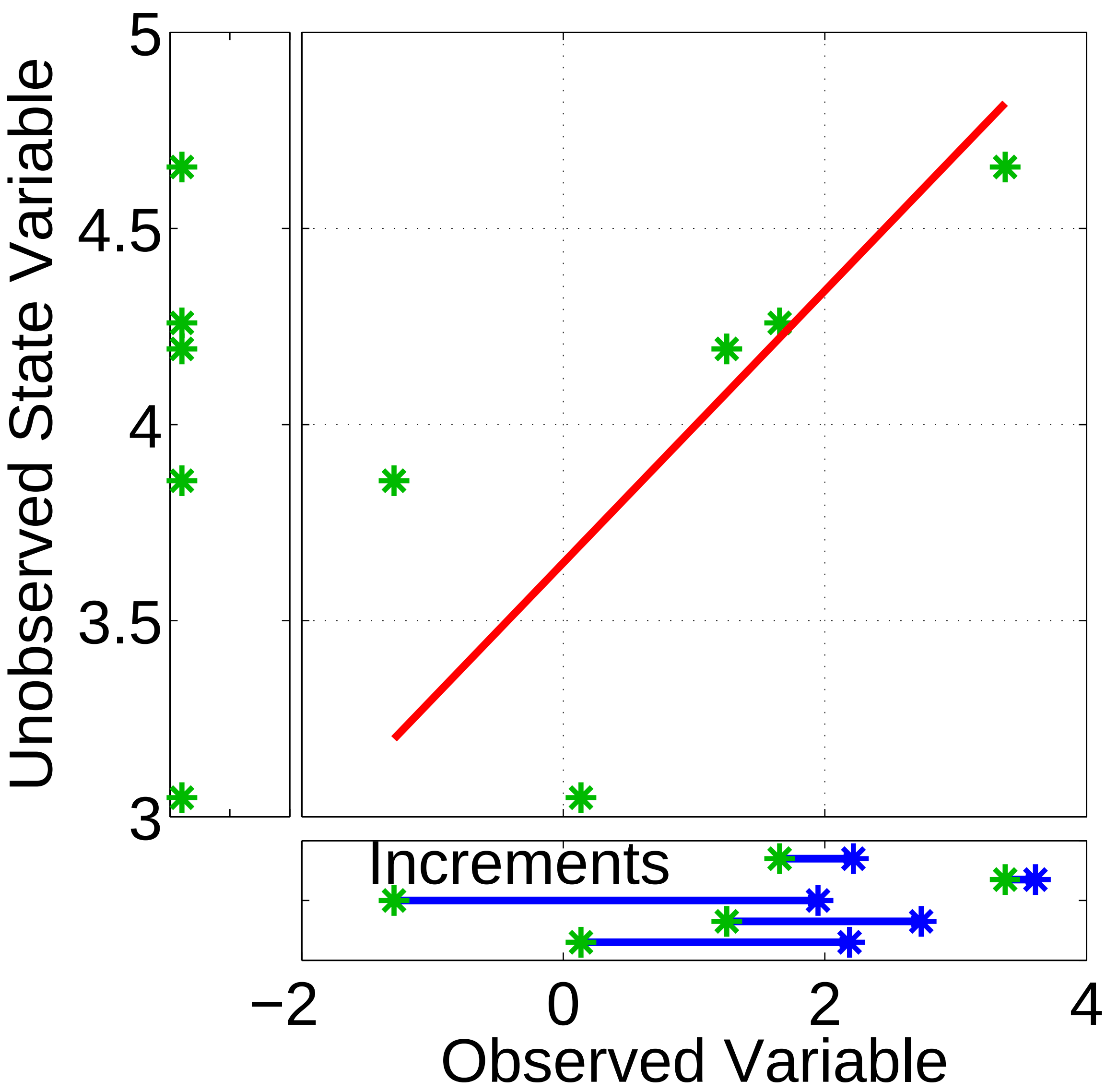
How should the unobserved variable be impacted?

First choice: least squares

Equivalent to linear regression.

Same as assuming binormal prior.

Ensemble filters: Updating additional prior state variables



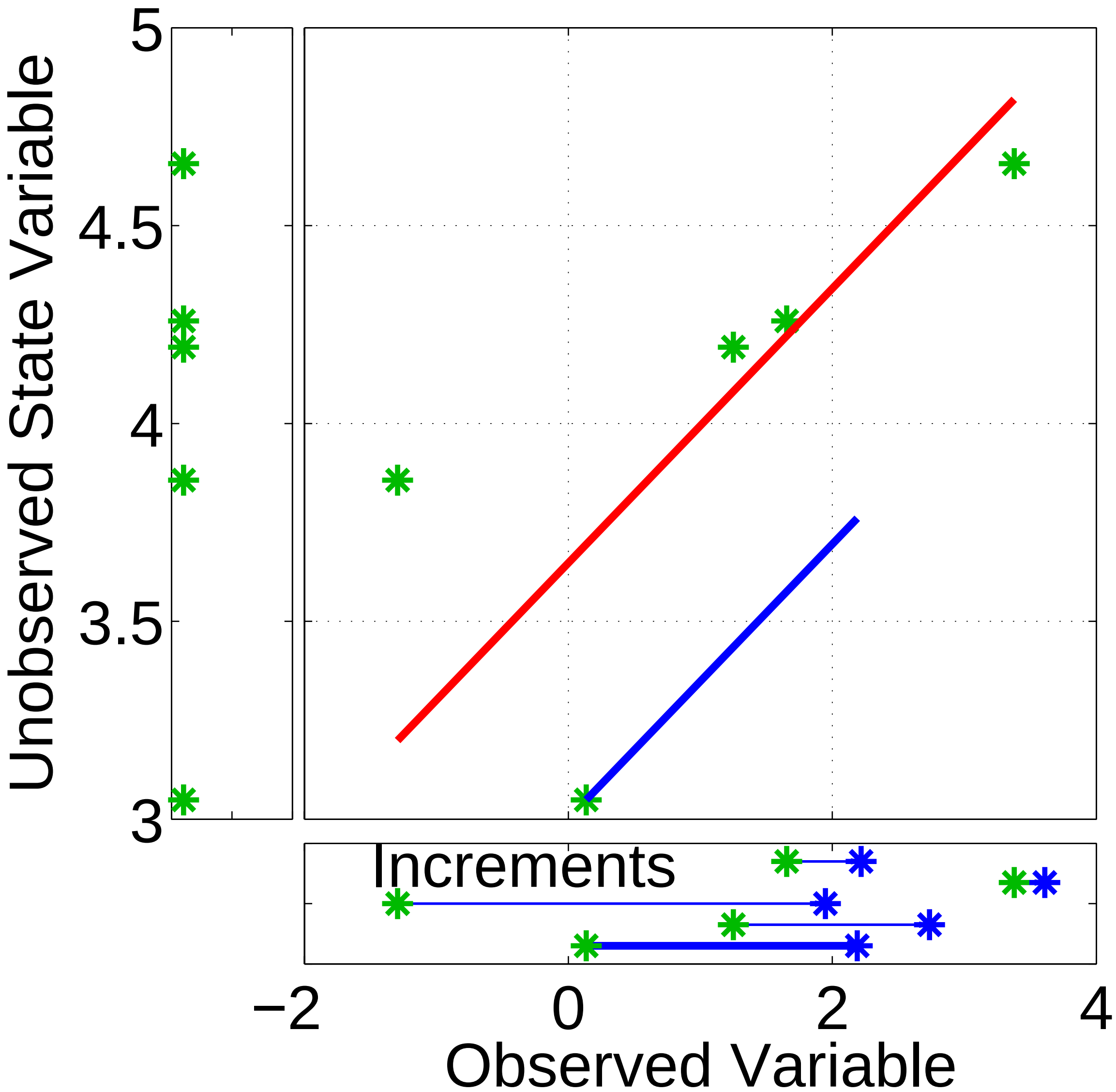
Have joint prior distribution of two variables.

How should the unobserved variable be impacted?

First choice: least squares

Begin by finding least squares fit.

Ensemble filters: Updating additional prior state variables

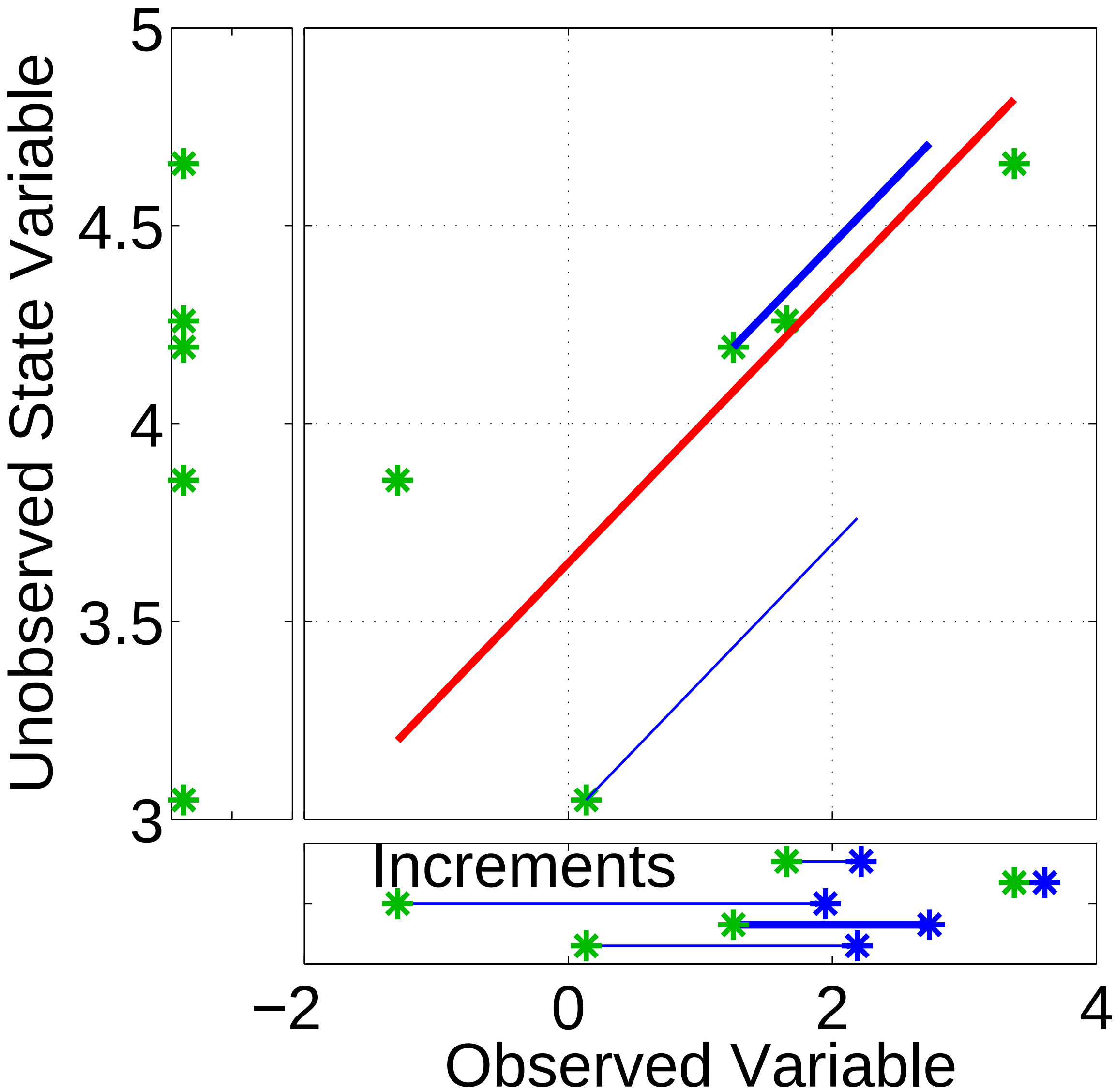


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

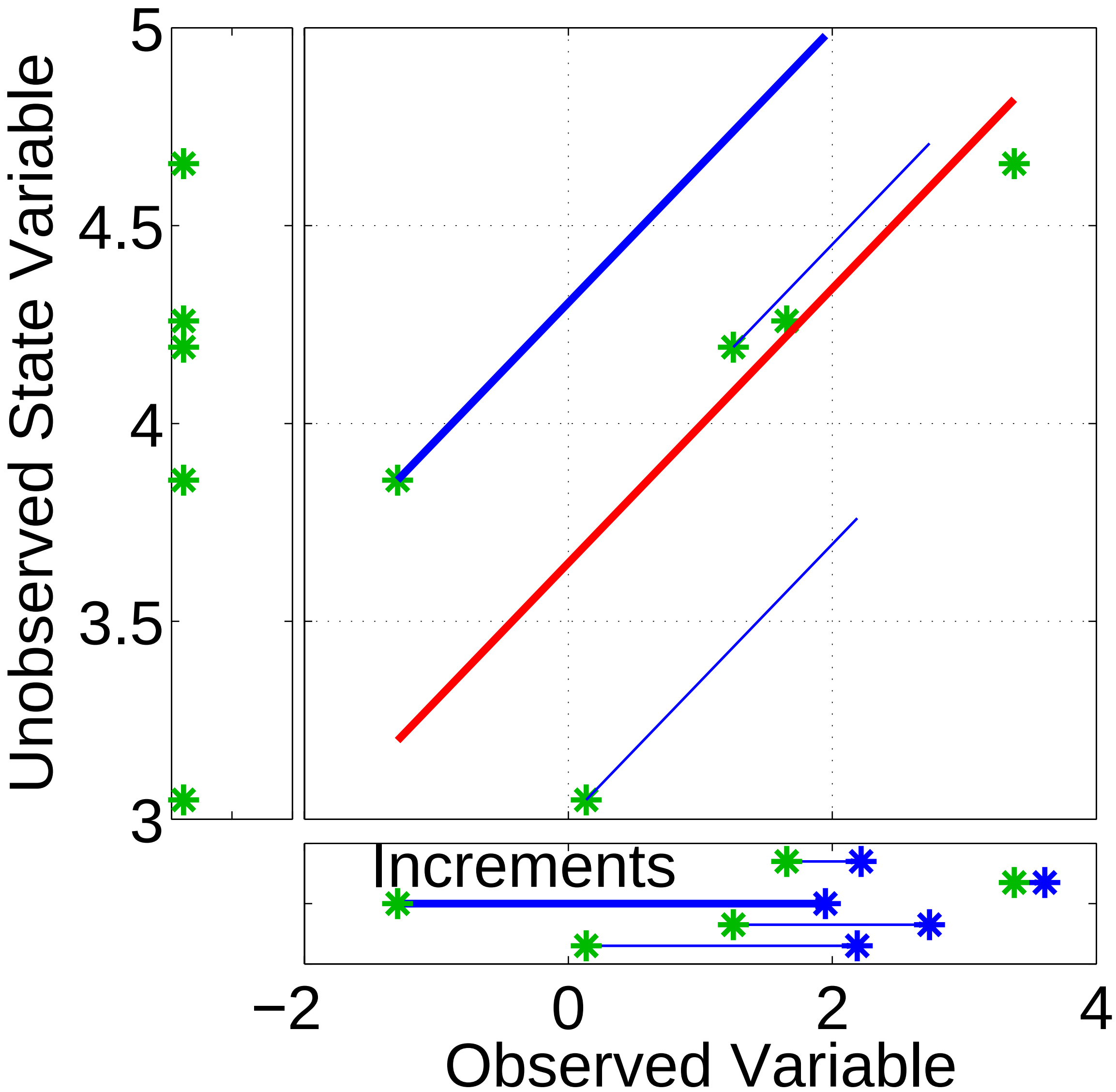


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

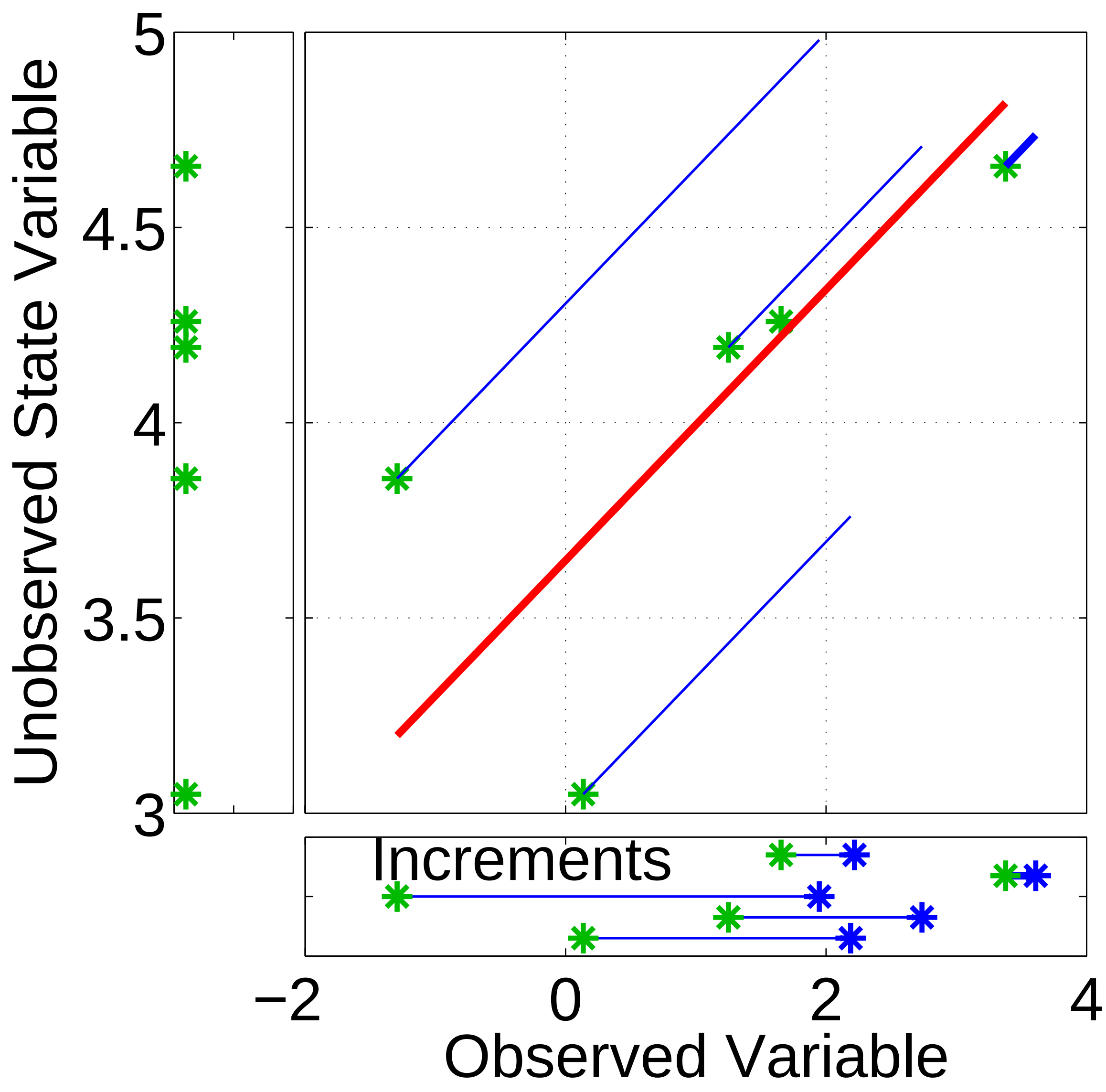


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

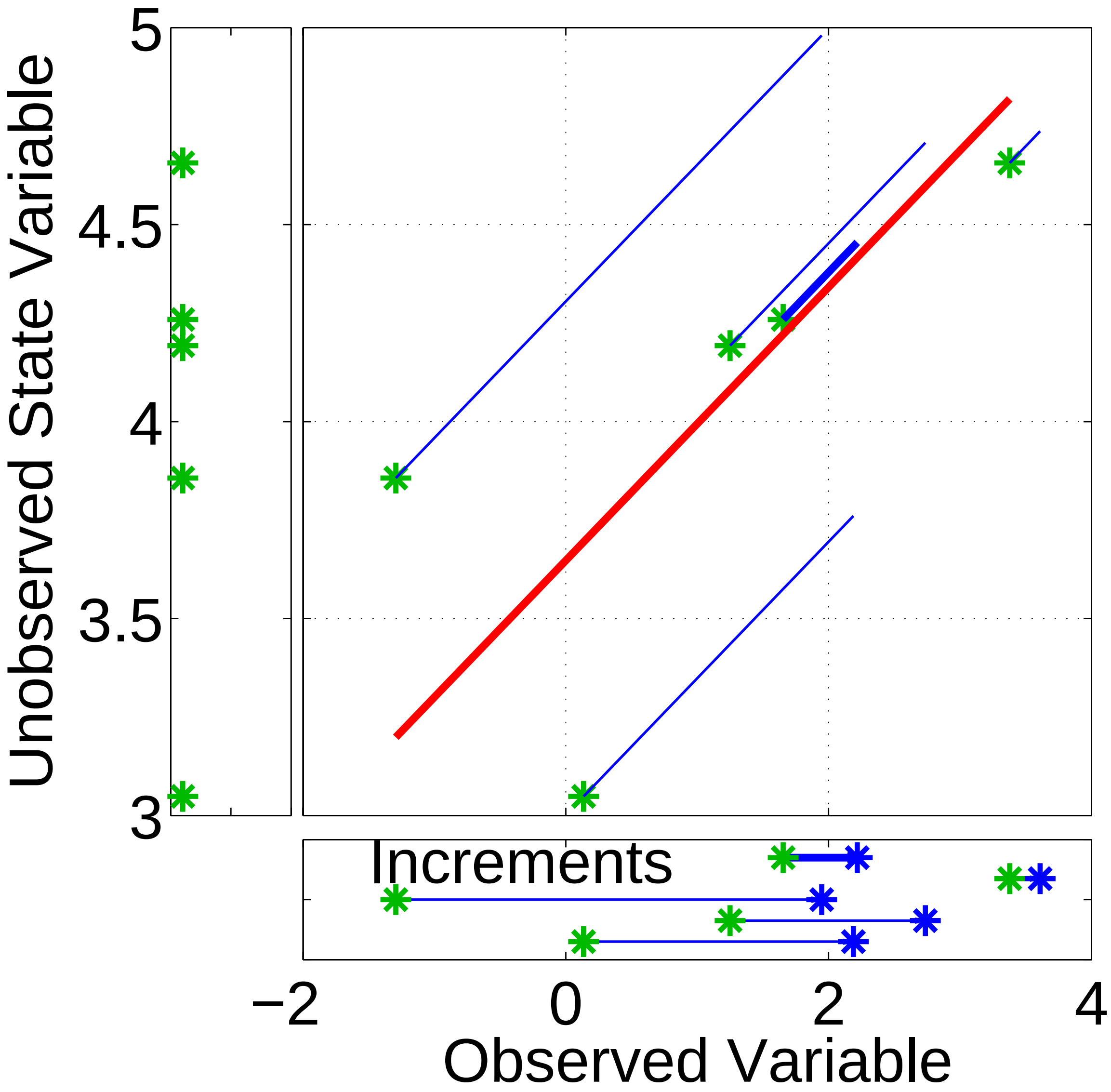


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables

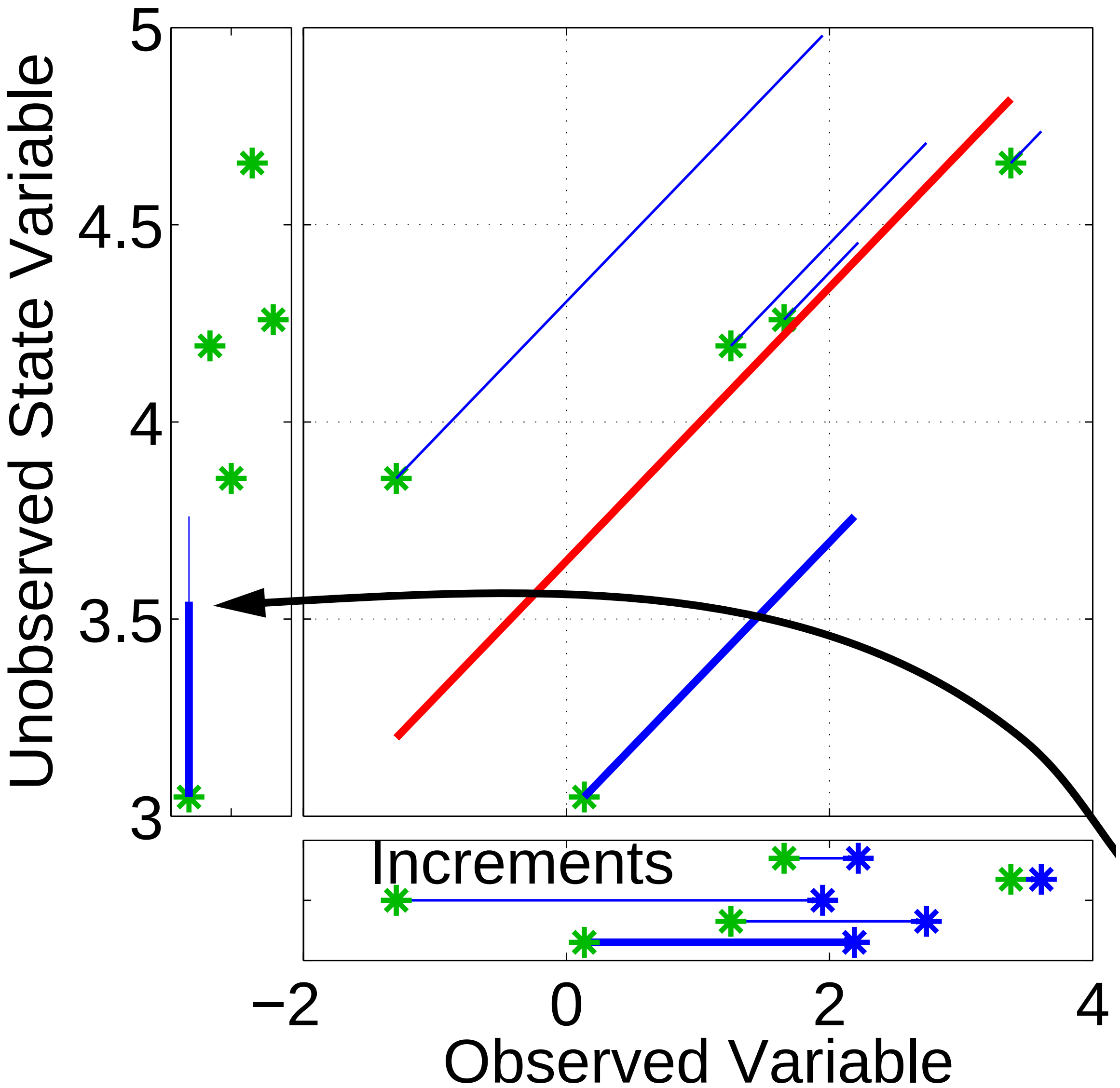


Have joint prior distribution of two variables.

Next, regress the observed variable increments onto increments for the unobserved variable.

Equivalent to first finding image of increment in joint space.

Ensemble filters: Updating additional prior state variables



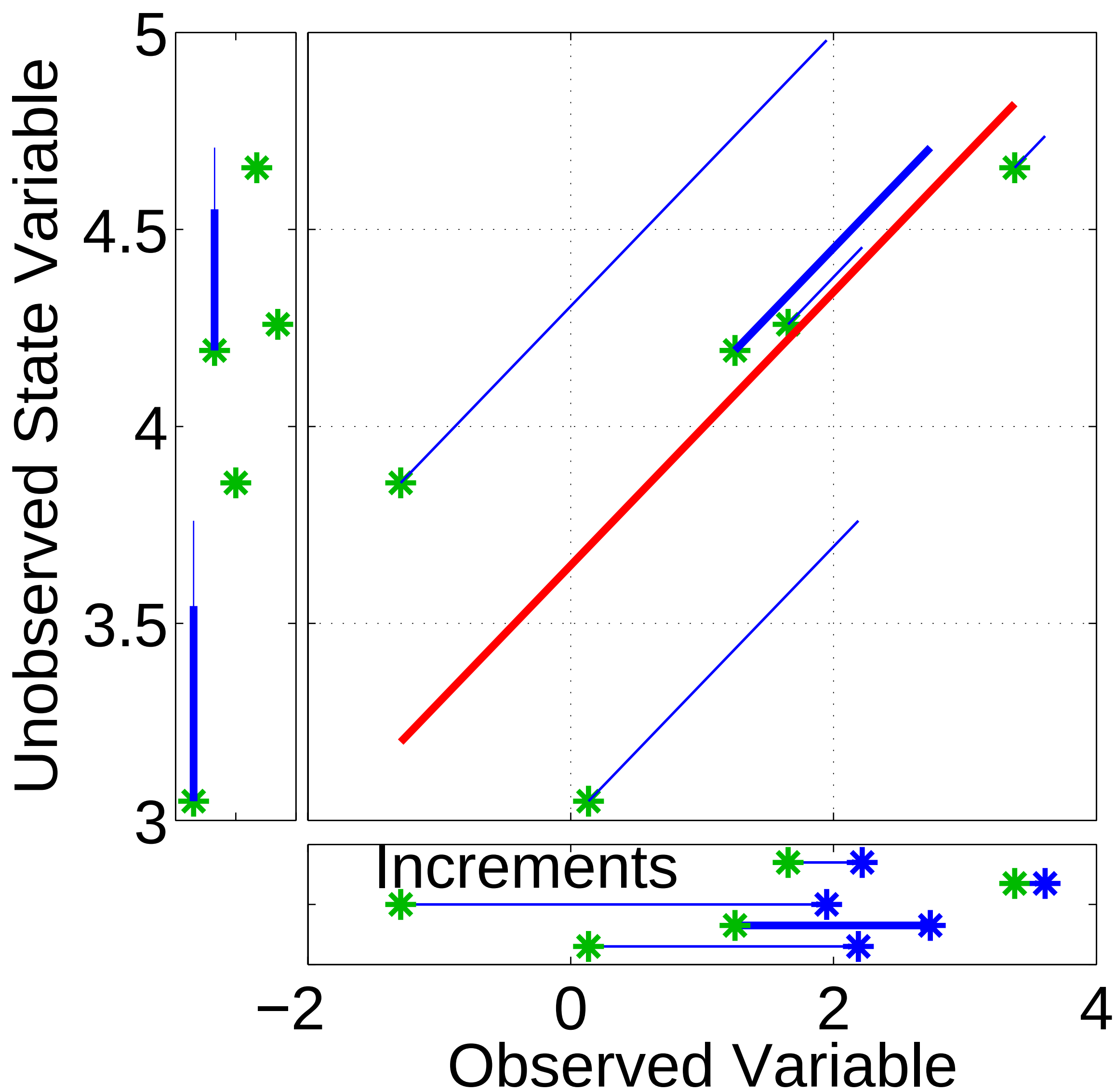
Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Finally, multiply by prior sample correlation.

Ensemble filters: Updating additional prior state variables



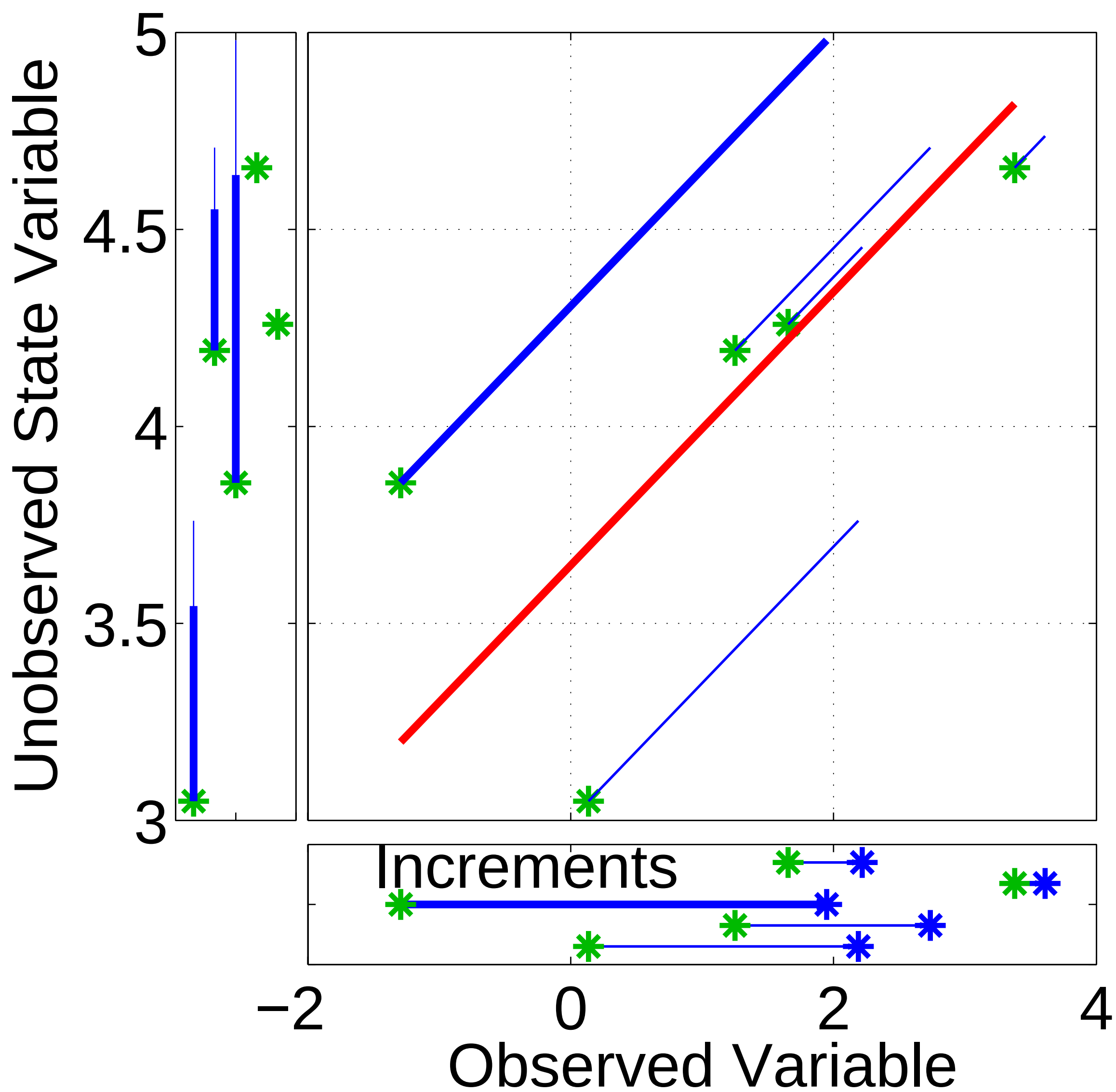
Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Finally, multiply by prior sample correlation.

Ensemble filters: Updating additional prior state variables



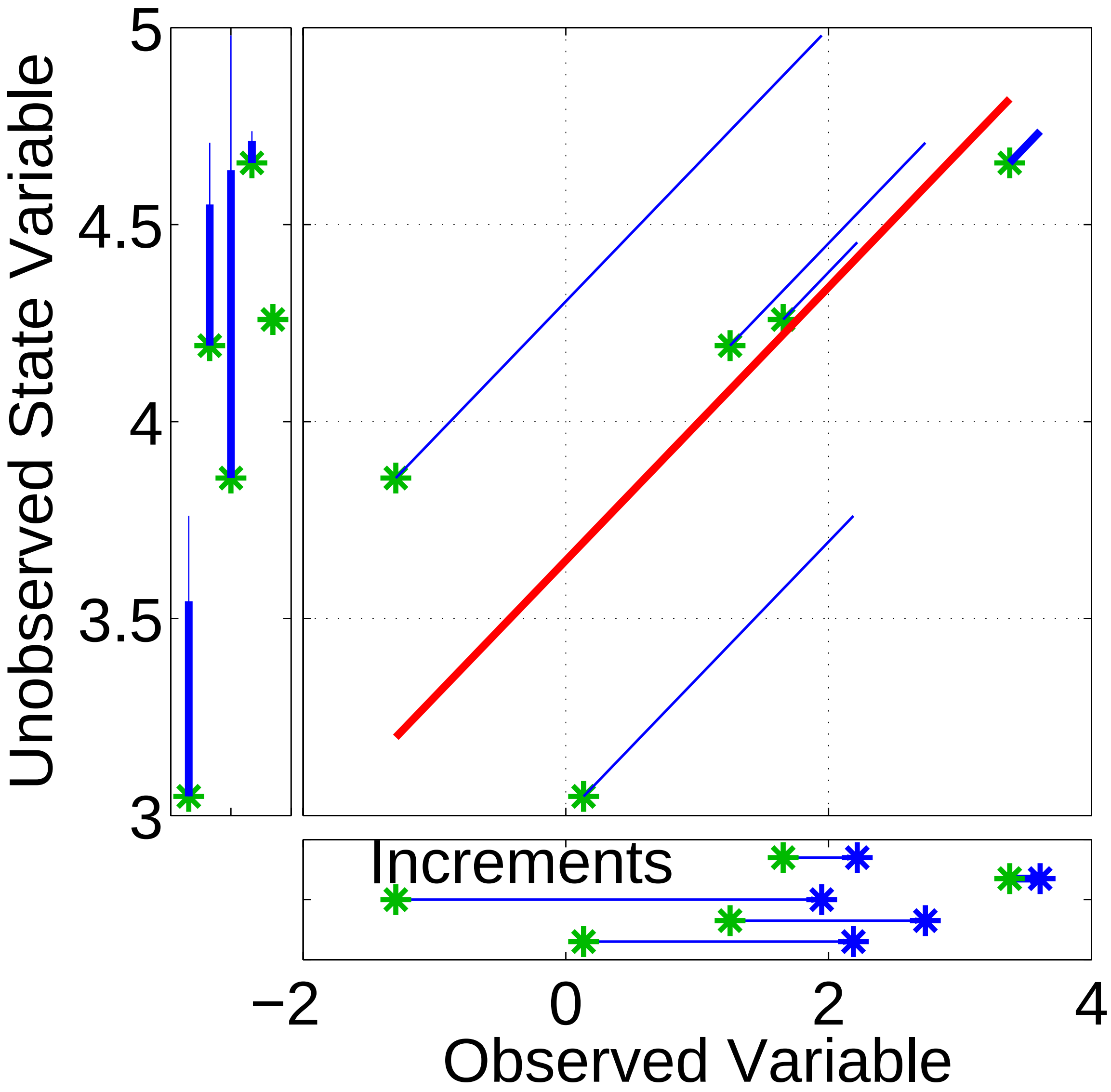
Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Finally, multiply by prior sample correlation.

Ensemble filters: Updating additional prior state variables



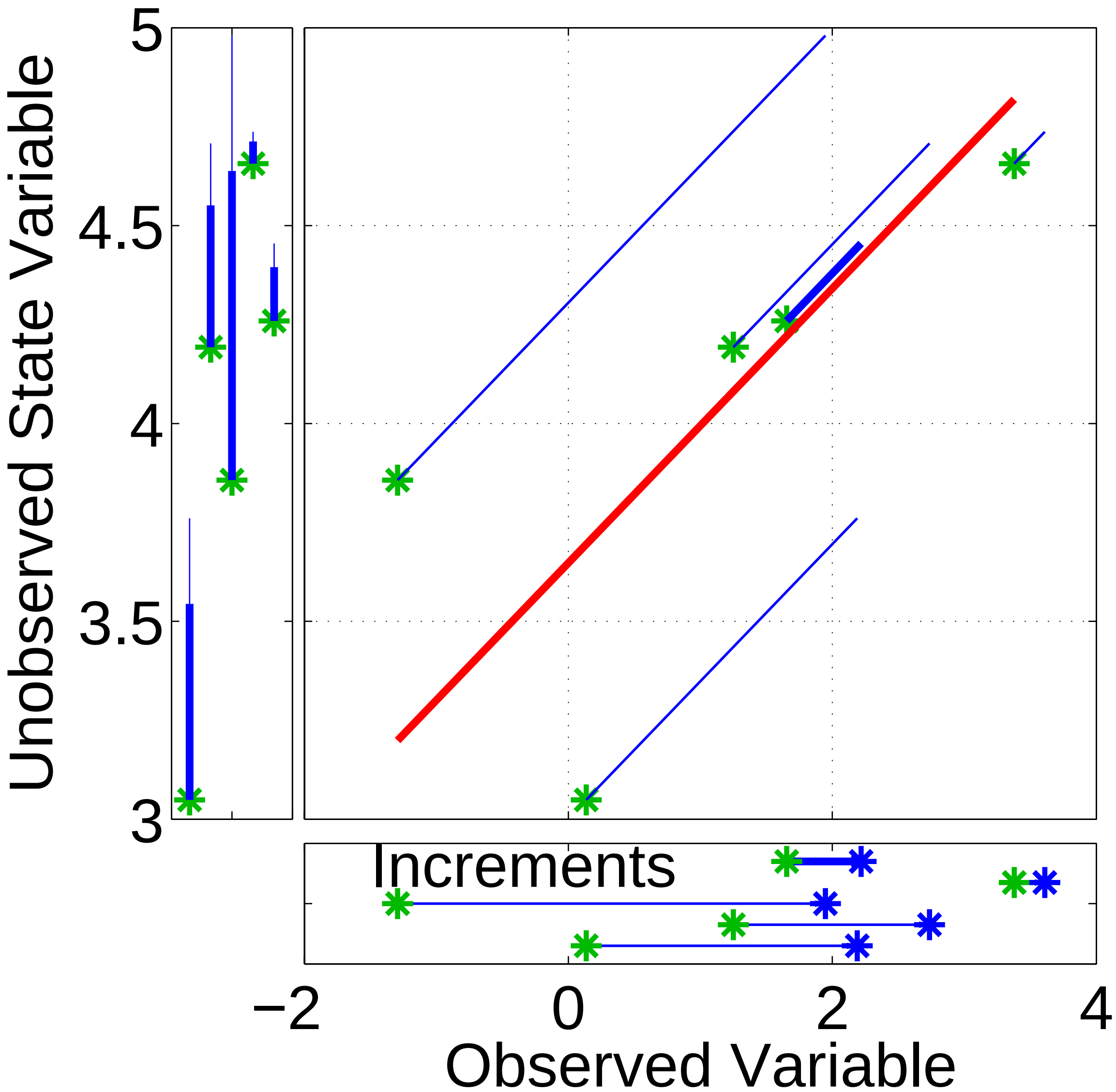
Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

Finally, multiply by prior sample correlation.

Ensemble filters: Updating additional prior state variables



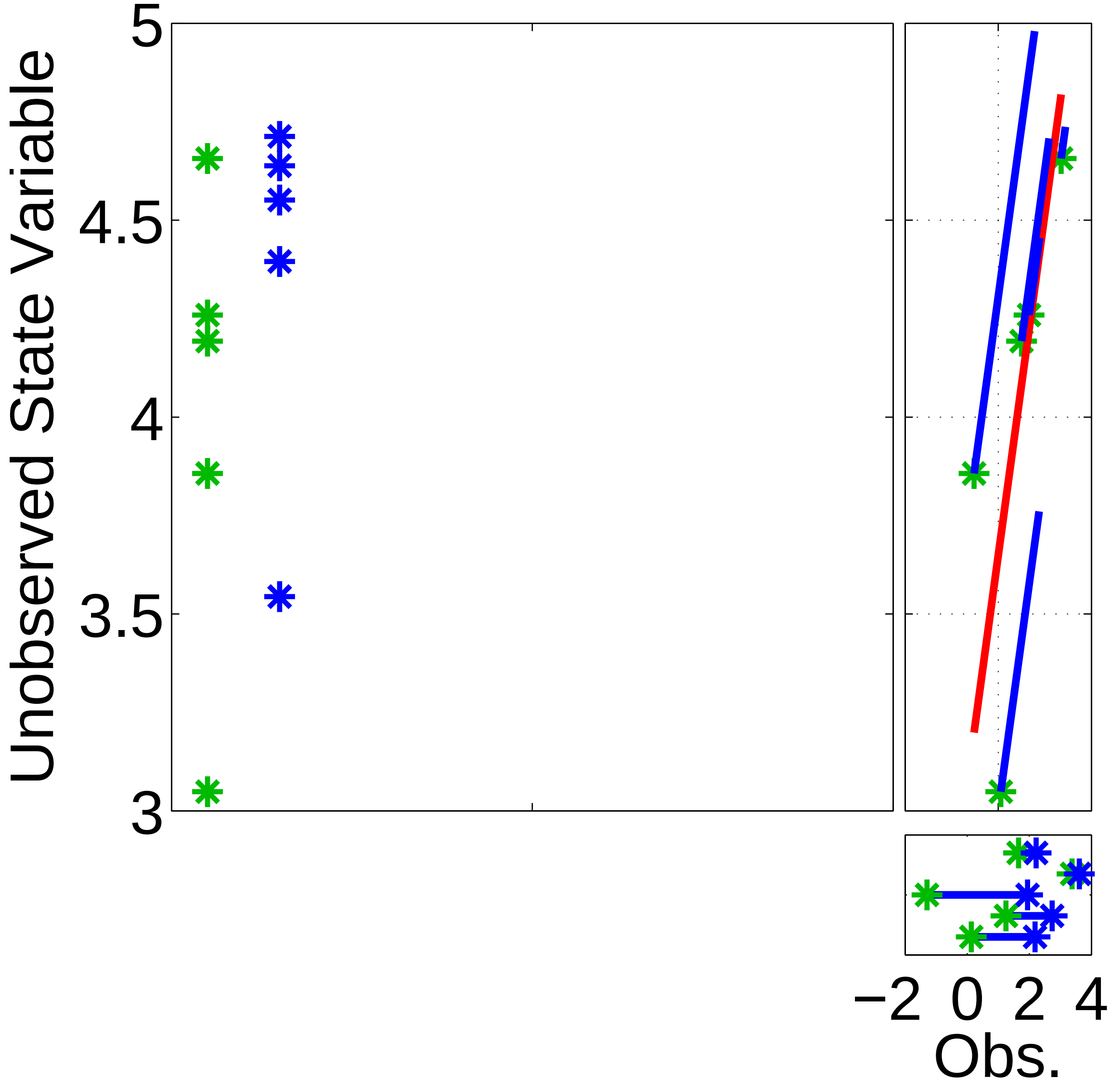
Have joint prior distribution of two variables.

Regression: Equivalent to first finding image of increment in joint space.

Then projecting from joint space onto unobserved priors.

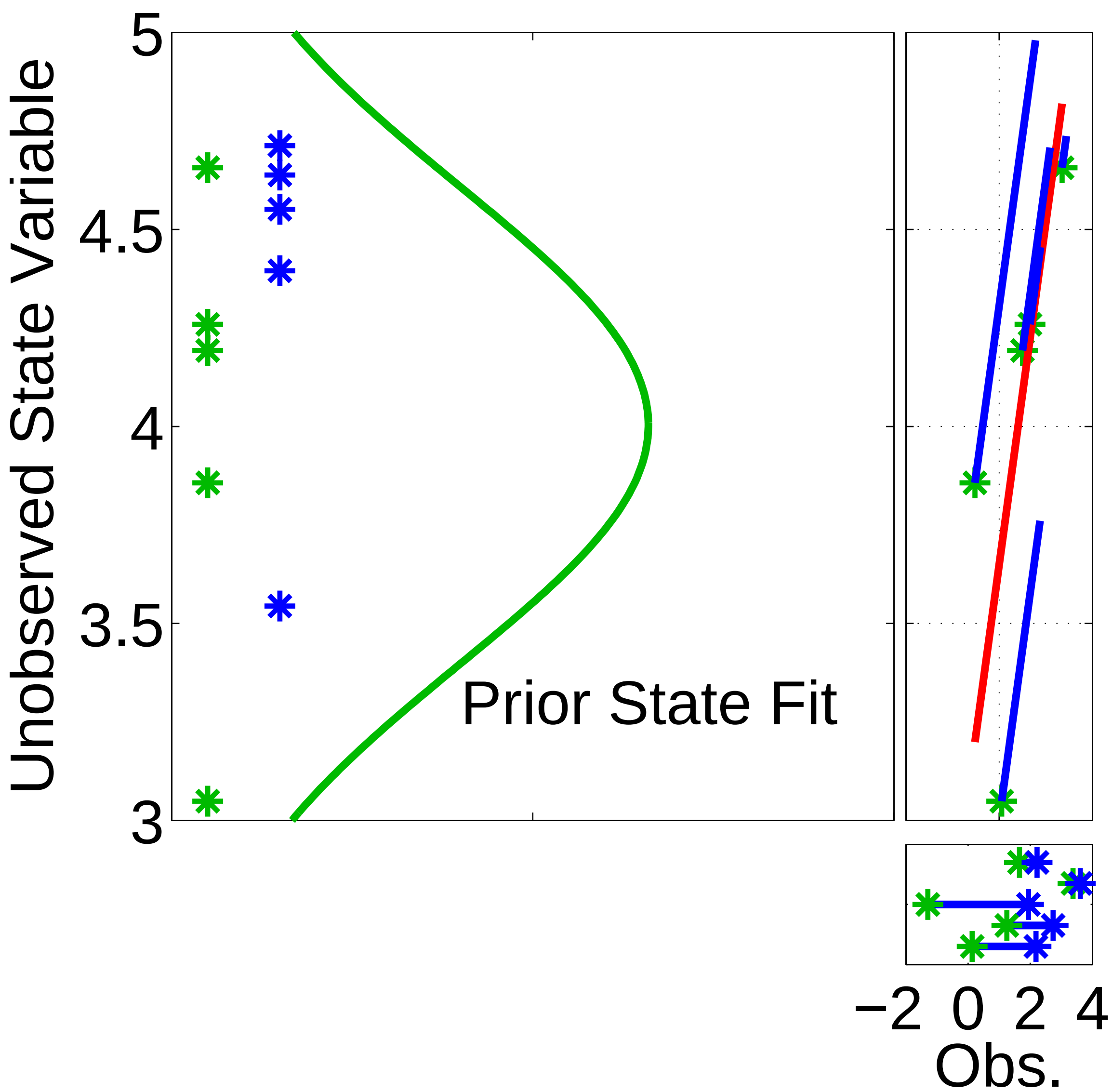
Finally, multiply by prior sample correlation.

Ensemble filters: Updating additional prior state variables



Now have an updated (posterior) ensemble for the unobserved variable.

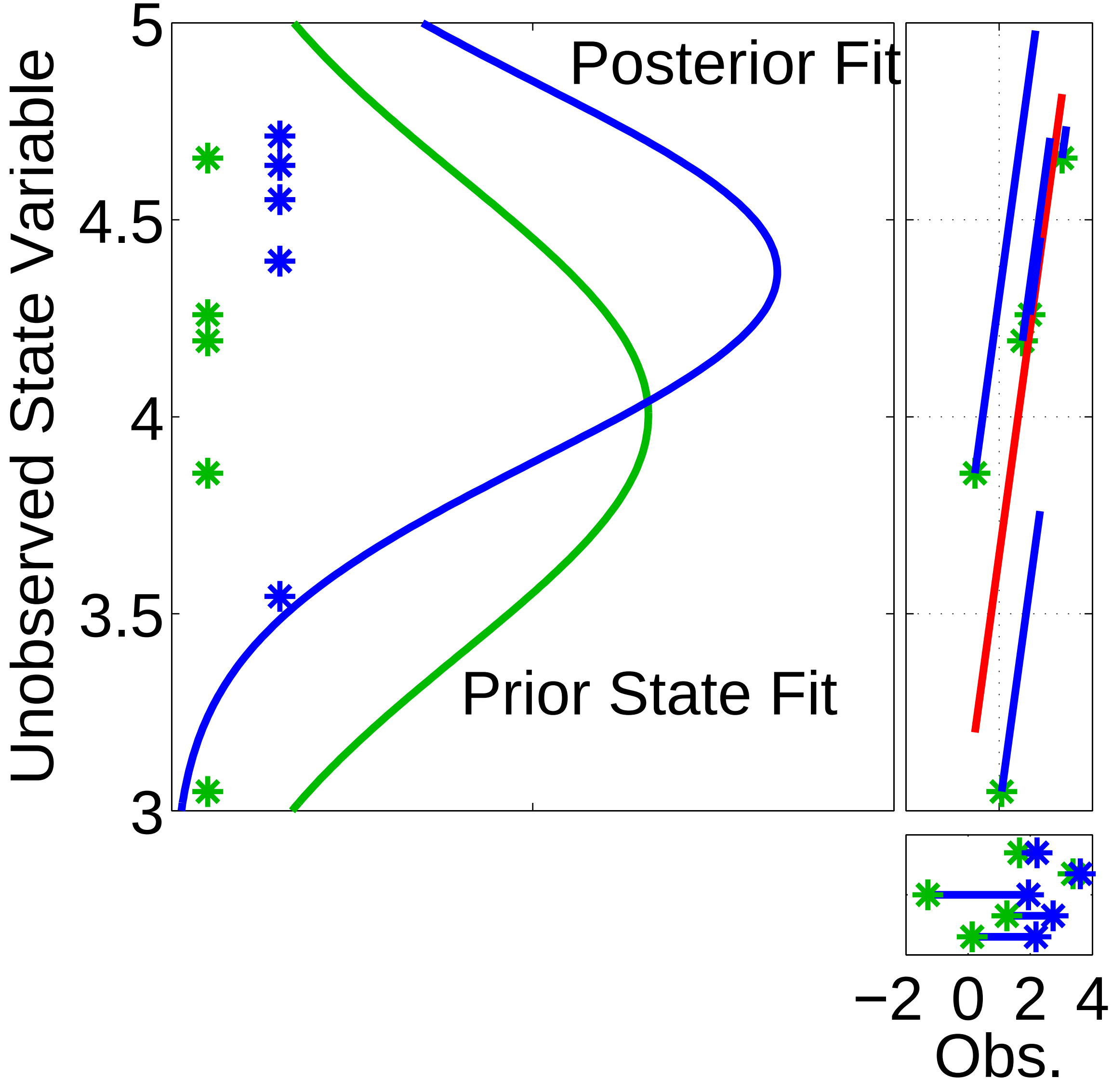
Ensemble filters: Updating additional prior state variables



Now have an updated (posterior) ensemble for the unobserved variable.

Fitting Gaussians shows that mean and variance have changed.

Ensemble filters: Updating additional prior state variables

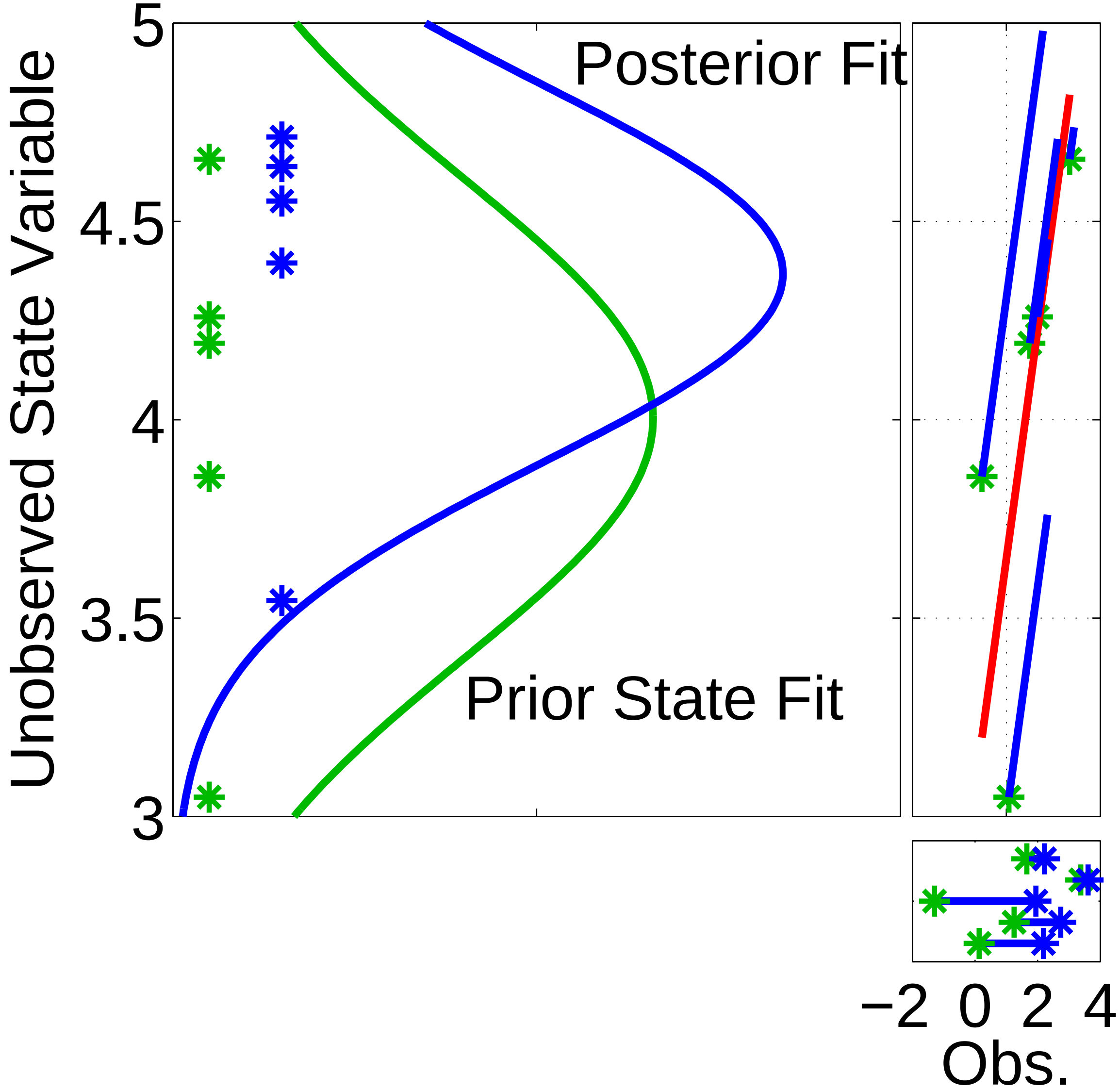


Now have an updated (posterior) ensemble for the unobserved variable.

Fitting Gaussians shows that mean and variance have changed.

Other features of the prior distribution may also have changed.

Ensemble filters: Updating additional prior state variables



CRITICAL POINT:

Since impact on unobserved variable is simply a linear regression, can do this INDEPENDENTLY for any number of unobserved variables!

Could also do many at once using matrix algebra as in traditional Kalman Filter.

Phase 3: Generalize to geophysical models and observations

Dynamical system governed by (stochastic) Difference Equation:

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

Observations at discrete times:

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

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:

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

Phase 3: Generalize to geophysical models and observations

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

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

Apply Bayes rule:

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

Noise is white in time (3) so:

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

Integrate numerator to get normalizing denominator:

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

Phase 3: Generalize to geophysical models and observations

Probability after new observation:

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

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.

Phase 3: Generalize to geophysical models and observations

Extending the state vector to joint state-observation vector.

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

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

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

Let z = [x, y] be the combined vector of state and observations.

Phase 3: Generalize to geophysical models and observations

NOW, we have a prior for each observation:

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

Phase 3: Generalize to geophysical models and observations

One more issue: how to deal with many observations in set y_k?

Let y_k be composed of s subsets of observations: $y_k = \{y_k^1, y_k^2, \ldots, y_k^s\}$

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.

Phase 3: Generalize to geophysical models and observations

One more issue: how to deal 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!

Applying an Ensemble Filter.

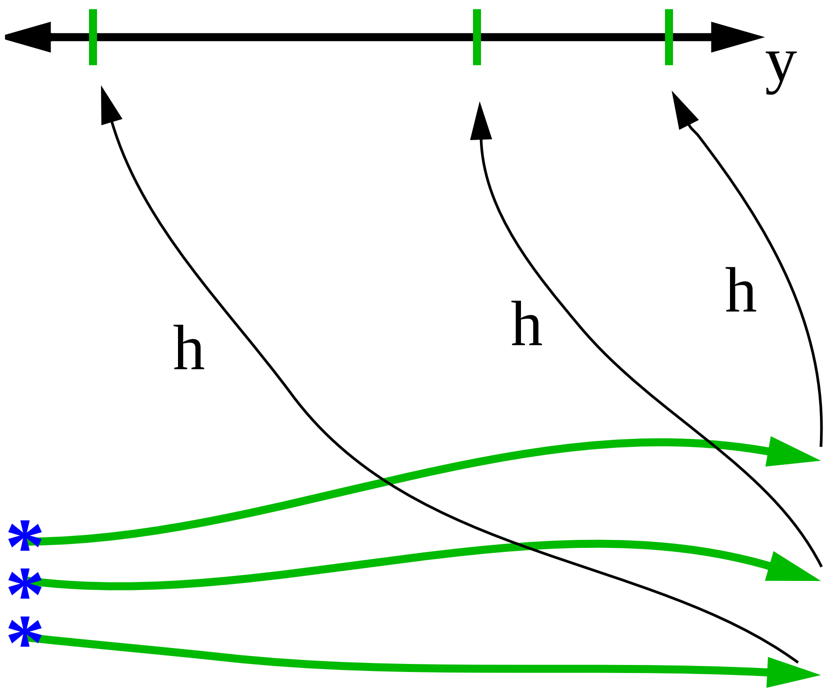
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis).

Ensemble state at time of next observation (prior).

Applying an Ensemble Filter.

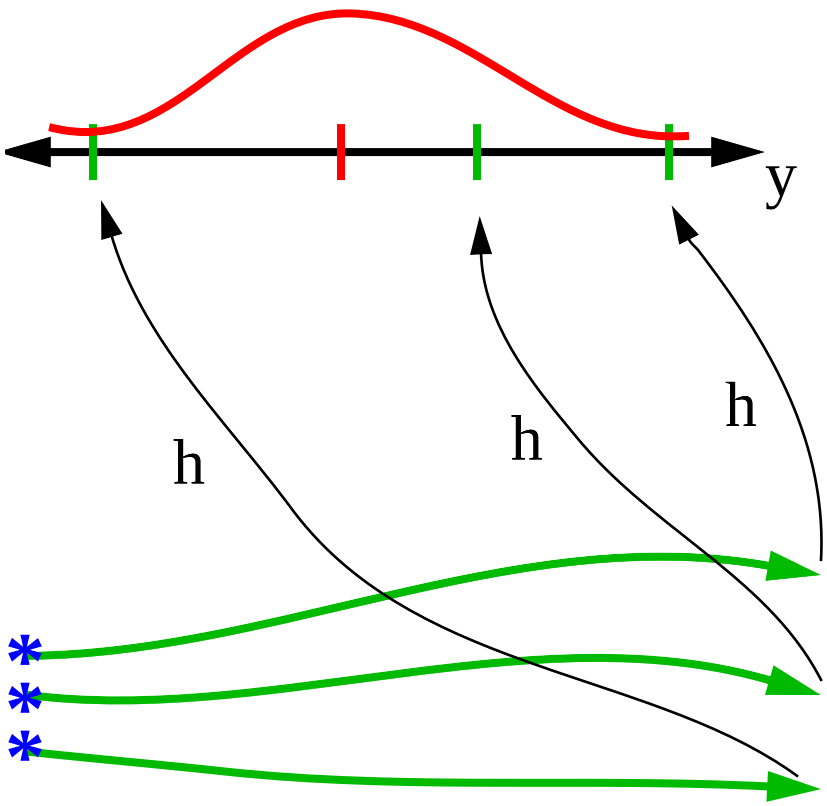
2. Get prior ensemble sample of observation, y=h(x), by applying forward operator h to each ensemble member.



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

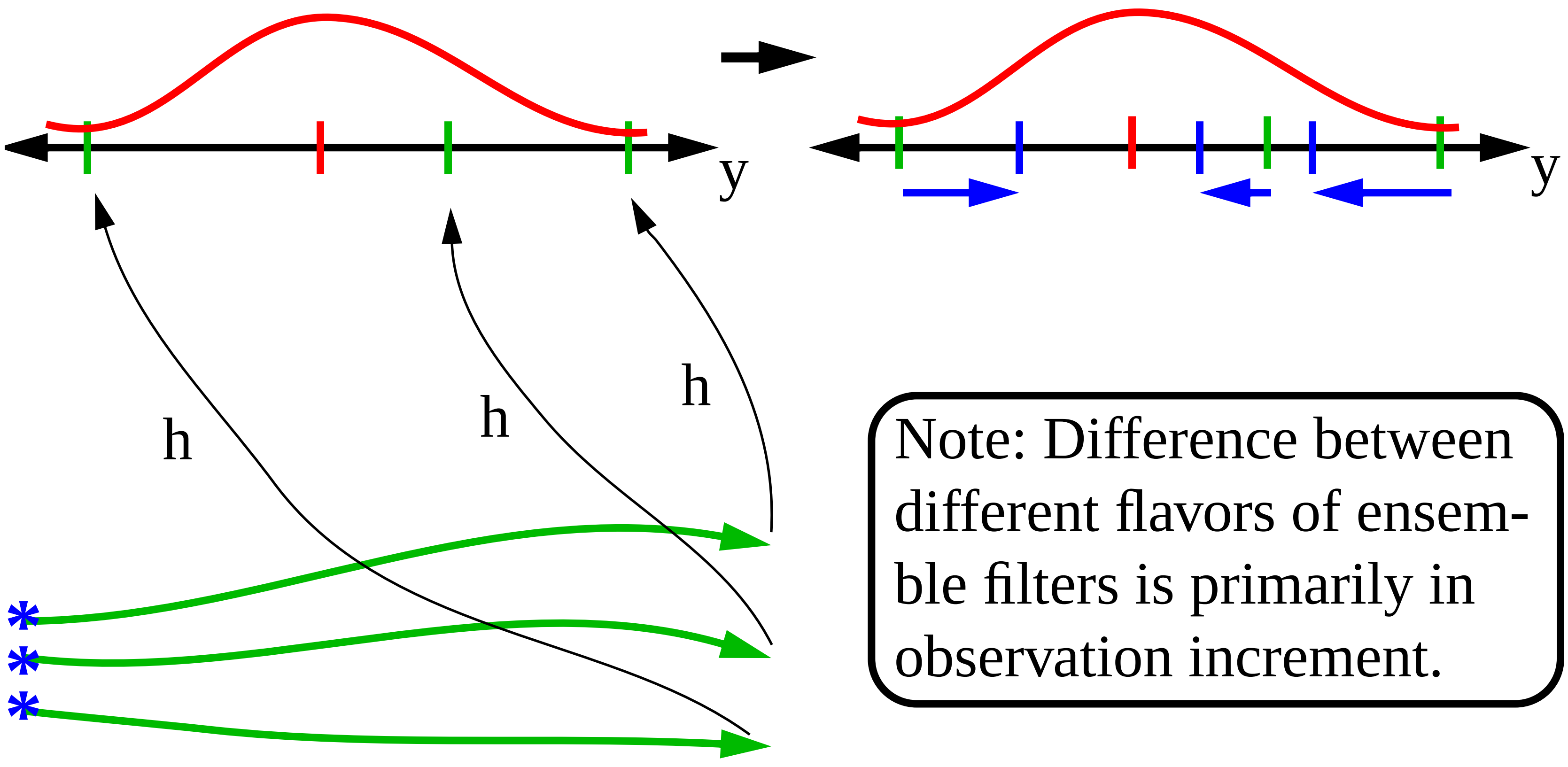
Applying an Ensemble Filter.

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



Applying an Ensemble Filter.

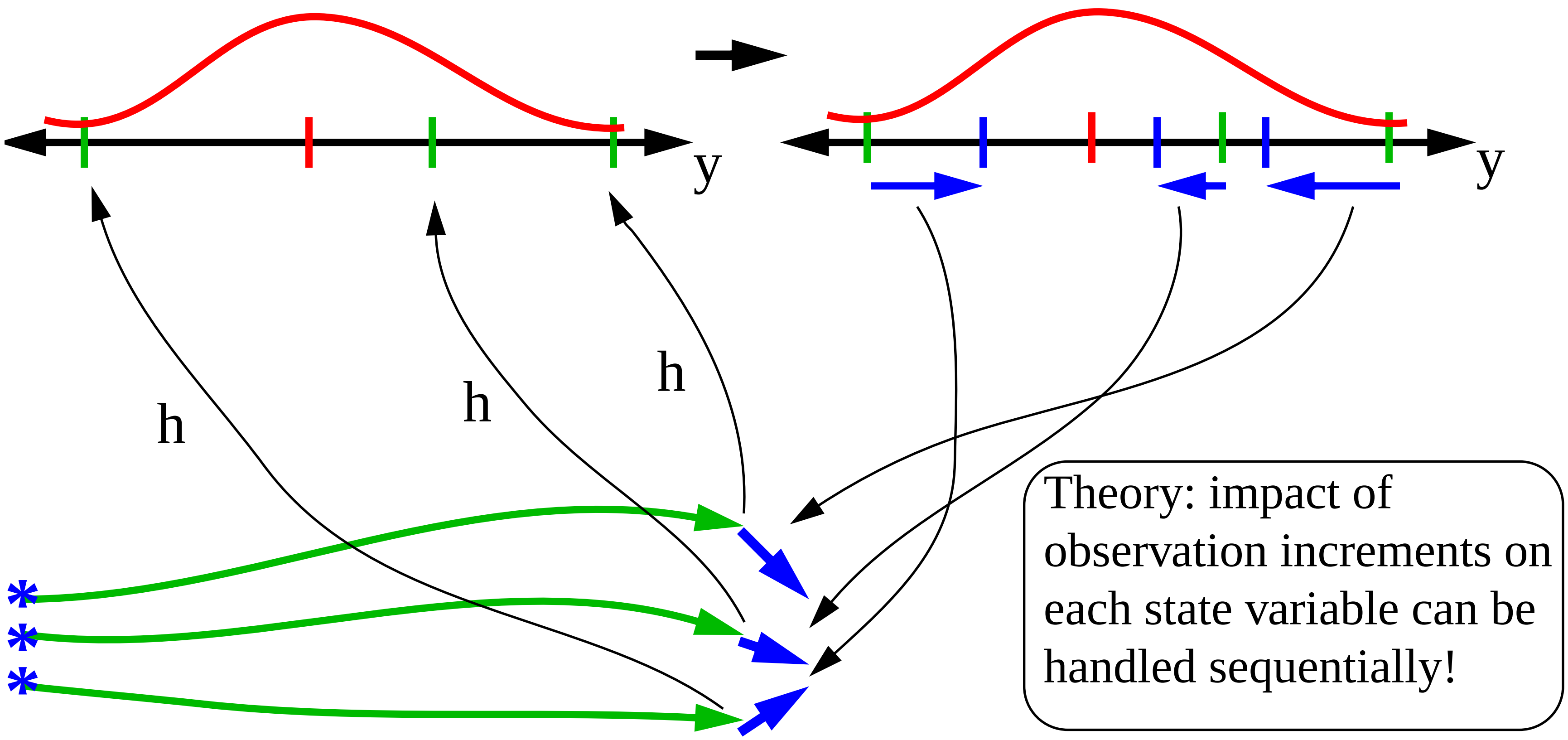
4. Find increment for each prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



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

Applying an Ensemble Filter.

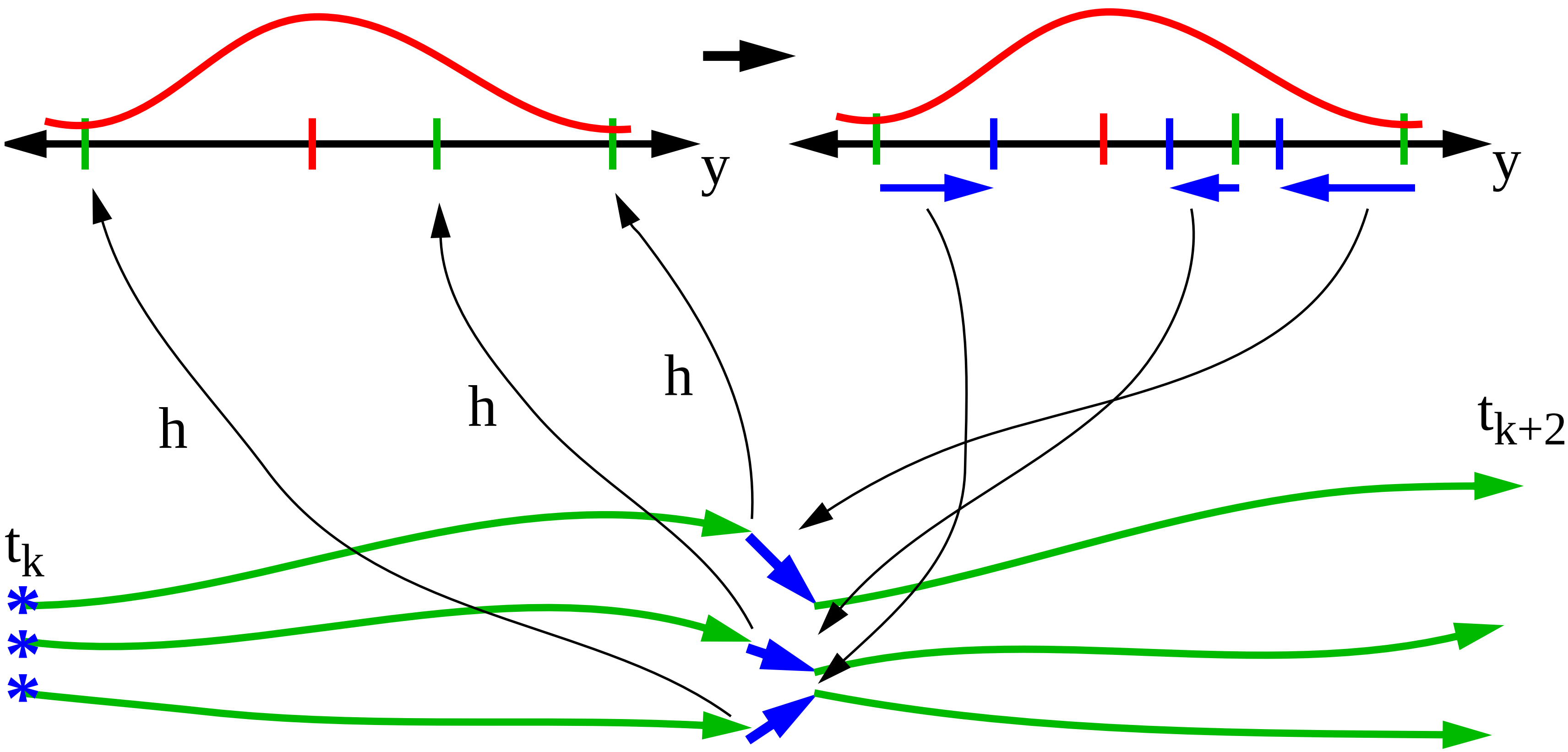
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



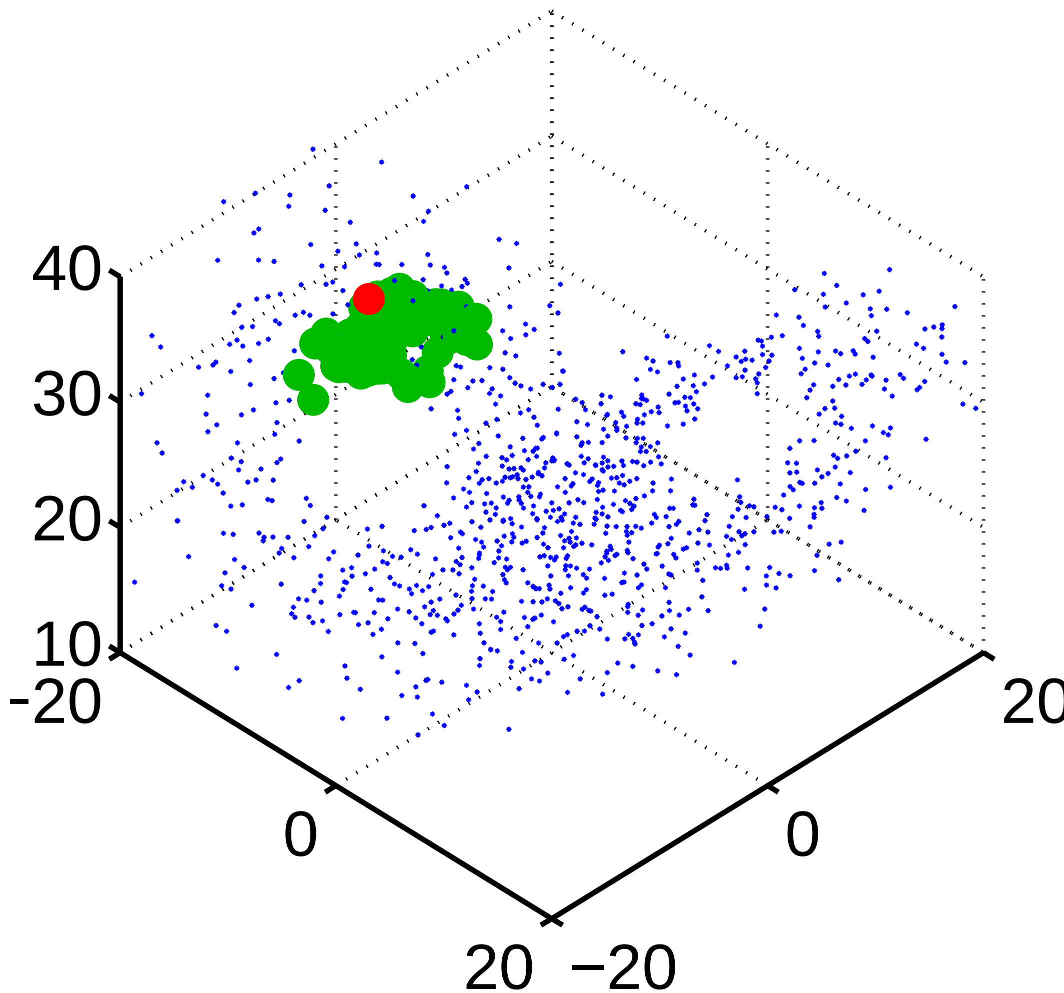
Theory: impact of observation increments on each state variable can be handled sequentially!

Applying an Ensemble Filter.

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



Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

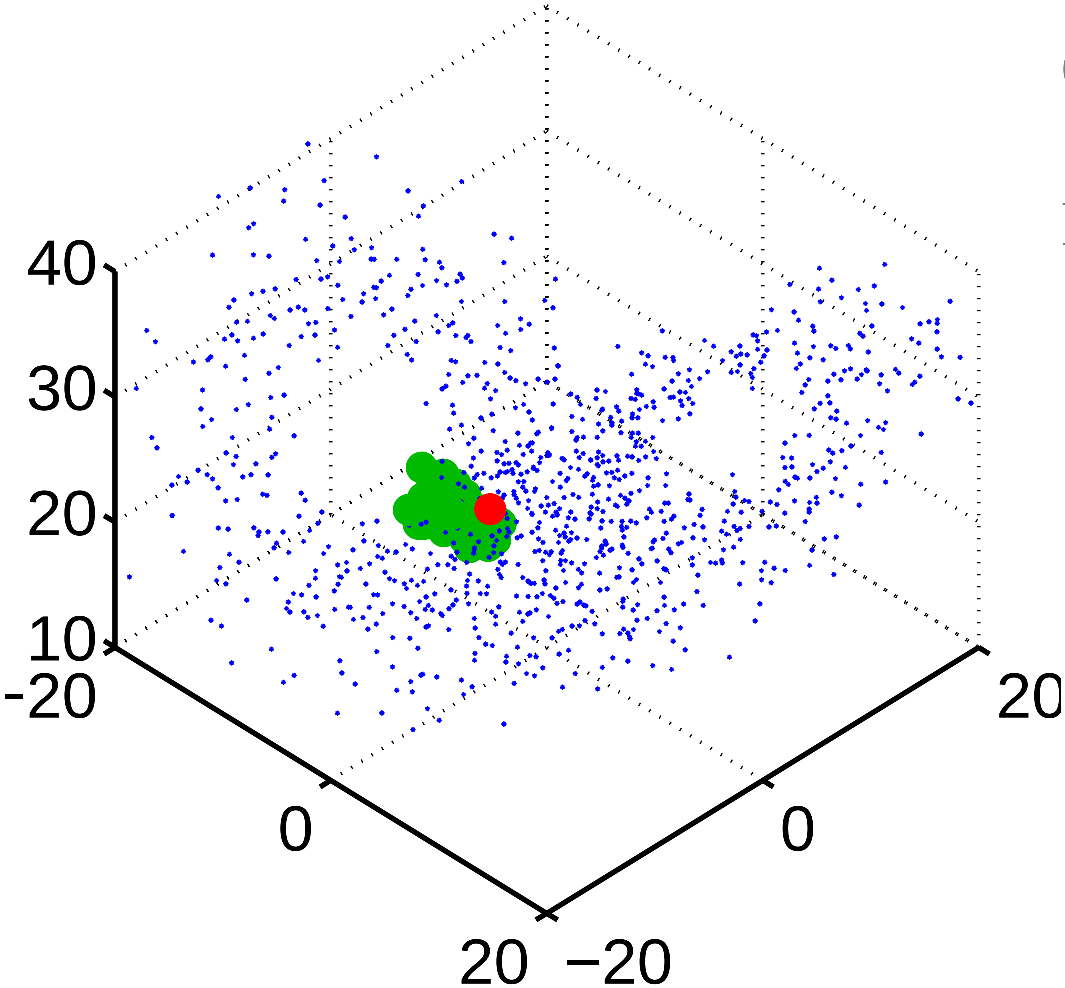
Prior ensemble in green.

Observing all three state variables.

Obs. error variance = 4.0.

4 20-member ensembles.

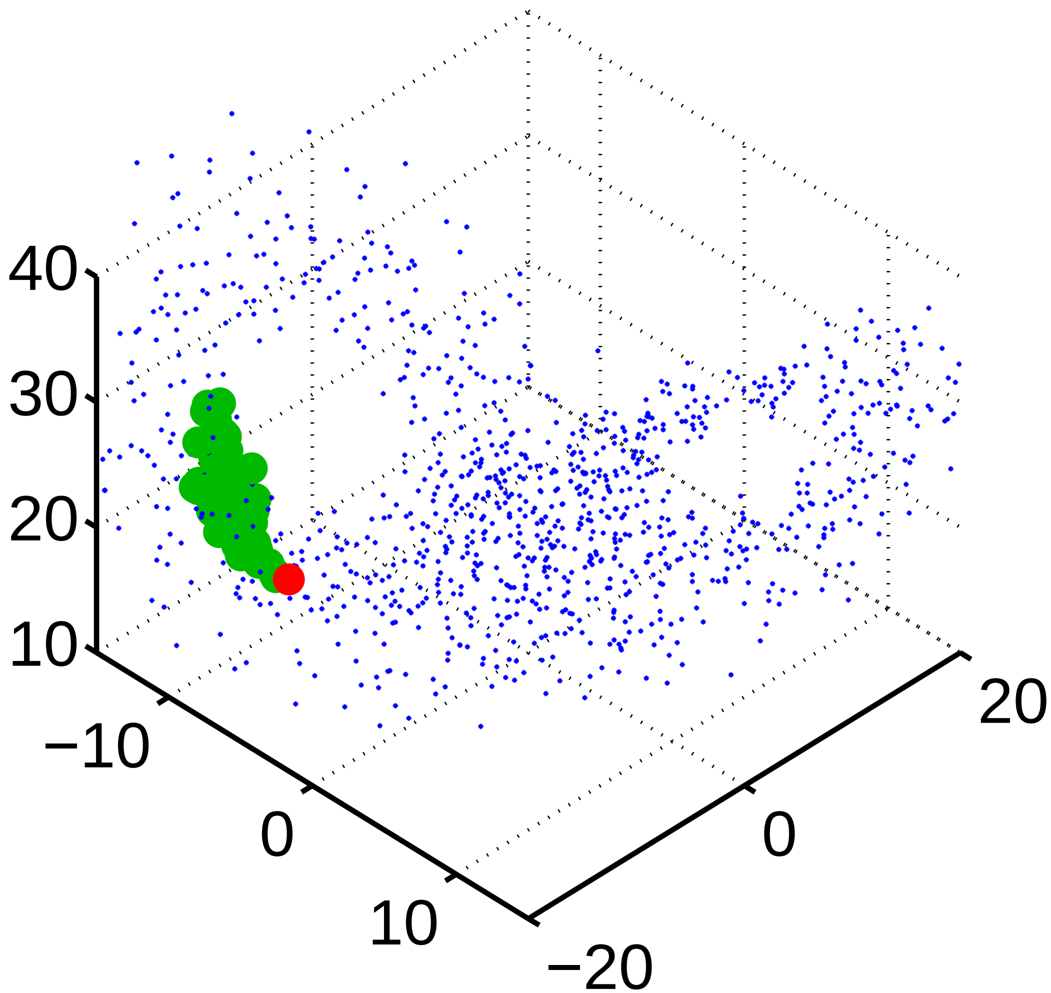
Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

Prior ensemble in green.

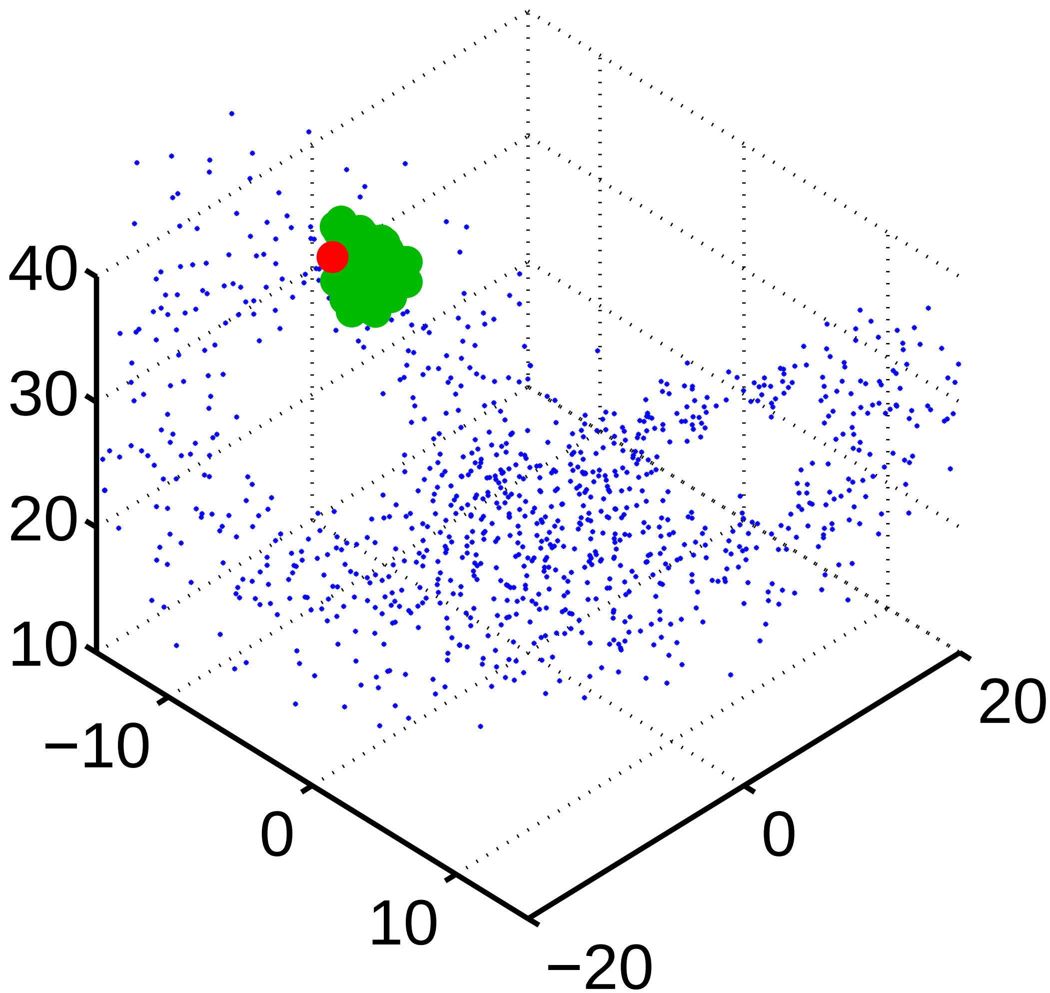
Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

Prior ensemble in green.

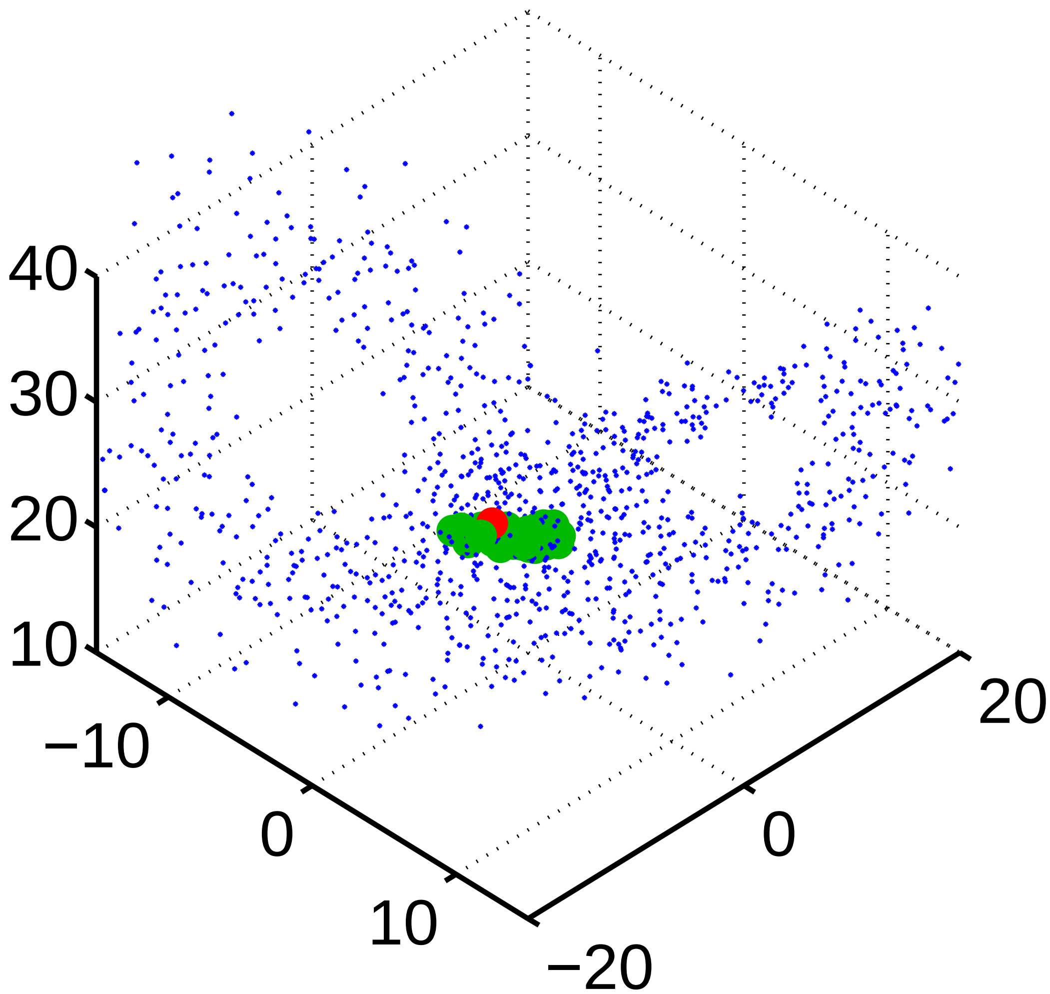
Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

Prior ensemble in green.

Some Fun with the Lorenz-63 3-Variable Chaotic Model.

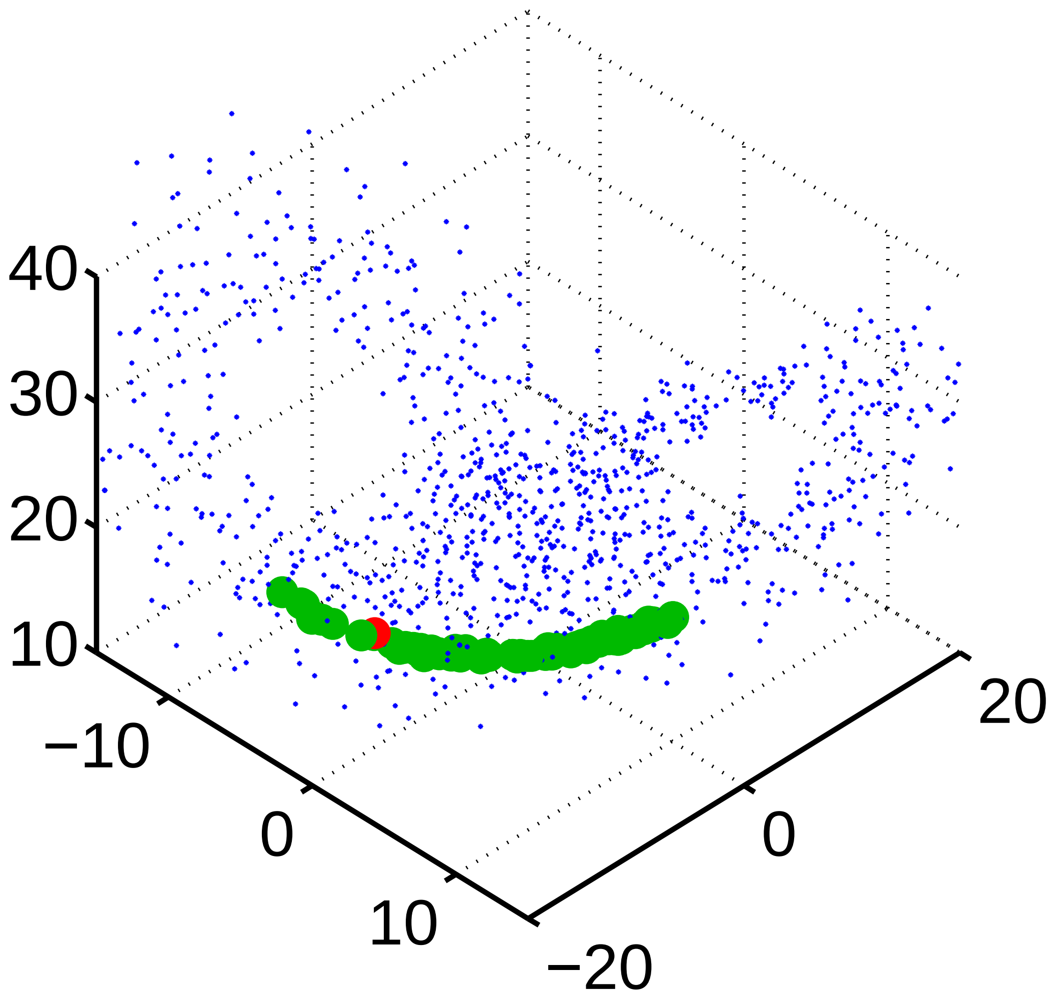


Observation in red.

Prior ensemble in green.

Ensemble is passing through unpredictable region.

Some Fun with the Lorenz-63 3-Variable Chaotic Model.

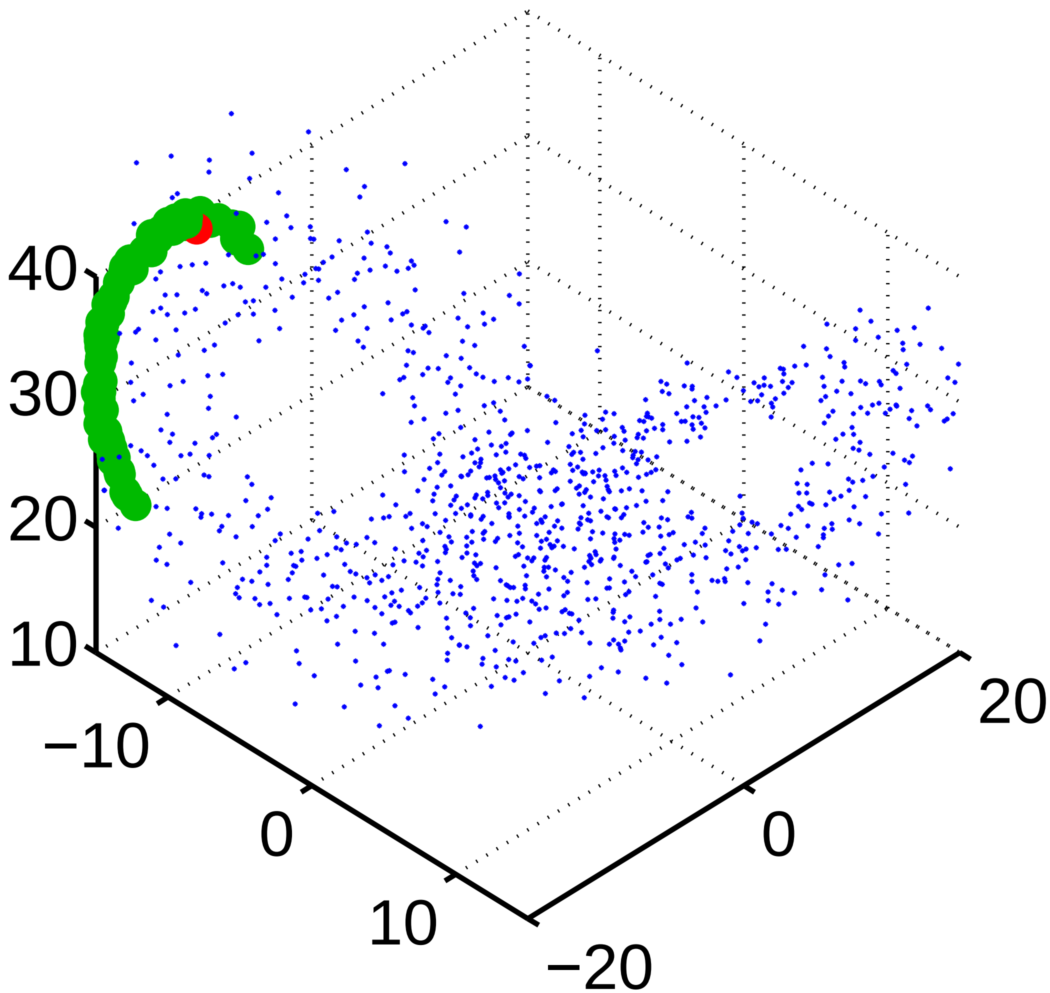


Observation in red.

Prior ensemble in green.

Part of ensemble heads for one lobe, the rest for the other.

Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

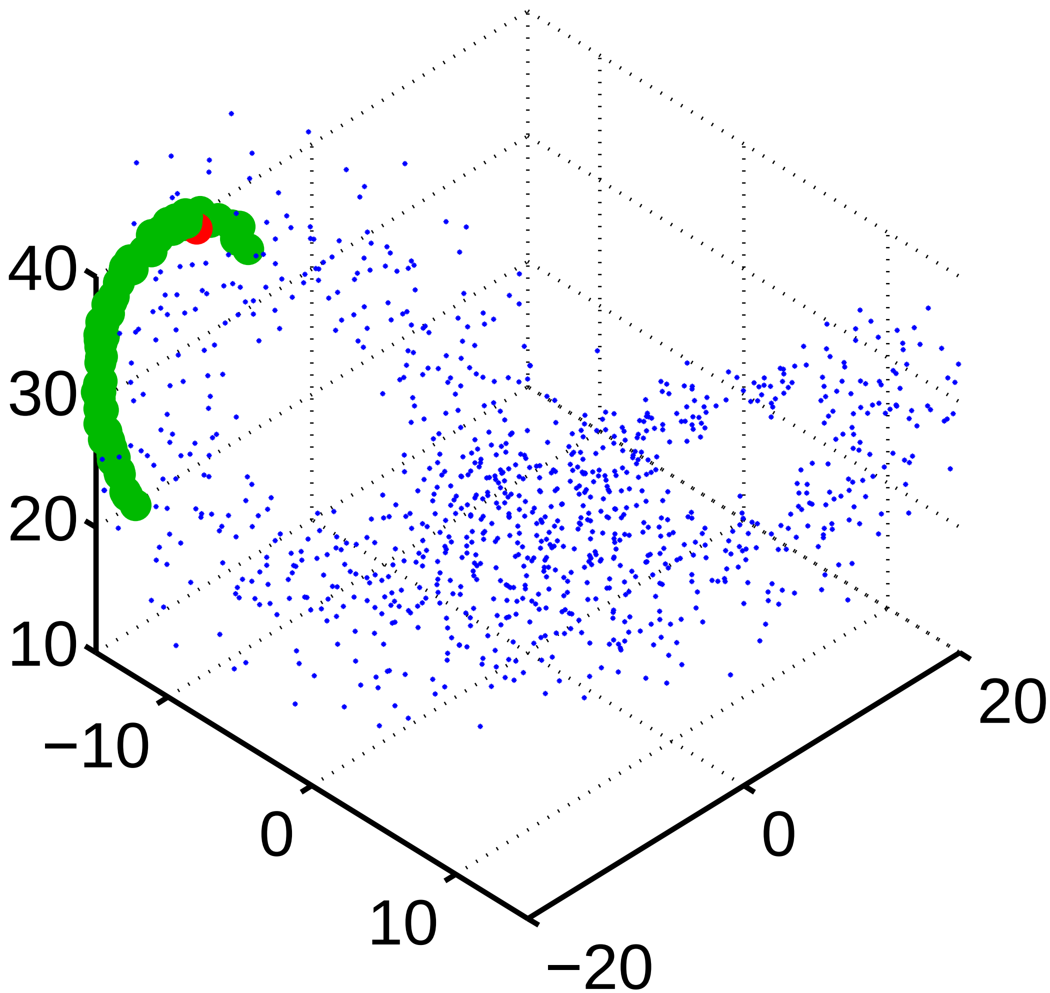
Prior ensemble in green.

The prior is not linear here.

Standard regression might be pretty bad.

.

Some Fun with the Lorenz-63 3-Variable Chaotic Model.



Observation in red.

Prior ensemble in green.

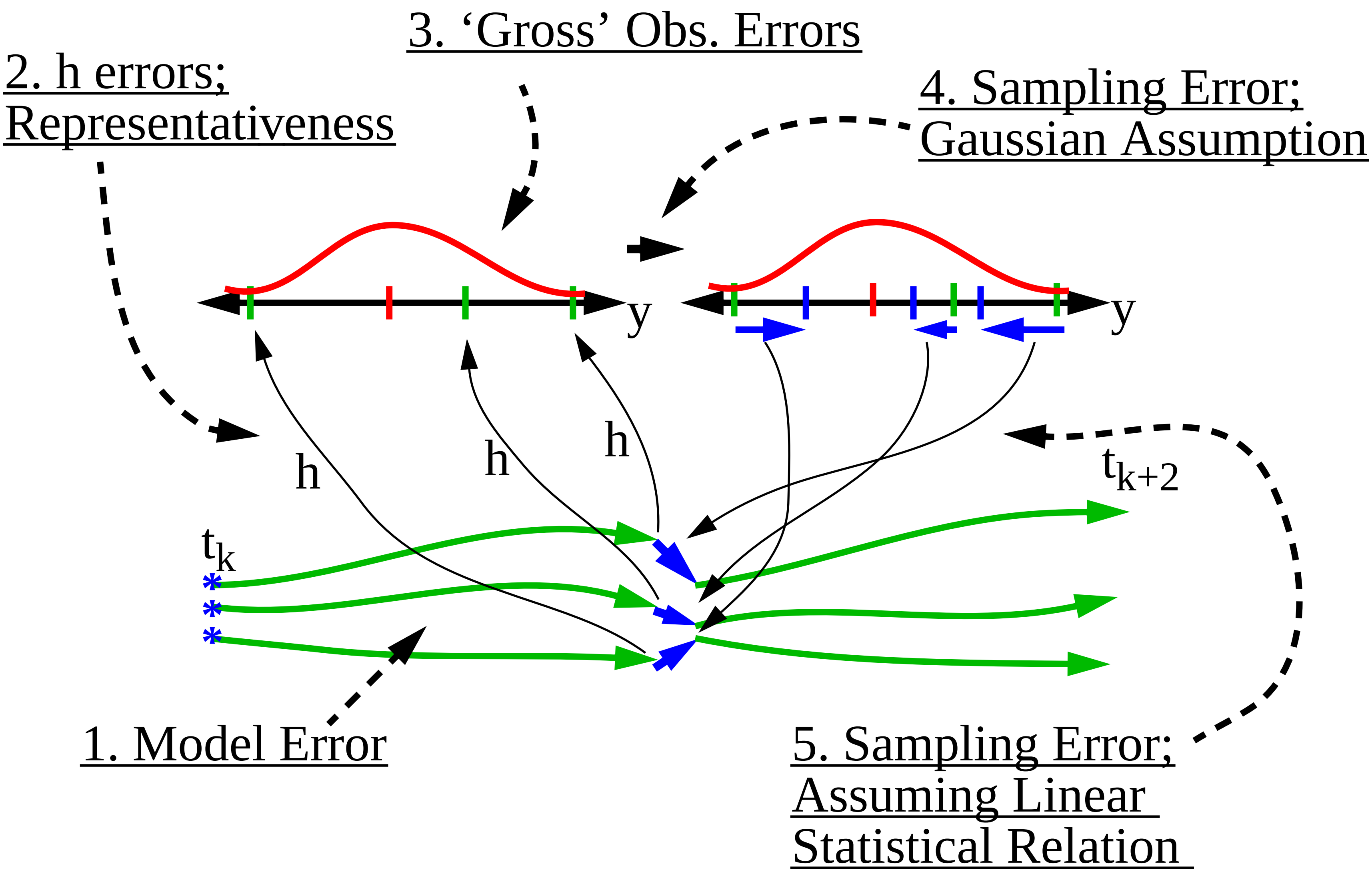
The prior is not linear here.

On the other hand...

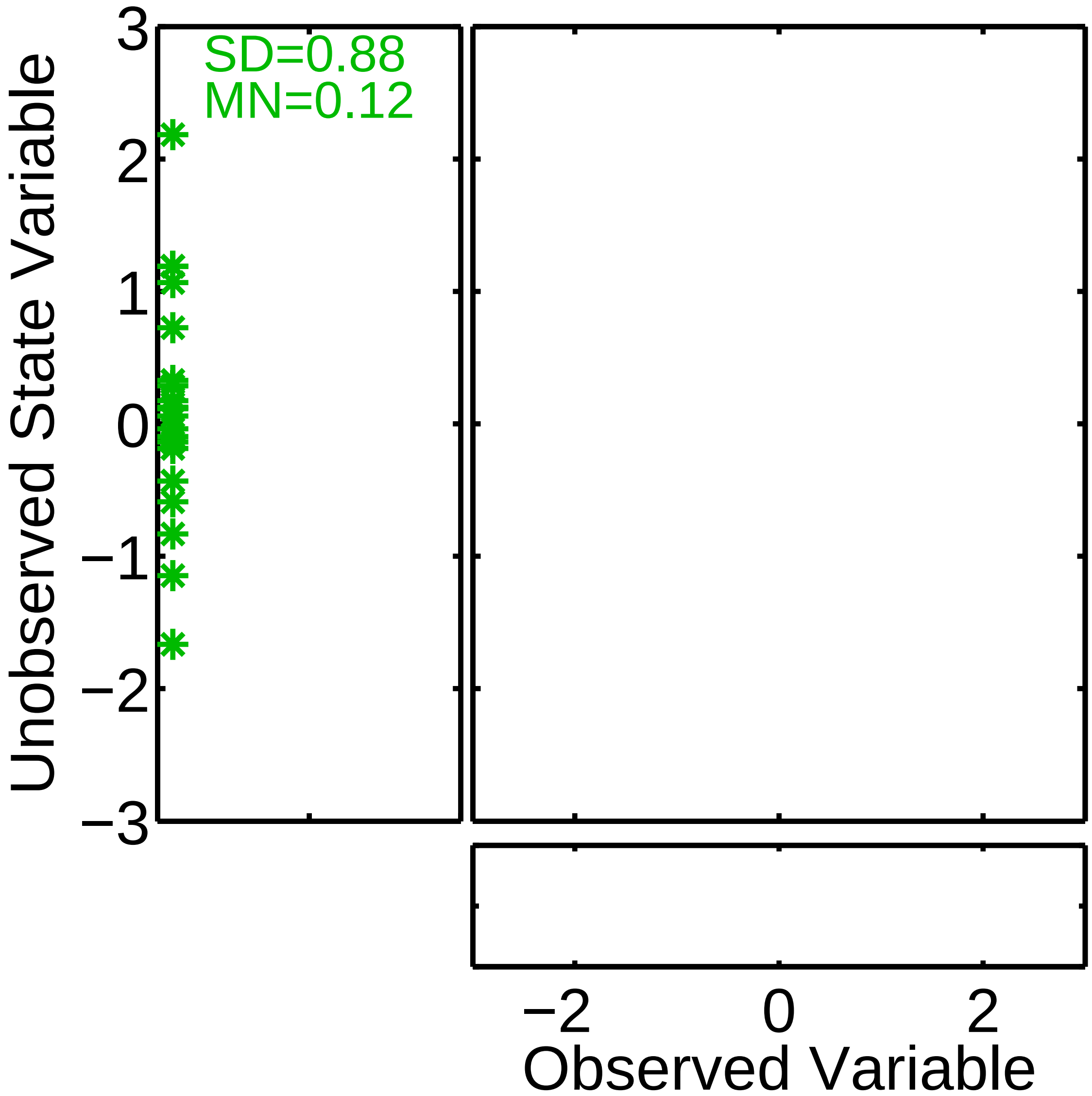
Hard to contrive examples this bad.

Behavior like this not apparent in real assimilations.

Basic filter implementation is subject to errors.



Regression sampling error and filter divergence



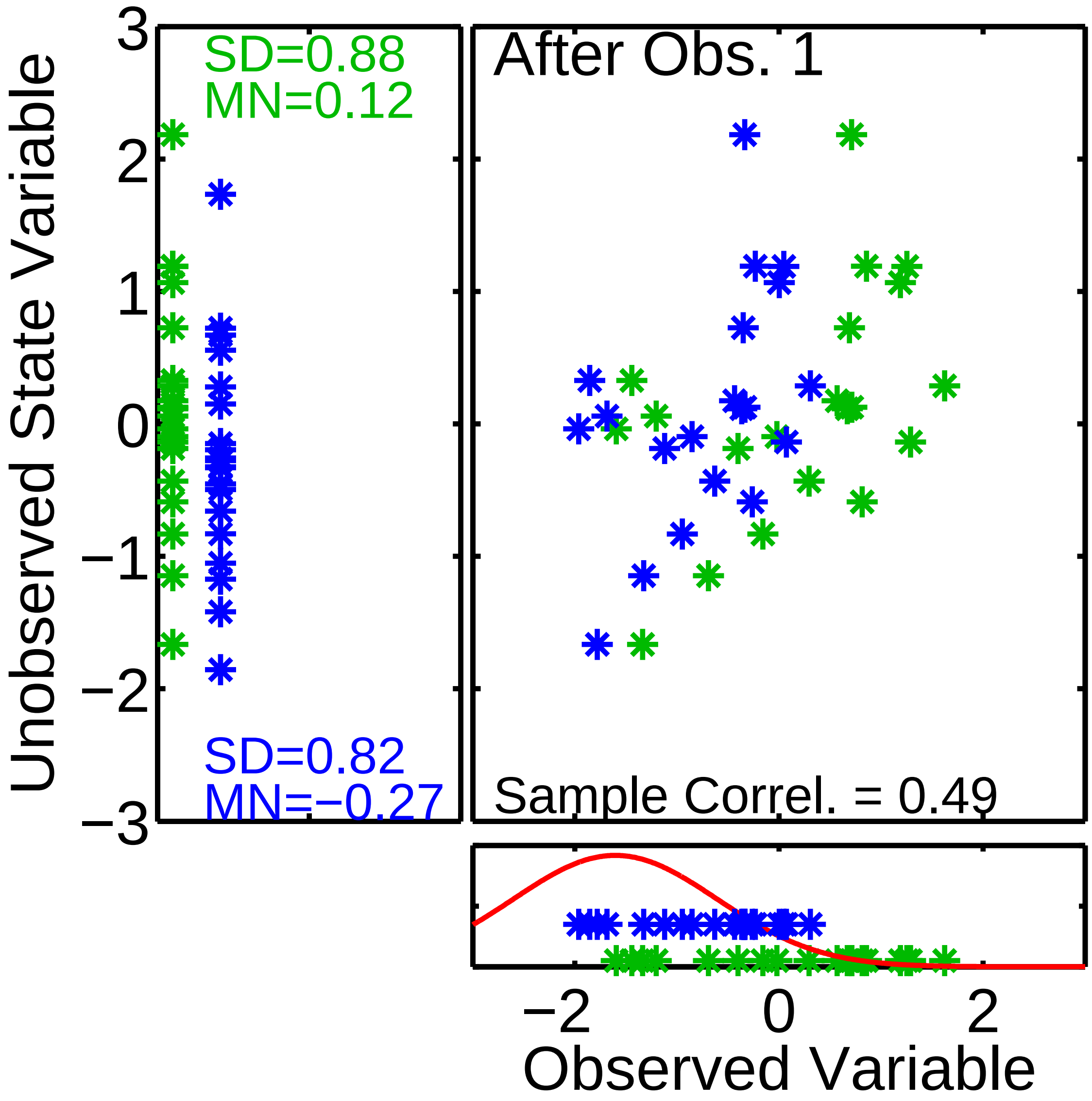
Suppose unobserved state variable is known to be unrelated to set of observed variables.

Unobserved variable should remain unchanged.

Let observations be of Antarctic wind velocity.

State variable is Park City temperature.

Regression sampling error and filter divergence

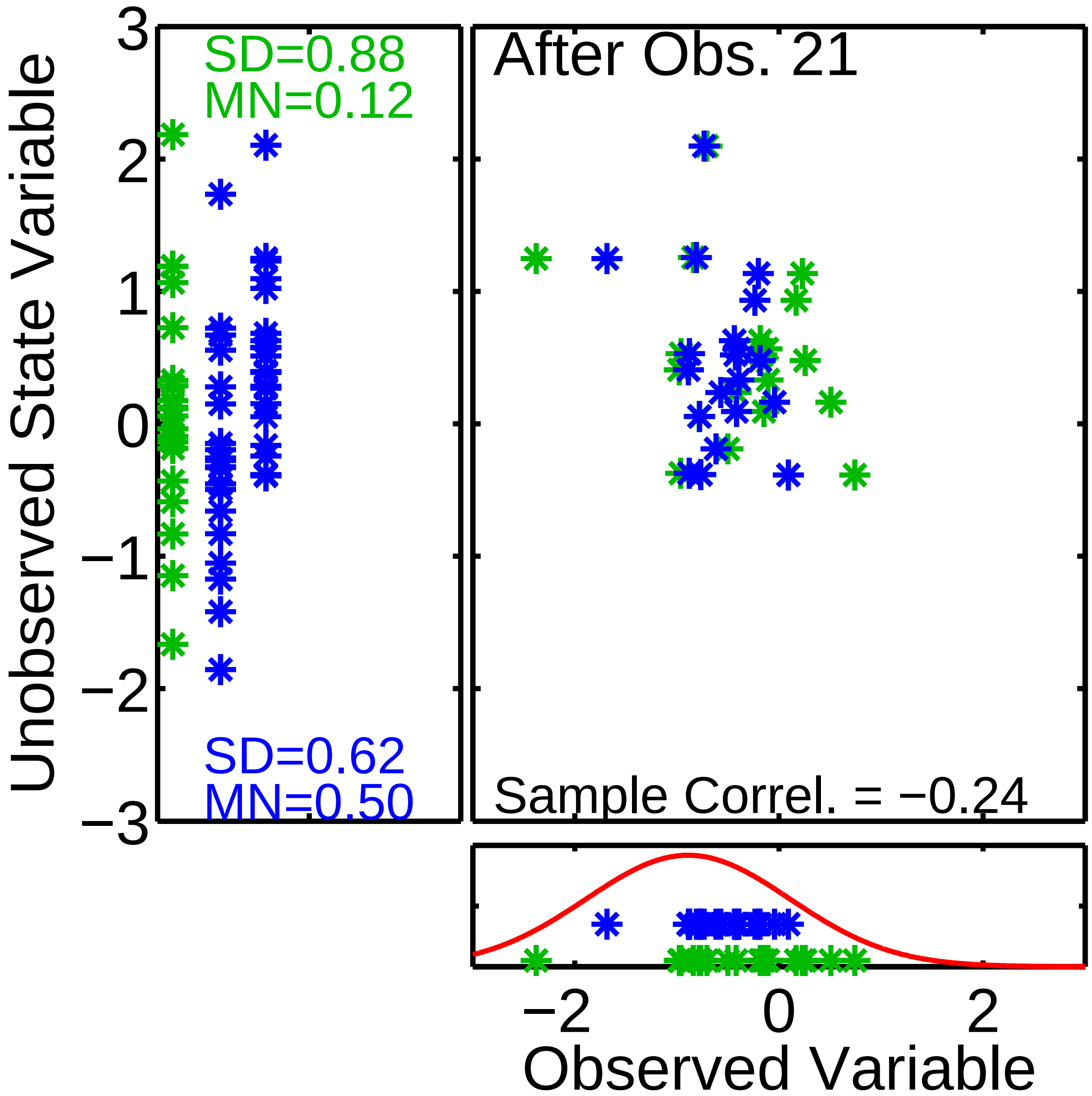


Suppose unobserved state variable is known to be unrelated to set of observed variables.

Finite samples from joint distribution will have non-zero correlation (expected |corr| = 0.19 for 20 samples).

After one observation, unobs. variable mean and S.D. change.

Regression sampling error and filter divergence

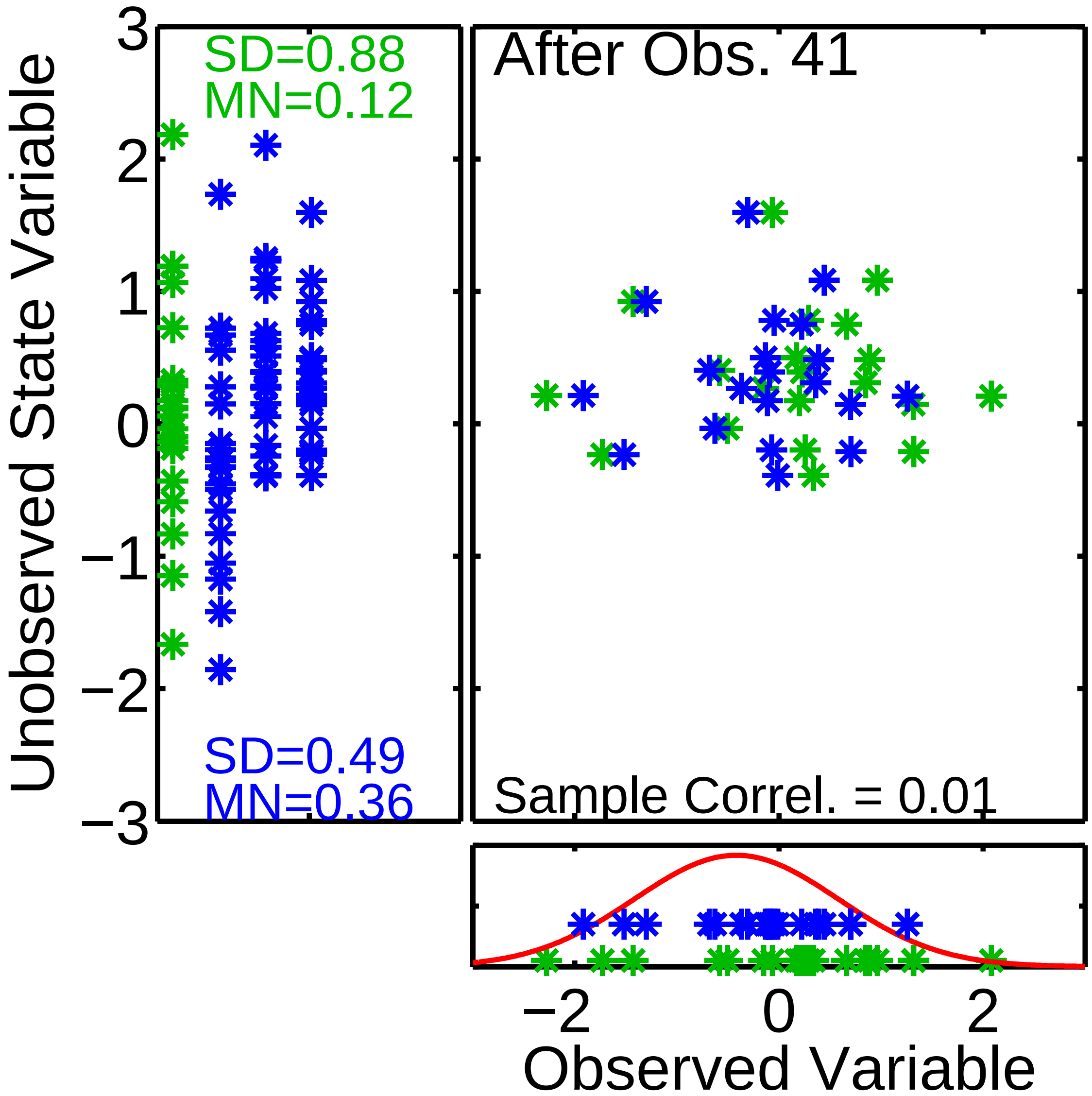


Suppose unobserved state variable is known to be unrelated to set of observed variables.

Unobserved variable should remain unchanged

Unobserved mean follows a random walk as more obs. are used.

Regression sampling error and filter divergence



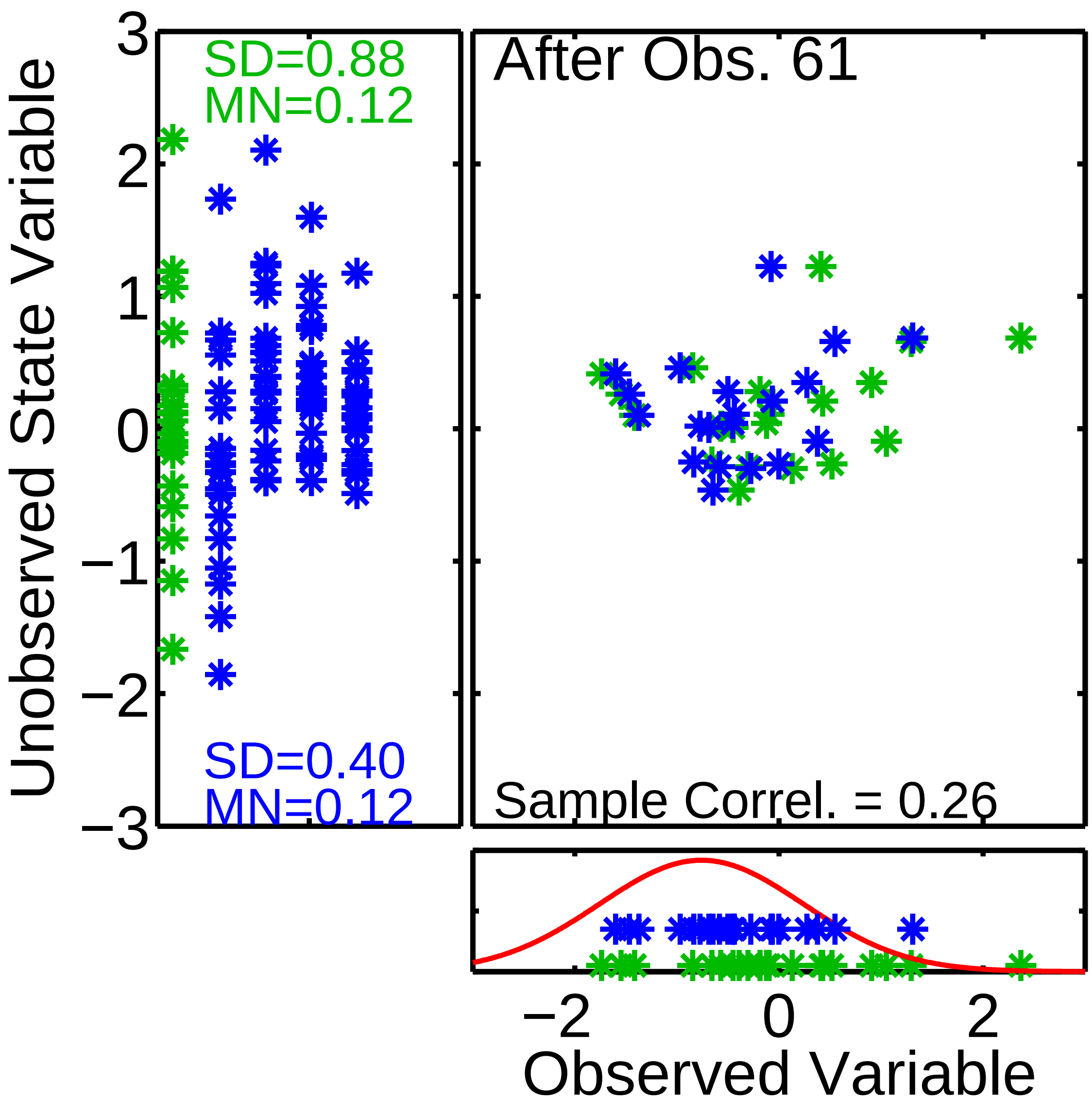
Suppose unobserved state variable is known to be unrelated to set of observed variables.

Unobserved variable should remain unchanged

Unobserved standard deviation is persistently decreased.

Expected change in |SD| is negative for any non-zero sample correlation!

Regression sampling error and filter divergence



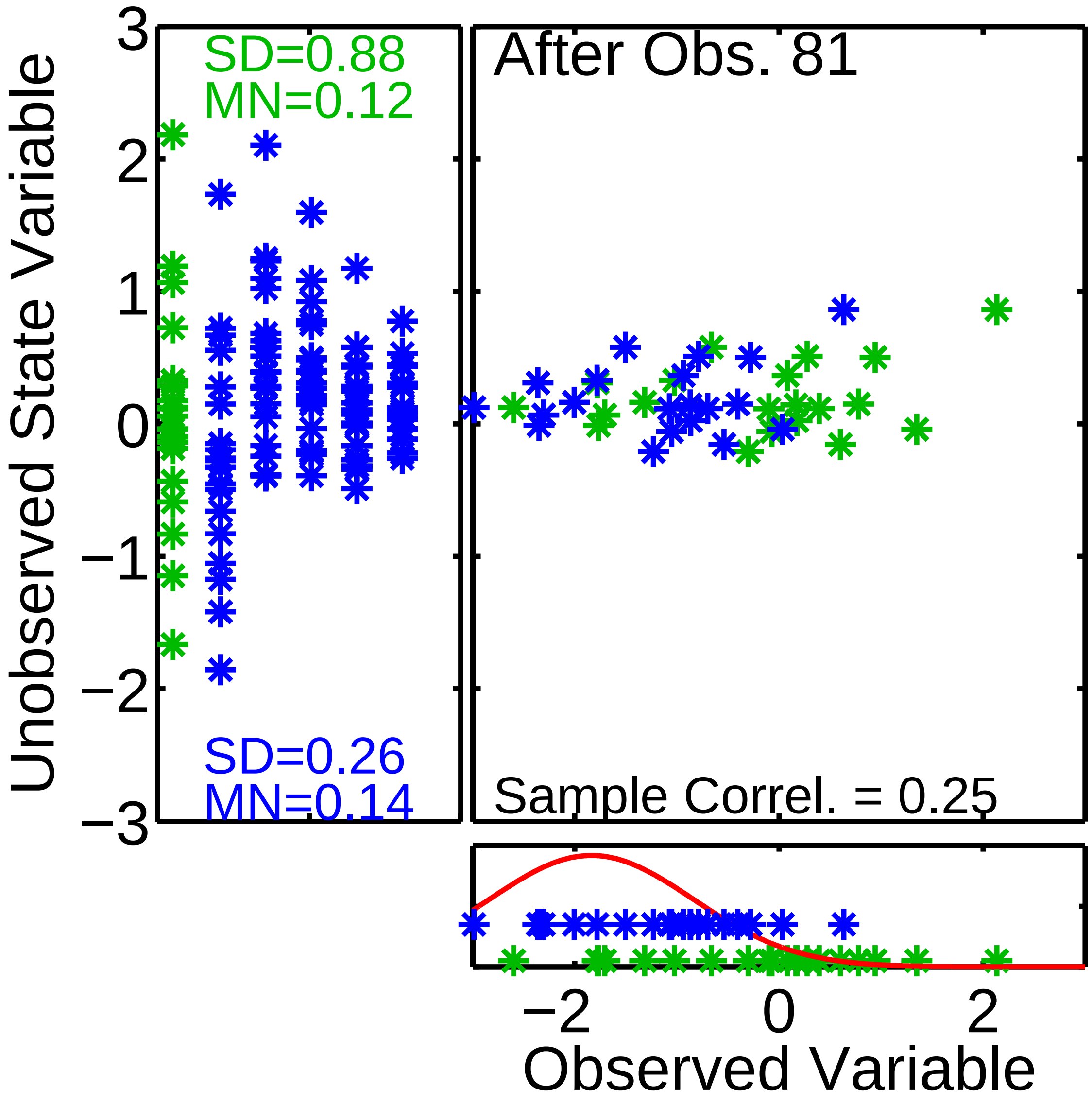
Suppose unobserved state variable is known to be unrelated to set of observed variables.

Unobserved variable should remain unchanged

Unobserved standard deviation is persistently decreased.

Expected change in |SD| is negative for any non-zero sample correlation!

Regression sampling error and filter divergence



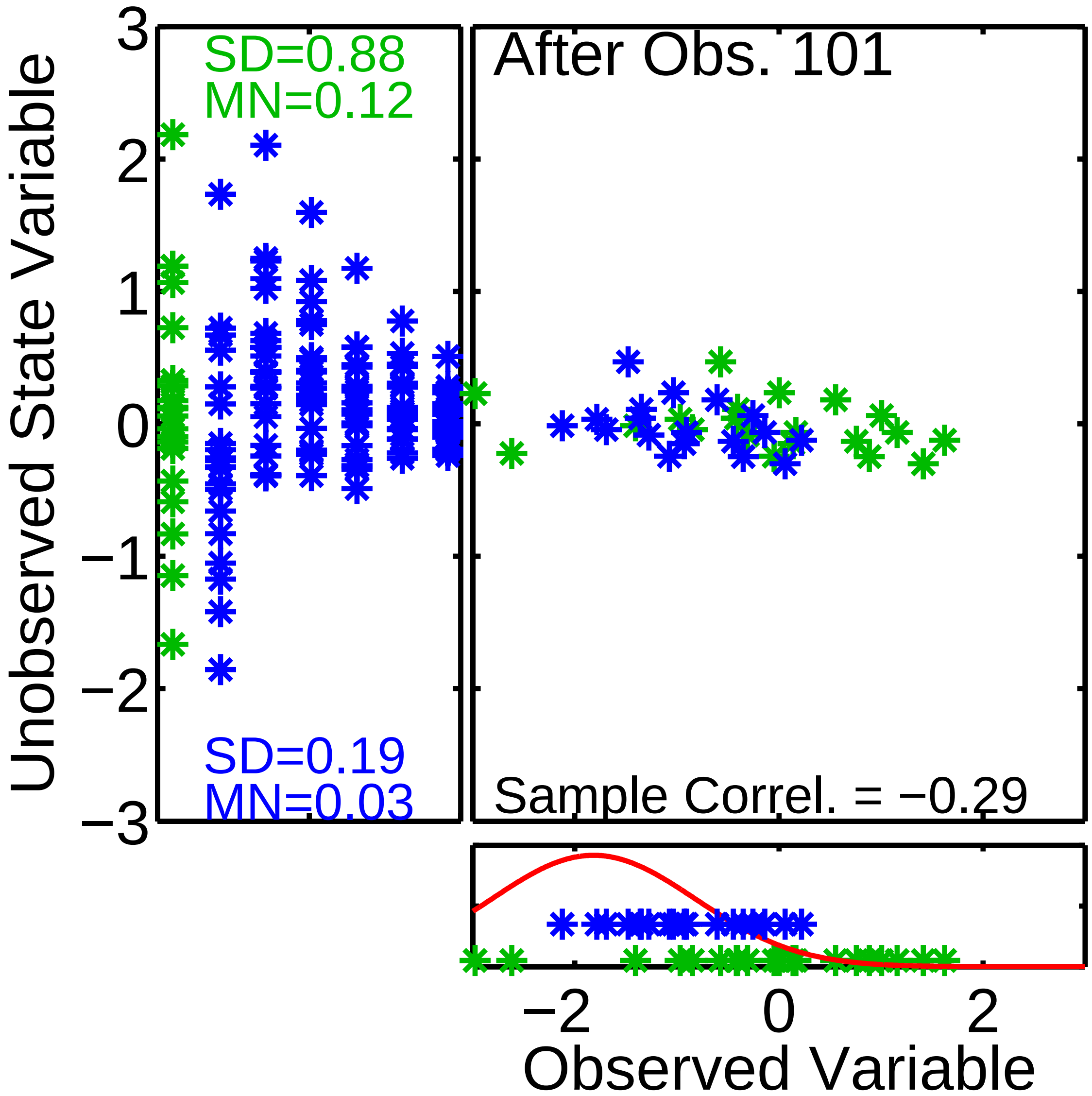
Suppose unobserved state variable is known to be unrelated to set of observed variables.

Unobserved variable should remain unchanged

Unobserved standard deviation is persistently decreased.

Expected change in |SD| is negative for any non-zero sample correlation!

Regression sampling error and filter divergence



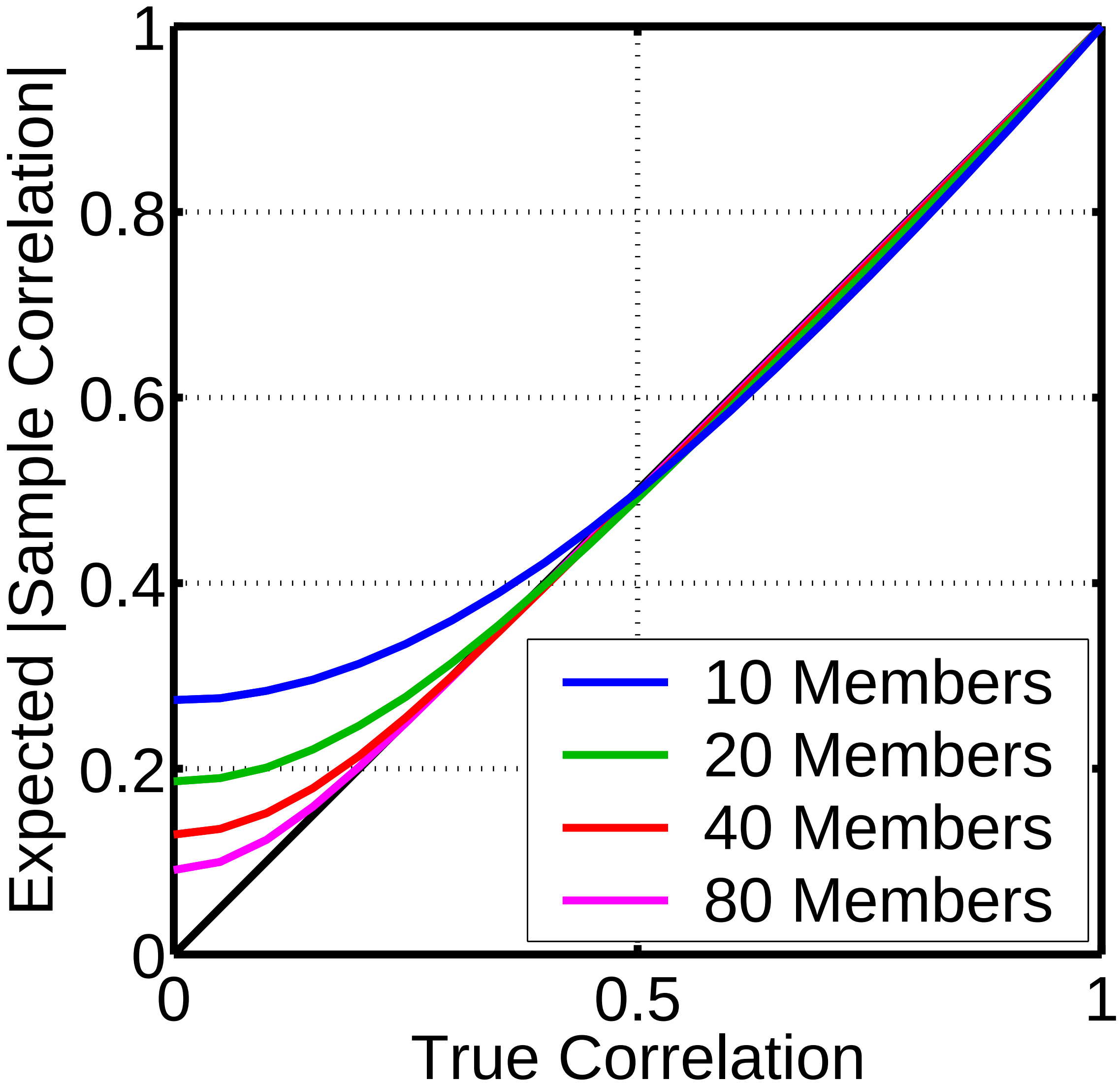
Suppose unobserved state variable is known to be unrelated to set of observed variables.

Estimates of unobs. become too confident

Give progressively less weight to any meaningful observations.

End result can be that meaningful obs. are essentially ignored.

Regression sampling error and filter divergence



Plot shows expected absolute value of sample correlation vs. true correlation.

Errors decrease with sample size and for large |real correlations|.

Ways to deal with regression sampling error:

1. Ignore it: if number of unrelated observations is small and there is some way of maintaining variance in priors.

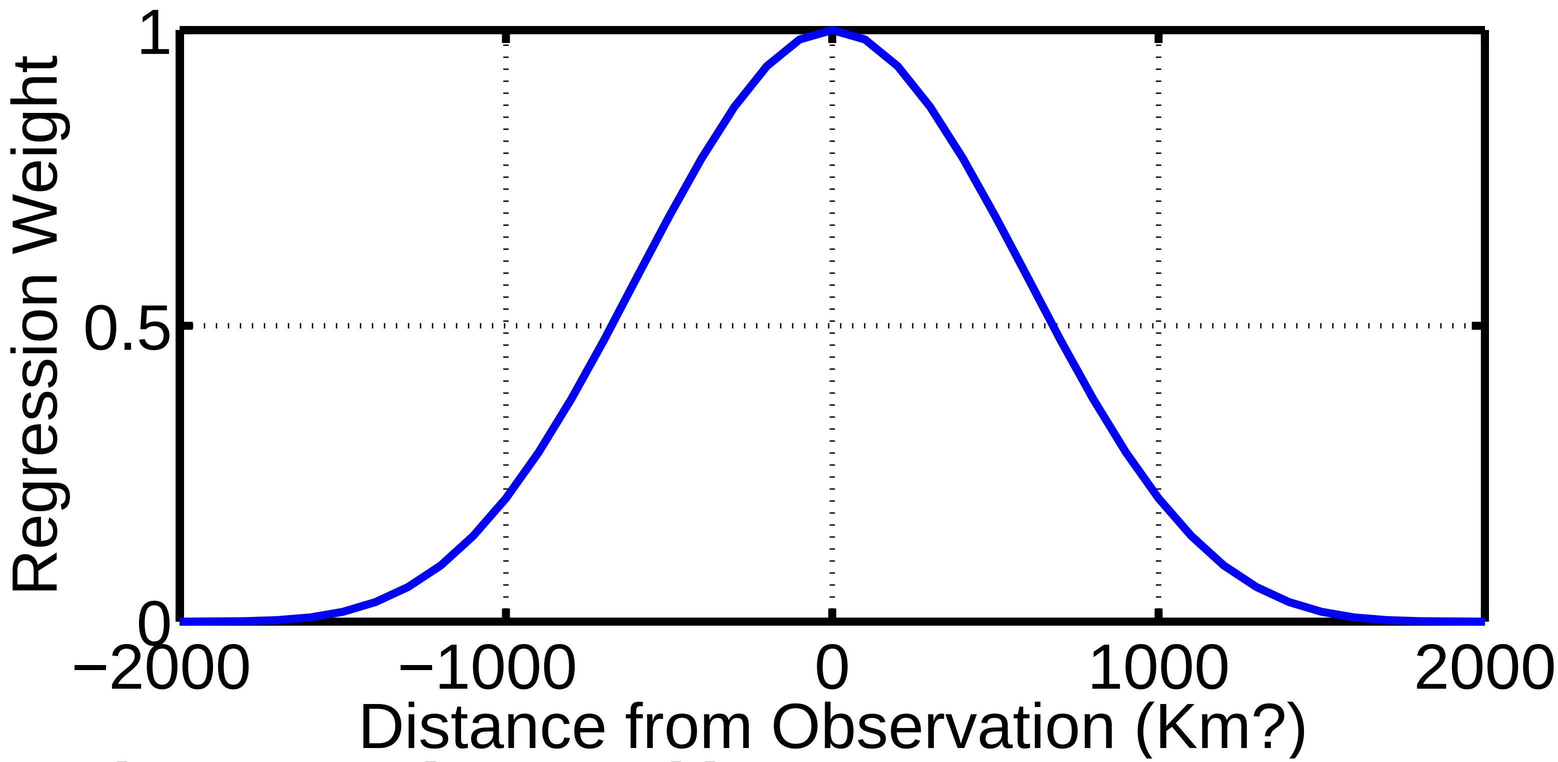
2. Use larger ensembles to limit sampling error.

3. Use additional *a priori* information about relation between observations and state variables.

4. Try to determine the amount of sampling error and correct for it.

<u>Ways to deal with regression sampling error:</u>

3. Use additional a priori information about relation between observations and state variables.



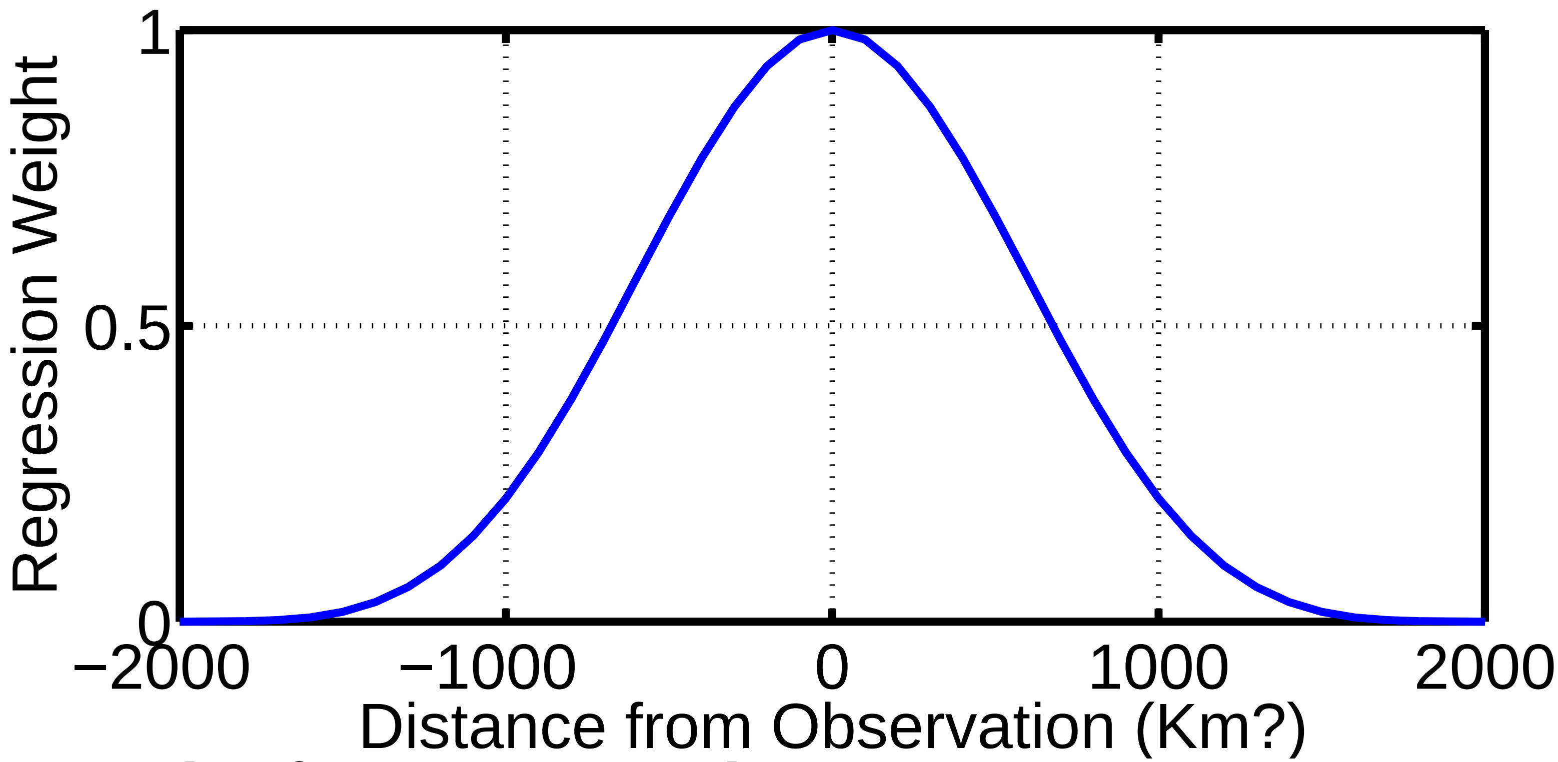
Atmospheric assimilation problems.

Weight regression as function of horizontal *distance* from observation.

Gaspari-Cohn: 5th order compactly supported polynomial.

Ways to deal with regression sampling error:

3. Use additional a priori information about relation between observations and state variables.



Can use other functions to weight regression.

Unclear what *distance* means for some obs./state variable pairs.

Referred to as LOCALIZATION.

Ways to deal with regression sampling error:

4. Try to determine the amount of sampling error and correct for it:

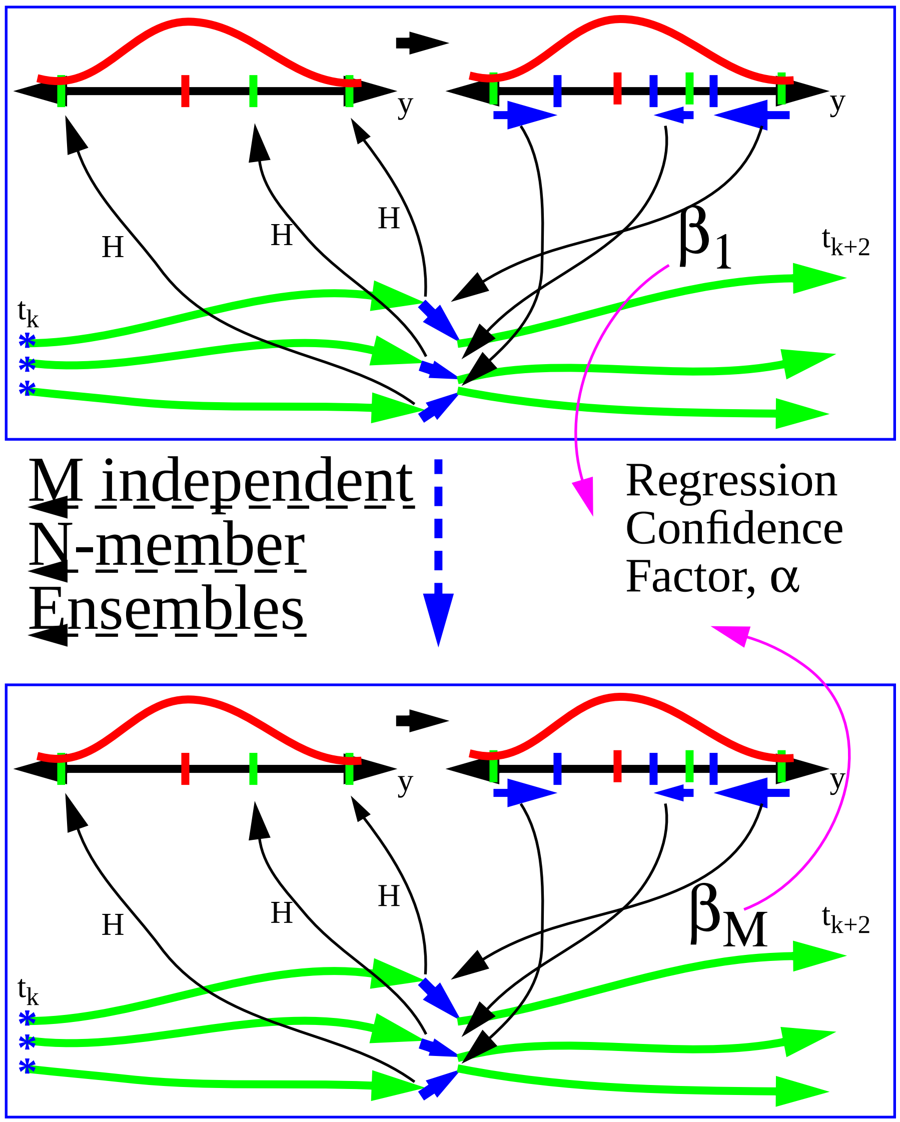
   A. Could weight regressions based on sample correlation.
      Limited success in tests.
      For small true correlations, can still get large sample correl.

   B. Do bootstrap with sample correlation to measure sampling error.
      Limited success.
      Repeatedly compute sample correlation with a sample removed.

   C. Use hierarchical Monte Carlo.
      Have a ‘sample’ of samples.
      Compute expected error in regression coefficients and weight.

<u>Ways to deal with regression sampling error:</u>

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



M groups of N-member ensembles.

Compute obs. increments for each group.

For given obs. / state pair:

1. Have M samples of regression coefficient, β.
2. Uncertainty in β implies state variable increments should be reduced.
3. Compute regression confidence factor, α.

Ways to deal with regression sampling error:

4C. Use hierarchical Monte Carlo: ensemble of ensembles.

Split ensemble into M independent groups.

For instance, 80 ensemble members becomes 4 groups of 20.

With M groups get M estimates of regression coefficient, β_i.

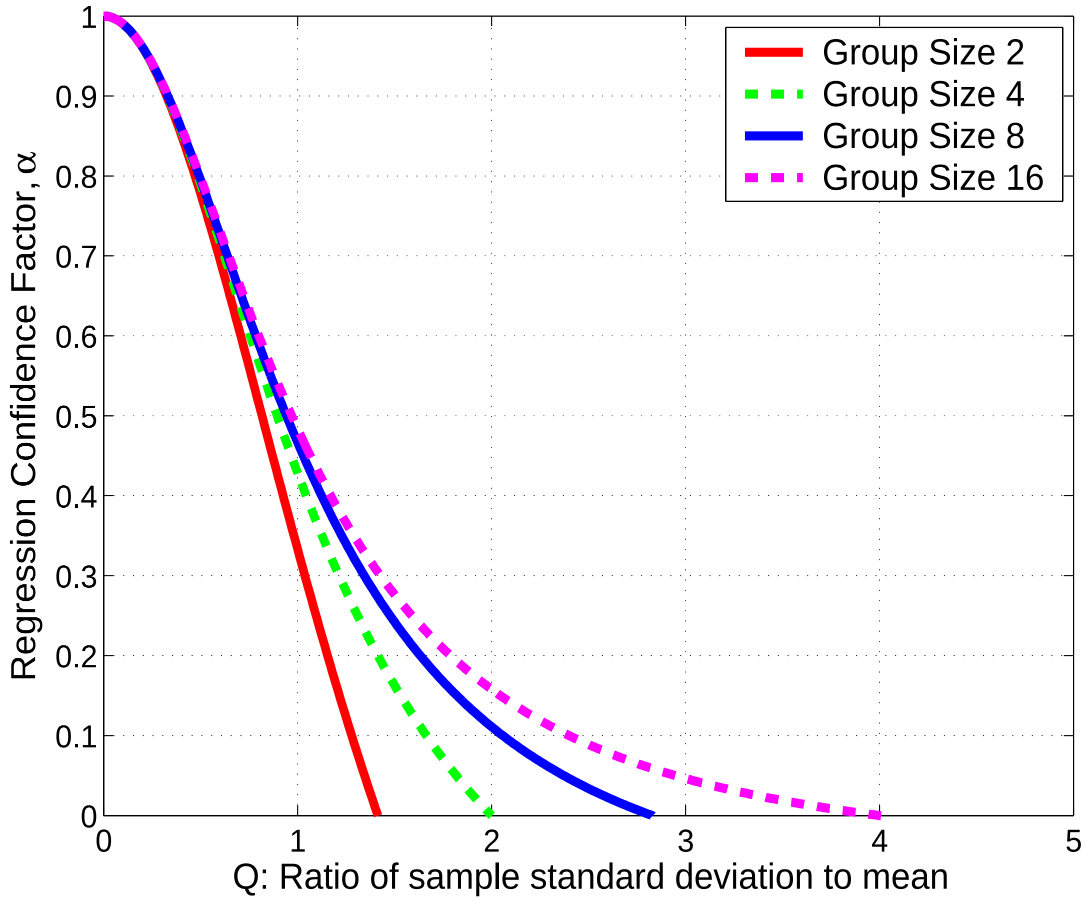
Find regression confidence factor α (weight) that minimizes:

$$\sqrt{\sum_{j=1}^{M} \sum_{i=1, i \neq j}^{M} [\alpha\beta_i - \beta_j]^2}$$

Minimizes RMS error in the regression (and state increments).

Ways to deal with regression sampling error:

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



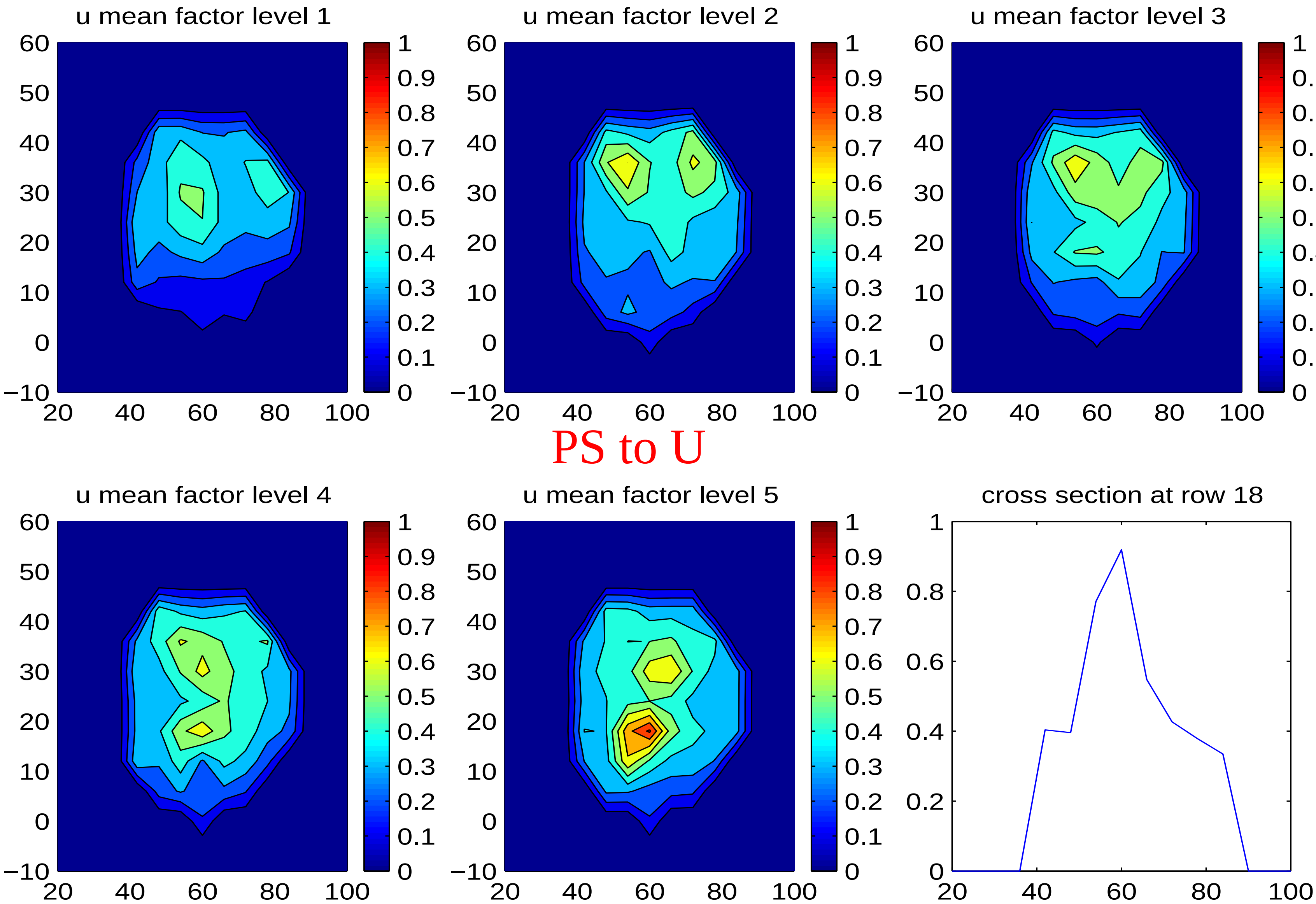
Weight regression by α.

If one has repeated observations, can generate sample mean or median statistics for α.

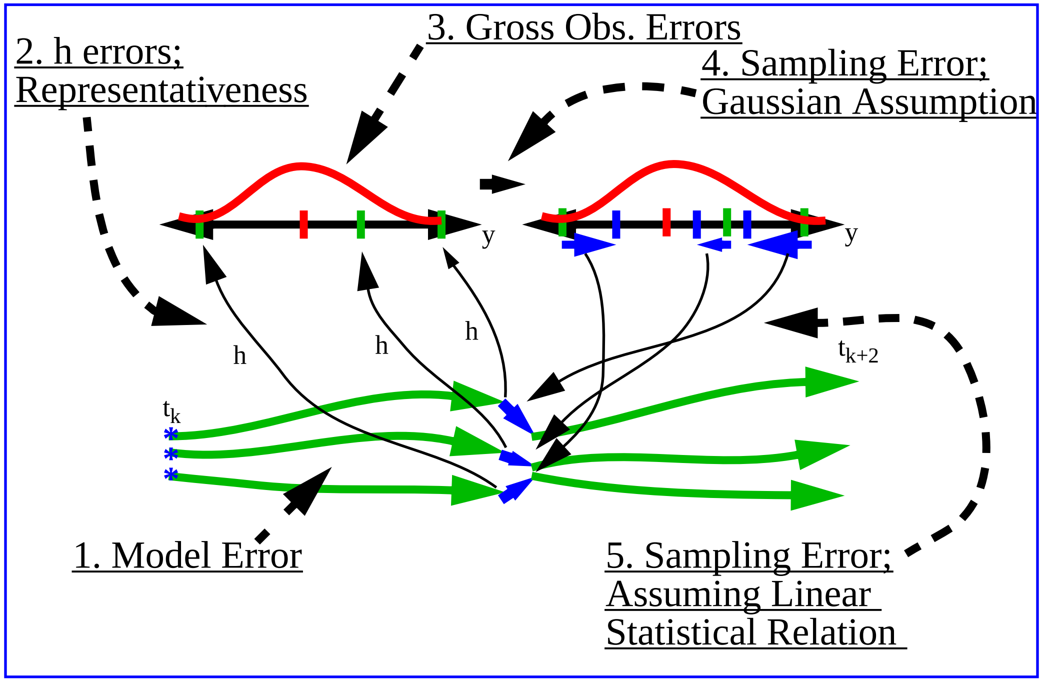
Mean α can be used in subsequent assimilations as a localization.

α is function of M and $Q = \Sigma_\beta / \bar{\beta}$ (sample SD /sample mean regression)

Localization in GCM can be very complex.   Surface Pressure Obs. at 20N, 60E



Dealing With Ensemble Filter Errors



Fix 1, 2, 3 independently HARD but ongoing.

Often, ensemble filters...

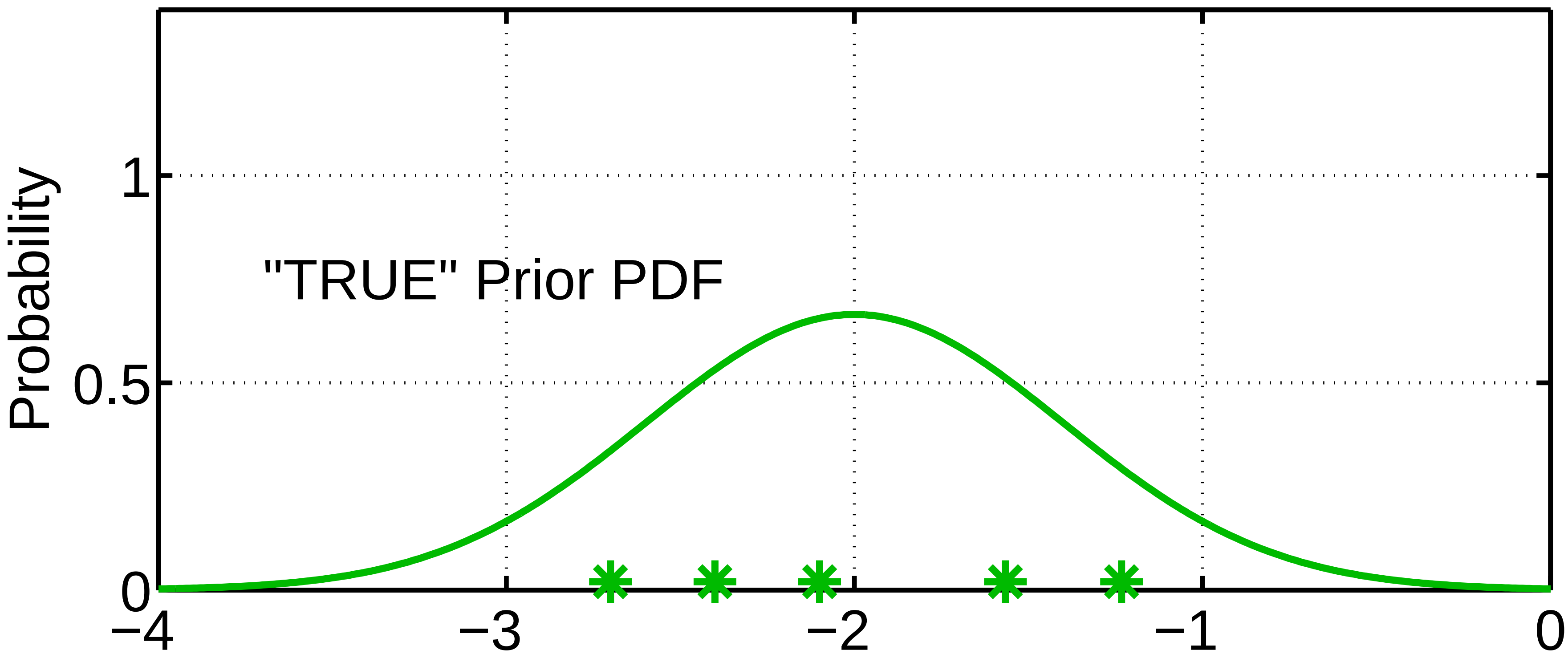
1-4: Covariance inflation, Increase prior uncertainty to give obs more impact.

5. 'Localization': only let obs. impact a set of 'nearby' state variables.

Often smoothly decrease impact to 0 as function of distance.

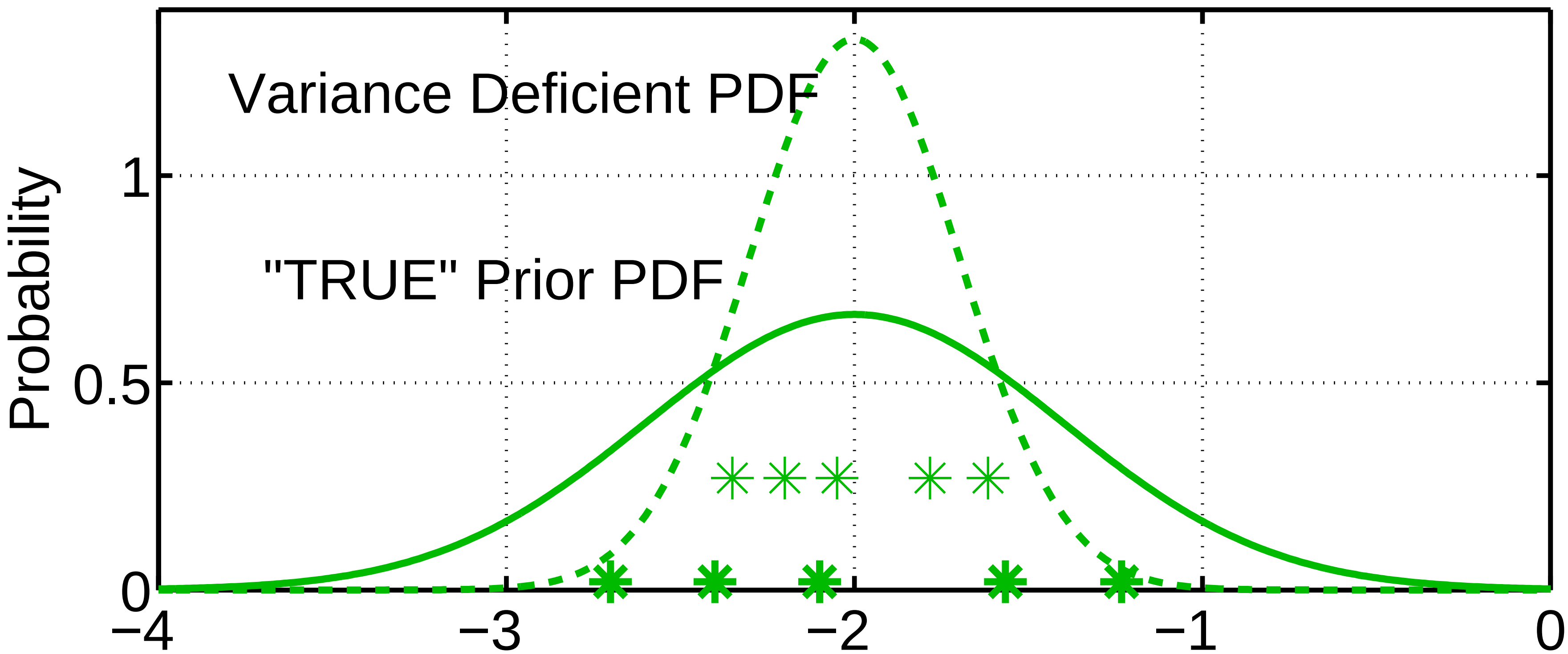
Model/Filter Error; Filter Divergence and Variance Inflation

1. History of observations and physical system => ‘true’ distribution.



Model/Filter Error; Filter Divergence and Variance Inflation

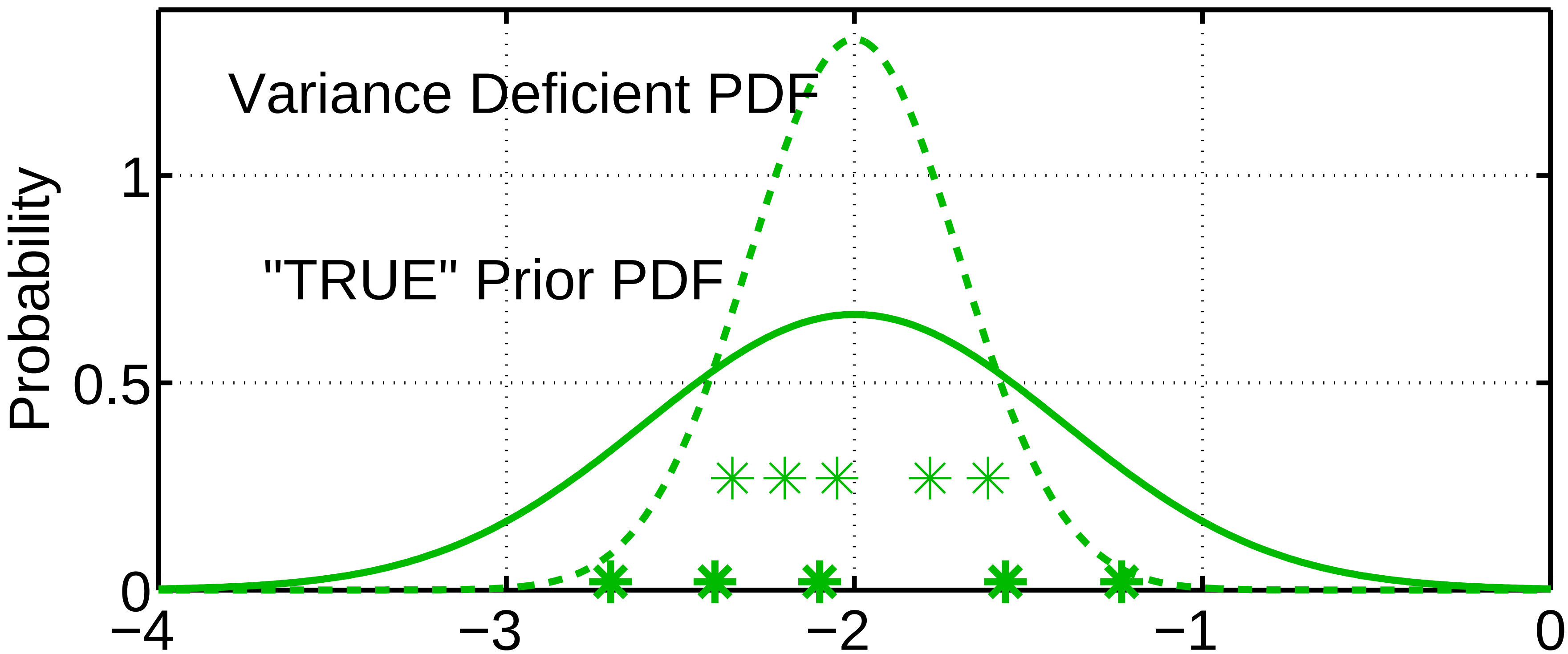
1. History of observations and physical system => ‘true’ distribution.
2. Sampling error, some model errors lead to insufficient prior variance.



3. Can lead to ‘filter divergence’: prior is too confident, obs. ignored

Model/Filter Error; Filter Divergence and Variance Inflation

1. History of observations and physical system => ‘true’ distribution.
2. Sampling error, some model errors lead to insufficient prior variance.

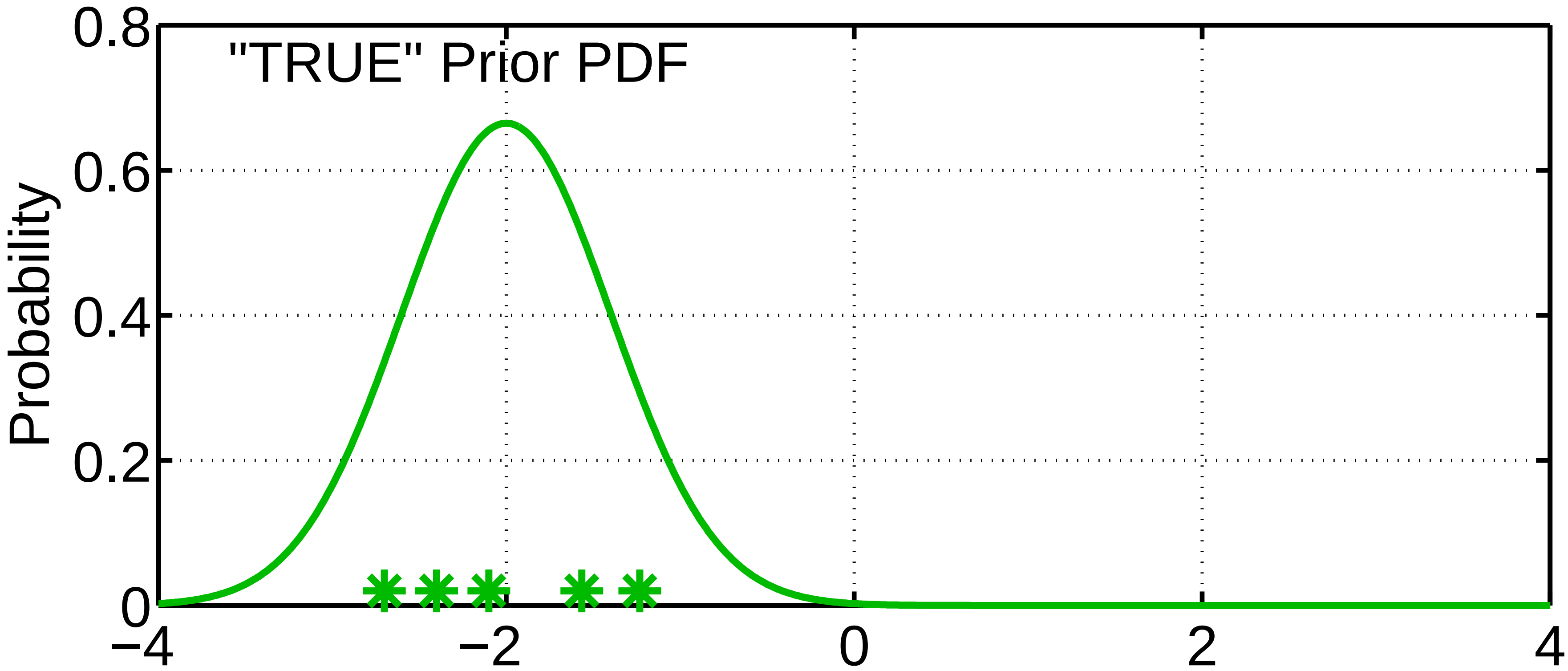


3. Naive solution is Variance inflation: just increase spread of prior
4. For ensemble member i, $inflate(x_i) = \sqrt{\lambda}(x_i - \bar{x}) + \bar{x}$.

Model/Filter Error; Filter Divergence and Variance Inflation

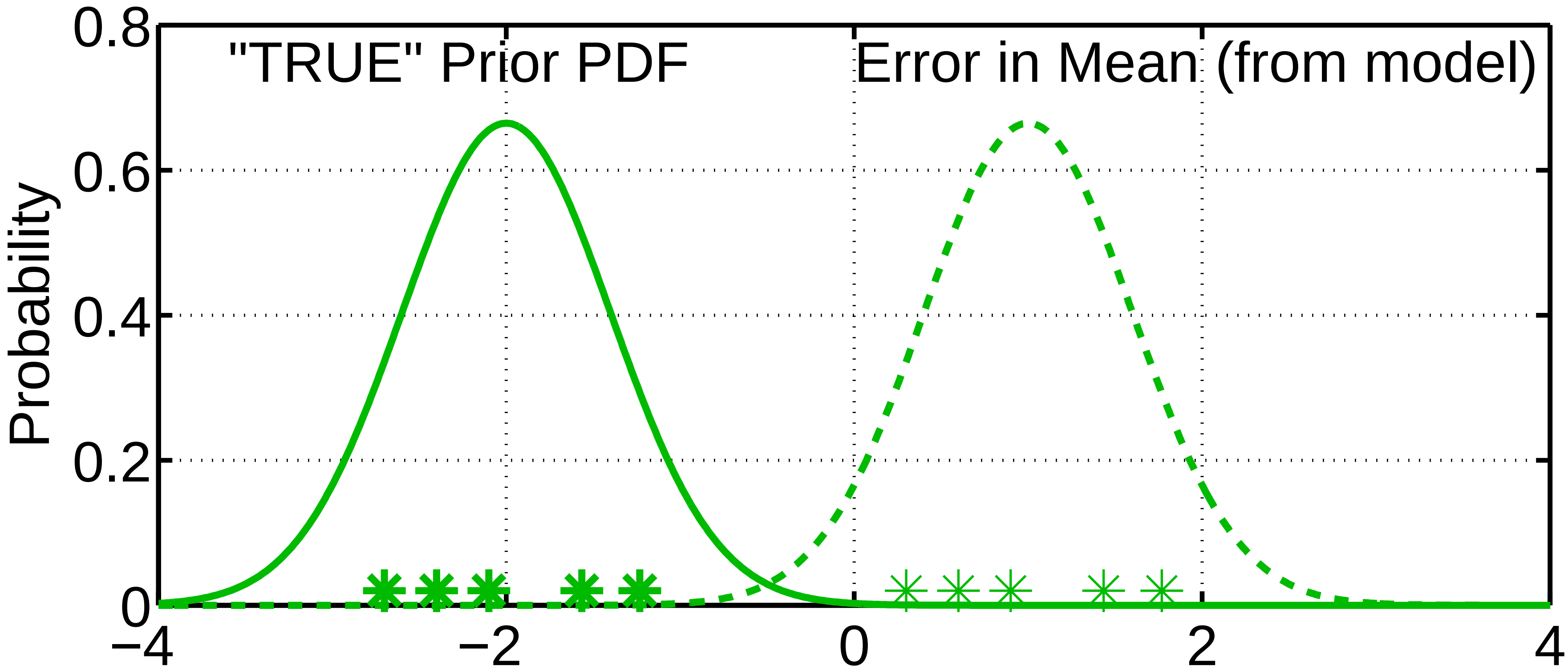
1. History of observations and physical system => ‘true’ distribution.

.



Model/Filter Error; Filter Divergence and Variance Inflation

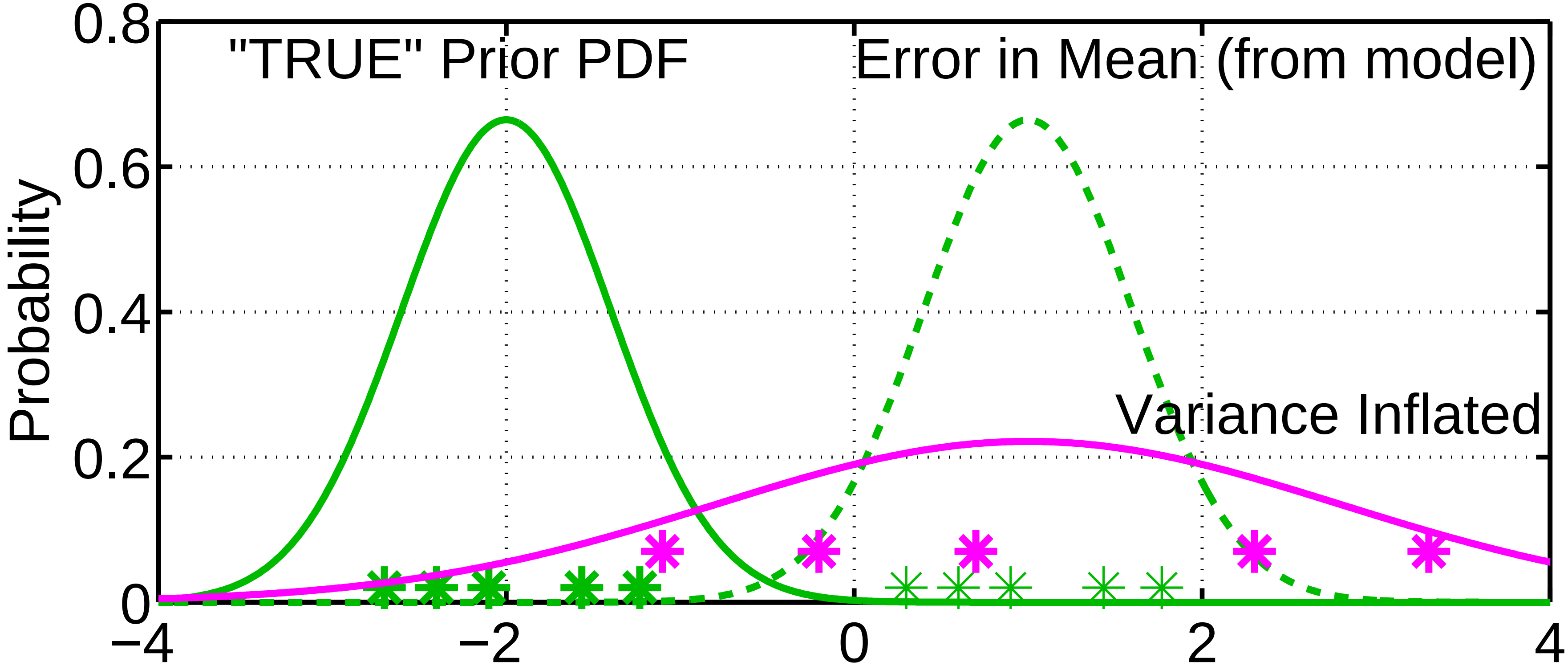
1. History of observations and physical system => ‘true’ distribution.
2. Most model errors also lead to erroneous shift in entire distribution.



3. Again, prior can be viewed as being TOO CERTAIN

Model/Filter Error; Filter Divergence and Variance Inflation

1. History of observations and physical system => ‘true’ distribution.
2. Most model errors also lead to erroneous shift in entire distribution.



3. Again, prior can be viewed as being TOO CERTAIN
4. Inflating can ameliorate this
5. Obviously, if we knew E(error), we’d correct for it directly

Physical Space Variance Inflation

Inflate all state variables by same amount before assimilation

Capabilities:

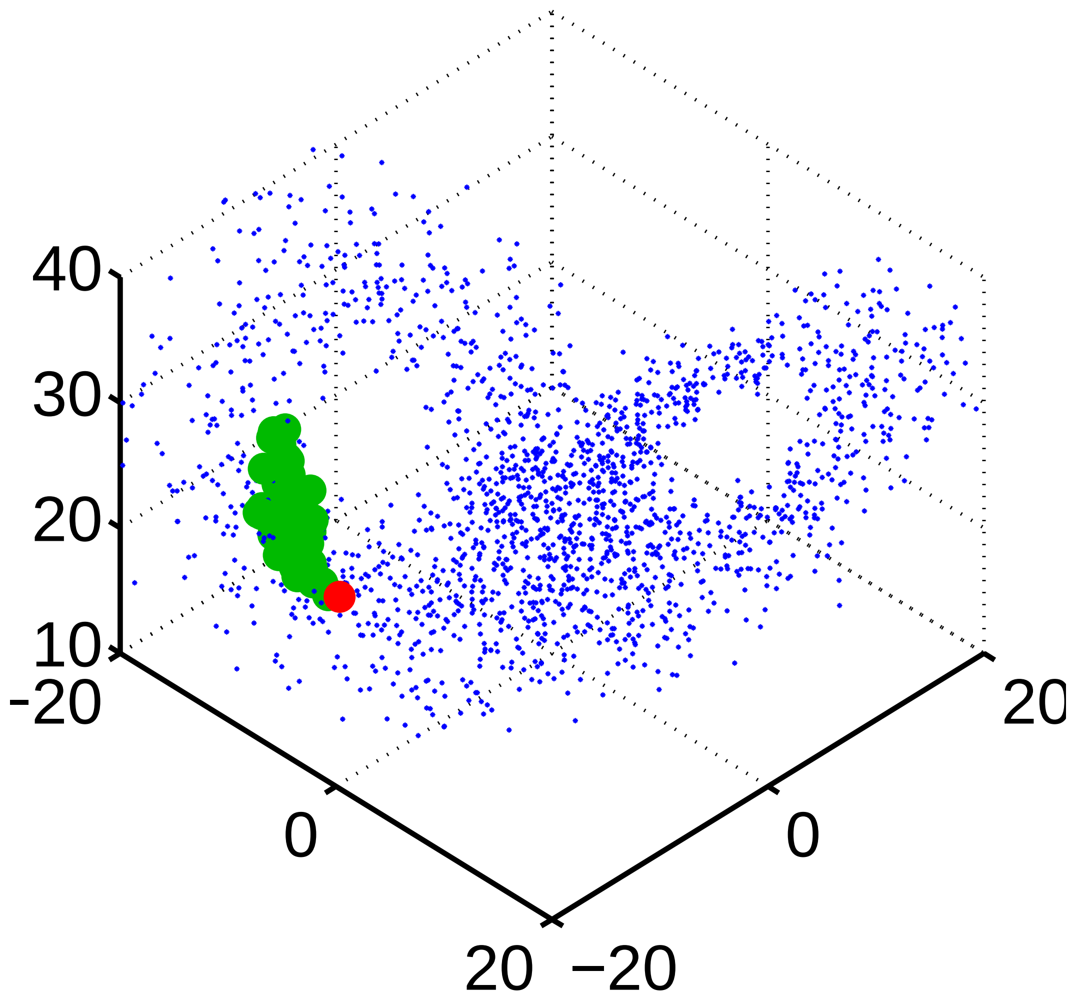
1. Can be very effective for a variety of models.
2. Can maintain linear balances.
3. Stays on local flat manifolds.
4. Simple and inexpensive.

Liabilities:

1. State variables not constrained by observations can ‘blow up’.
   For instance unobserved regions near the top of AGCMs.
2. Magnitude of λ normally selected by trial and error.

Physical space covariance inflation in Lorenz-63

Observation outside prior: danger of filter divergence

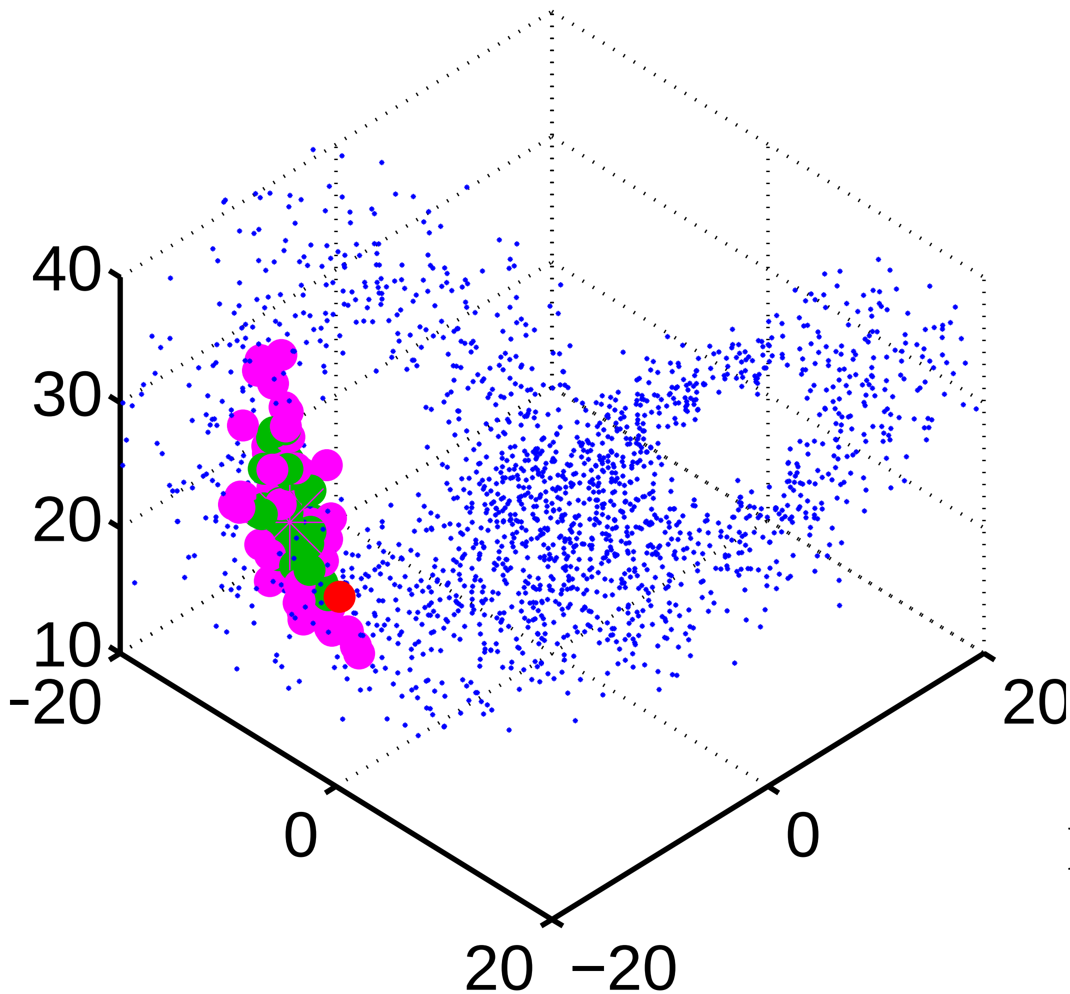


Observation in red.

Prior ensemble in green.

Physical space covariance inflation in Lorenz-63

After inflating, observation is in prior cloud: filter divergence avoided



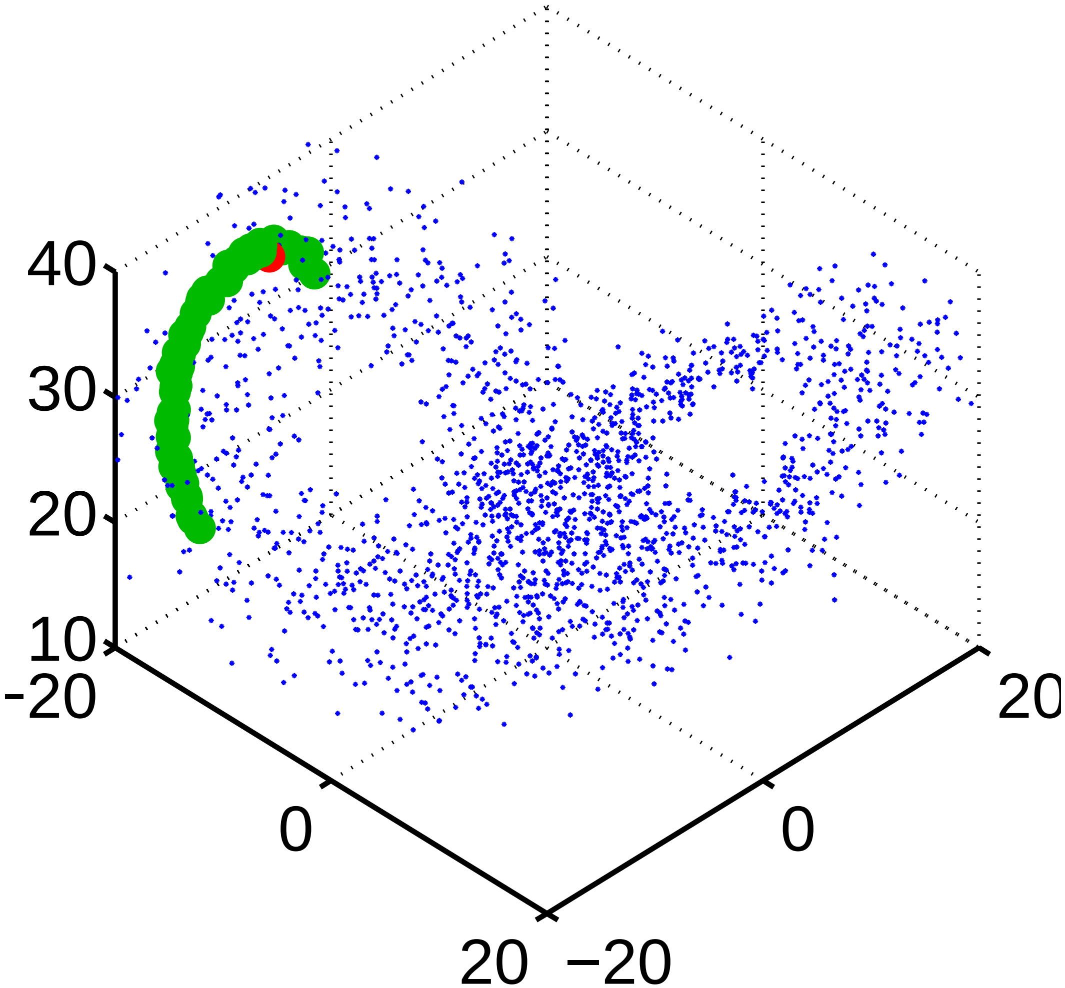
Observation in red.

Prior ensemble in green.

Inflated ensemble in magenta.

Physical space covariance inflation in Lorenz-63

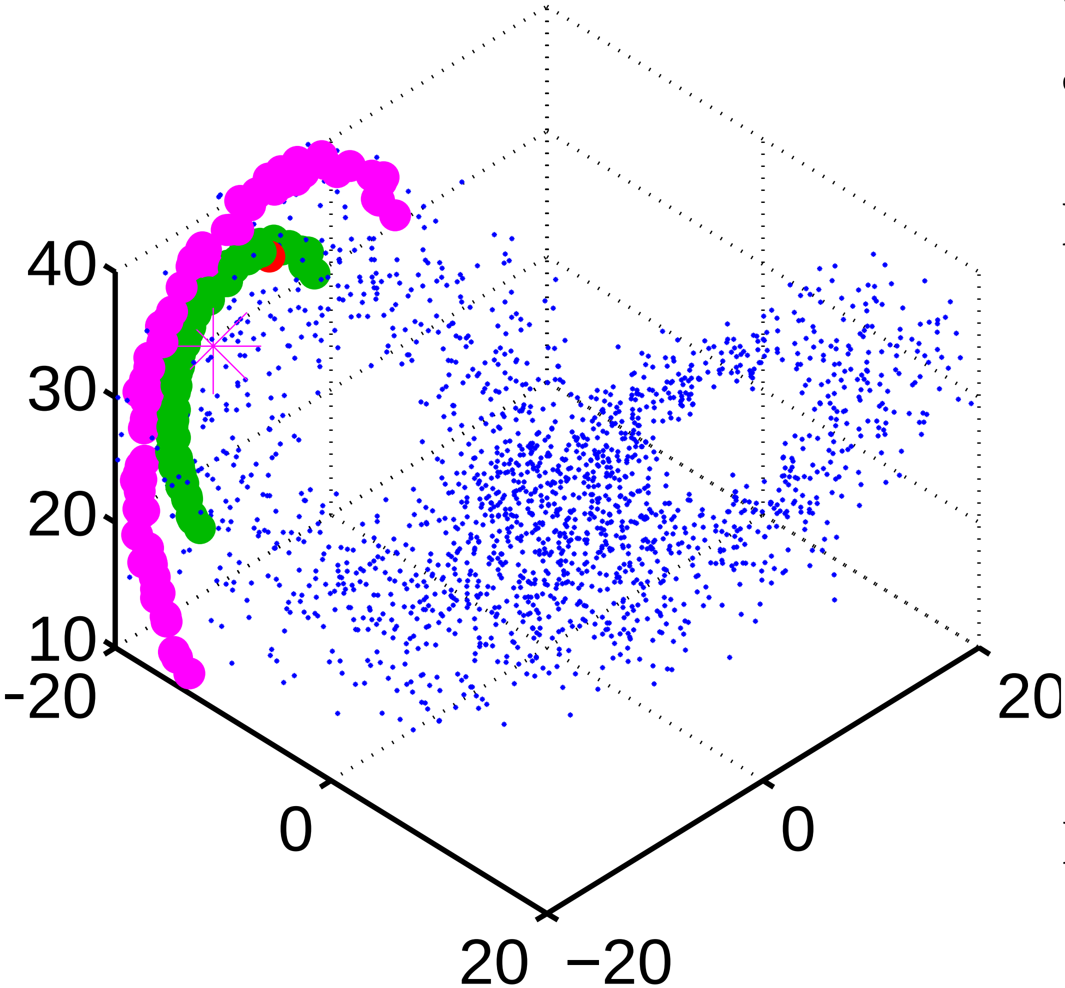
Prior distribution is significantly 'curved'



Observation in red.

Prior ensemble in green.

Inflated prior outside attractor. Posterior will also be off attractor.



Can lead to transient off-attractor behavior or...

Model ‘blow-up’.

Observation in red.

Prior ensemble in green.

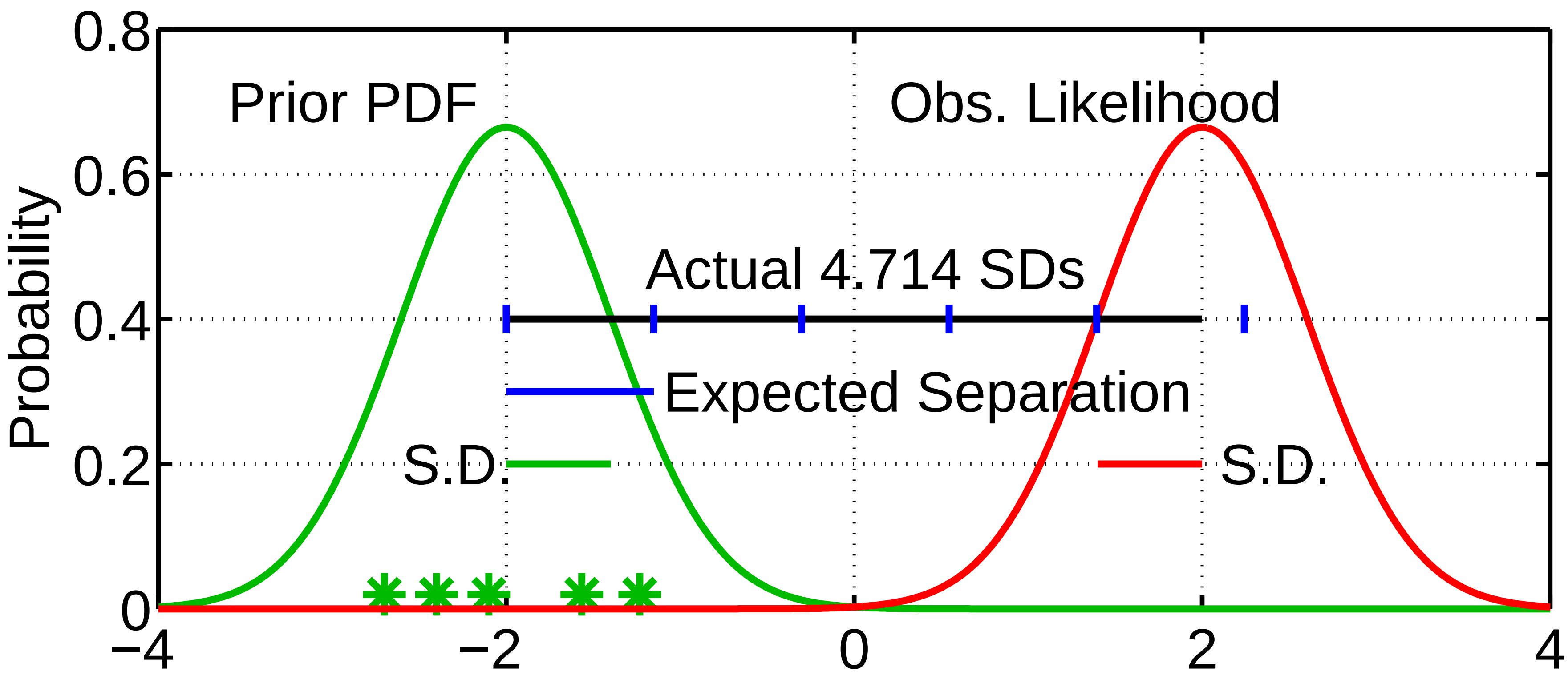
Inflated ensemble in magenta.

Adjunct Algorithms Developed for DART Can Tolerate Errors.

1. Adaptive Error Tolerant Filters.
    - Automatically detect error in assimilation system.
    - Add uncertainty when model disagrees with observations.
    - Can deal with LARGE model, observation, and filter error.

Variance inflation for Observations: An Adaptive Error Tolerant Filter

1. For observed variable, have estimate of prior-observed inconsistency



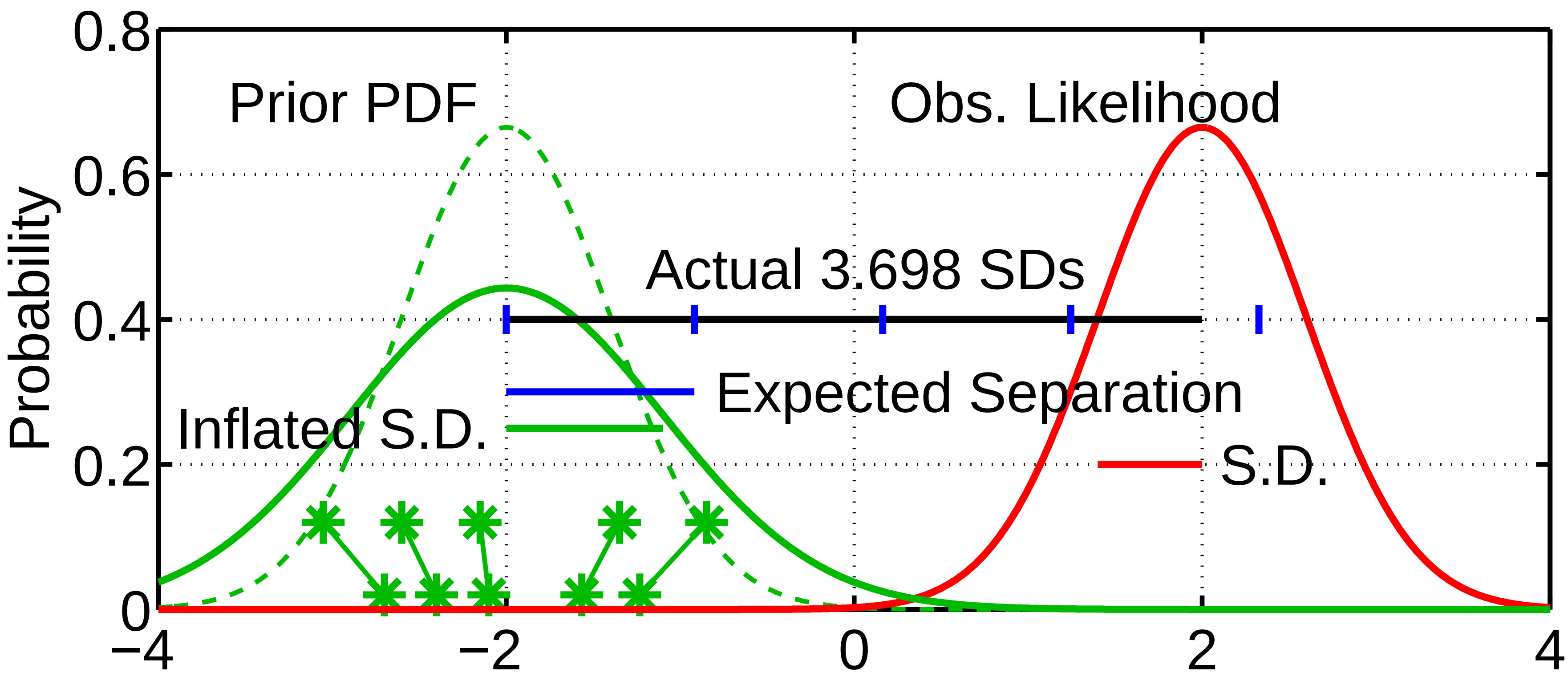
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

Assumes that prior and observation are supposed to be unbiased.

Is it model error or random chance?

Variance inflation for Observations: An Adaptive Error Tolerant Filter

1. For observed variable, have estimate of prior-observed inconsistency



2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

3. Inflating increases expected separation.

Increases ‘apparent’ consistency between prior and observation.

Adjunct Algorithms Developed for DART Can Tolerate Errors.

1. Adaptive Error Tolerant Filters.
   Automatically detect error in assimilation system.
   Add uncertainty when model disagrees with observations.
   Can deal with LARGE model, observation, and filter error.

2. Hierarchical filters detect and avoid small ensemble sampling errors.
   Ensemble of ensembles for tuning period.
   Limit impact of observations as required.
   Eliminate unnecessary calculation.

Can apply filters without tuning to large problems.

Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

    Initial conditions for forecasts.

    Includes time evolution and ‘balances’.
    High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

    Estimate value of existing or planned observations.
    Design observing systems that provide increased information.

3. Improve model of physical system.

    Evaluate model systematic errors.
    Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
    Select appropriate values for model parameters.

DART/CAM NWP Assimilation: January, 2003

Model: CAM 3.1 T85L26.

Initialized from a climatological distribution (huge spread).

Observations: **Radiosondes, ACARS, Satellite Winds.**

Subset of observations used in NCAR/NCEP reanalysis.

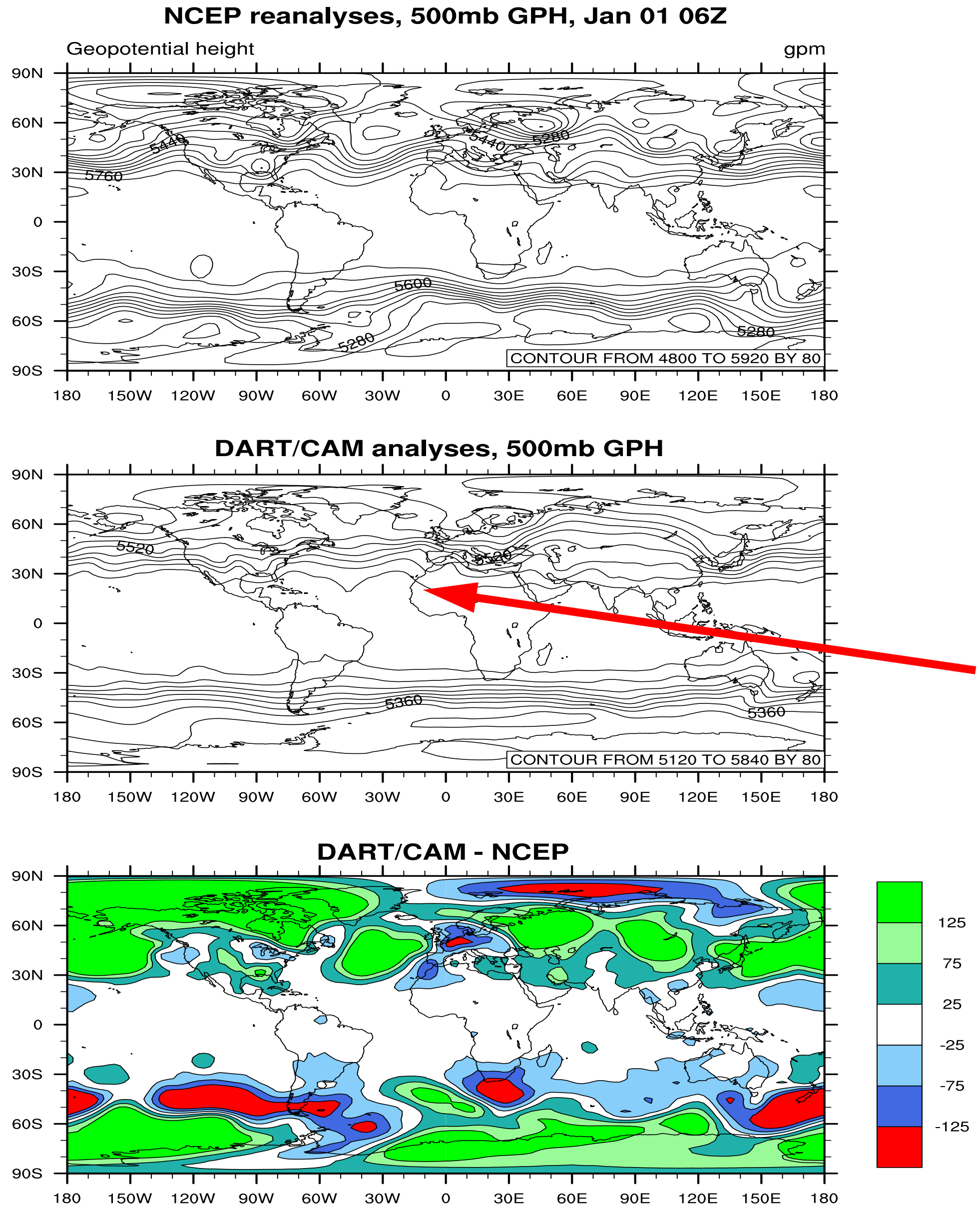
Compare to NCEP operational, T254L64, uses radiances.

After 6 hours.

NCEP

DART/CAM

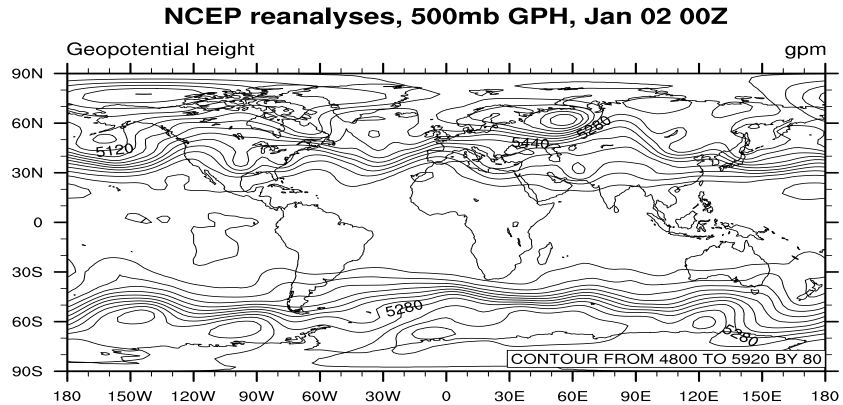
Difference.



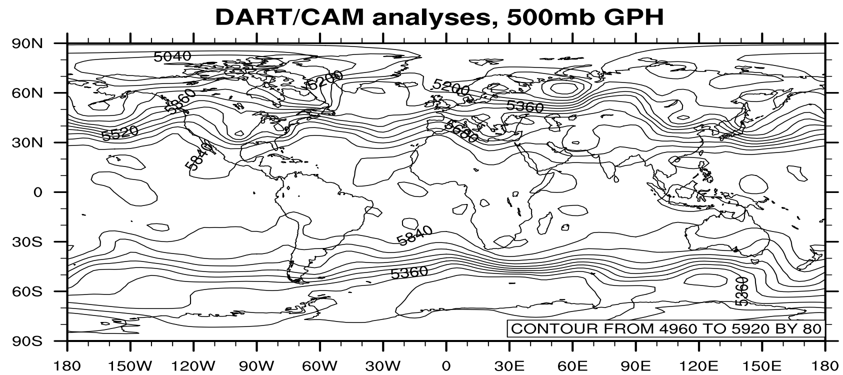
CAM
starts with
climatology!
Nearly zonal.

After 1 day.

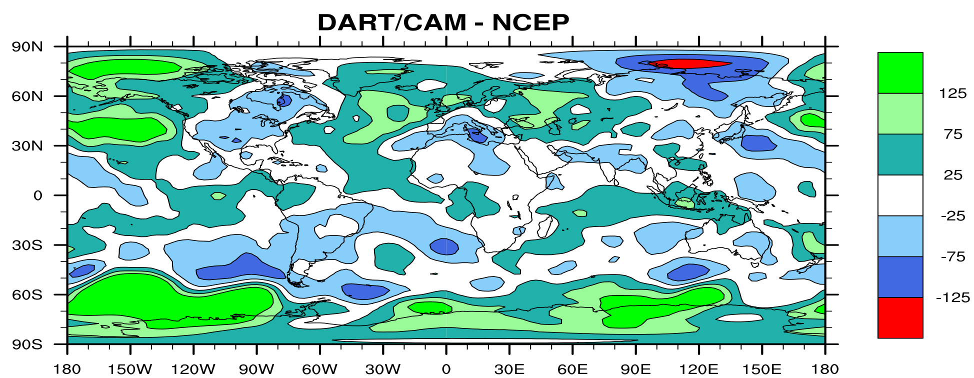
NCEP



DART/CAM

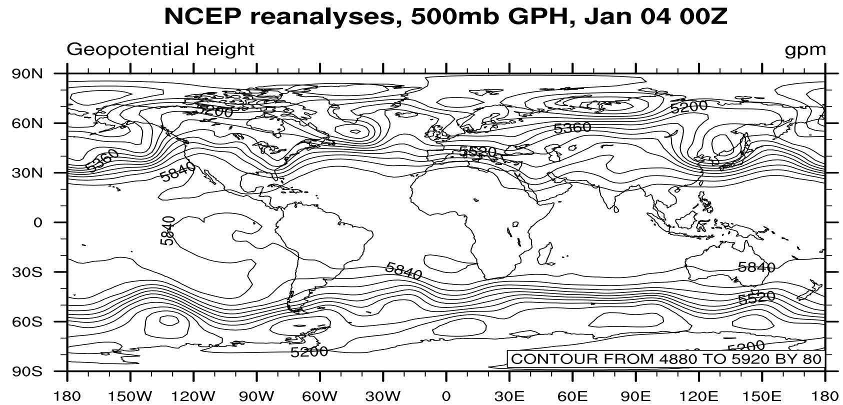


Difference.

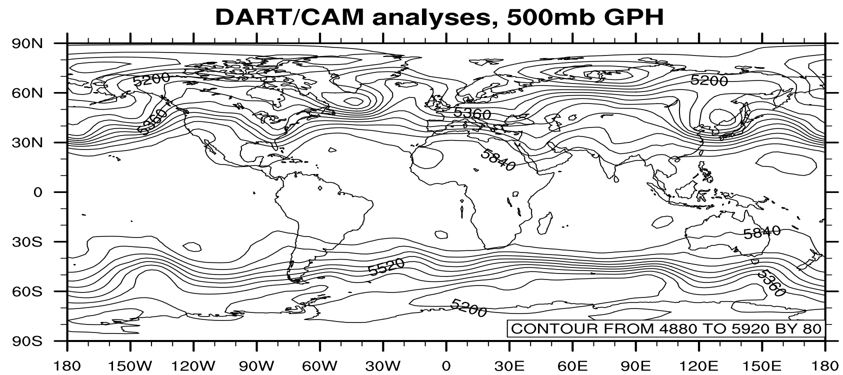


After 3 days.

NCEP

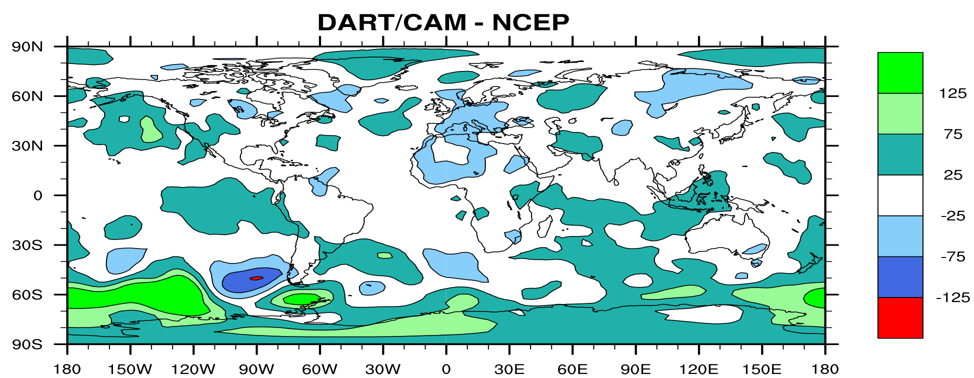


DART/CAM



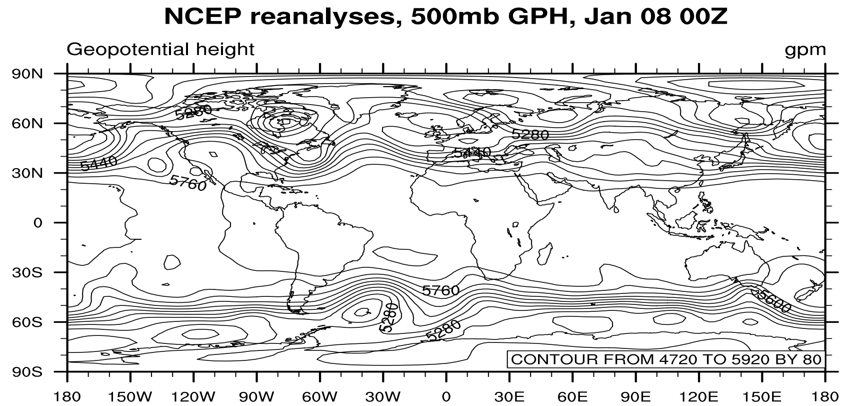
CAM gains zonal structure.

Difference.

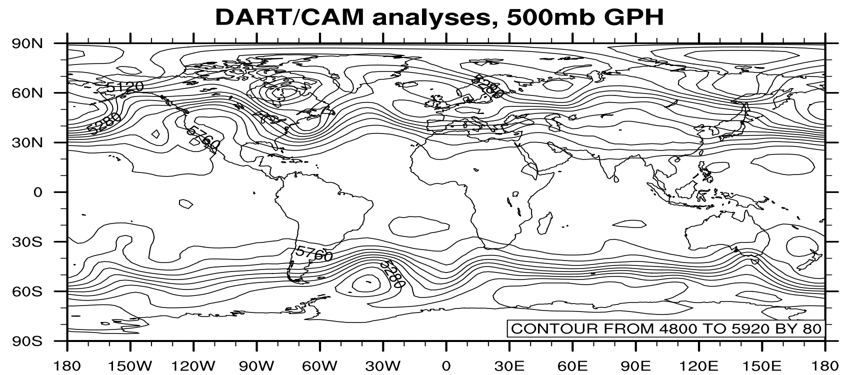


After 7 days.

NCEP

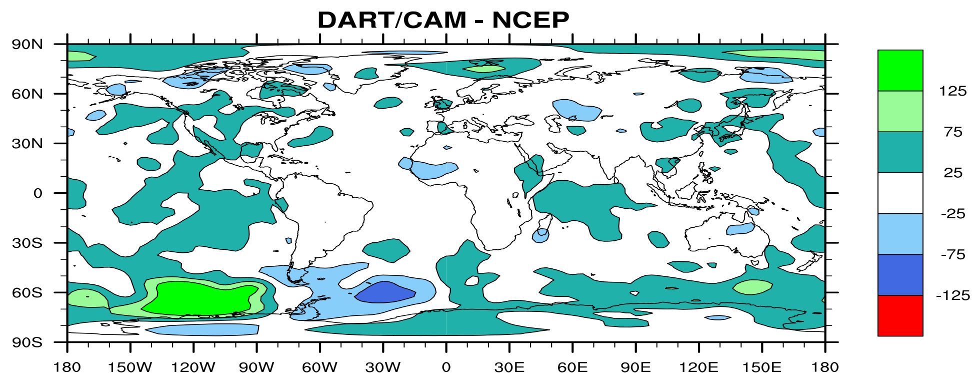


DART/CAM

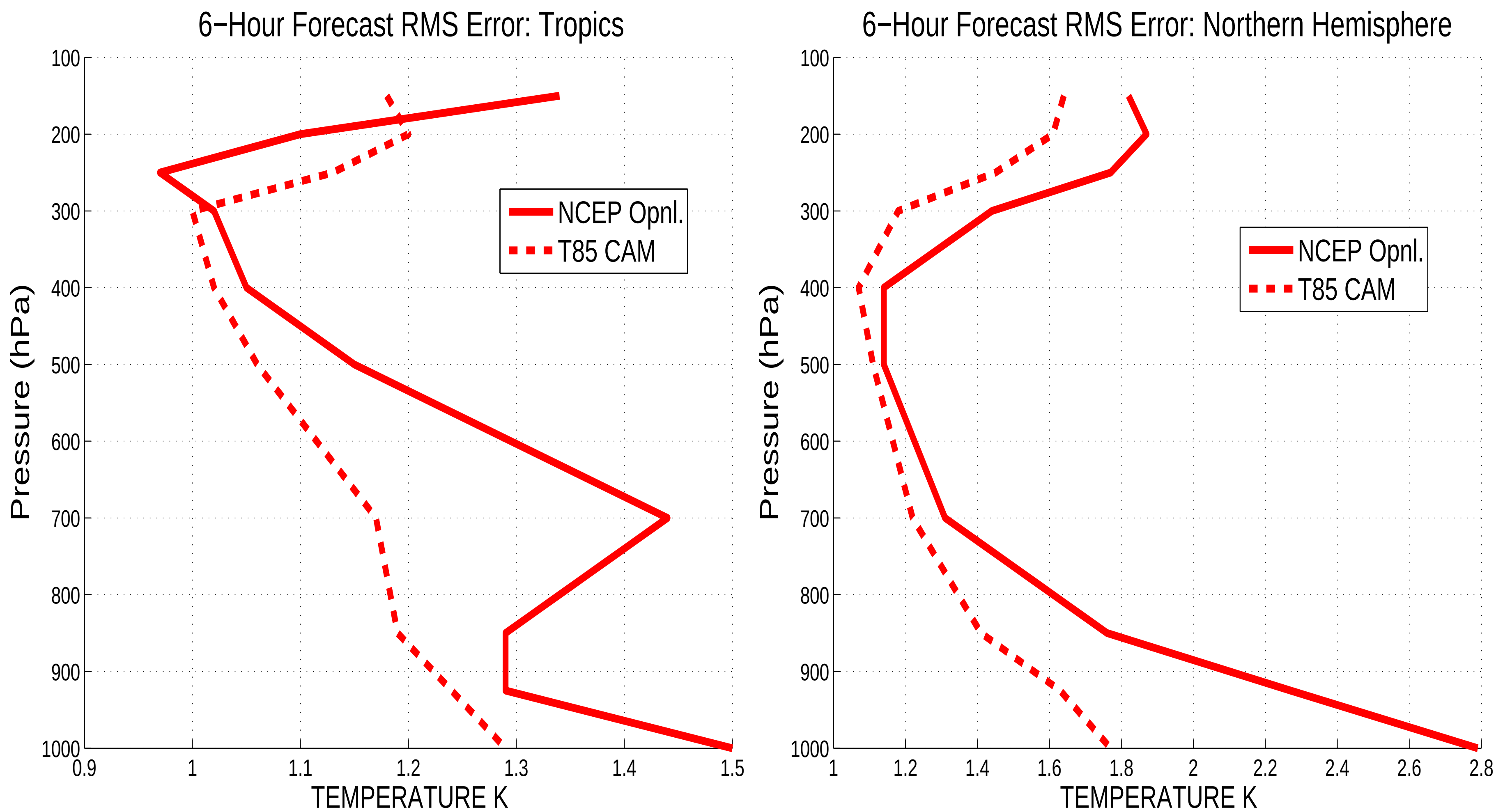


NH converged. SH poorly observed.

Difference.



6-Hour Forecast Observation Space Temperature RMS



Tropics **Northern Hemisphere**

DART/CAM competitive with operational NWP system.

Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

Initial conditions for forecasts.
Includes time evolution and ‘balances’.
High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

Estimate value of existing or planned observations.
Design observing systems that provide increased information.

3. Improve model of physical system.

Evaluate model systematic errors.
Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
Select appropriate values for model parameters.

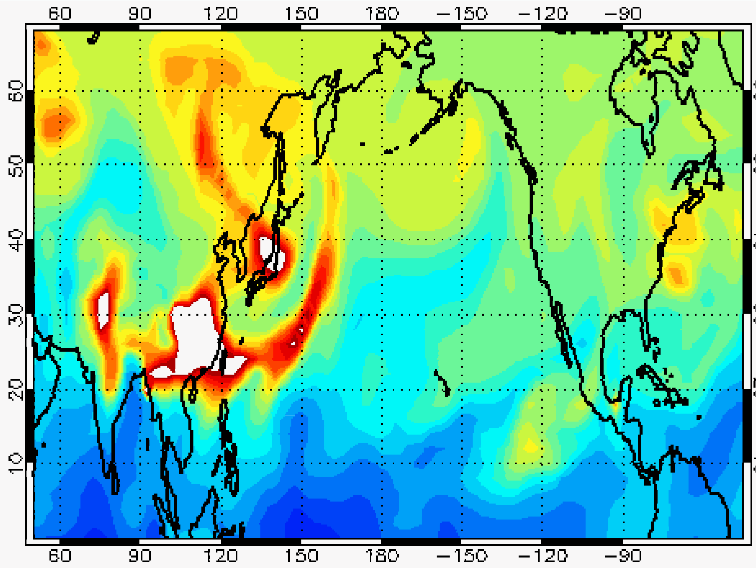
High-quality analysis of CO in Finite Volume CAM-CHEM model.

Assimilate standard observations plus MOPITT CO observations.

Work by Ave Arellano and Peter Hess supported by Kevin Raeder.

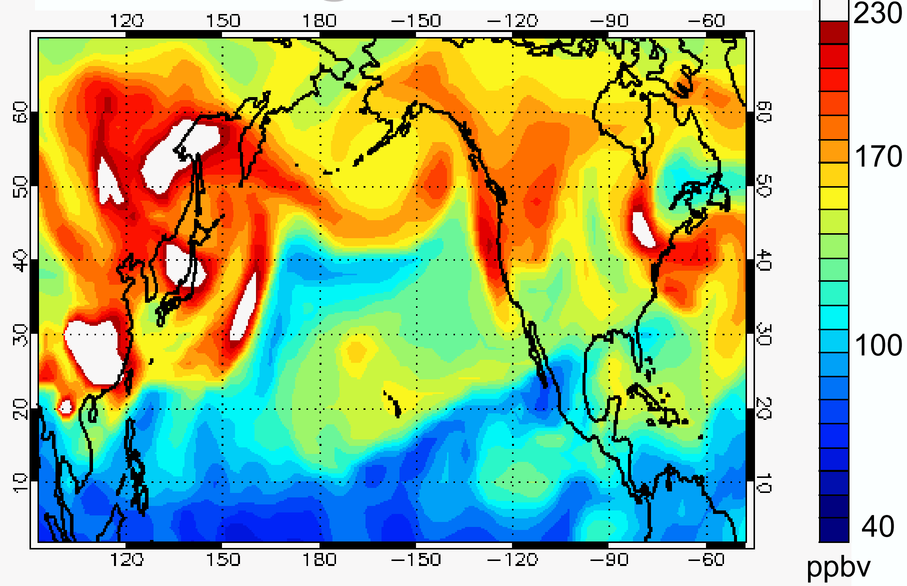
Impact of Assimilation in Modeled CO

No Assimilation @700 hPa 041706 18Z



Assimilating MOPITT CO provides important constraints to regional CO distribution in the troposphere.

Assimilation @700 hPa 041706 18Z



Suggests the utility of assimilation in providing better initial/boundary conditions to regional CO forecasts.

Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

    Initial conditions for forecasts.
    Includes time evolution and ‘balances’.
    High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

    Estimate value of existing or planned observations.
    Design observing systems that provide increased information.

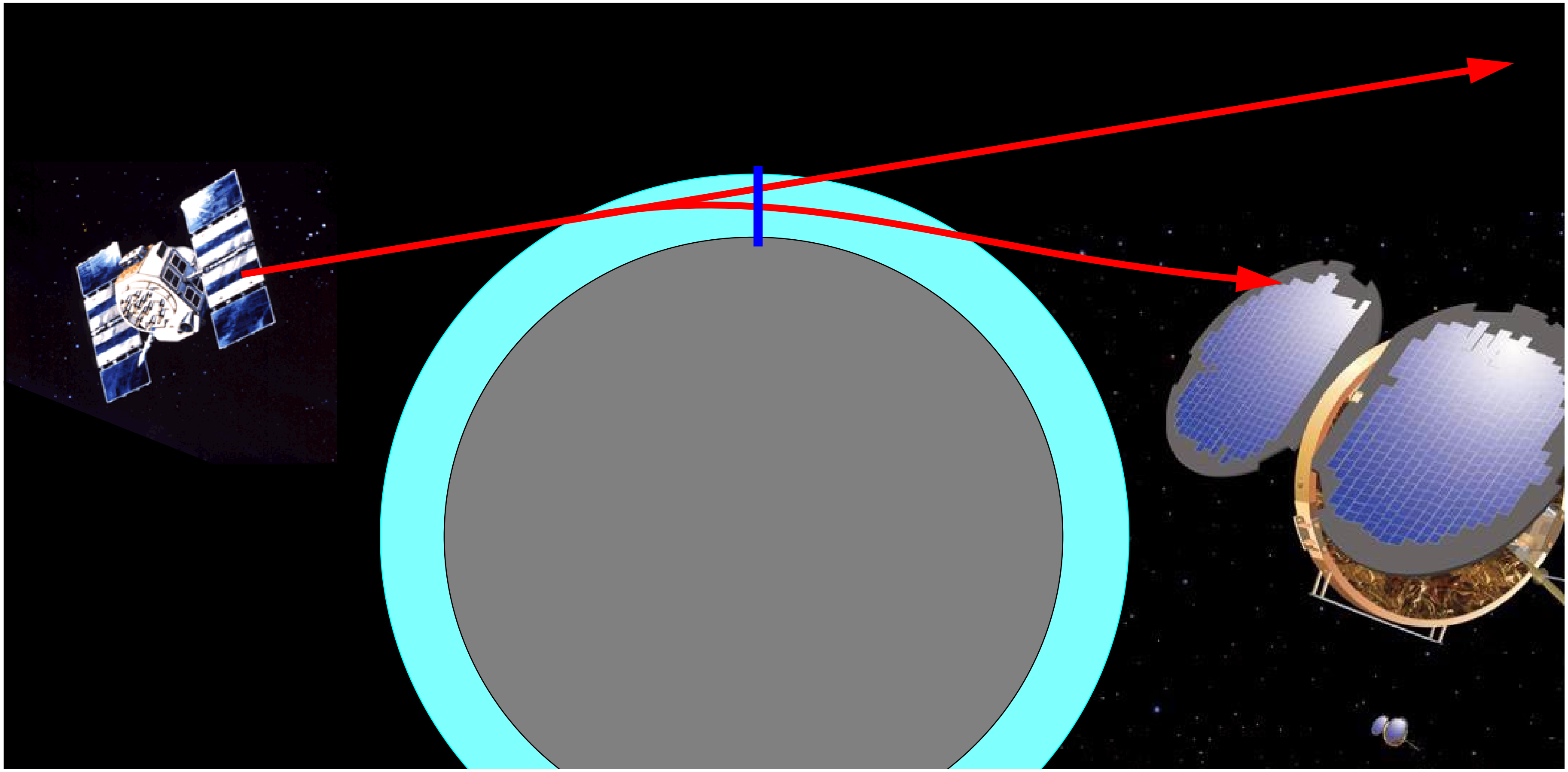
3. Improve model of physical system.

    Evaluate model systematic errors.
    Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
    Select appropriate values for model parameters.

Assimilating GPS Radio Occultation Observations in WRF

Assimilated as refractivity along beam path.

Complicated function of T, Q, P and ionospheric electric field.

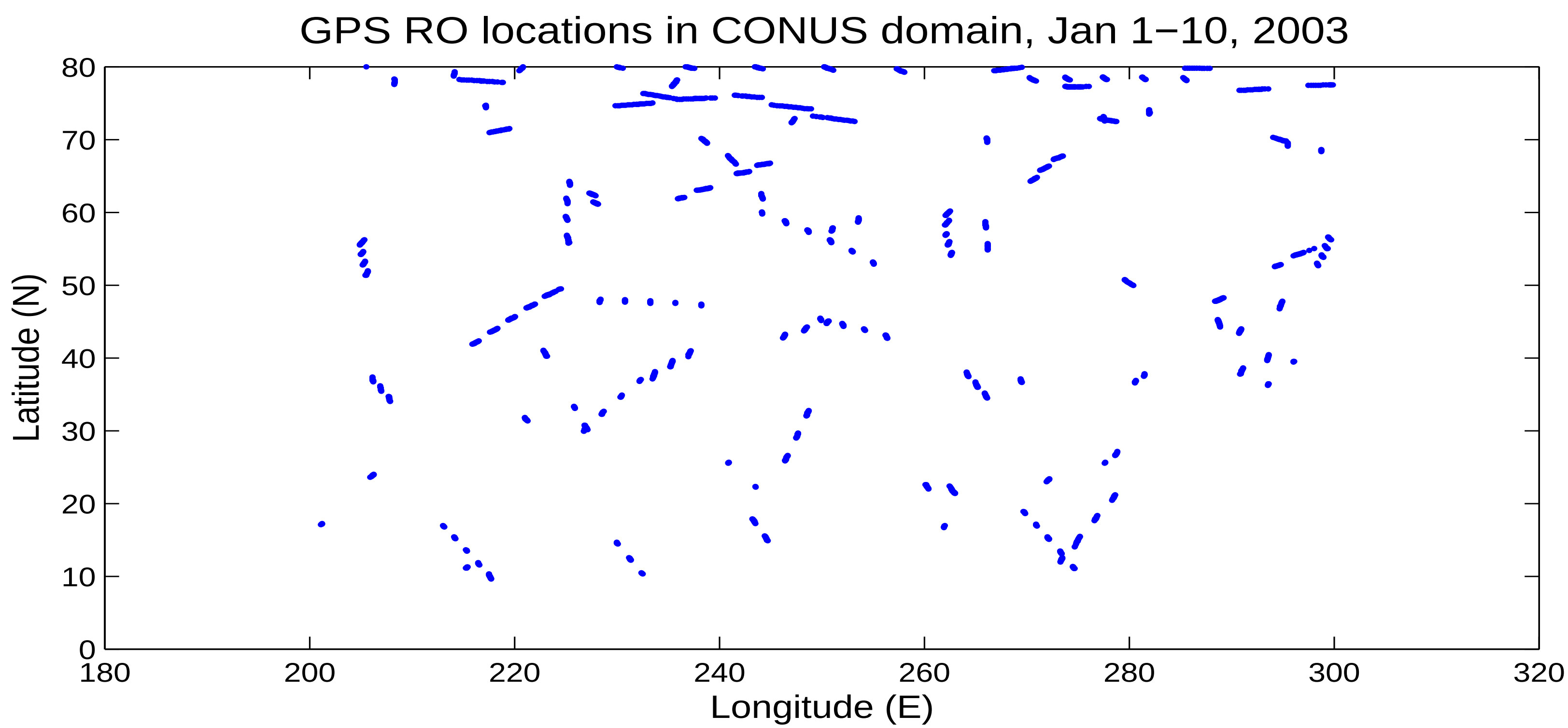


Get a sounding as GPS satellite sets relative to low earth satellite.

Assimilating GPS Radio Occultation Observations in WRF

Weather Research and Forecasting Model.
Regional Weather Prediction model.
Configured for CONUS domain, 50 km grid.

Several hundred profiles available from CHAMP satellite.



Assimilating GPS Radio Occultation Observations in WRF

Evaluating Impact of GPS Observations.

Case 1: Assimilate radiosondes EXCEPT those close to GPS profiles.
Case 2: Also assimilate GPS profiles.

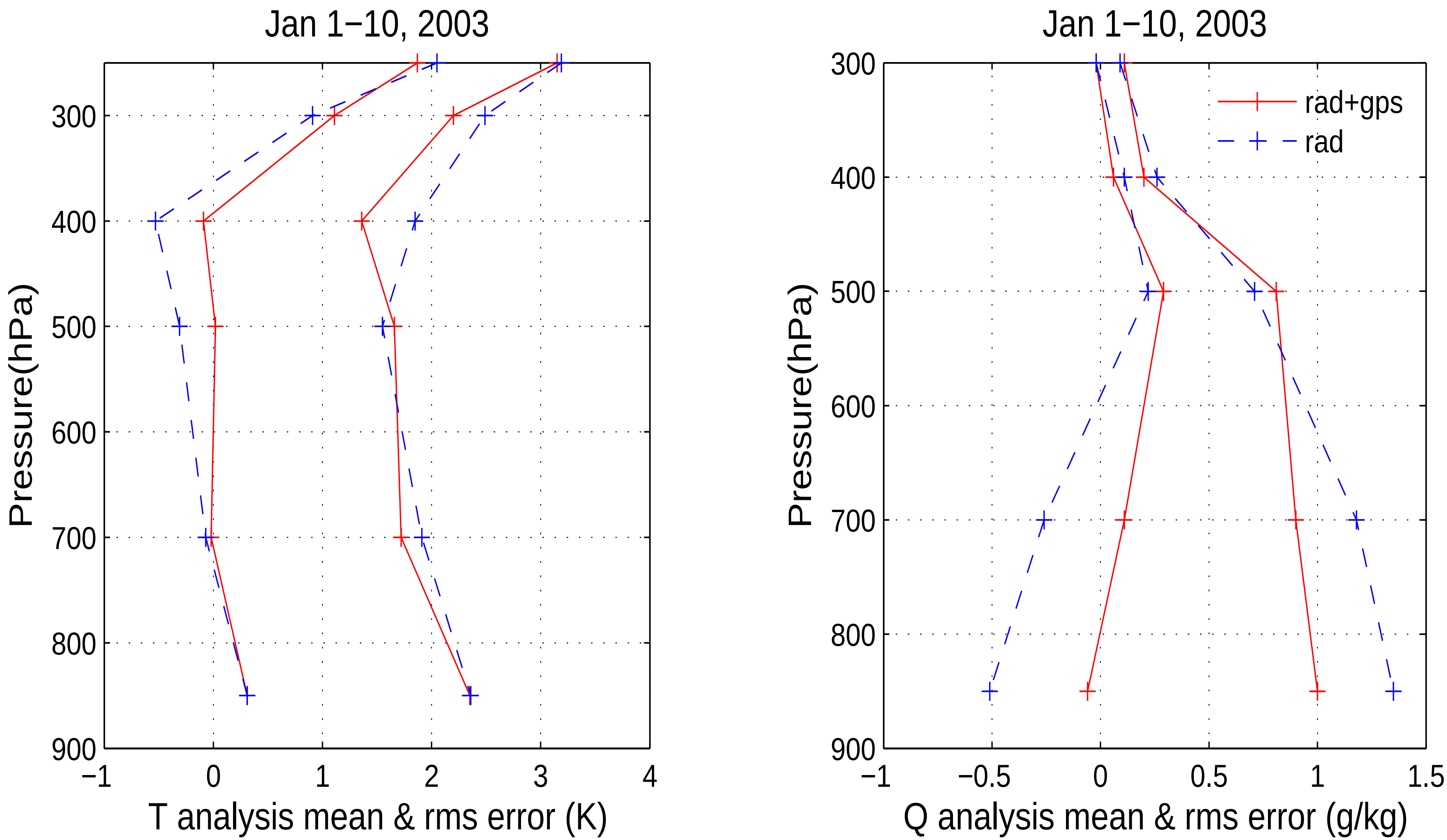
Look at reduction in error from close (unused) radiosonde profiles.

NOTE: Identical code allows assimilation in CAM, GFDL, GFS...

GPS Radio Occultation Impact on T and Q Errors in WRF

Each plot displays bias (left pair) and RMS (right pair).

Red curves include GPS: reduced bias and RMS.



Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

Initial conditions for forecasts.
Includes time evolution and ‘balances’.
High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

Estimate value of existing or planned observations.
Design observing systems that provide increased information.

3. Improve model of physical system.

Evaluate model systematic errors.
Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
Select appropriate values for model parameters.

Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

Initial conditions for forecasts.
Includes time evolution and ‘balances’.
High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

Estimate value of existing or planned observations.
Design observing systems that provide increased information.

3. Improve model of physical system.

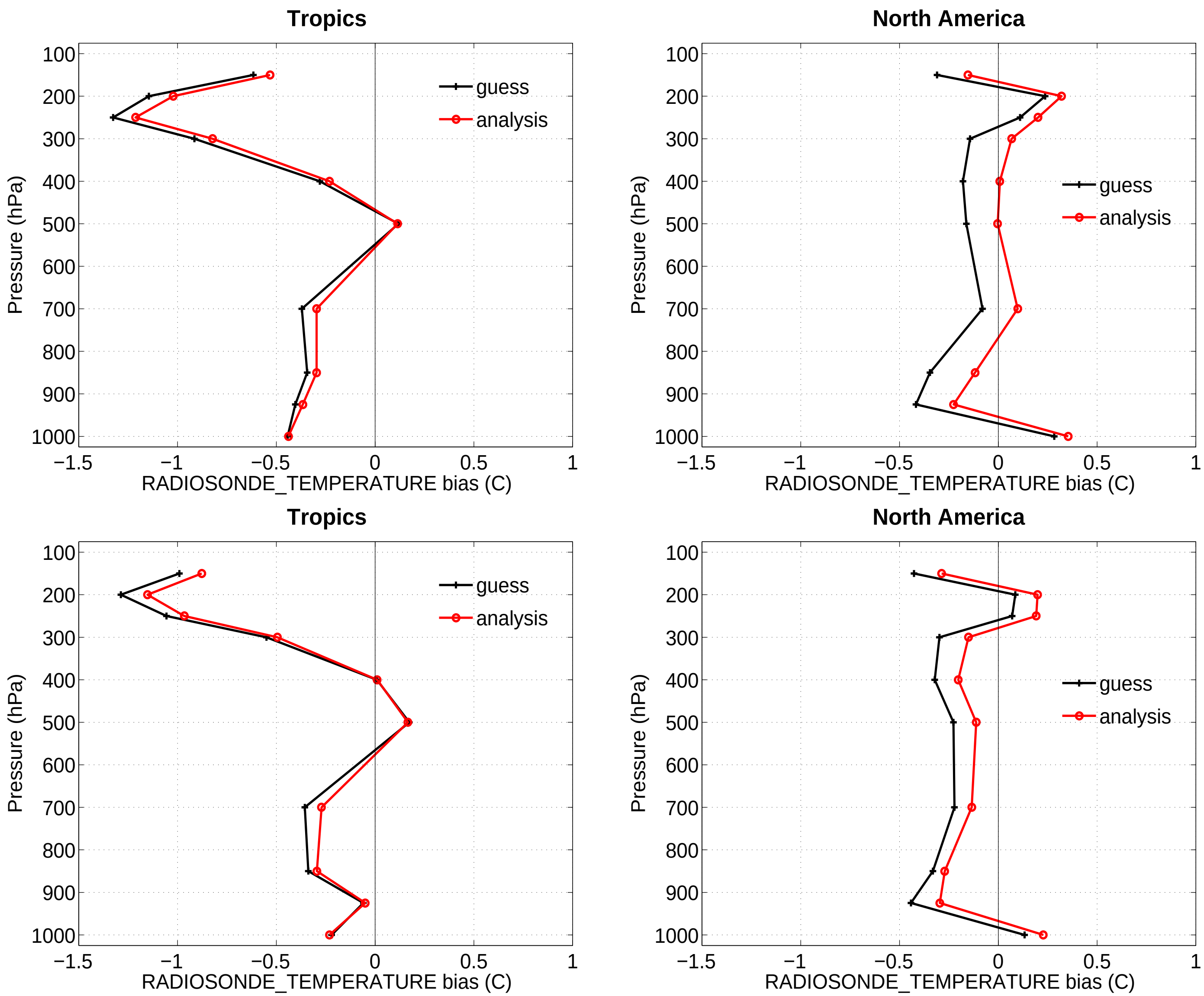
Evaluate model systematic errors.
Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
Select appropriate values for model parameters.

Example of low-resolution assimilation comparisons.

CAM spectral vs. FV for January, 2003: Temperature Bias

Spectral T21

Finite Volume 2x2.5



Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

Initial conditions for forecasts.
Includes time evolution and ‘balances’.
High quality analyses (re-analyses).

2. Get better estimates of observing system error characteristics.

Estimate value of existing or planned observations.
Design observing systems that provide increased information.

3. Improve model of physical system.

Evaluate model systematic errors.
Forward/backward sensitivity analysis (adjoint/linear tangent proxy).
Select appropriate values for model parameters.

Ensemble Sensitivity Analysis

Can compute correlation (covariance) between ANY forecast or analysis quantity and ALL other forecast and analysis quantities or functions thereof at any time lag.

Can get same information as unlimited number of adjoint and linear tangent integrations over arbitrary periods.

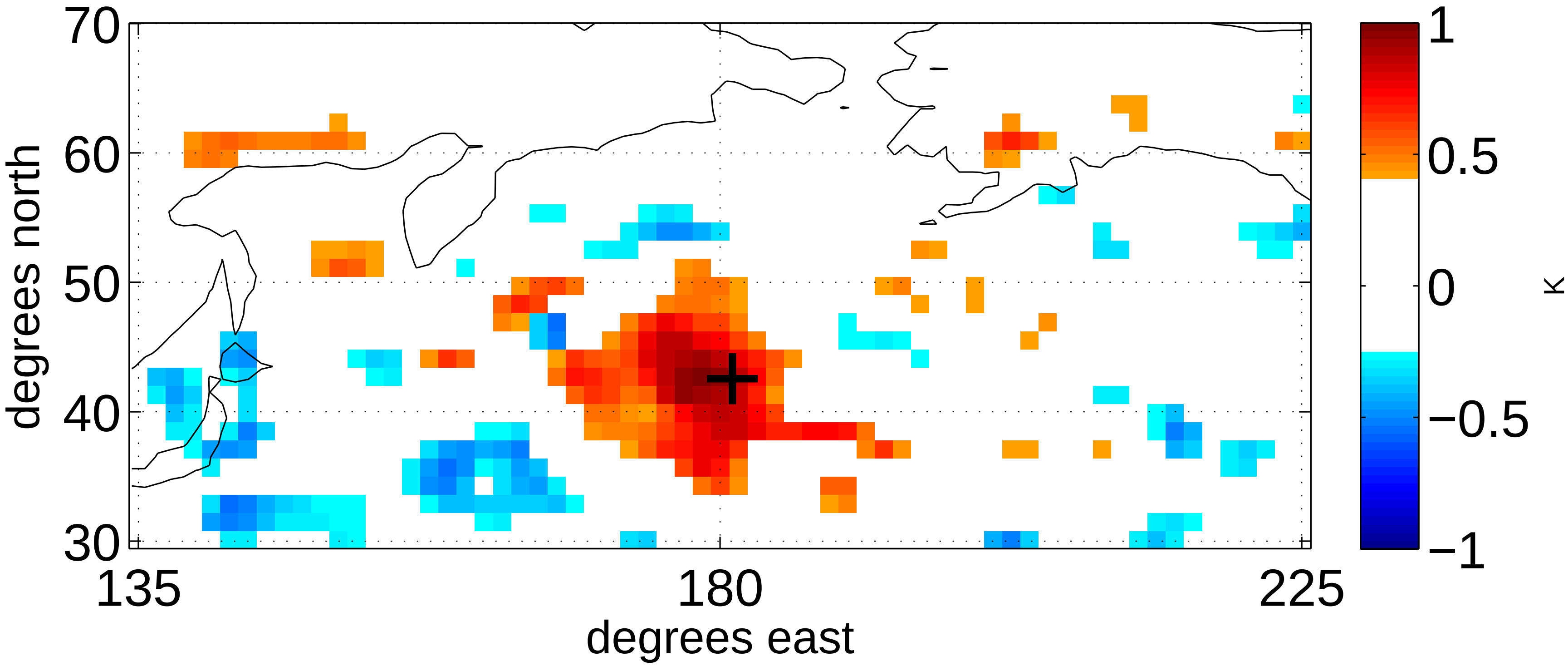
Explore relations between variables, observations, or functions thereof.

Example 1: Base point is 500 hPa mid-latitude temperature.
Look at impact on evolution of 500hPa temperatures.

Similar to linear tangent integration.
Significant correlations from 20 member T85 ensemble.

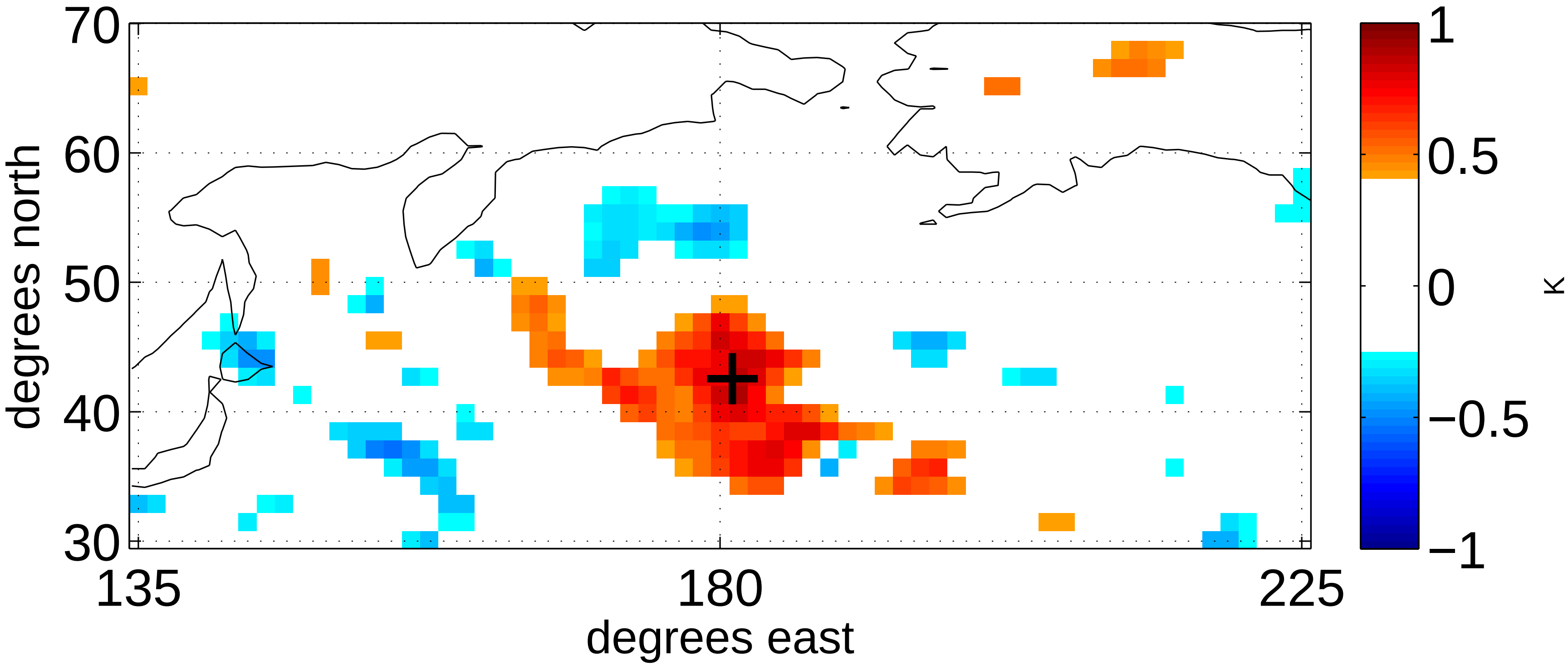
Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 00 hours: 500 hPa Temperature to 500 hPa Temperature



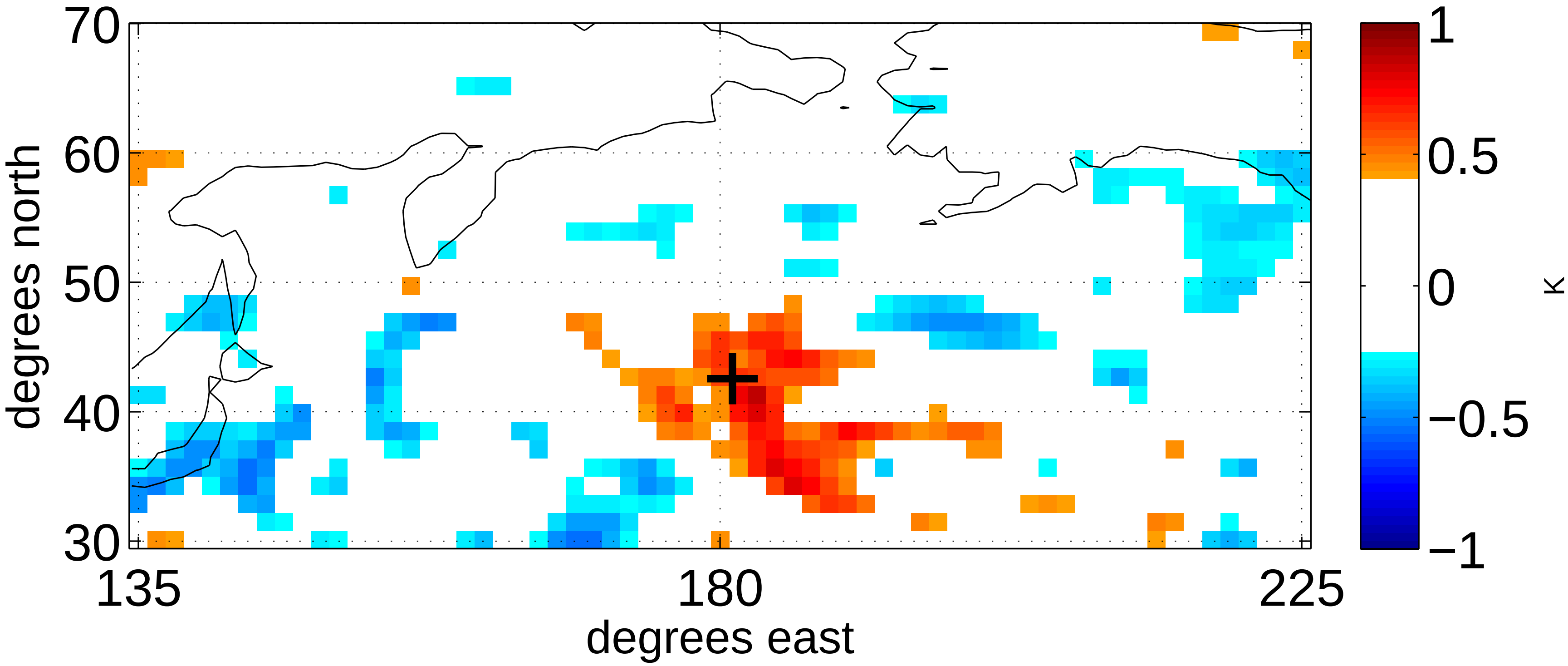
Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 06 hours: 500 hPa Temperature to 500 hPa Temperature



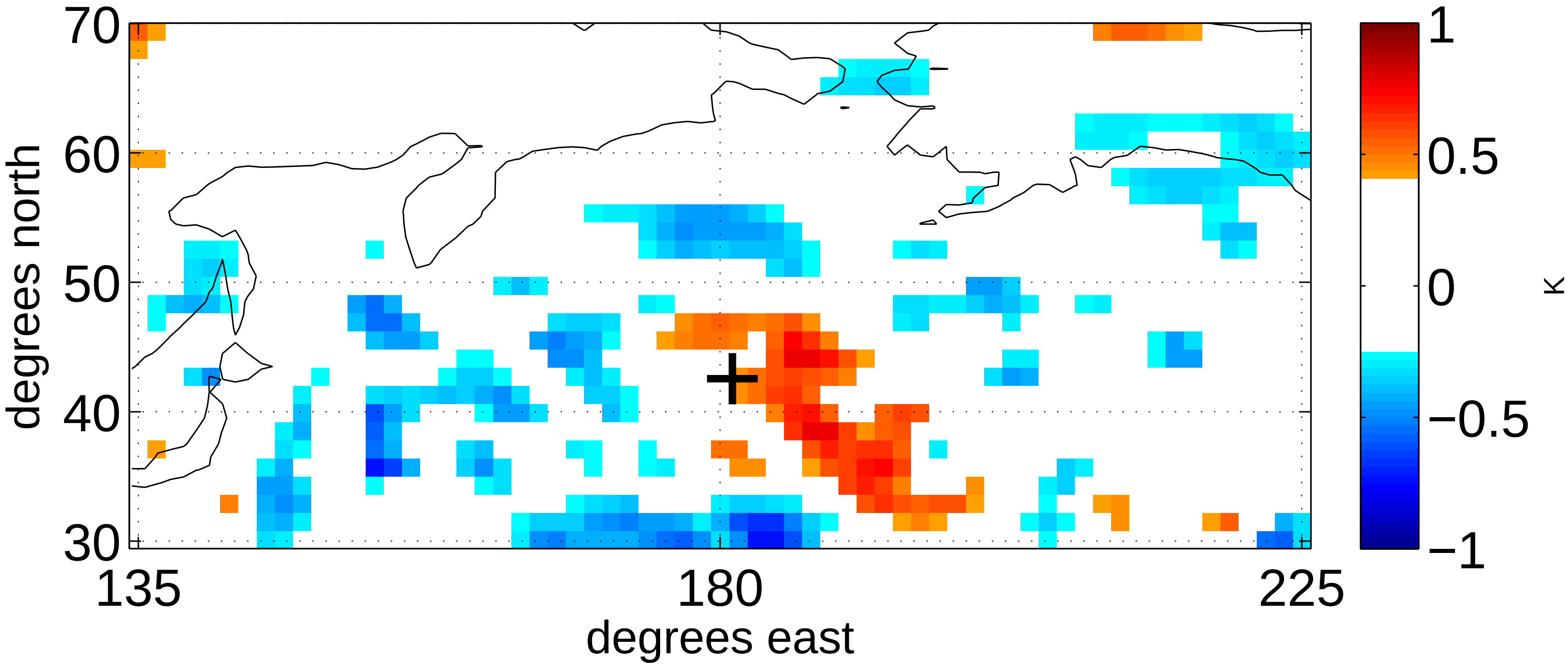
Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 12 hours: 500 hPa Temperature to 500 hPa Temperature



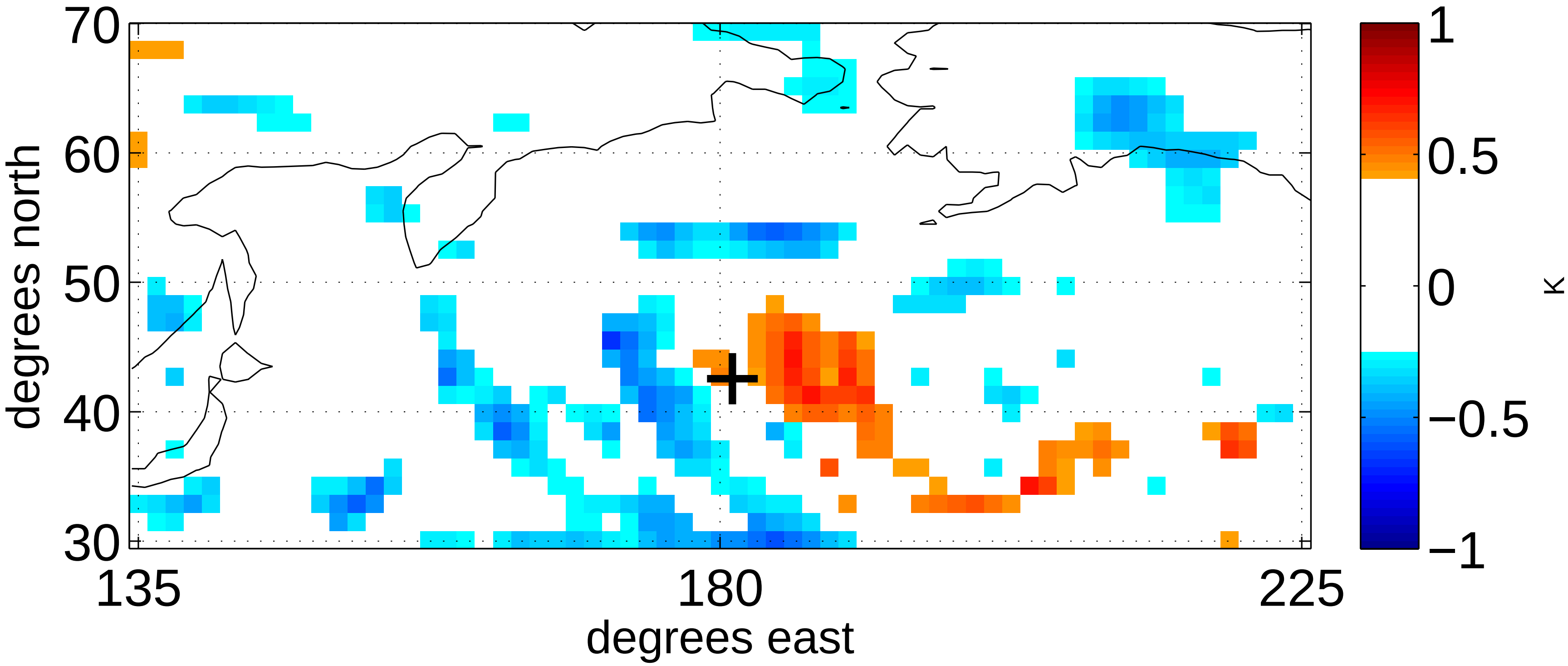
Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 18 hours: 500 hPa Temperature to 500 hPa Temperature



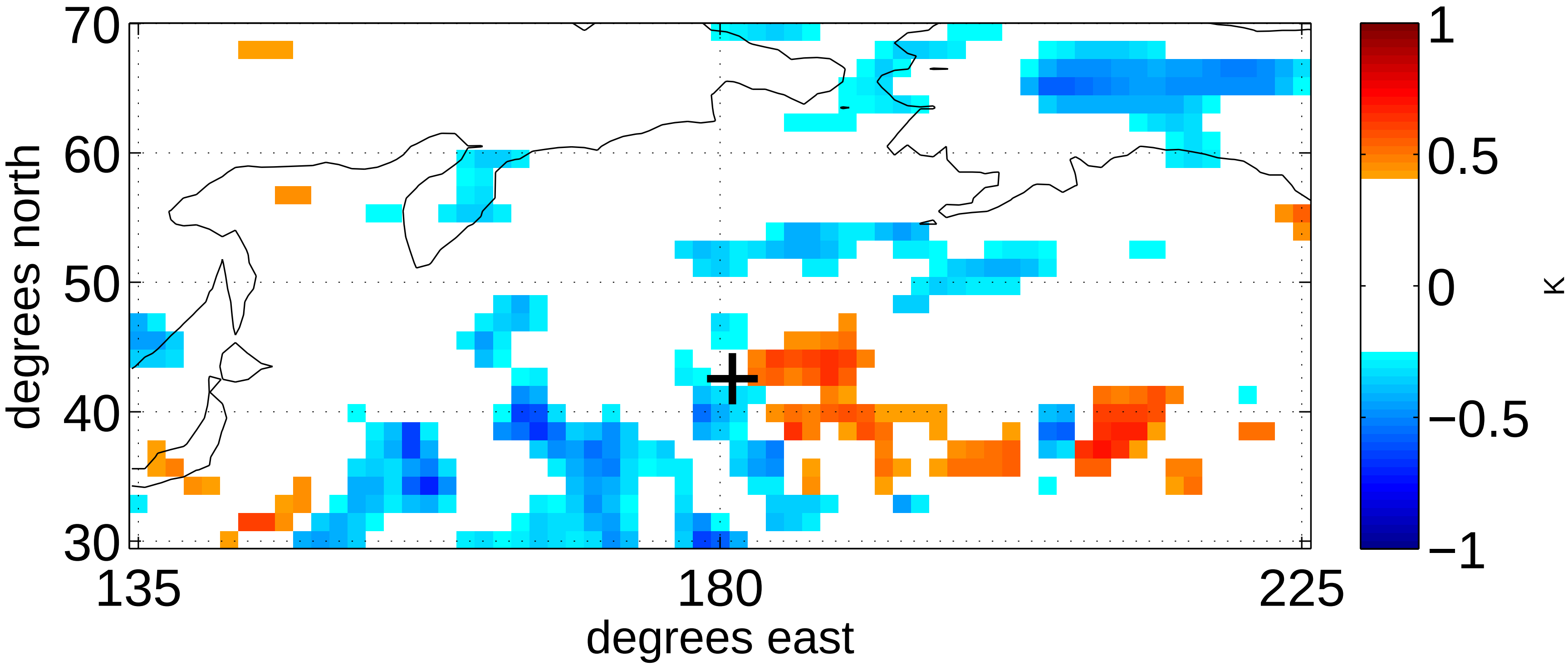
Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 24 hours: 500 hPa Temperature to 500 hPa Temperature



Forward in Time Sensitivity (Linear Tangent equivalent)

Time lag 30 hours: 500 hPa Temperature to 500 hPa Temperature



Ensemble Sensitivity Analysis

Can compute correlation (covariance) between ANY forecast or analysis quantity and ALL other forecast and analysis quantities or functions thereof.

Can get same information as unlimited number of adjoint and linear tangent integrations over arbitrary periods.

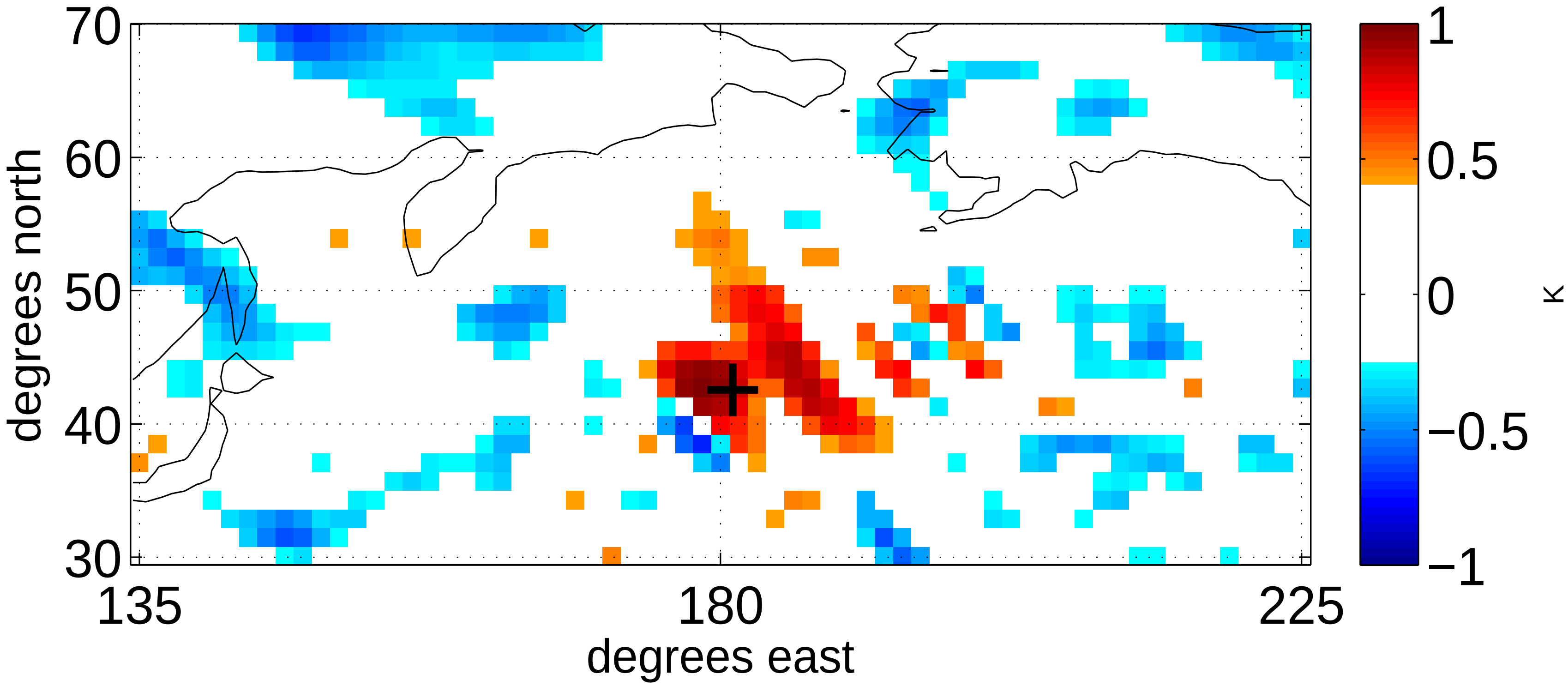
Explore relations between variables, observations, or functions thereof.

Example 2: Base point is 500 hPa mid-latitude zonal velocity.
Look at impact of previous 500 hPa temperature.

Compare to an adjoint integration.

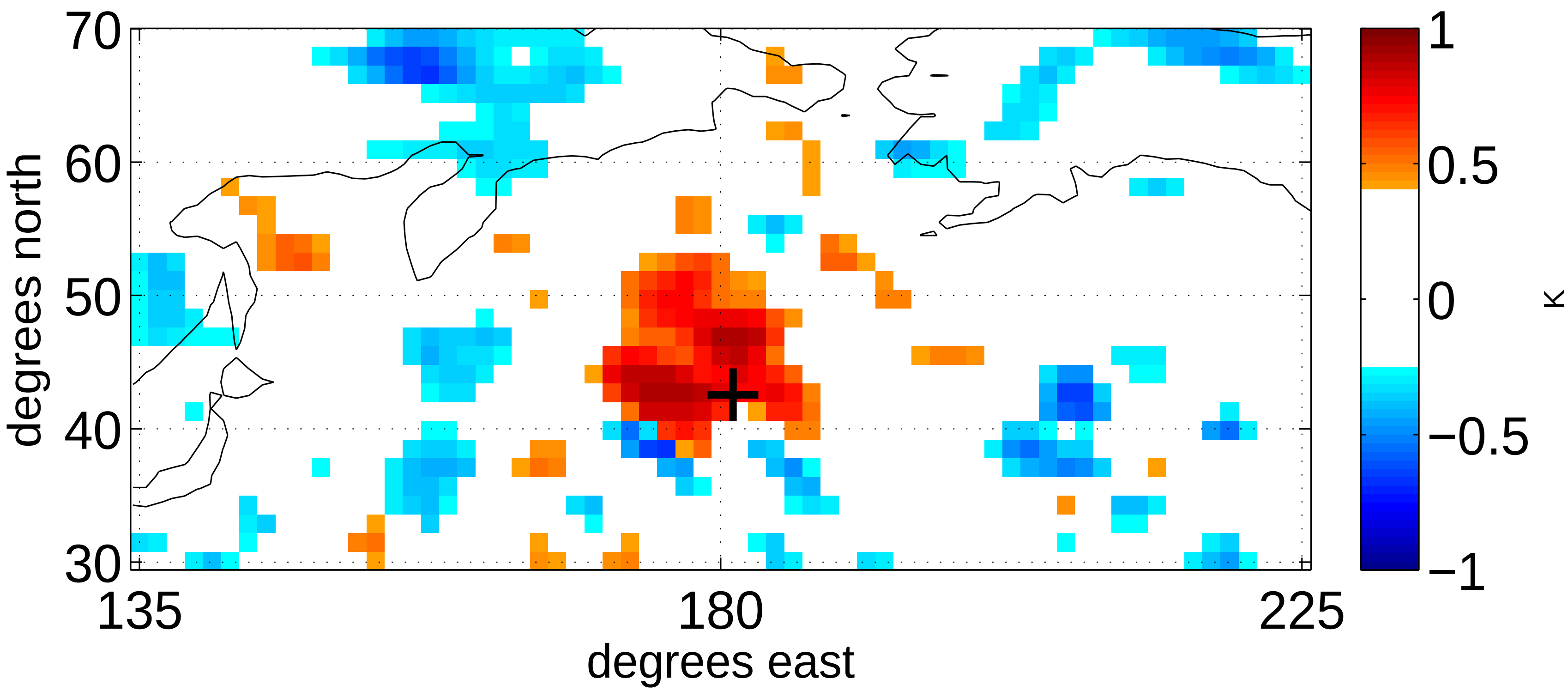
Backward in Time Sensitivity (Adjoint equivalent)

Time lag -00 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



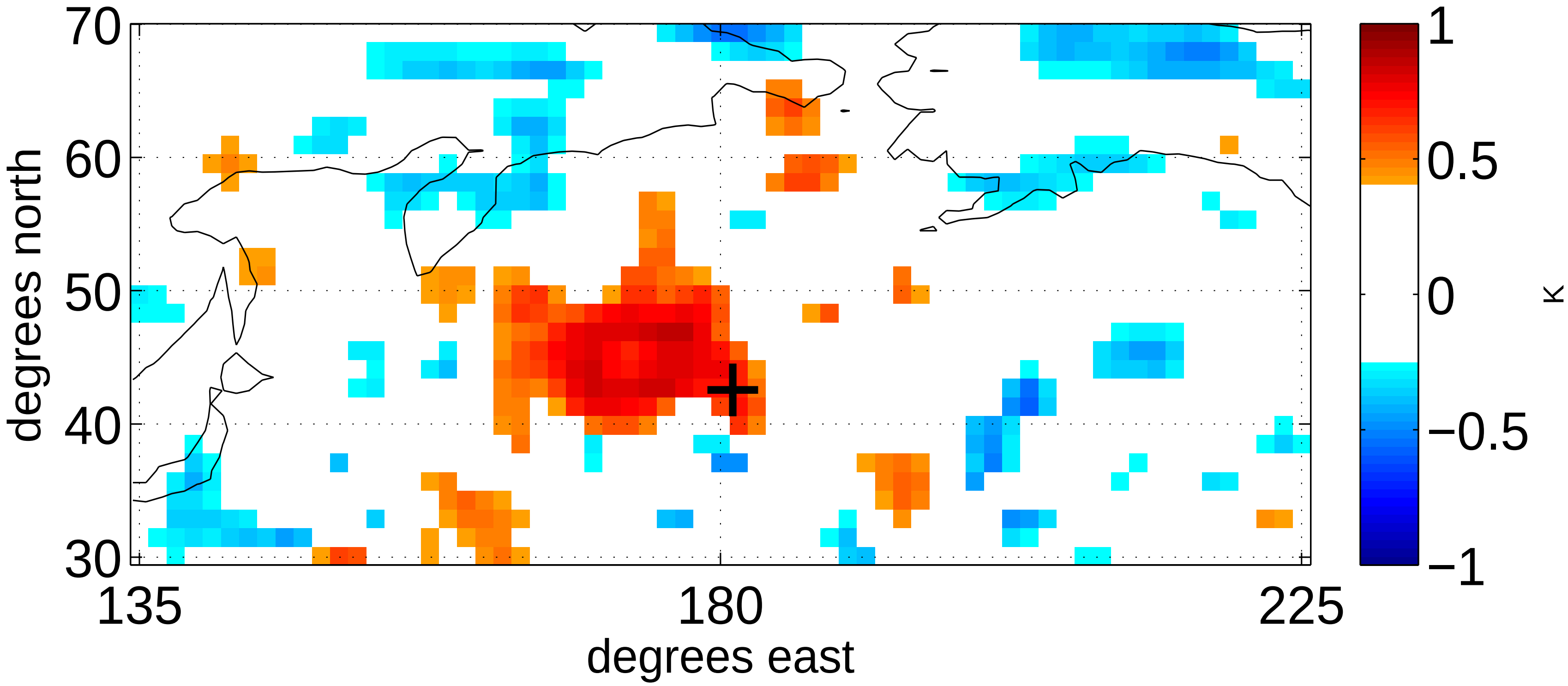
Backward in Time Sensitivity (Adjoint equivalent)

Time lag -06 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



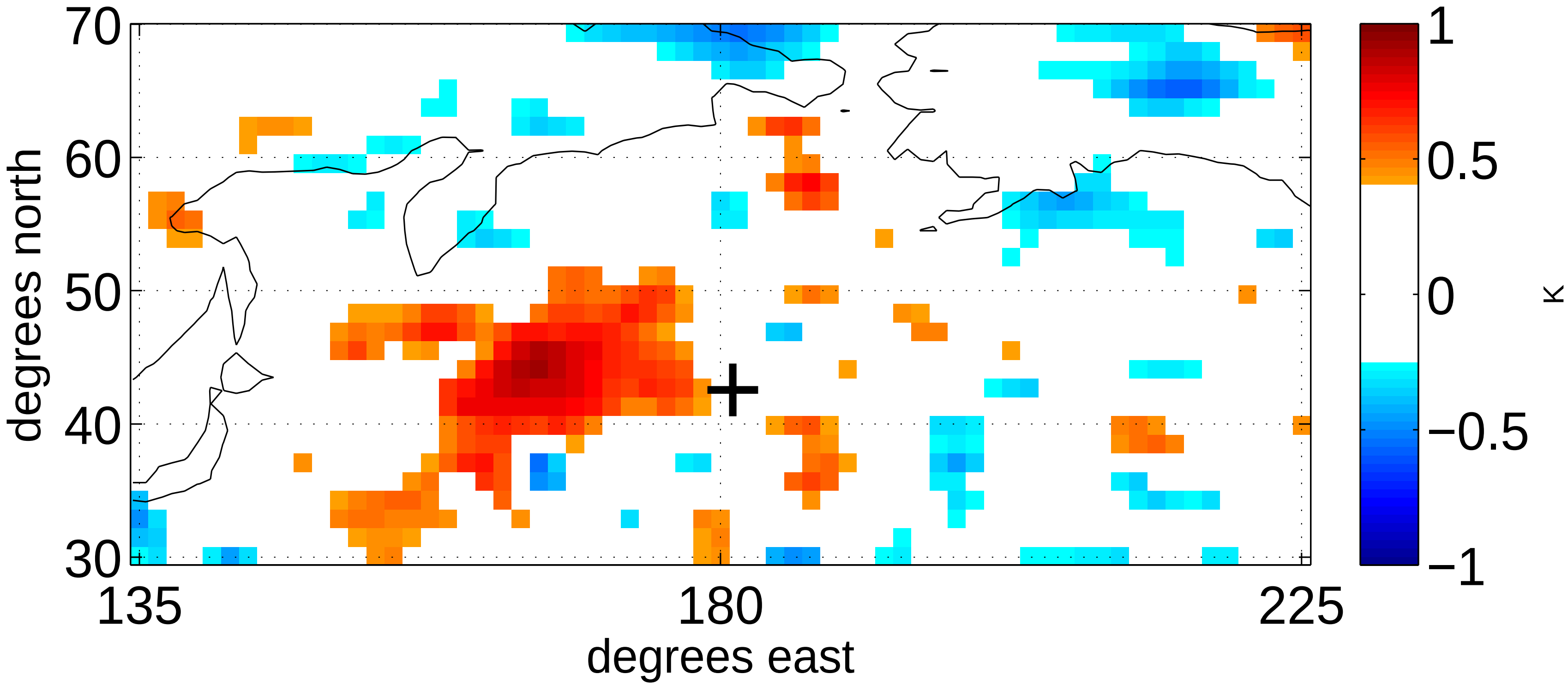
Backward in Time Sensitivity (Adjoint equivalent)

Time lag -12 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



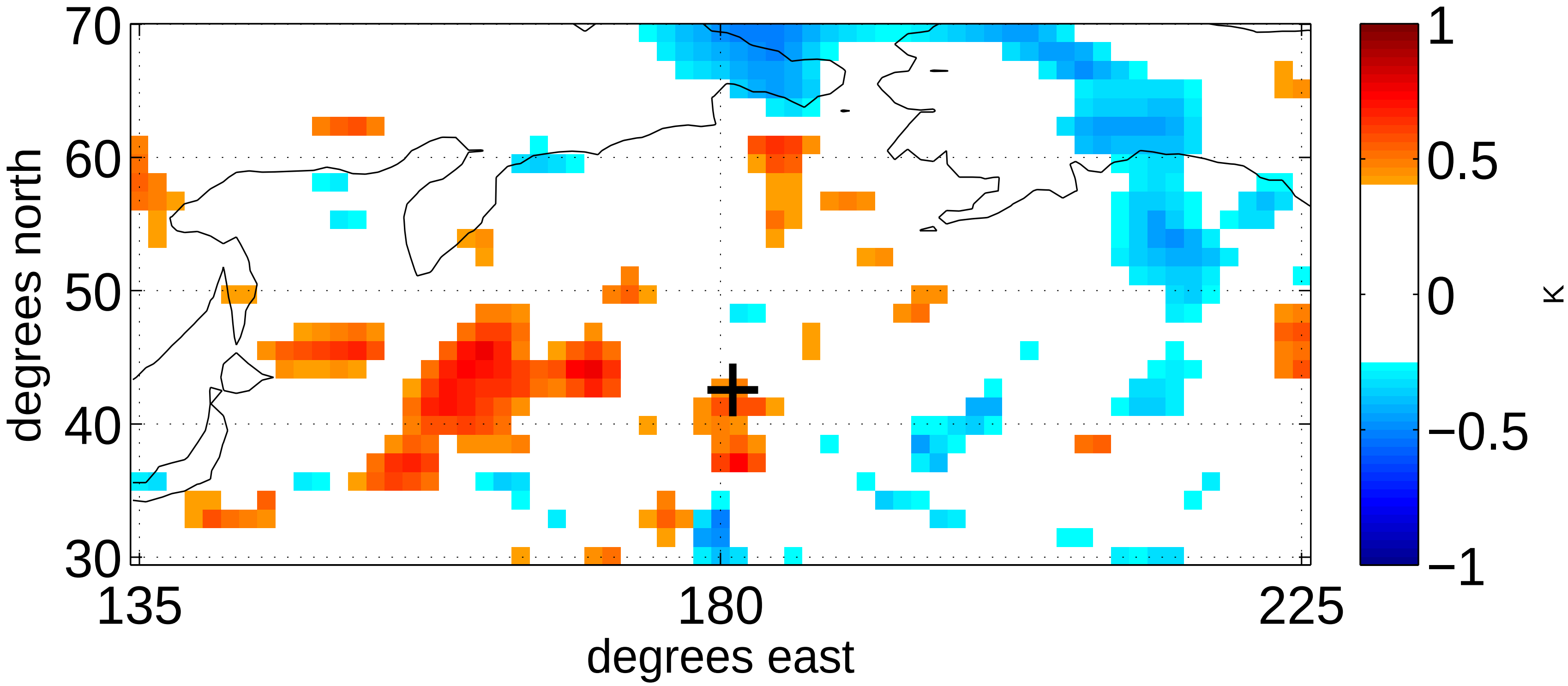
Backward in Time Sensitivity (Adjoint equivalent)

Time lag -18 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



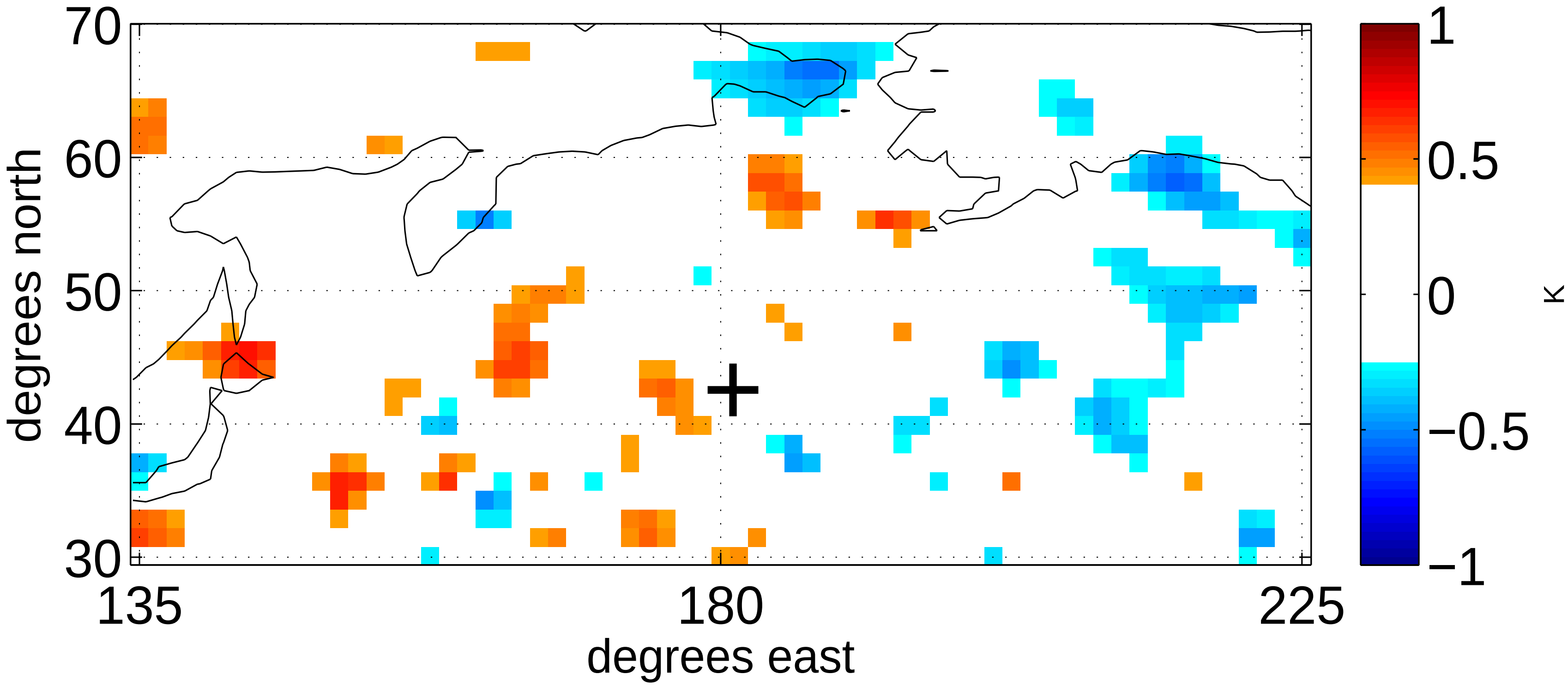
Backward in Time Sensitivity (Adjoint equivalent)

Time lag -24 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



Backward in Time Sensitivity (Adjoint equivalent)

Time lag -30 hours: 500 hPa Zonal Velocity to 500 hPa Temperature



Assimilation increases information about all three pieces:

1. Get an improved estimate of state of physical system.

Initial conditions for forecasts.
Includes time evolution and ‘balances’.
High quality analyses (re-analyses).

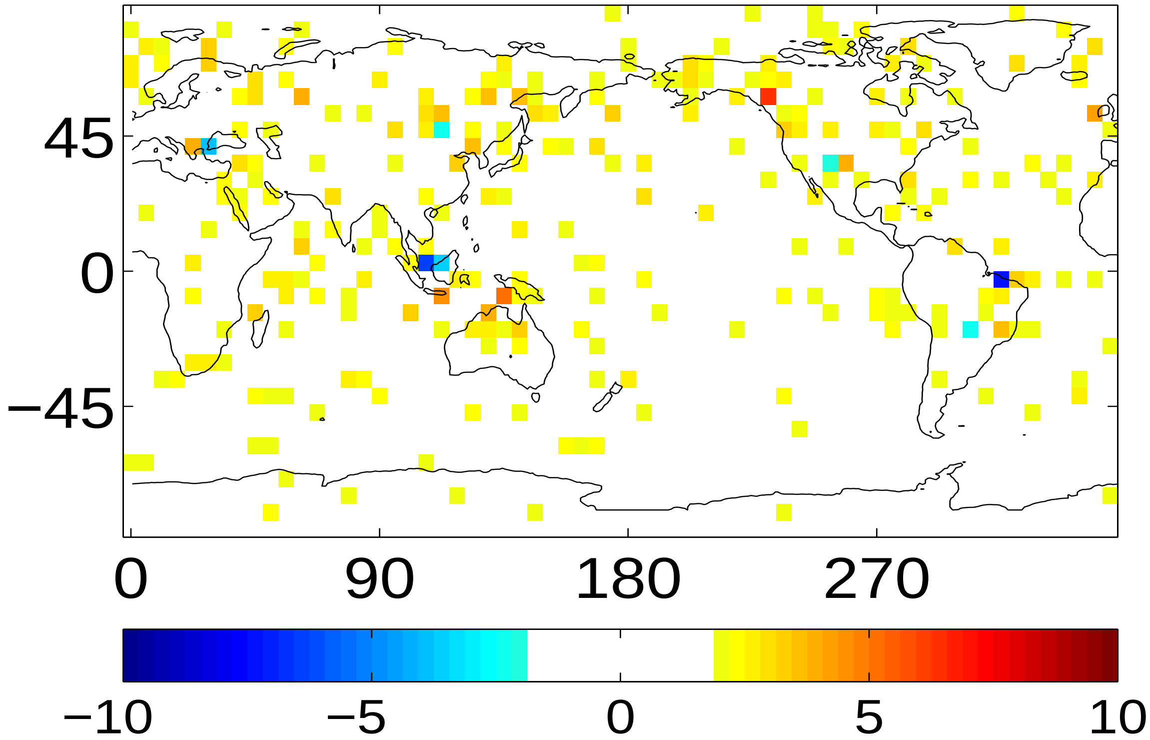
2. Get better estimates of observing system error characteristics.

Estimate value of existing or planned observations.
Design observing systems that provide increased information.

3. Improve model of physical system.

Evaluate model systematic errors.
Forward and backward sensitivity analysis (adjoint and linear tangent replacement).
Select appropriate values for model parameters.

Climate Model Parameter Estimation via Ensemble Data Assimilation.



T21 CAM assimilation of gravity wave drag efficiency parameter.

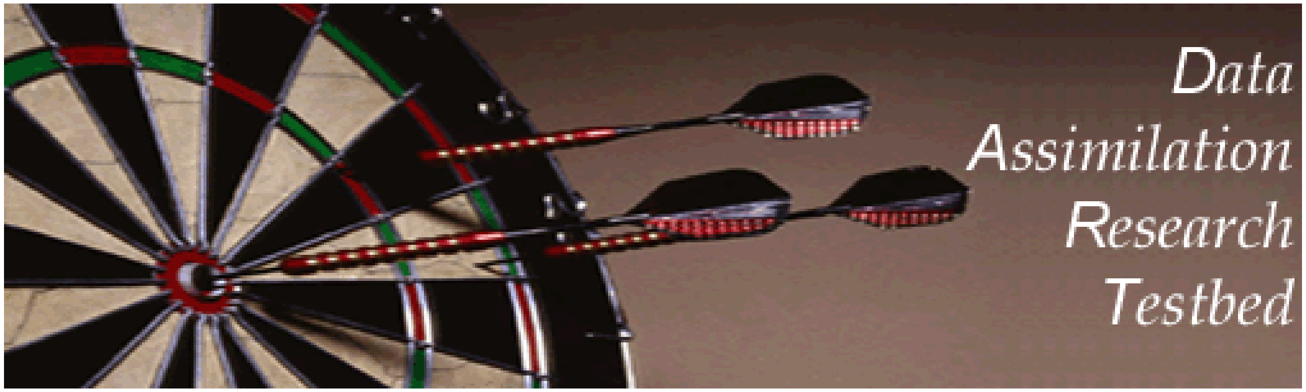
Oceanic values are noise (should be 0).

0< efficiency< ~4 suggested by modelers.

Positive values over NH land expected.
Problem: large negative values over tropical land near convection.
May reduce wind bias in tropical troposphere, but for ‘Wrong Reason’.

Assimilation tries to use free parameter to fix ALL model problems

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/.