Ensemble Data Assimilation for Climate System Component Models

Jeffrey Anderson

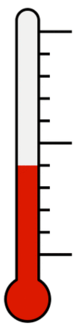
National Center for Atmospheric Research

In collaboration with:

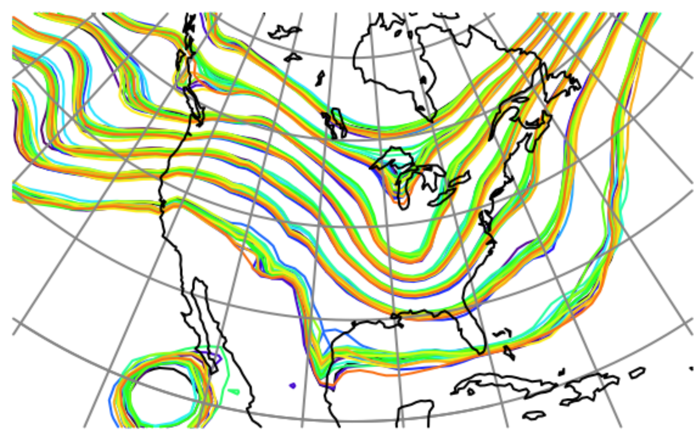
Alicia Karspeck, Kevin Raeder, Tim Hoar, Nancy Collins

What is Data Assimilation?

Observations combined with a Model forecast…



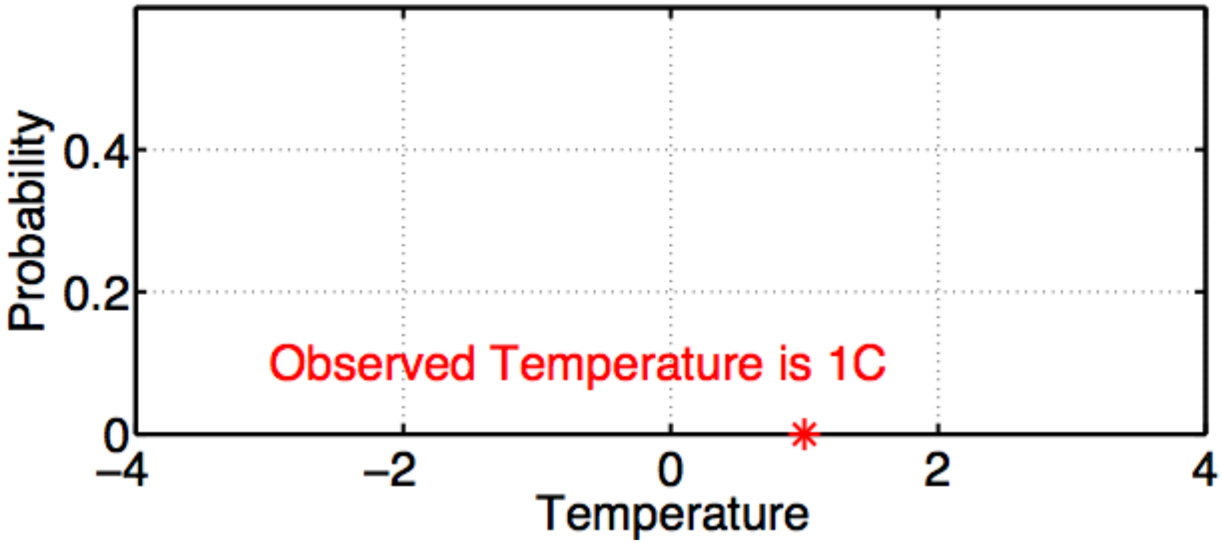
+



…to produce an analysis
(best possible estimate).

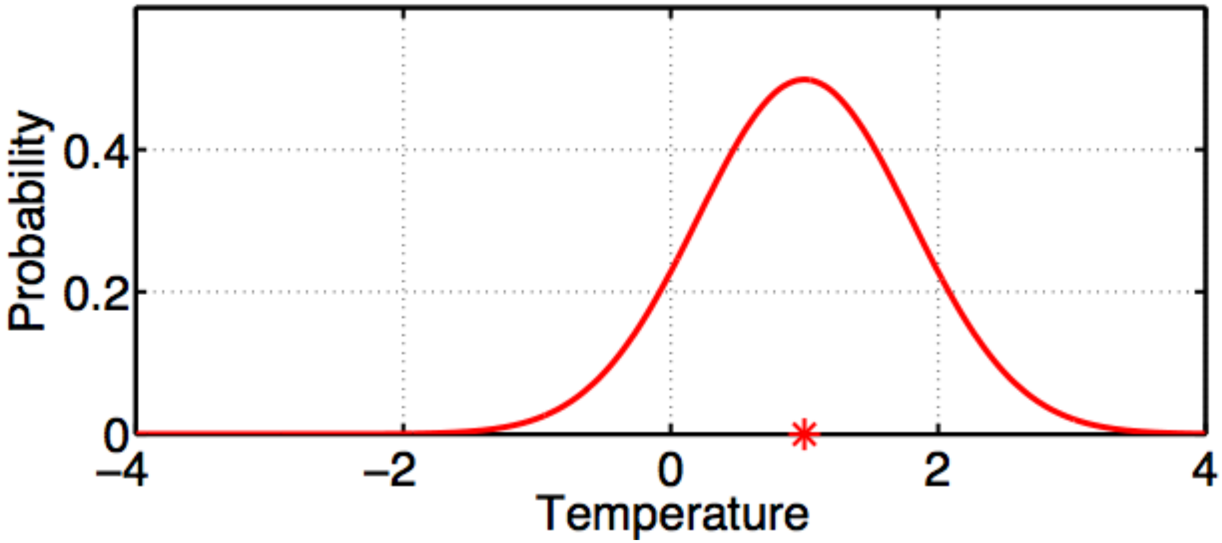
Example: Estimating the Temperature Outside

An observation has a value (*),



Example: Estimating the Temperature Outside

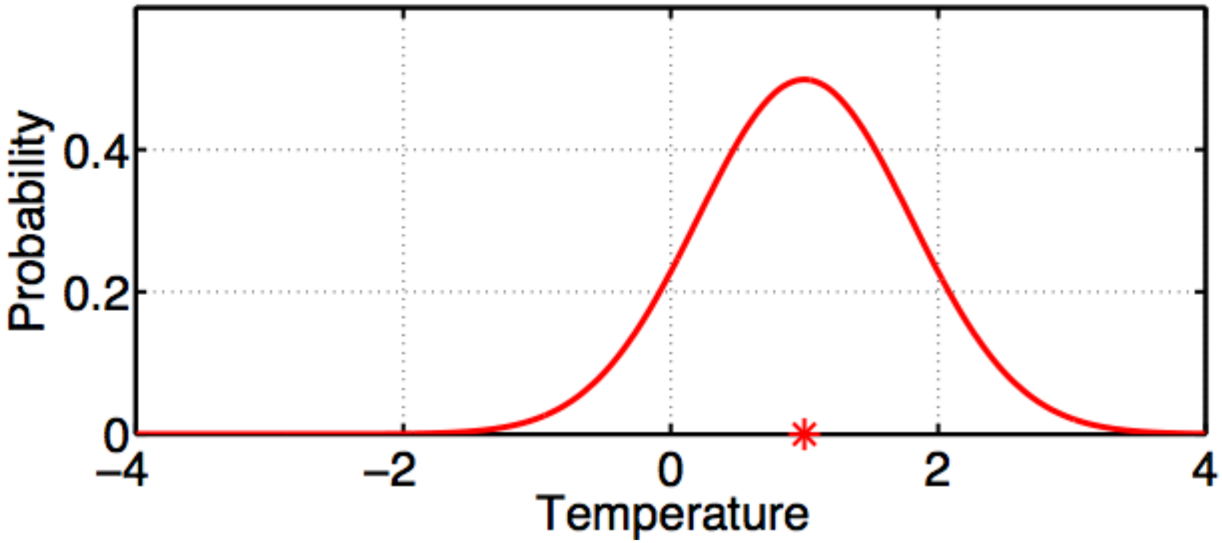
An observation has a value (*),

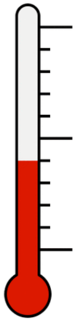


and an error distribution (red curve) that is associated with the instrument.

Example: Estimating the Temperature Outside

Thermometer outside measures 1C.

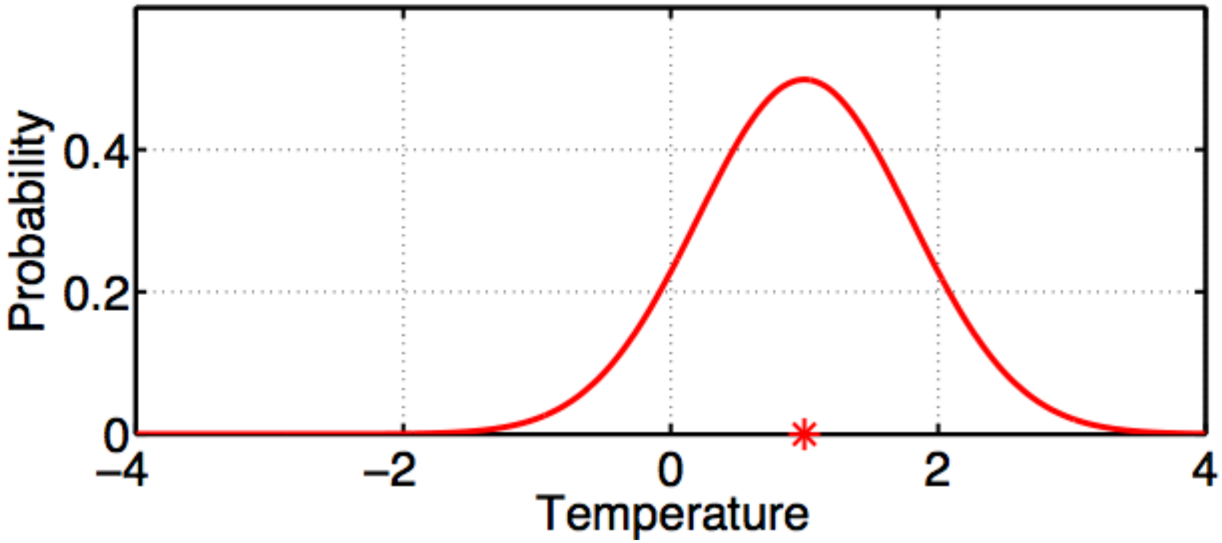


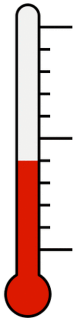


Instrument builder says thermometer is unbiased with +/- 0.8C gaussian error.

Example: Estimating the Temperature Outside

Thermometer outside measures 1C.

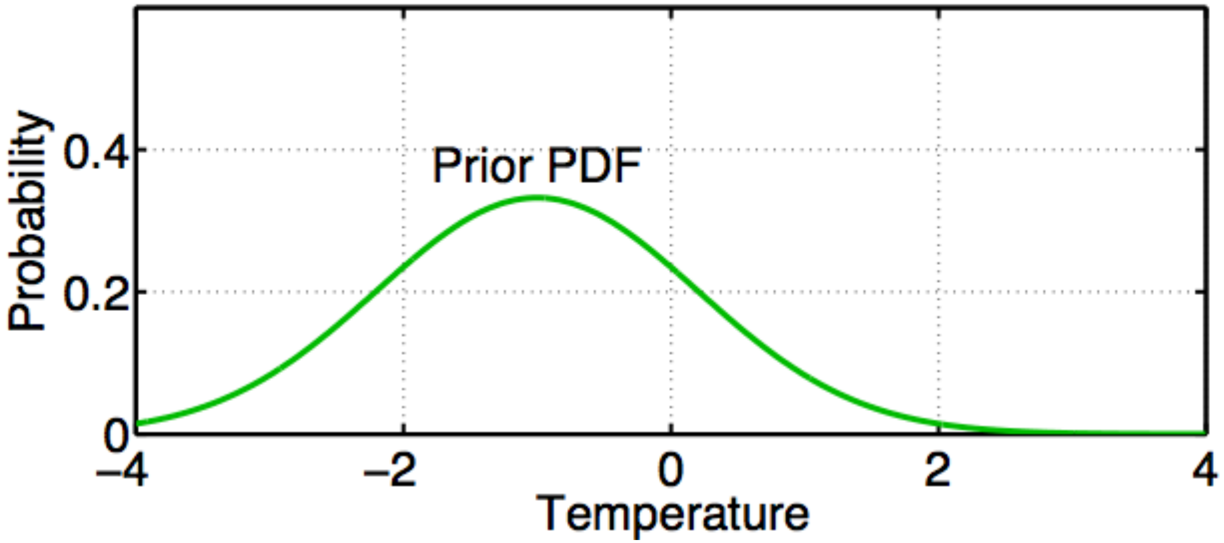




The red plot is $P(T \mid T_o)$, probability of temperature given that T_o was observed.

Example: Estimating the Temperature Outside

We also have a prior estimate of temperature.



The green curve is $P(T \mid C)$; probability of temperature given all available prior information C.

Example: Estimating the Temperature Outside

Prior information C can include:

1. Observations of things besides T;
2. Model forecast made using observations at earlier times;
3. *A priori* physical constraints (T > -273.15C);
4. Climatological constraints (-30C < T < 40C).

Combining the Prior Estimate and Observation

Bayes Theorem:

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{Normalization}$$

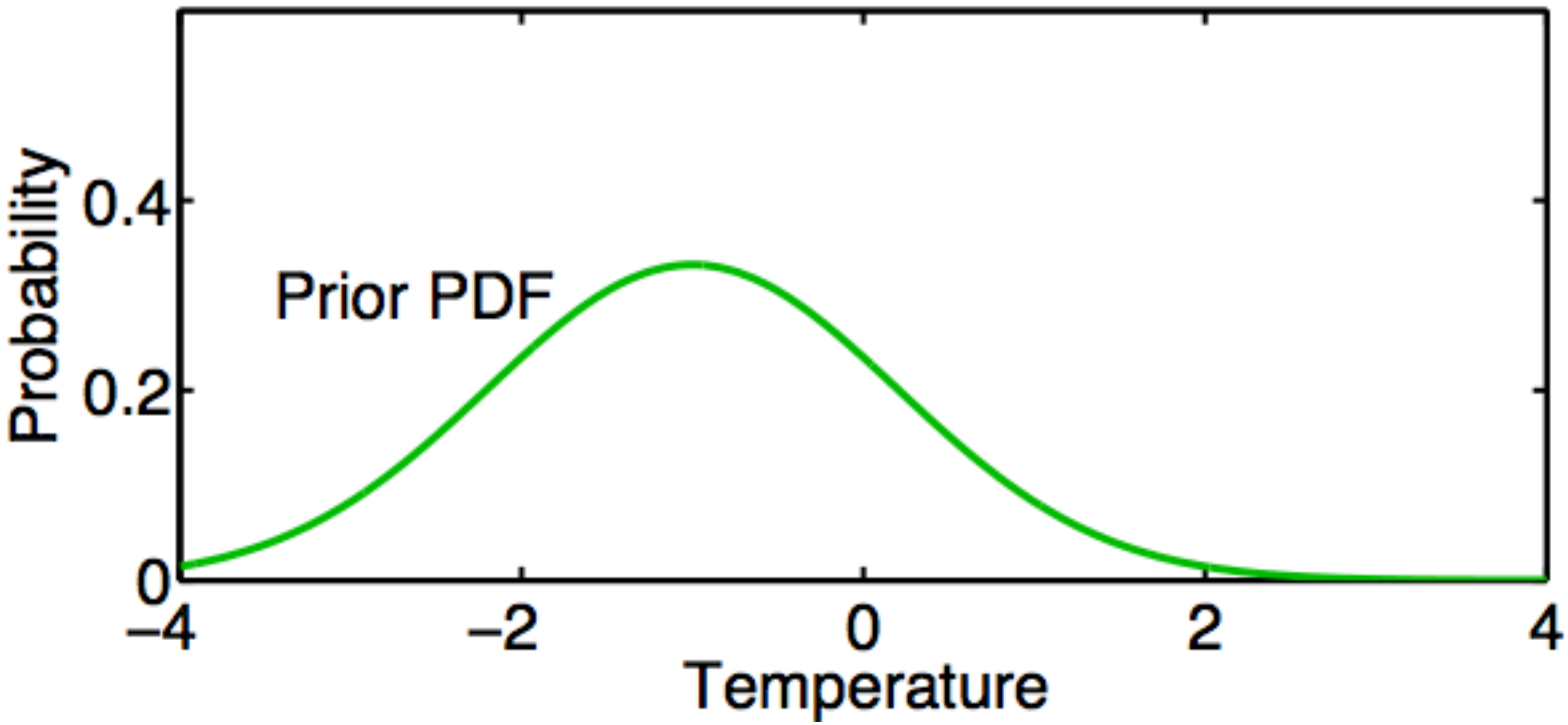
Prior

Posterior: Probability of T given observations and Prior. Also called update or analysis.

Likelihood: Probability that T_o is observed if T is true value and given prior information C.

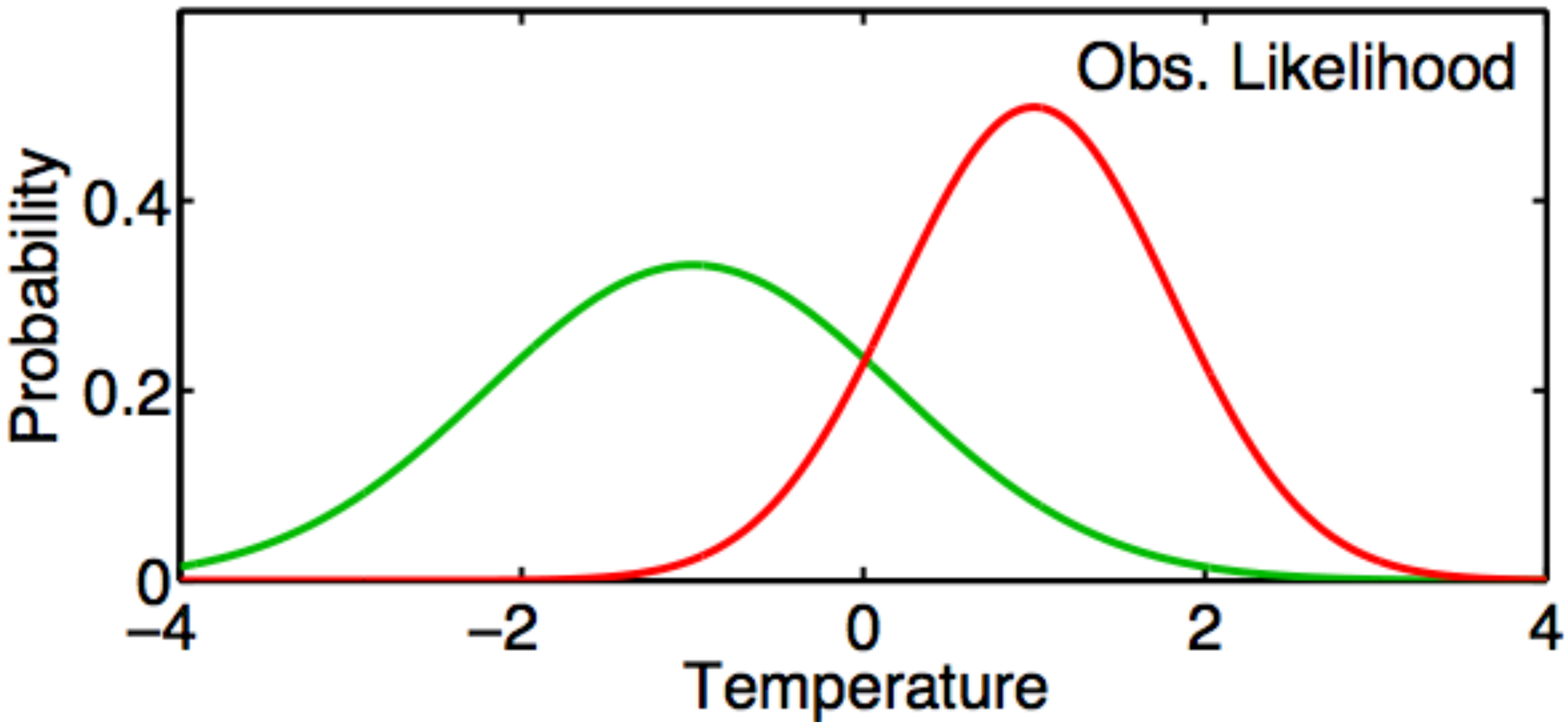
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



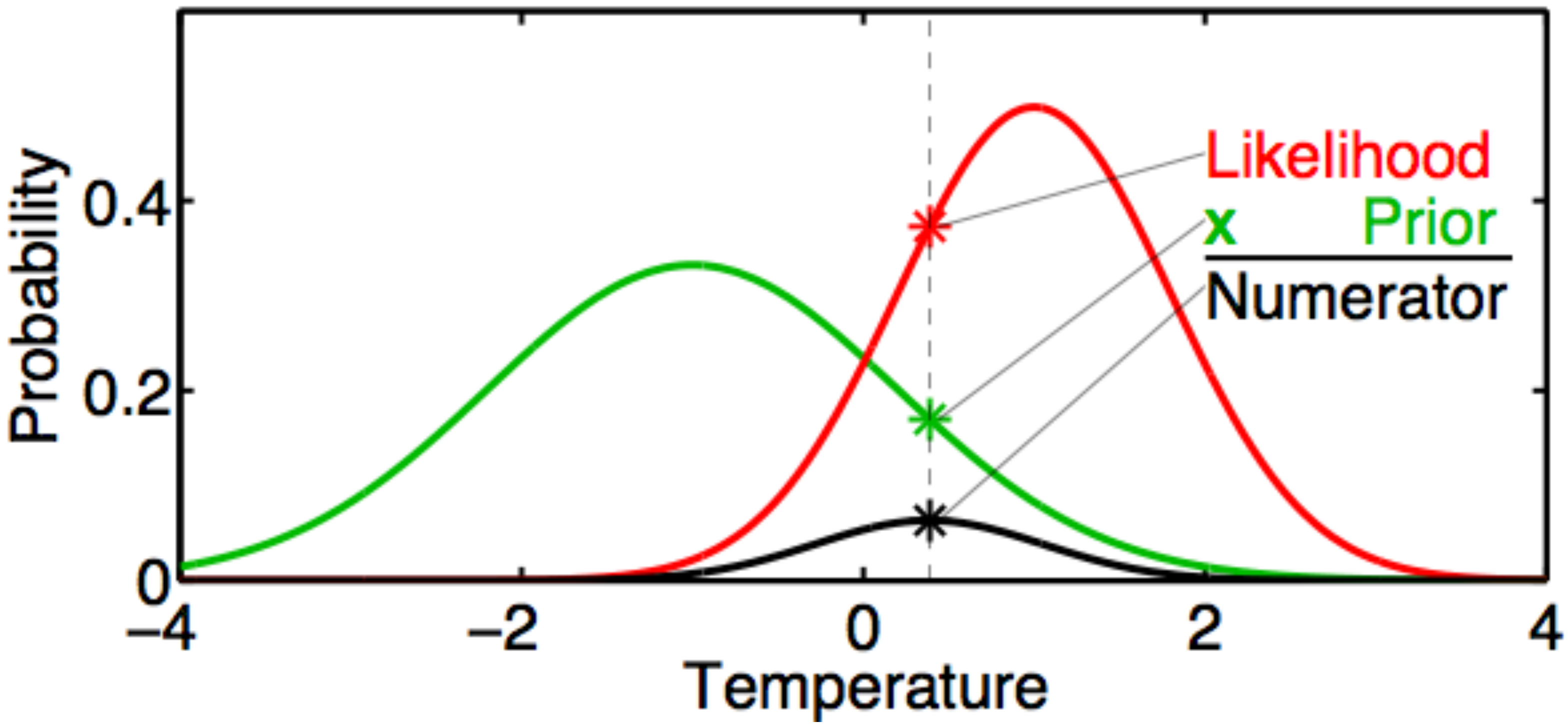
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



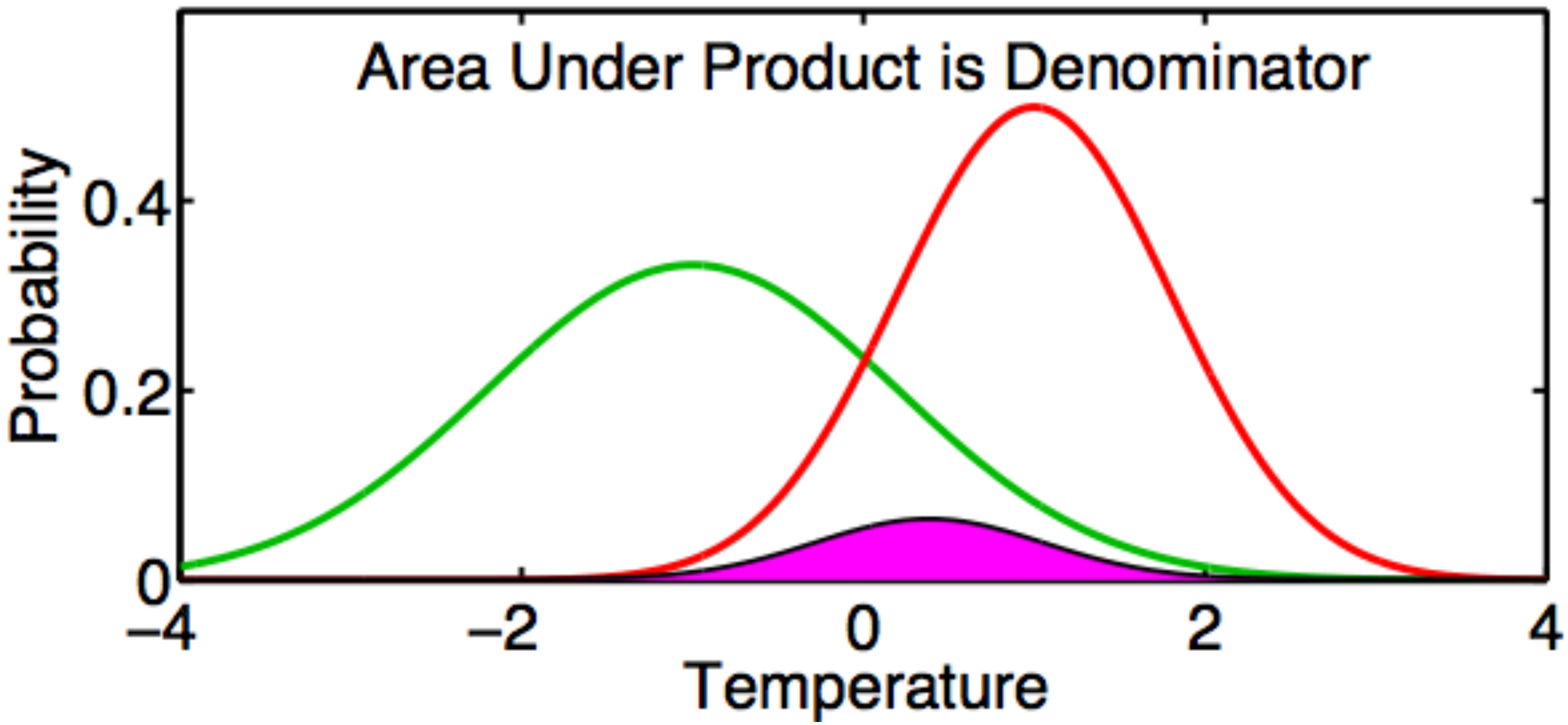
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



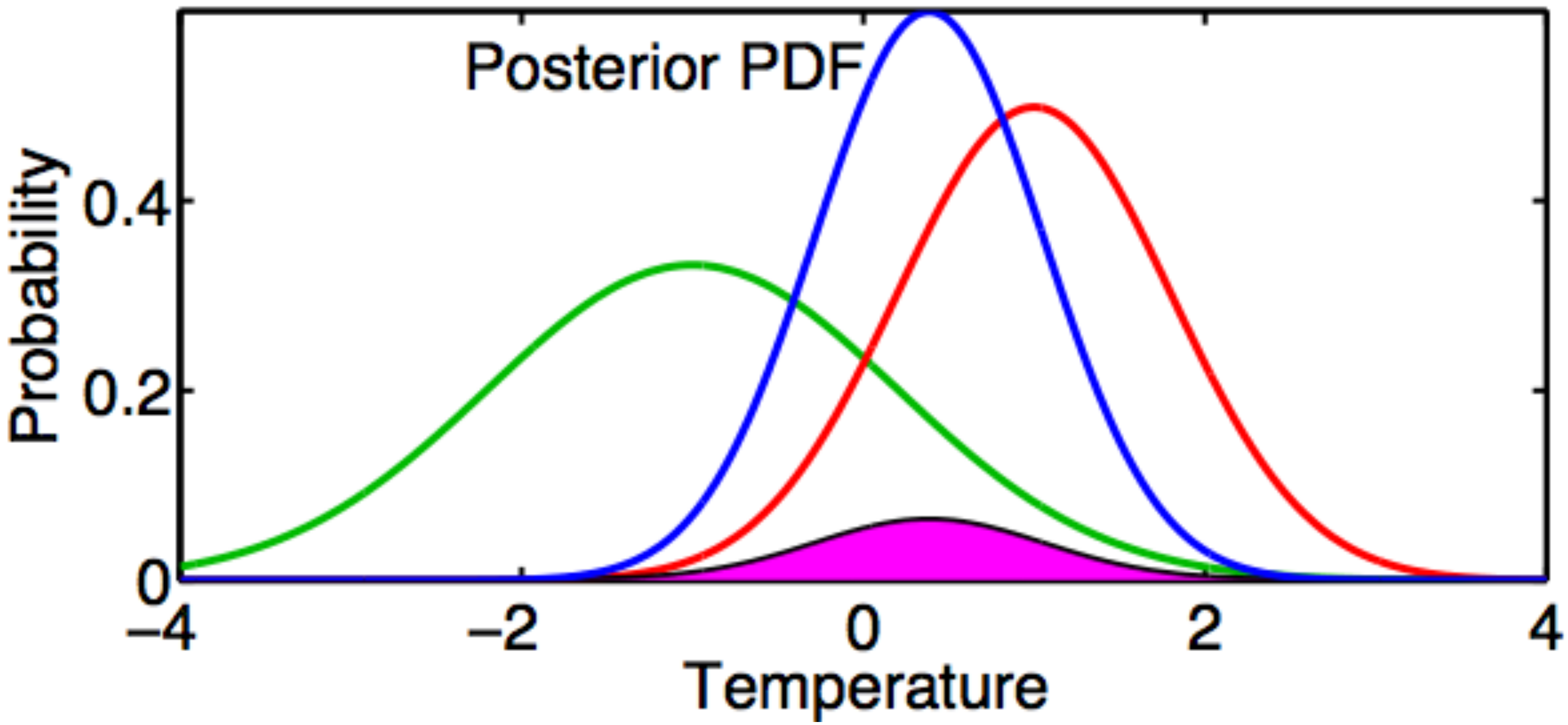
Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



Combining the Prior Estimate and Observation

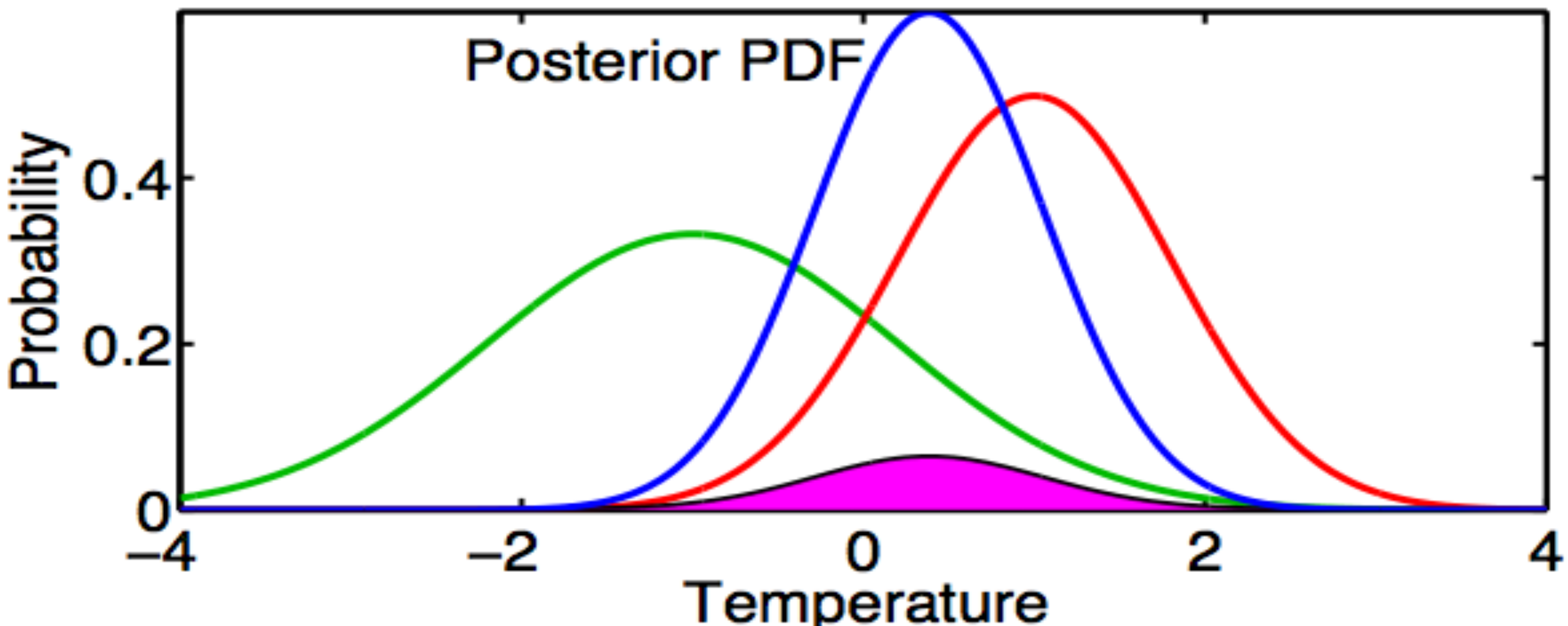
$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$



Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\, P(T \mid C)}{normalization}$$

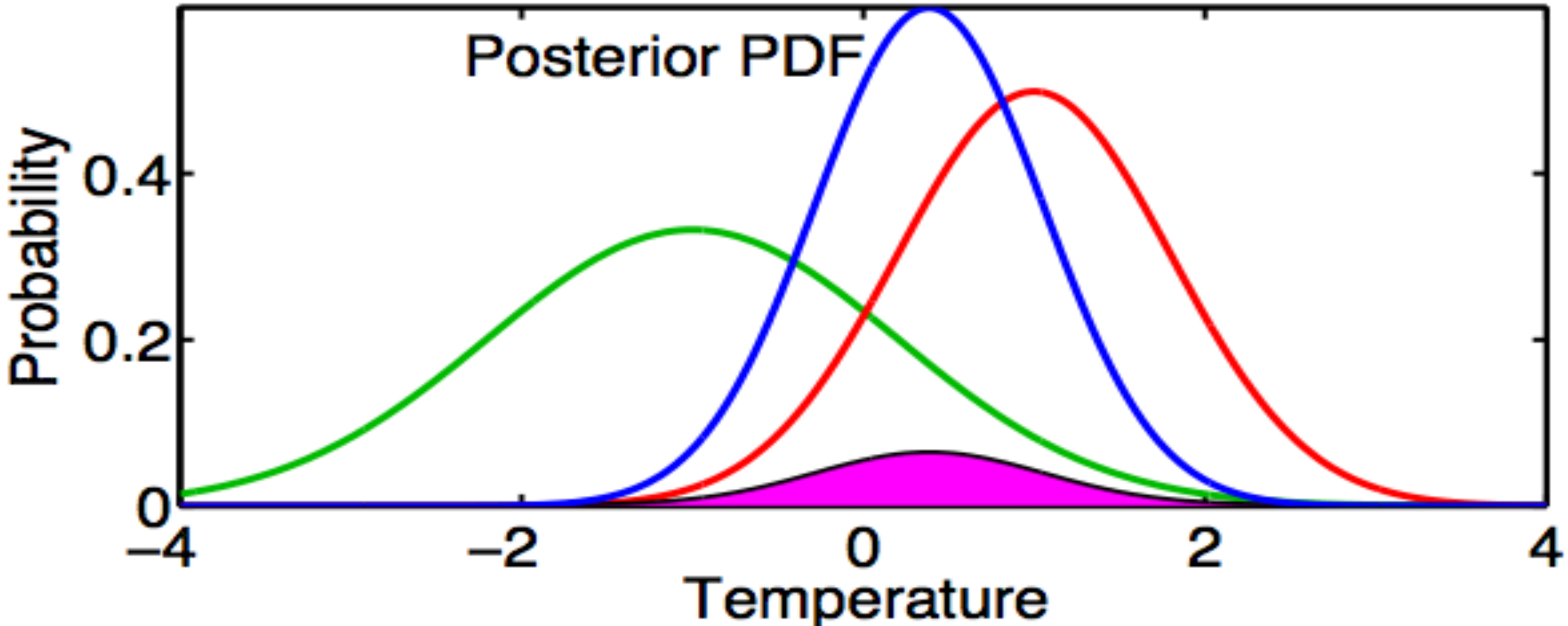
Generally no analytic solution for Posterior.



Combining the Prior Estimate and Observation

$$P(T \mid T_o, C) = \frac{P(T_o \mid T, C)\,P(T \mid C)}{normalization}$$

Gaussian Prior and Likelihood -> Gaussian Posterior



Combining the Prior Estimate and Observation

For Gaussian prior and likelihood…

Prior $$P(T \mid C) = Normal(T_p, \sigma_p)$$

Likelihood $$P(T_o \mid T, C) = Normal(T_o, \sigma_o)$$

Then, Posterior $$P(T \mid T_o, C) = Normal(T_u, \sigma_u)$$

With

$$\sigma_u = \sqrt{\left(\sigma_p^{-2} + \sigma_o^{-2}\right)^{-1}}$$

$$T_u = \sigma_u^2\left[\sigma_p^{-2} T_p + \sigma_o^{-2} T_o\right]$$

The One-Dimensional Kalman Filter

1. Suppose we have a linear forecast model L

   A. If temperature at time $t_1 = T_1$, then temperature at $t_2 = t_1 + \Delta t$ is $T_2 = L(T_1)$

   B. Example: $T_2 = T_1 + \Delta t T_1$

The One-Dimensional Kalman Filter

1. Suppose we have a linear forecast model L.

   A. If temperature at time $t_1 = T_{1,}$ then
      temperature at $t_2 = t_1 + \Delta t$ is $T_2 = L(T_1)$.

   B. Example: $T_2 = T_1 + \Delta t T_1$.

2. If posterior estimate at time t_1 is *Normal*$(T_{u,1}, \sigma_{u,1})$ then
   prior at t_2 is *Normal*$(T_{p,2}, \sigma_{p,2})$.

   $T_{p,2} = T_{u,1}, + \Delta t T_{u,1}$

   $\sigma_{p,2} = (\Delta t + 1)\, \sigma_{u,1}$

The One-Dimensional Kalman Filter

1. Suppose we have a linear forecast model L.

   A. If temperature at time $t_1 = T_1$, then
      temperature at $t_2 = t_1 + \Delta t$ is $T_2 = L(T_1)$.

   B. Example: $T_2 = T_1 + \Delta t T_1$.

2. If posterior estimate at time t_1 is *Normal*($T_{u,1}$, $\sigma_{u,1}$) then
   prior at t_2 is *Normal*($T_{p,2}$, $\sigma_{p,2}$).

3. Given an observation at t_2 with distribution *Normal*(t_o, σ_o)
   the likelihood is also *Normal*(t_o, σ_o).

The One-Dimensional Kalman Filter

1. Suppose we have a linear forecast model L.

   A. If temperature at time $t_1 = T_{1,}$ then temperature at $t_2 = t_1 + \Delta t$ is $T_2 = L(T_1)$.

   B. Example: $T_2 = T_1 + \Delta t T_1$.

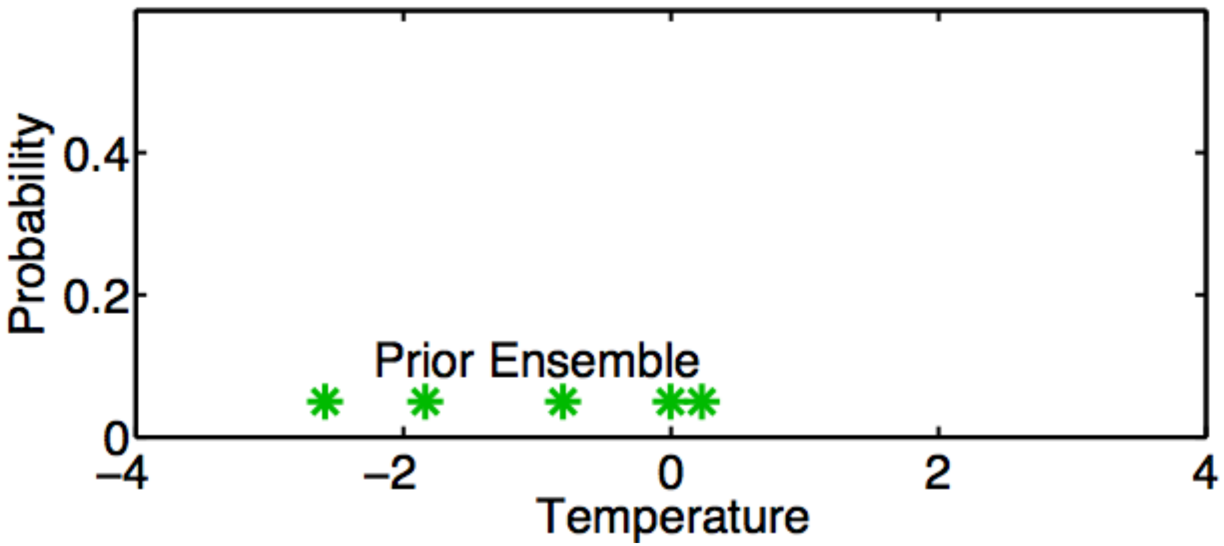
2. If posterior estimate at time t_1 is *Normal*$(T_{u,1}, \sigma_{u,1})$ then prior at t_2 is *Normal*$(T_{p,2}, \sigma_{p,2})$.

3. Given an observation at t_2 with distribution *Normal*(t_o, σ_o) the likelihood is also *Normal*(t_o, σ_o).

4. The posterior at t_2 is *Normal*$(T_{u,2}, \sigma_{u,2})$ where $T_{u,2}$ and $\sigma_{u,2}$ come from page 17.

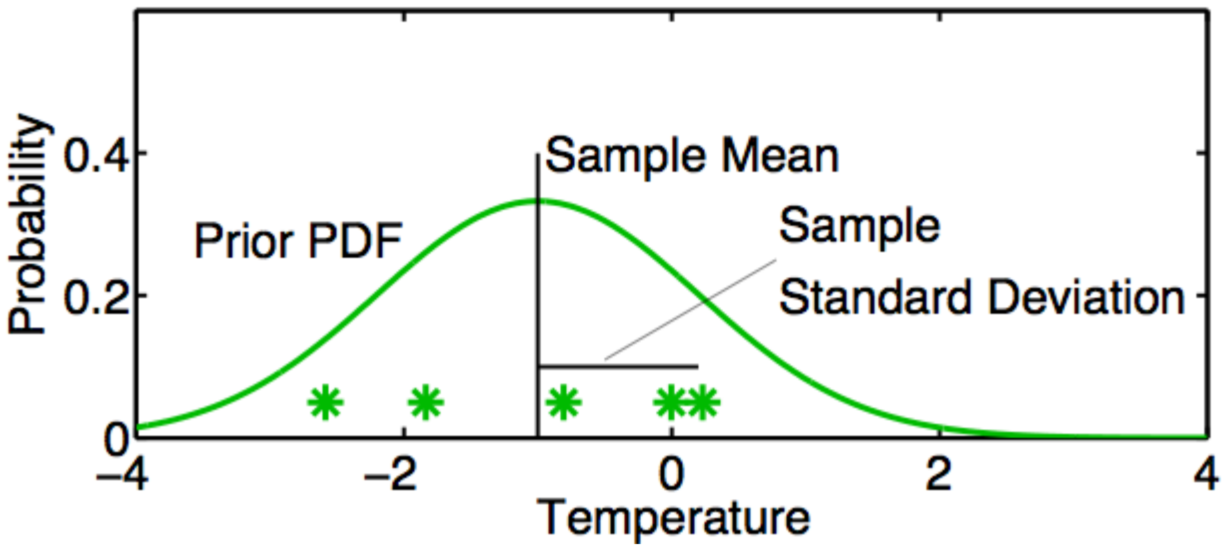
A One-Dimensional Ensemble Kalman Filter

Represent a prior pdf by a sample (ensemble) of N values:



A One-Dimensional Ensemble Kalman Filter

Represent a prior pdf by a sample (ensemble) of N values:



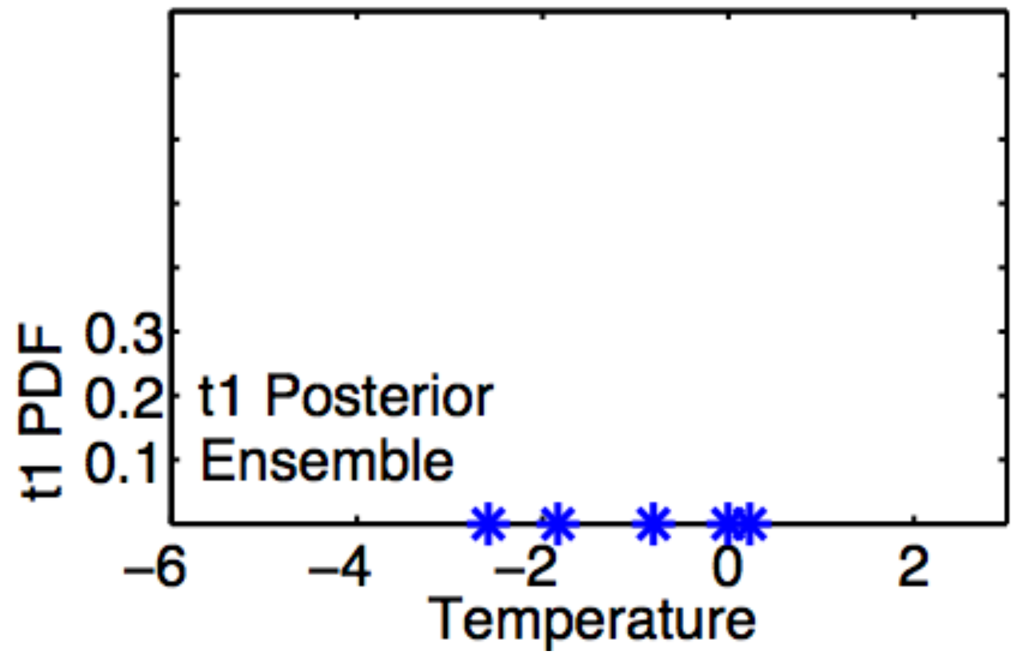
Use sample mean $\overline{T} = \sum_{n=1}^{N} T_n / N$

and sample standard deviation $\sigma_T = \sqrt{\sum_{n=1}^{N} \left(T_n - \overline{T}\right)^2 / (N-1)}$

to determine a corresponding continuous distribution $Normal\left(\overline{T}, \sigma_T\right)$

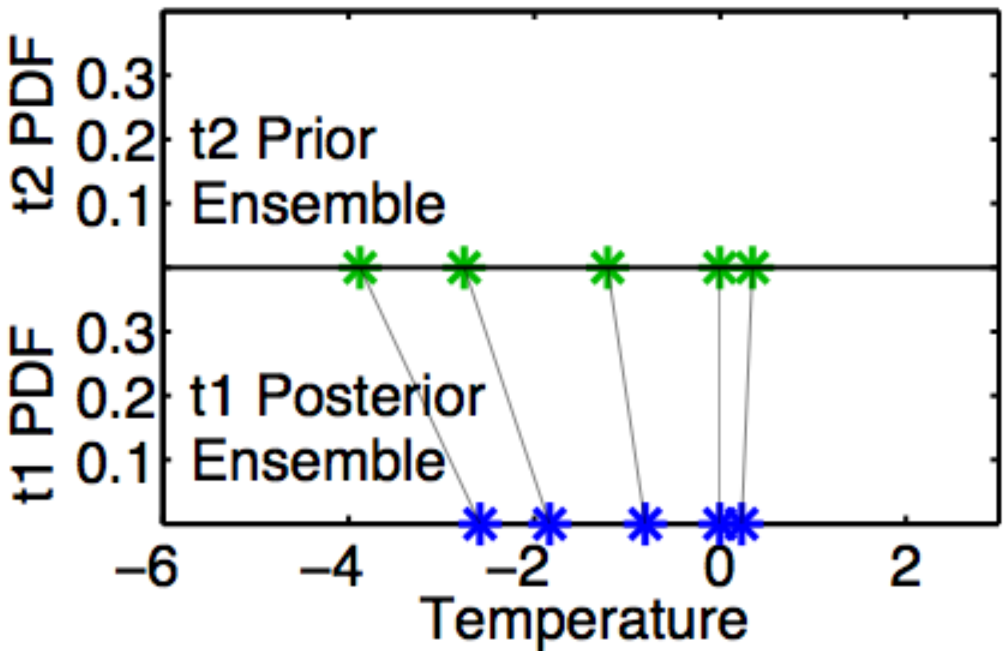
A One-Dimensional Ensemble Kalman Filter: Model Advance

If posterior ensemble at time t_1 is $T_{1,n}$, $n = 1, \ldots, N$



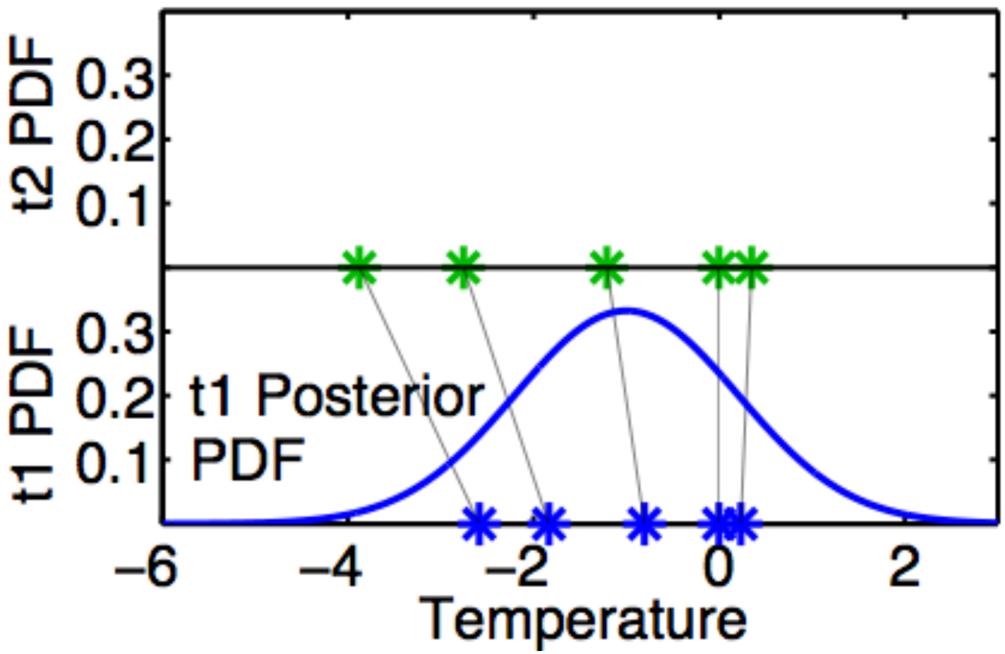
A One-Dimensional Ensemble Kalman Filter: Model Advance

If posterior ensemble at time t_1 is $T_{1,n}$, $n = 1, \ldots, N$,
advance each member to time t_2 with model, $T_{2,n} = L(T_{1,n})$ $n = 1, \ldots, N$.



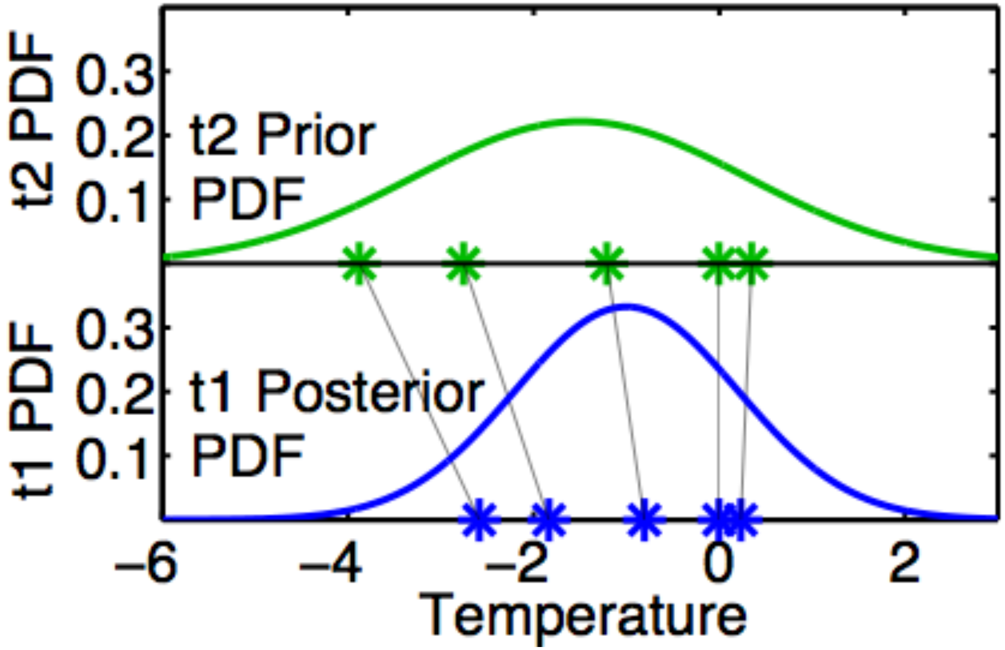
A One-Dimensional Ensemble Kalman Filter: Model Advance

Same as advancing continuous pdf at time t_1 …

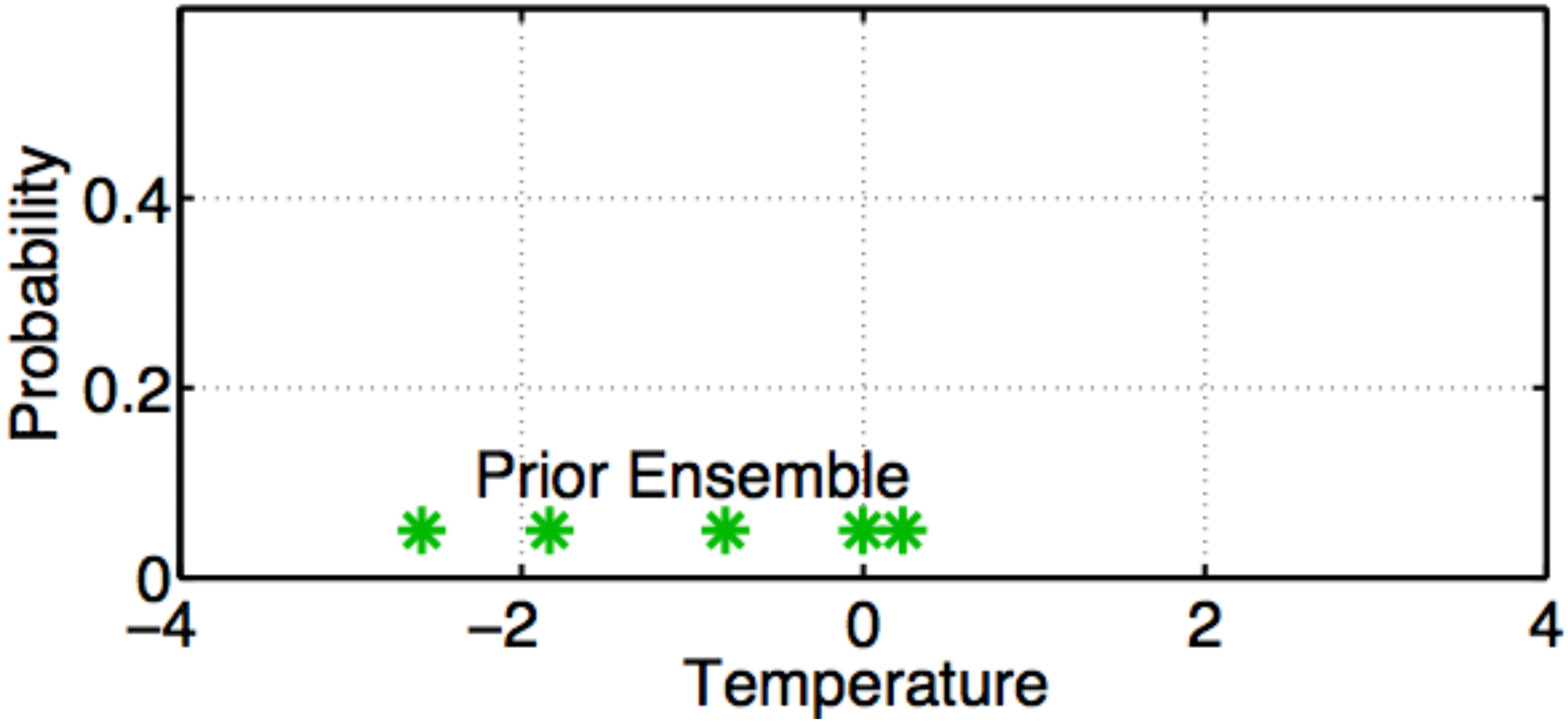


A One-Dimensional Ensemble Kalman Filter: Model Advance

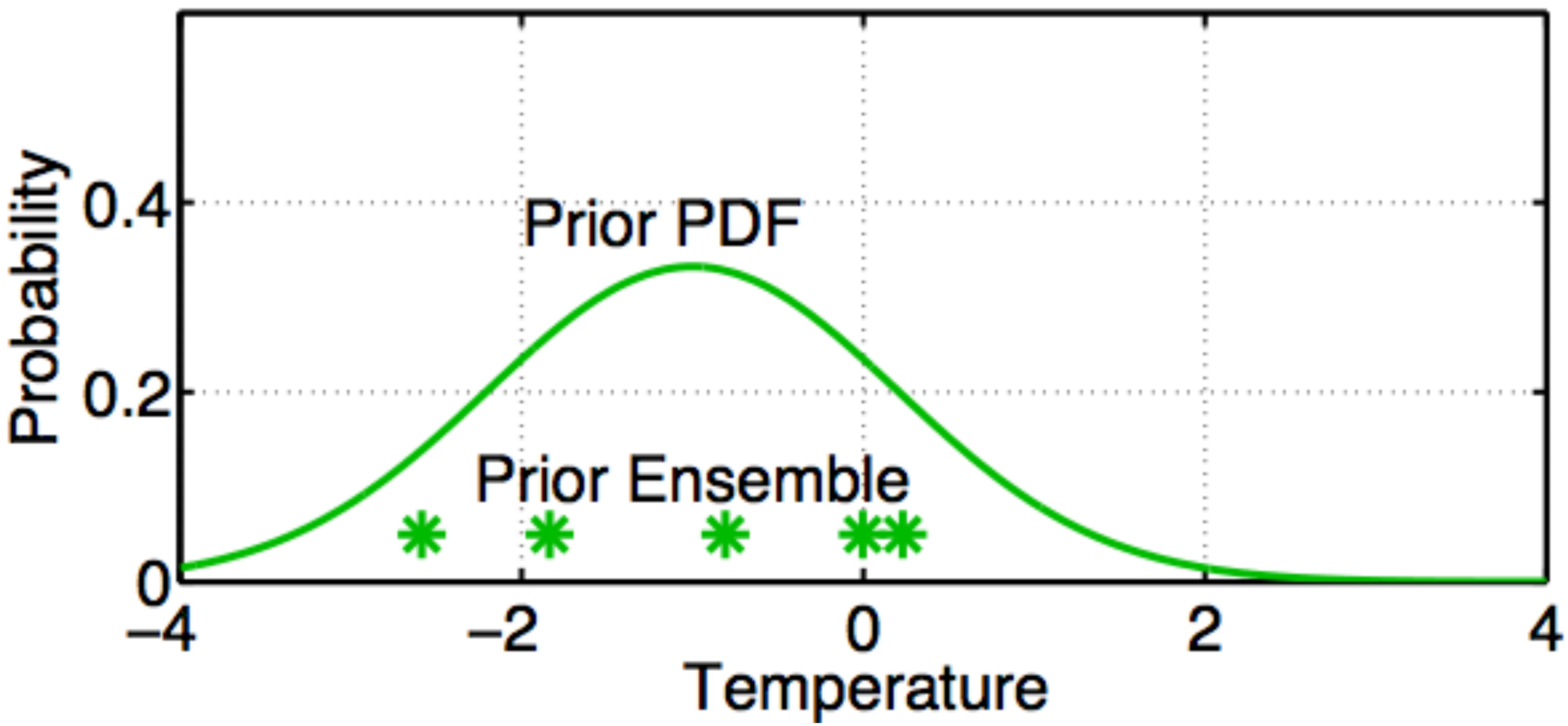
Same as advancing continuous pdf at time t_1 to time t_2 with model L.



A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation

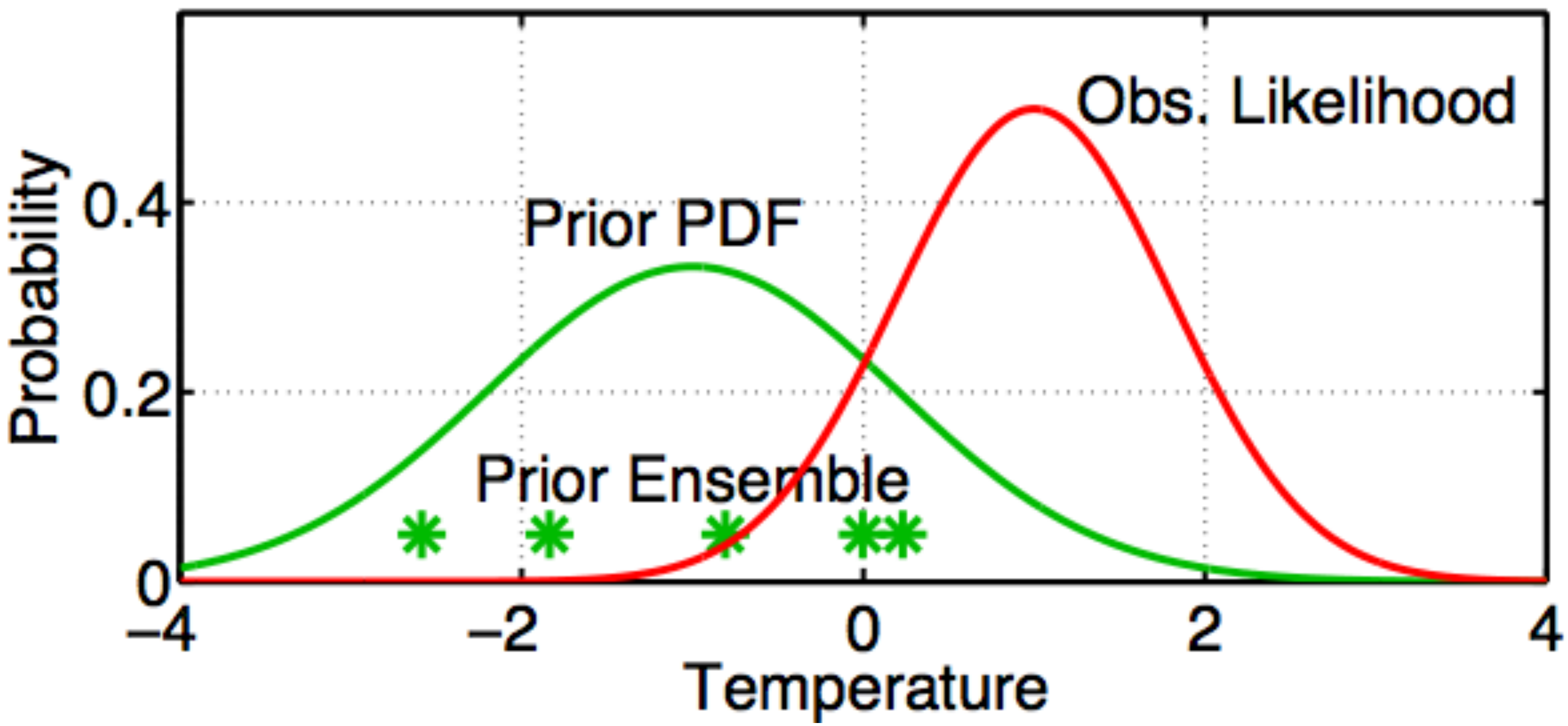


A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



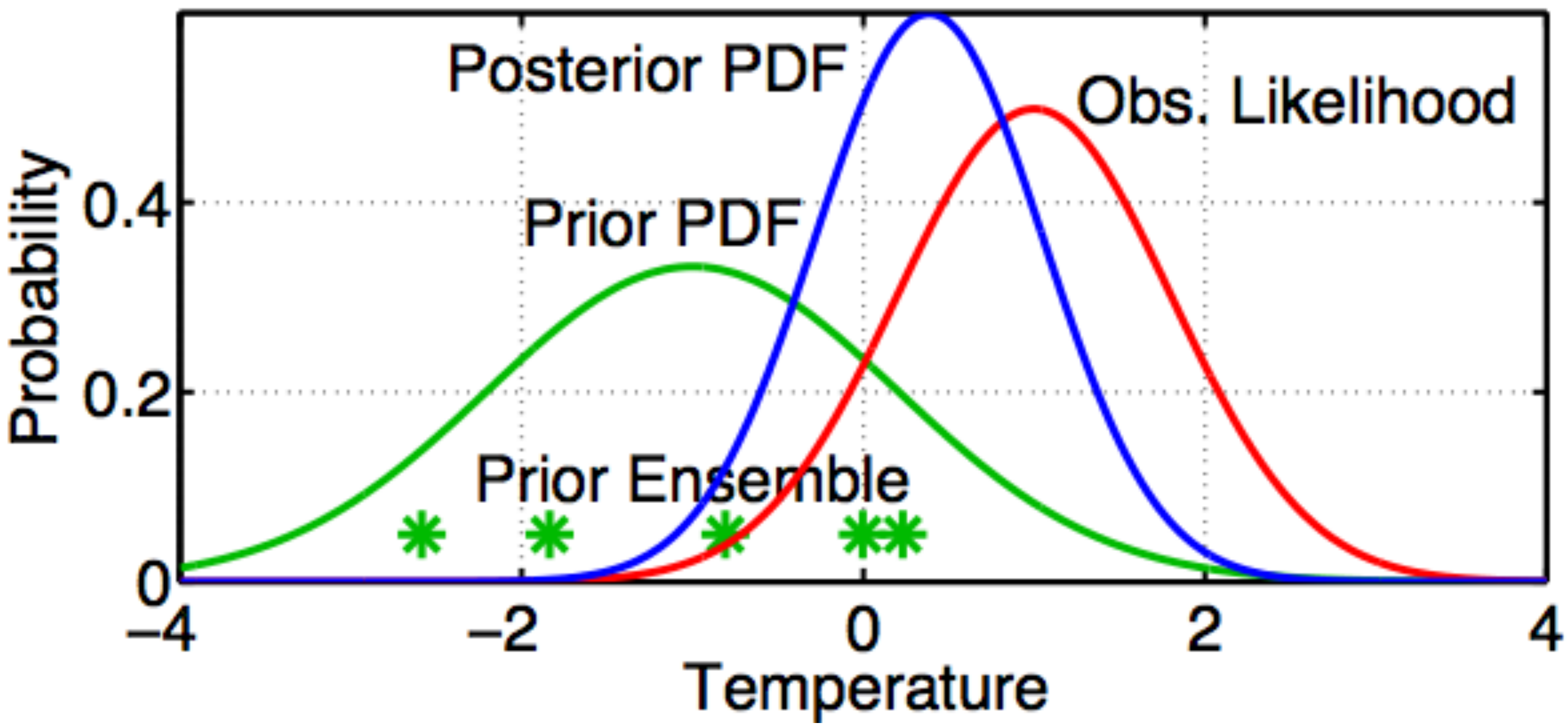
Fit a Gaussian to the sample.

A One-Dimensional Ensemble Kalman Filter:
Assimilating an Observation



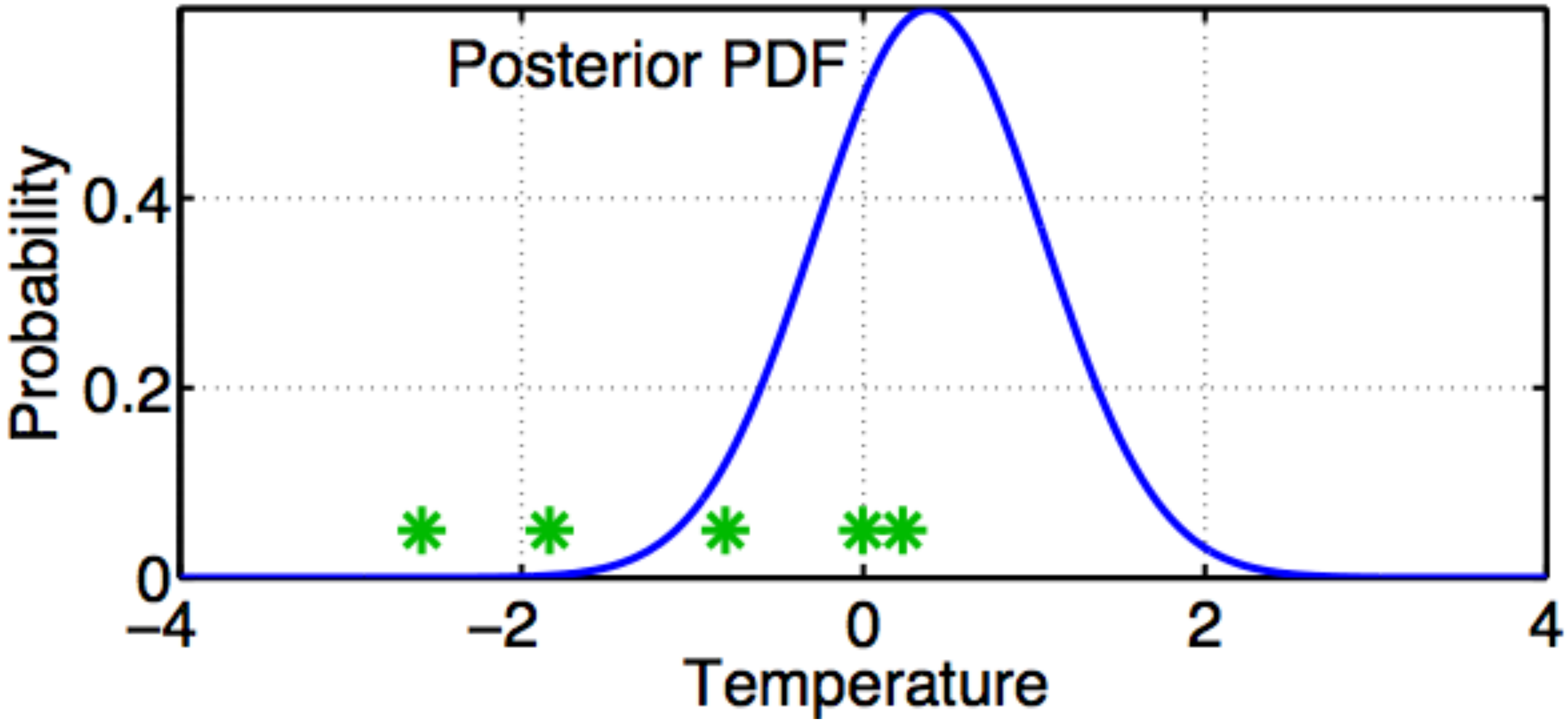
Get the observation likelihood.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



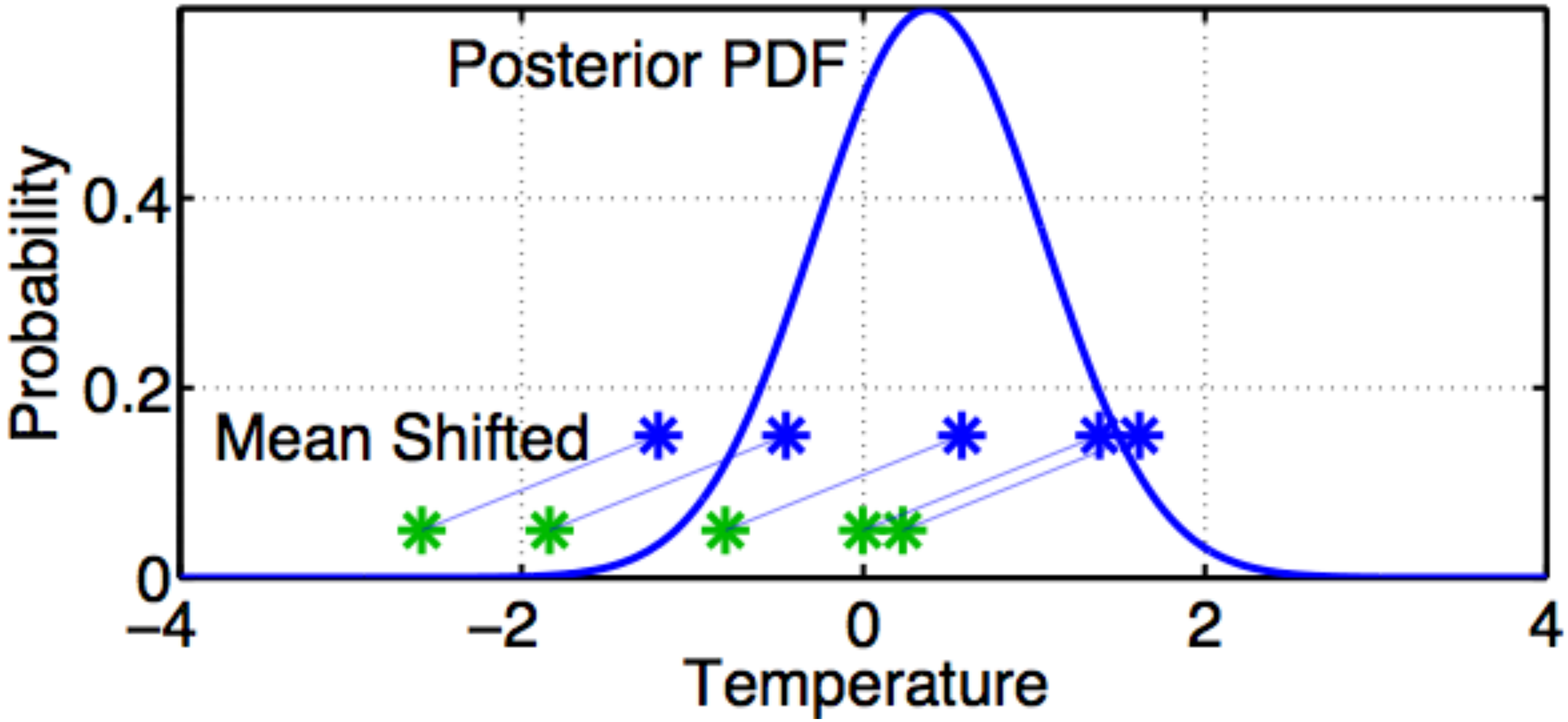
Compute the continuous posterior PDF.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



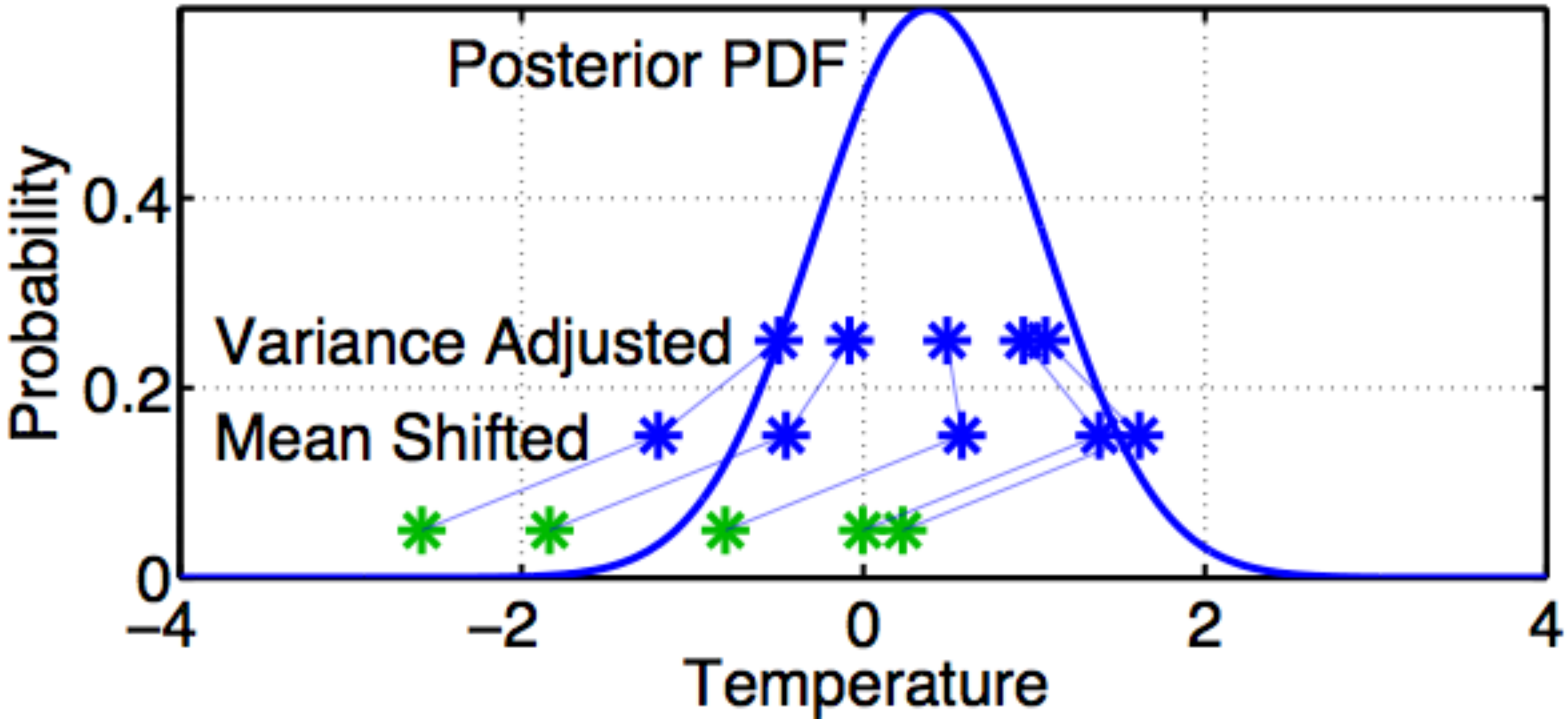
Use a deterministic algorithm to ‘adjust’ the ensemble.

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



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

A One-Dimensional Ensemble Kalman Filter: Assimilating an Observation



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

Second, linearly contract to have the exact variance of the posterior.

Sample statistics are identical to Kalman filter.

Initial Comments on (Ensemble) Kalman Filter

- KF optimal for linear model, gaussian likelihood.
- In KF, only mean and variance have meaning.
- The deterministic Ensemble KF gives identical mean, variance.

- The original Ensemble KF uses a Monte Carlo algorithm for the observation impact; has sampling error.

- Ensemble allows computation of many other statistics.
- What do they mean? Not entirely clear.

- Example: Kurtosis. Completely constrained by initial ensemble.
  It is problem specific whether this is even defined!

Multivariate Kalman Filter

Product of d-dimensional normals is normal

$$N(\mu_1,\Sigma_1)N(\mu_2,\Sigma_2) = cN(\mu,\Sigma)$$

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

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

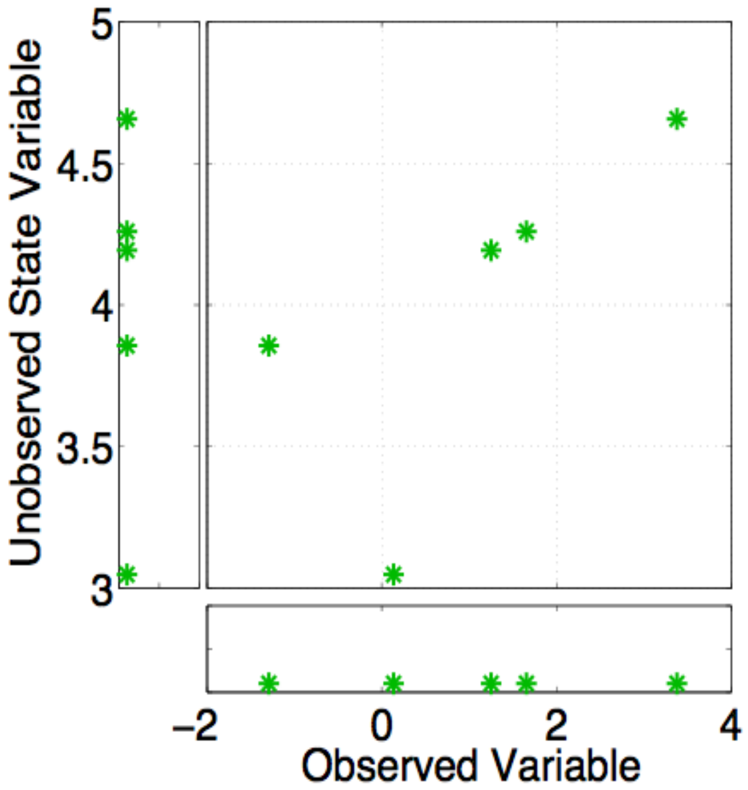
Weight: $c = \left[(2\pi)^{d/2}\left|\Sigma_1 + \Sigma_2\right|^{1/2}\right]^{-1} \exp\left\{-1/2\left[(\mu_2-\mu_1)^T(\Sigma_1+\Sigma_2)^{-1}(\mu_2-\mu_1)\right]\right\}$

Weight normalizes away.

Multivariate Ensemble Kalman Filter

- So far, we have an observation likelihood for single variable.
- Suppose the model prior has additional variables.
- KF equivalent to linear regression to update additional variables.
- Need ensemble size > d to represent d-dimensional normal.

Ensemble filters: Updating additional prior state variables

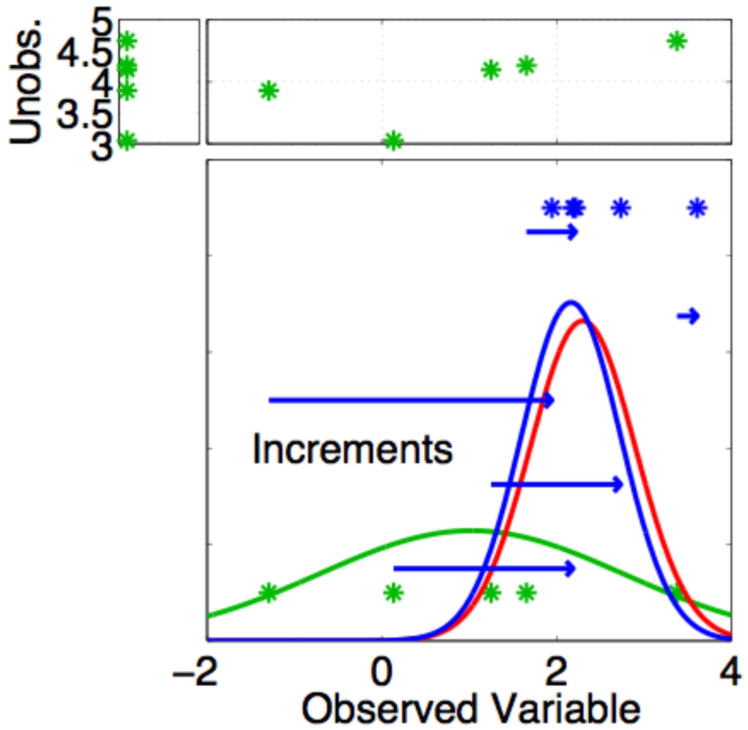


Assume that all we know is prior joint distribution.

One variable is observed.

What should happen to the unobserved variable?

Ensemble filters: Updating additional prior state variables

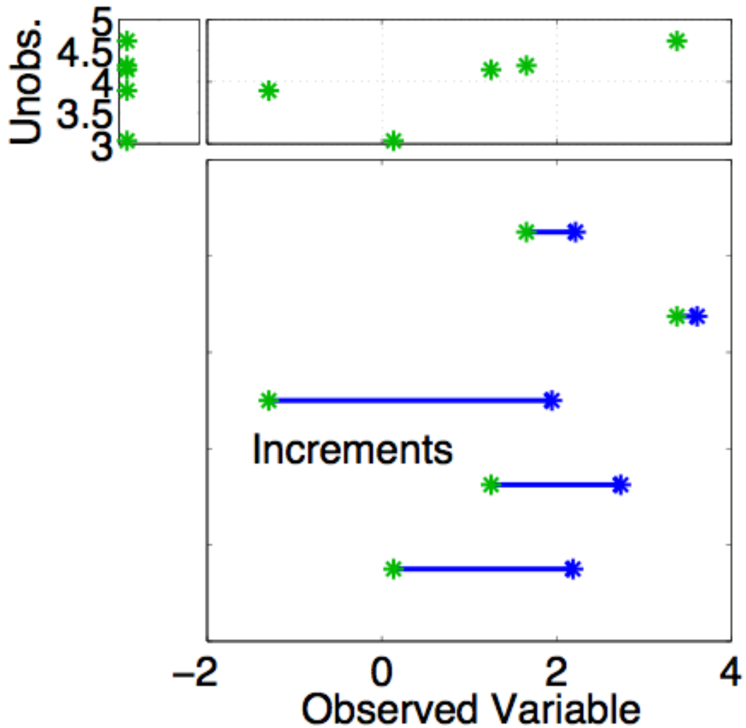


Assume that all we know is prior joint distribution.

One variable is observed.

Compute increments for prior ensemble members of observed variable.

Ensemble filters: Updating additional prior state variables

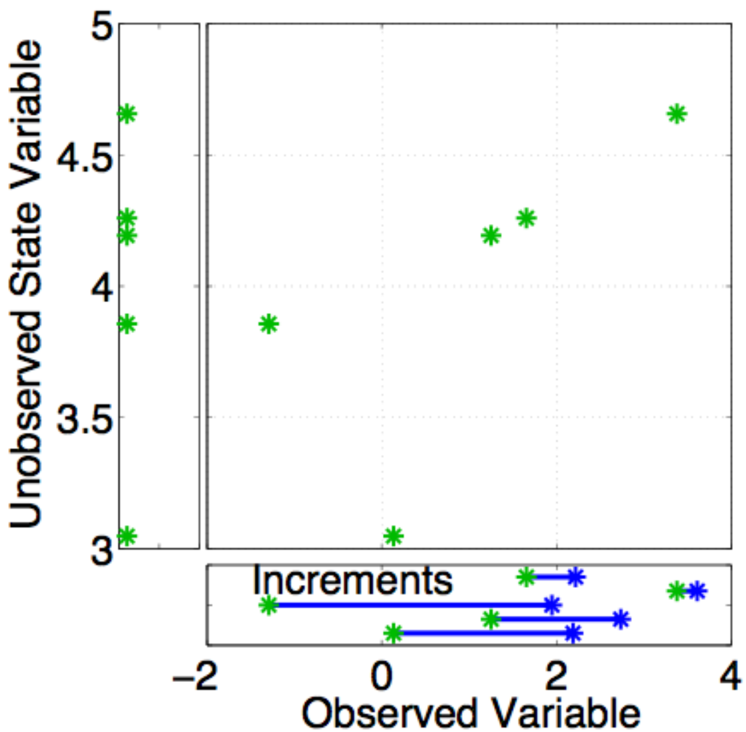


Assume that all we know is prior joint distribution.

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.

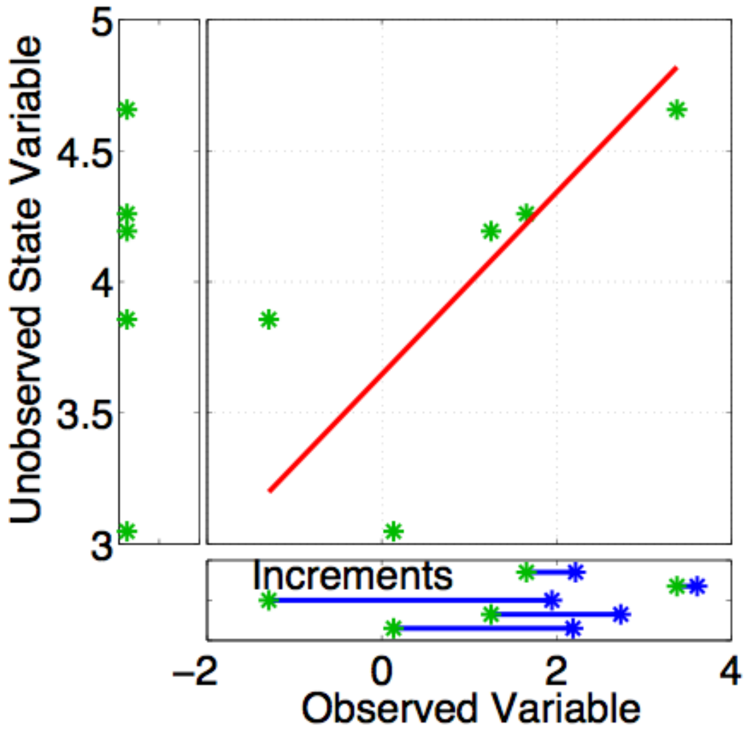
How should the unobserved variable be impacted?

First choice: least squares.

Equivalent to linear regression.

Same as assuming binormal prior.

Ensemble filters: Updating additional prior state variables



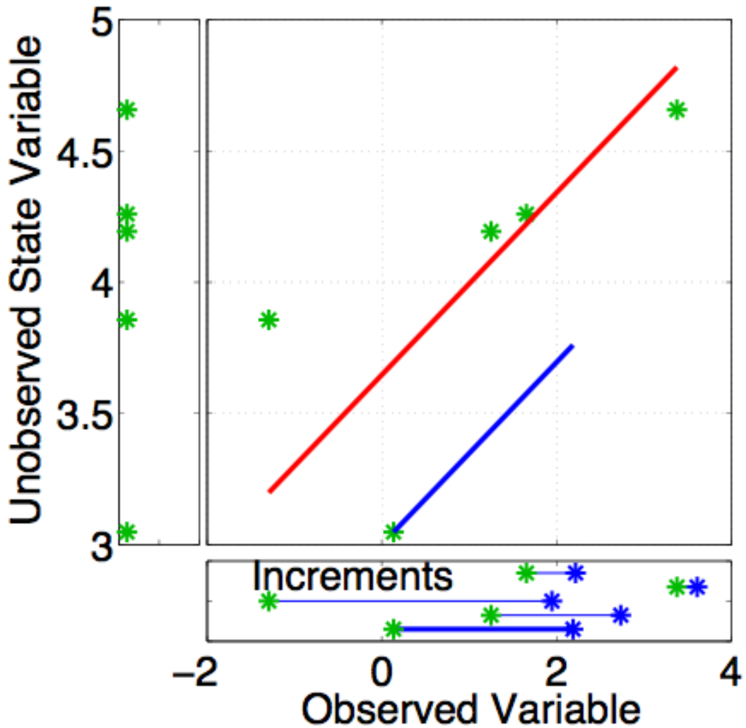
Have joint prior distribution of two variables.

How should the unobserved variable be impacted?

First choice: least squares.

Begin by finding least squares fit.

Ensemble filters: Updating additional prior state variables

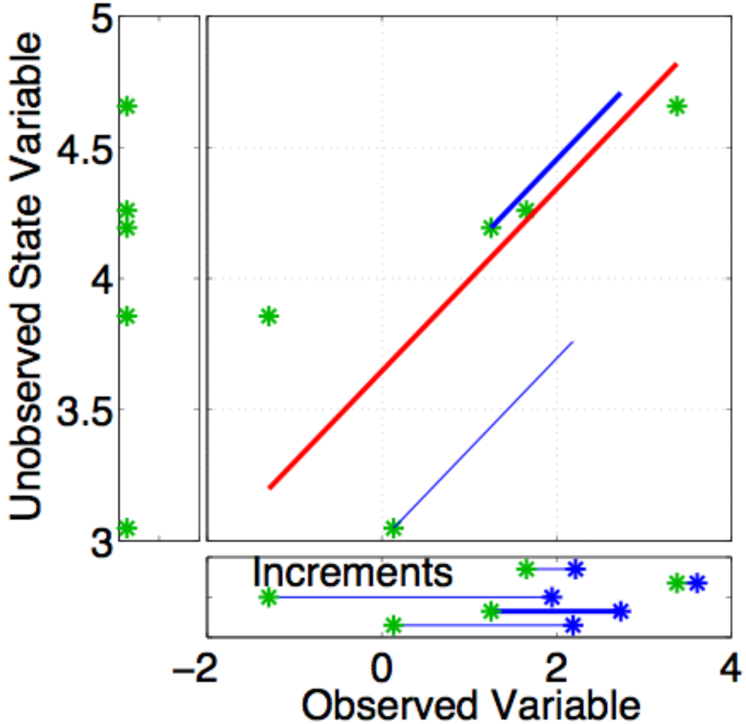


Have joint prior distribution of two variables.

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

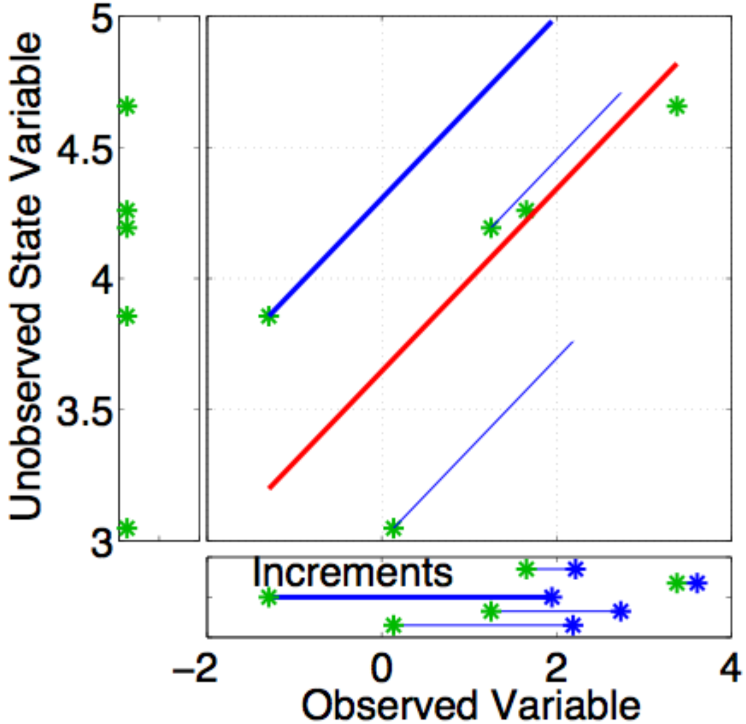


Have joint prior distribution of two variables.

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

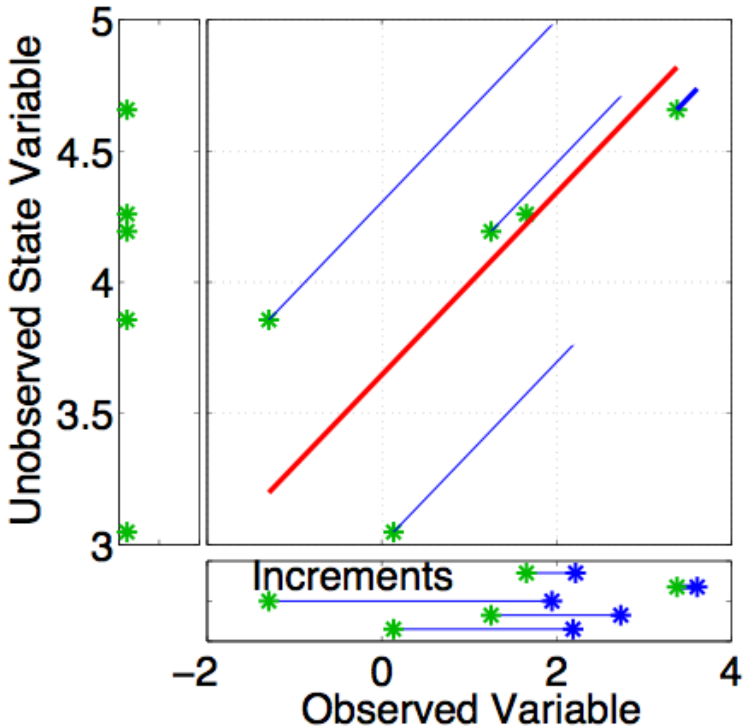


Have joint prior distribution of two variables.

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

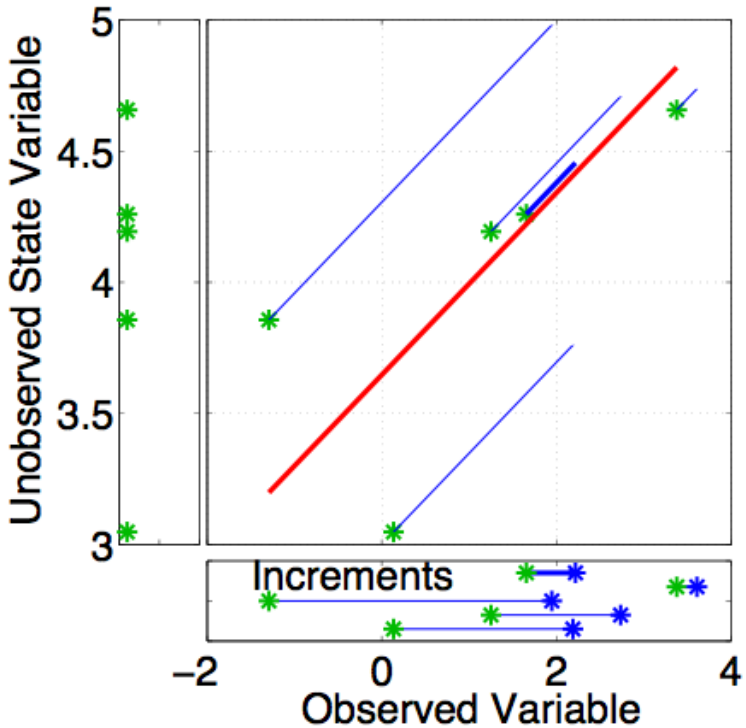


Have joint prior distribution of two variables.

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

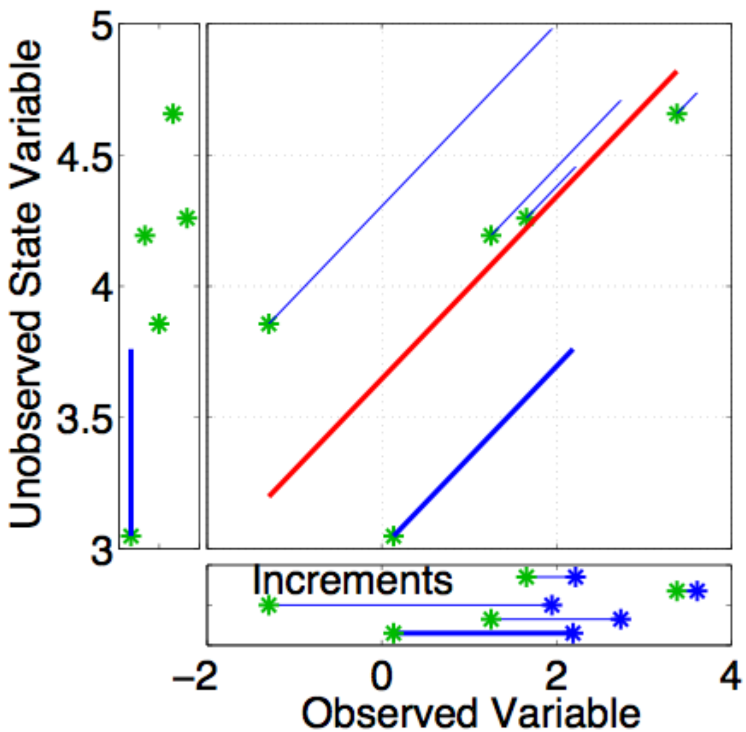


Have joint prior distribution of two variables.

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

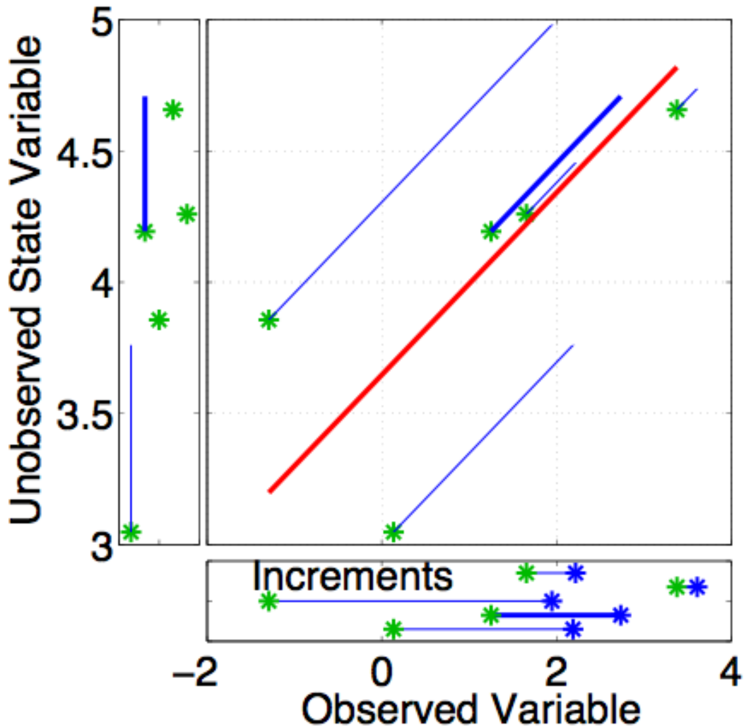


Have joint prior distribution of two variables.

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

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

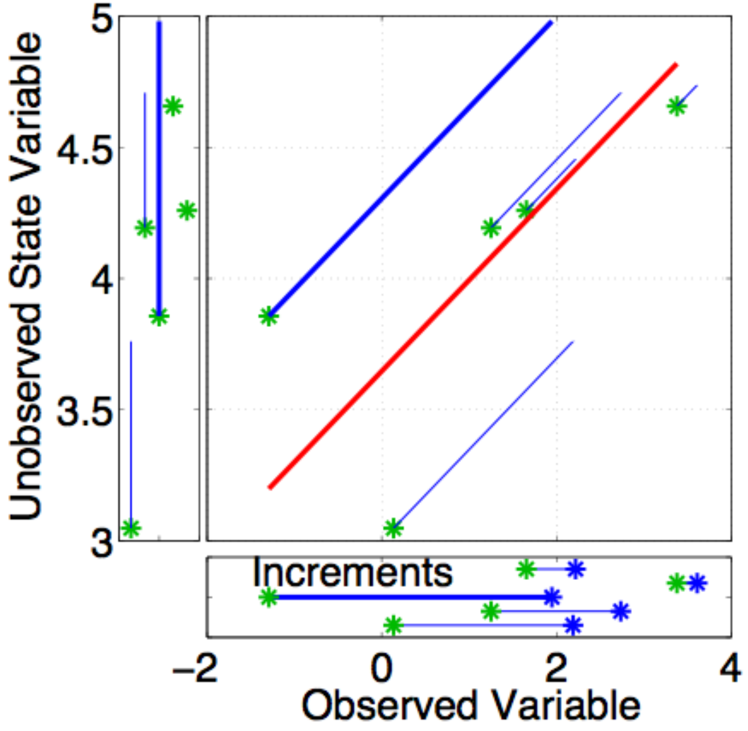


Have joint prior distribution of two variables.

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

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

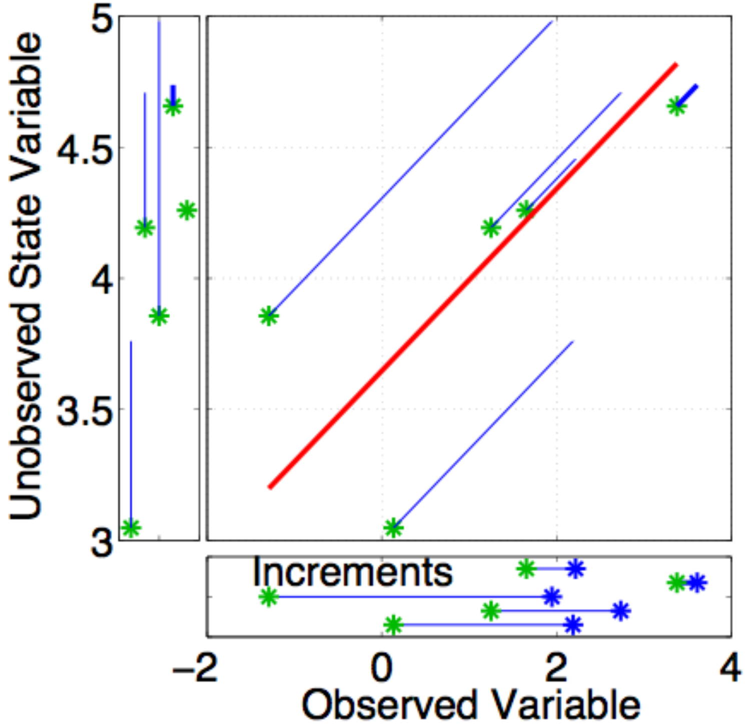


Have joint prior distribution of two variables.

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

Then projecting from joint space onto unobserved priors.

Ensemble filters: Updating additional prior state variables

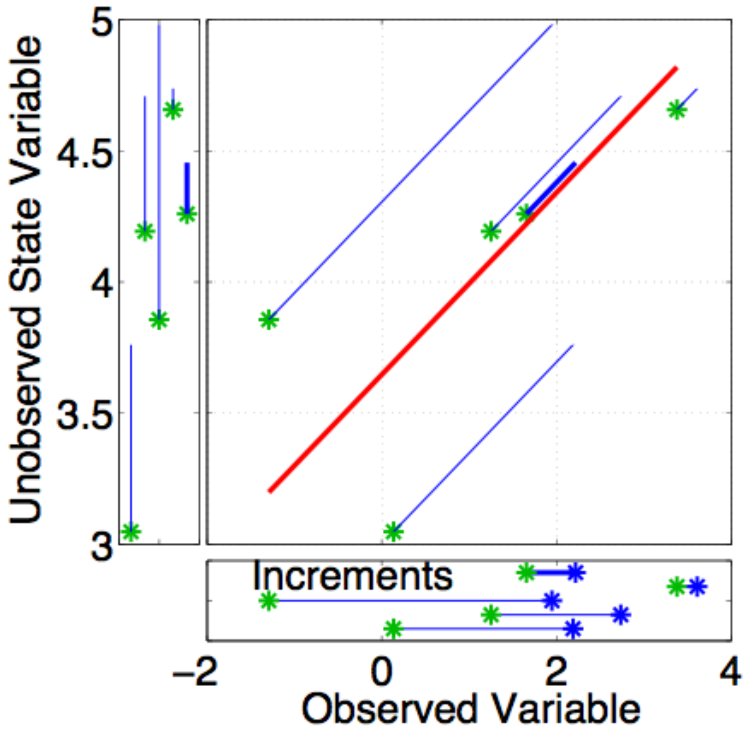


Have joint prior distribution of two variables.

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

Then projecting from joint space onto unobserved priors.

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.

Comments on multidimensional (Ensemble) Kalman Filter

- KF optimal for linear model, gaussian likelihood.
- The deterministic Ensemble KF gives identical mean, covariance with sufficiently large ensemble size.

- Basic ensemble filter fails for ensemble too small.

- For nonlinear model, non-gaussian likelihood, all bets are off.
- Both deterministic and stochastic Ensemble KFs become Monte Carlo algorithms.

Ensemble Filter for Large Geophysical Models

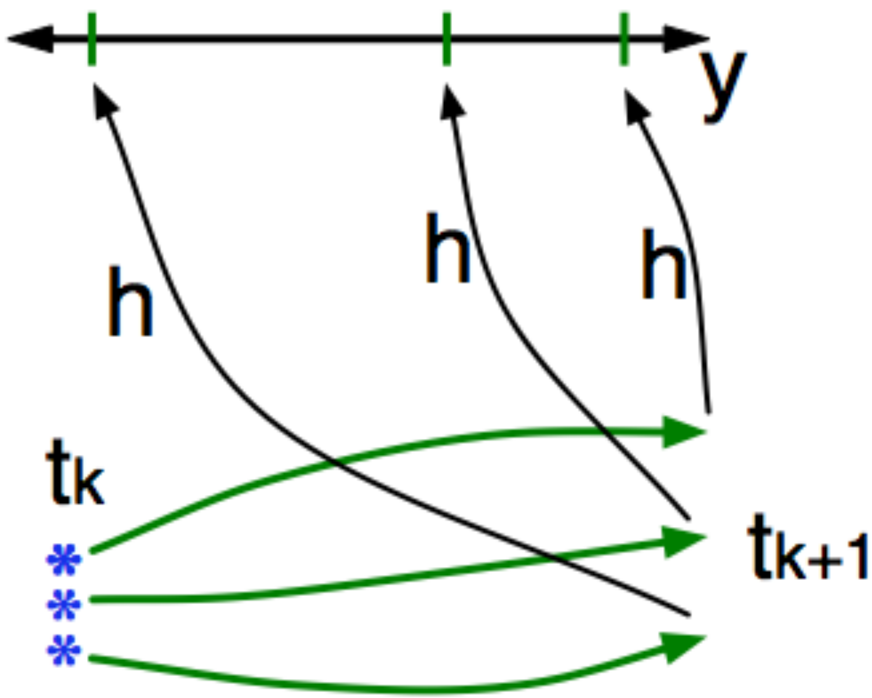
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Filter for Large Geophysical Models

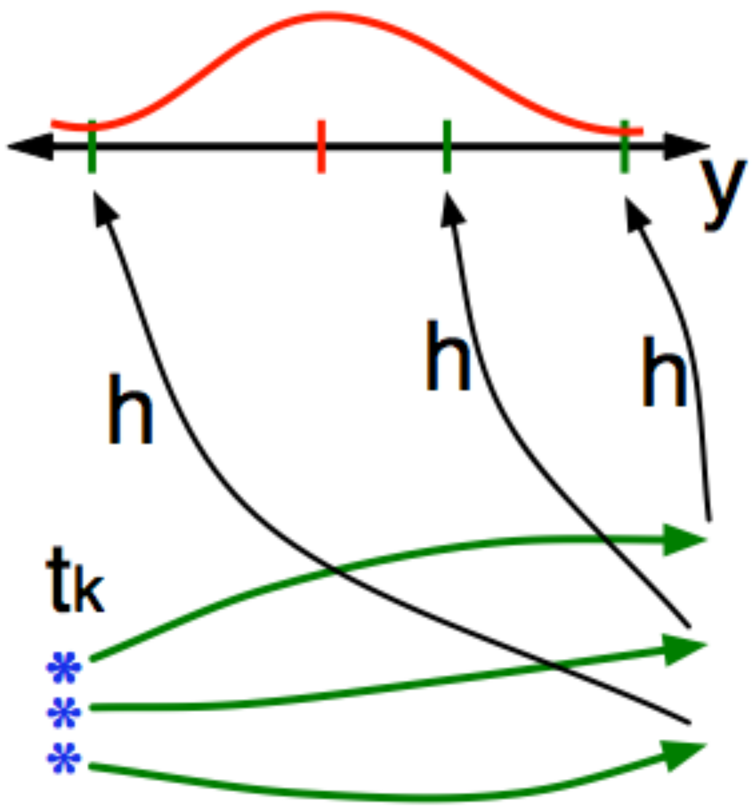
2. Get prior ensemble sample of observation, $y = h(x)$, by applying forward operator **h** to each ensemble member.



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

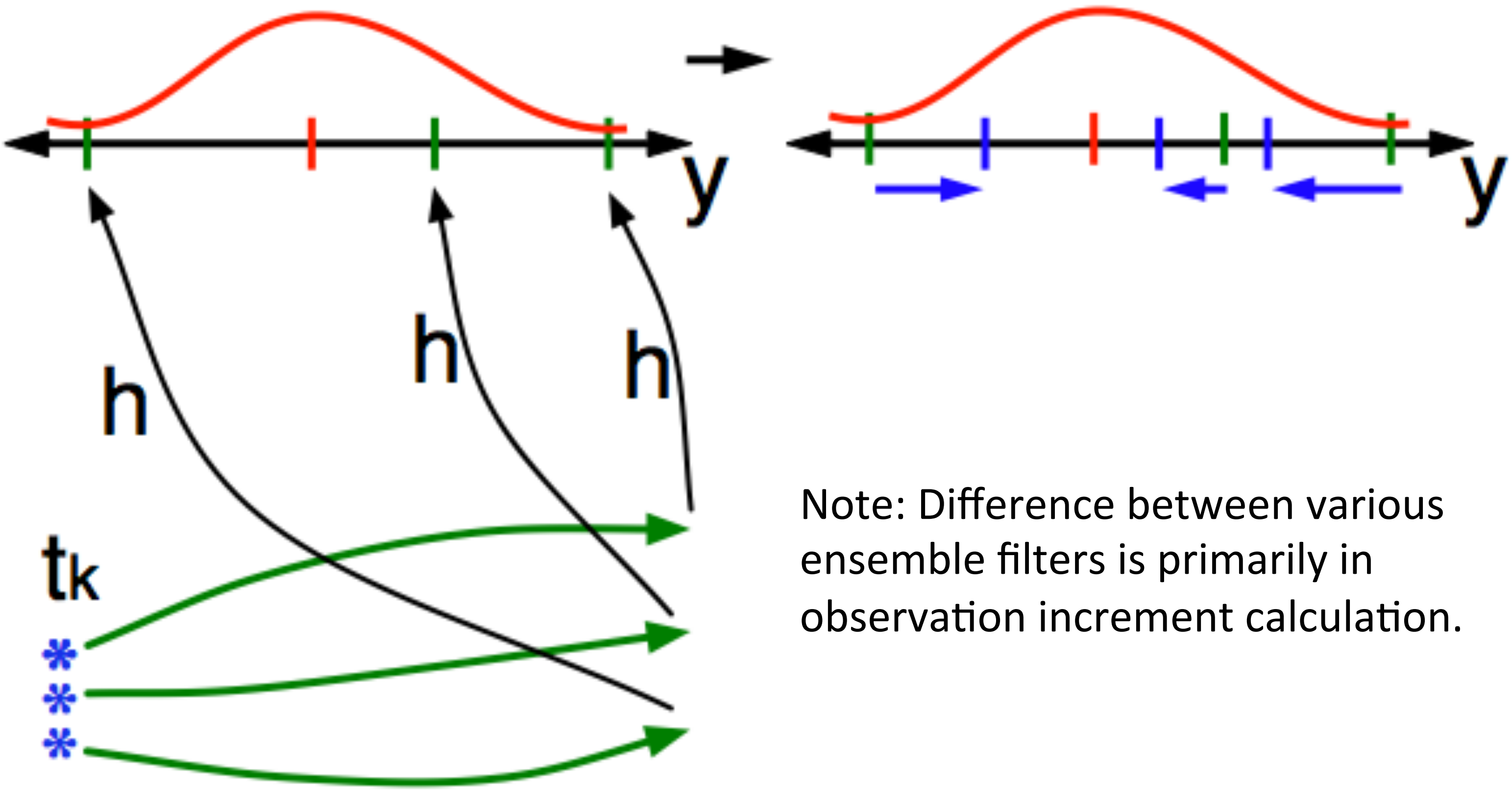
Ensemble Filter for Large Geophysical Models

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



Ensemble Filter for Large Geophysical Models

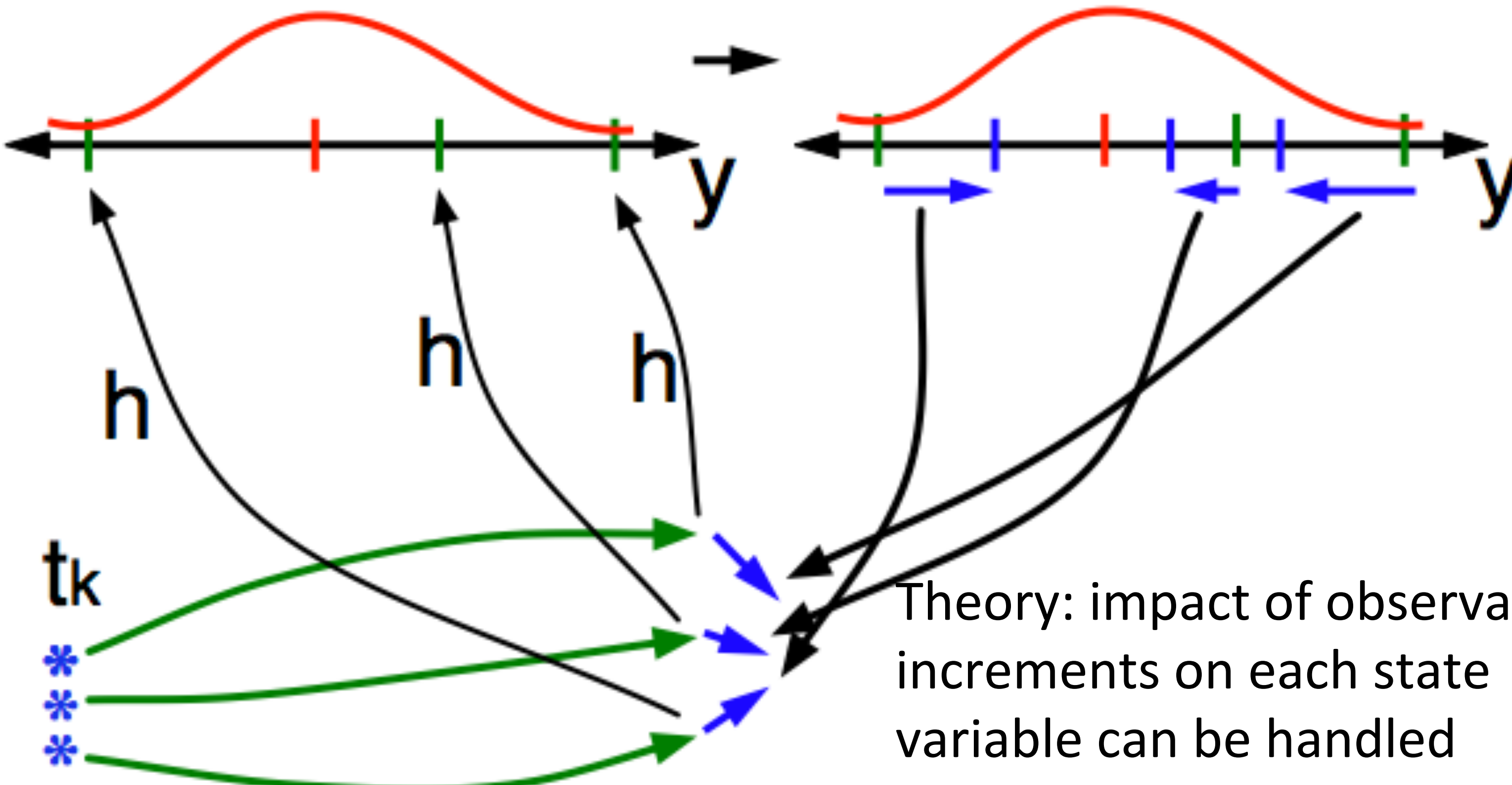
4. Find the increments for the prior observation ensemble
(this is a scalar problem for uncorrelated observation errors).



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

Ensemble Filter for Large Geophysical Models

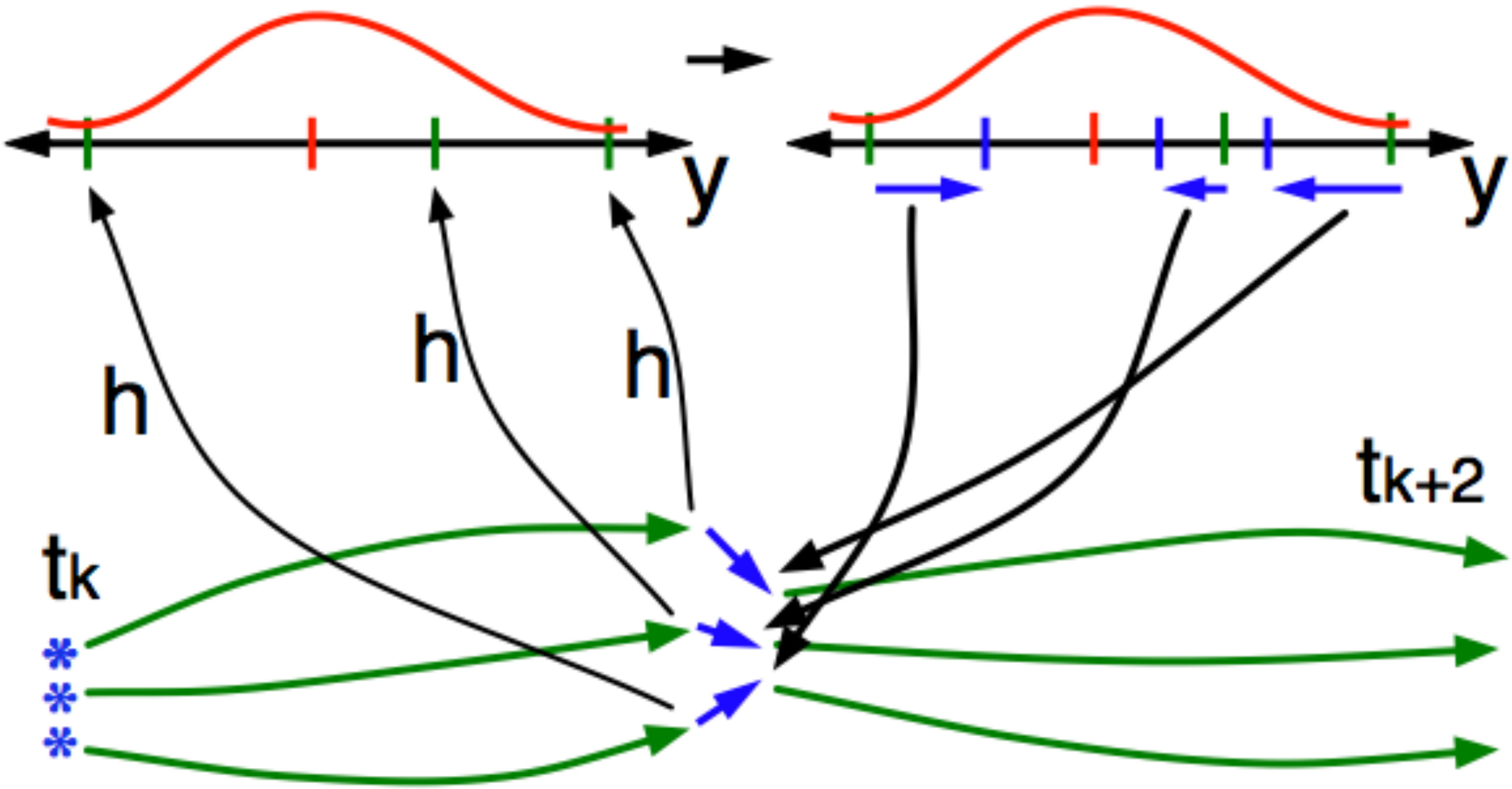
5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



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

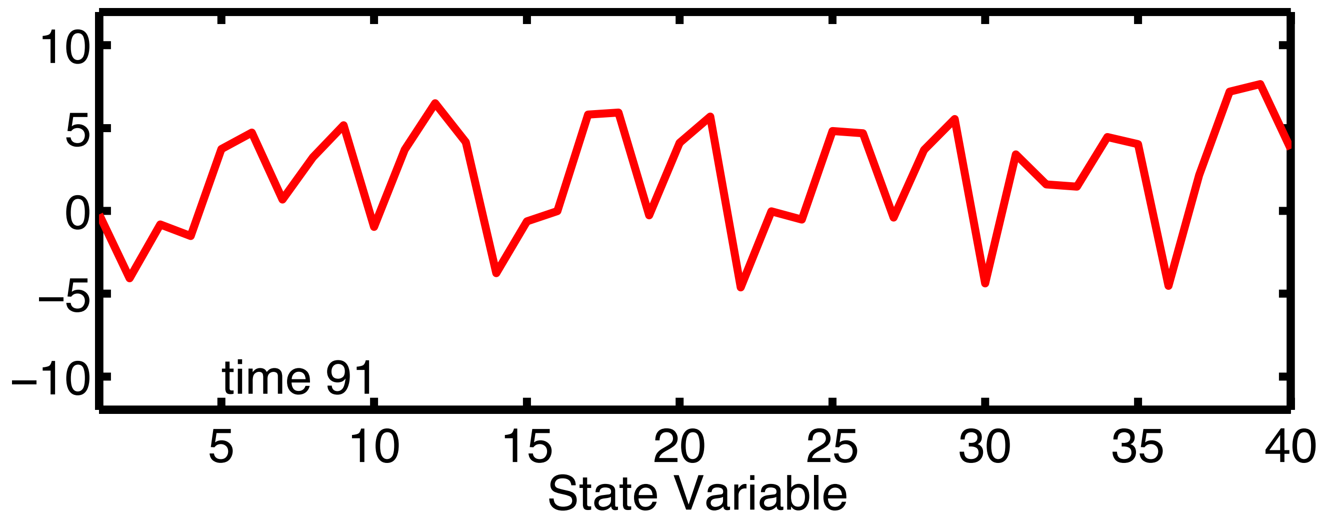
Ensemble Filter for Large Geophysical Models

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



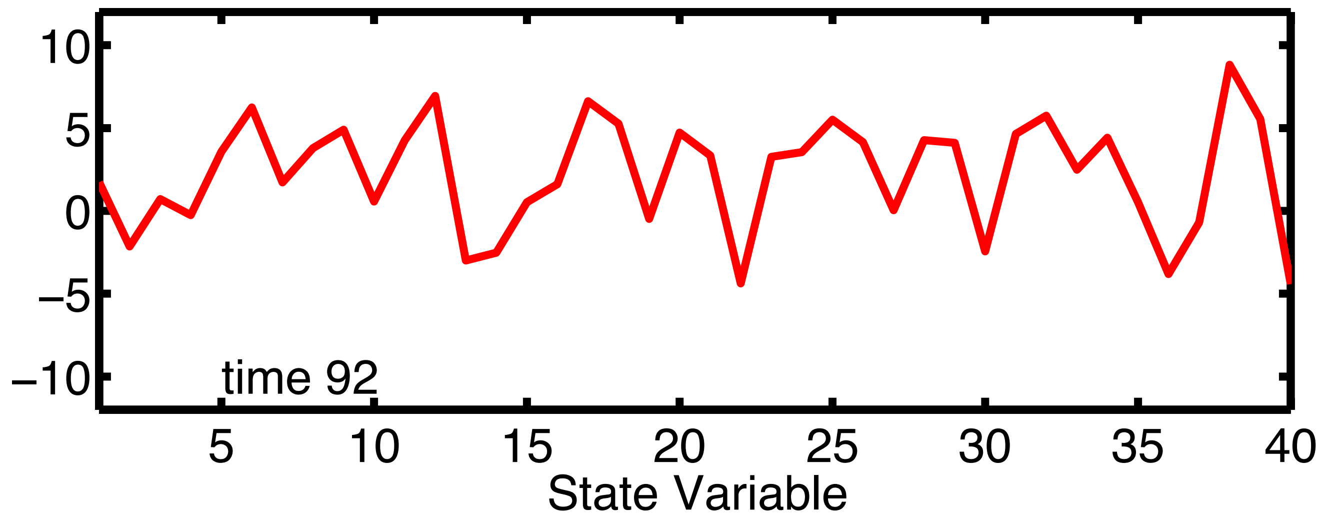
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



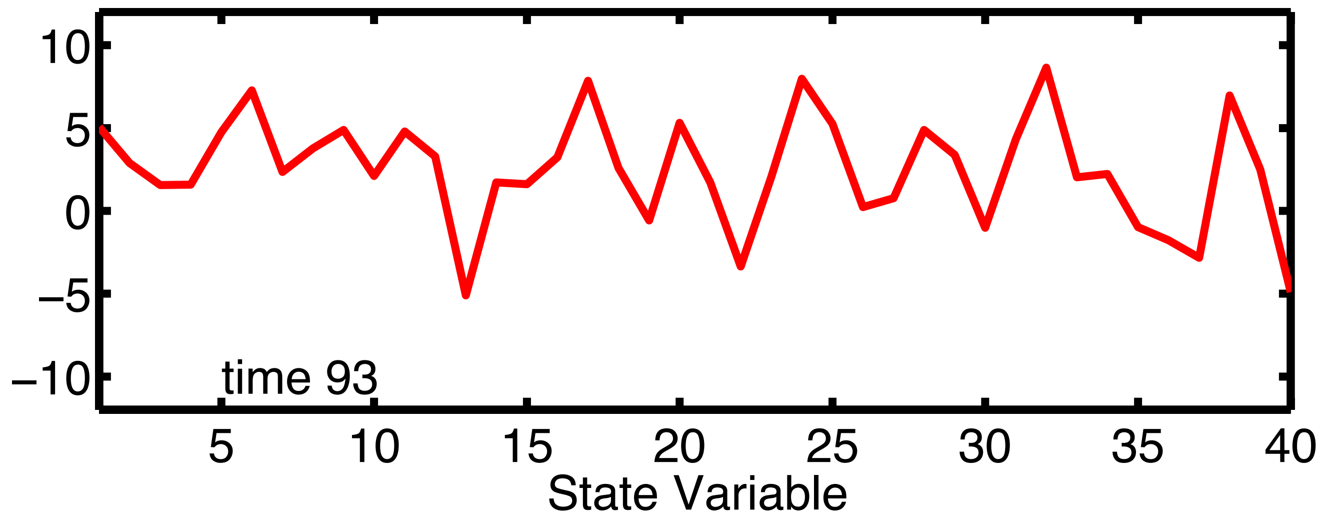
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



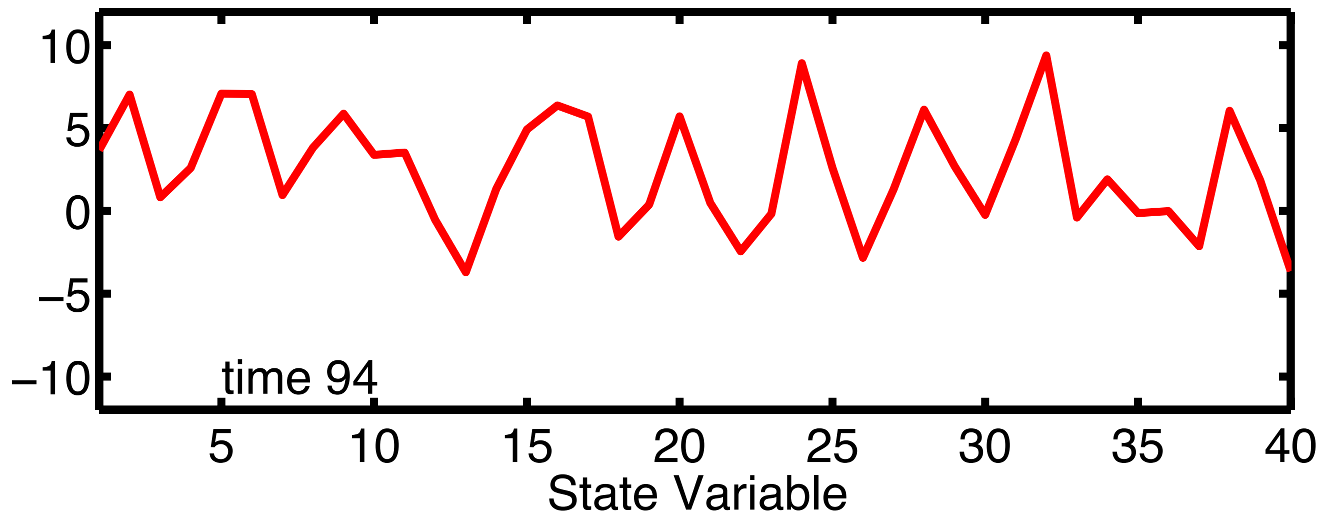
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts ‘something’ like synoptic weather around a latitude band.



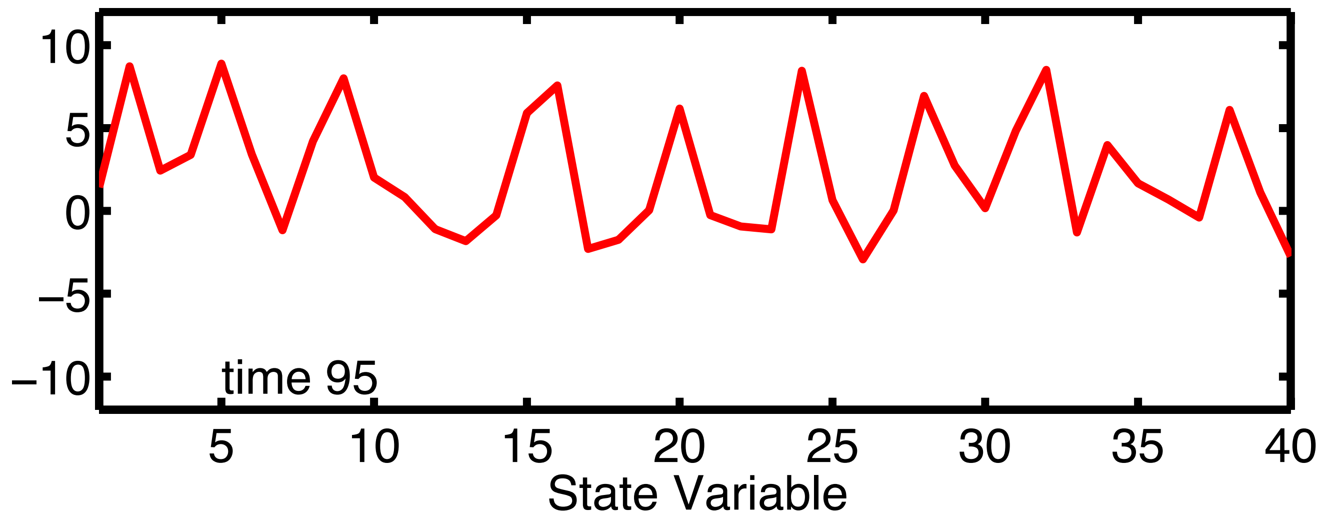
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



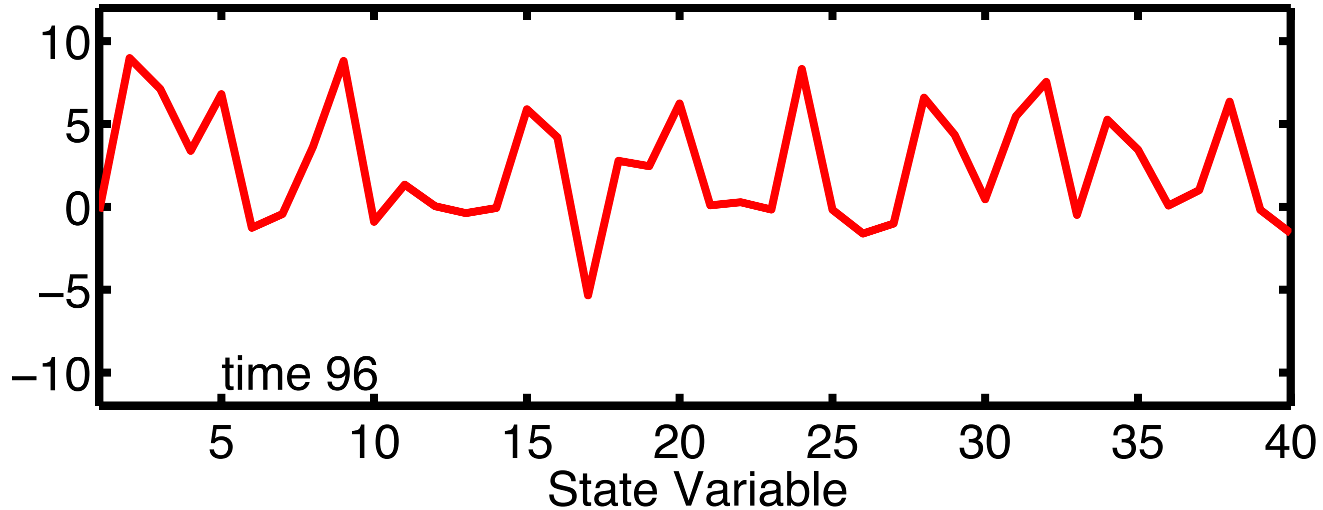
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



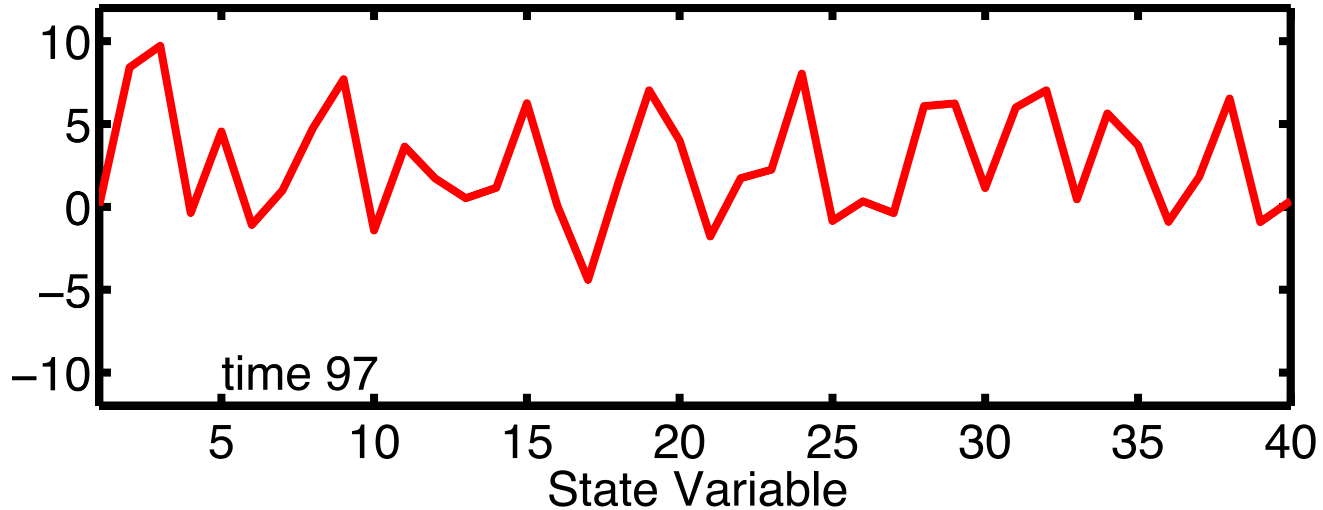
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



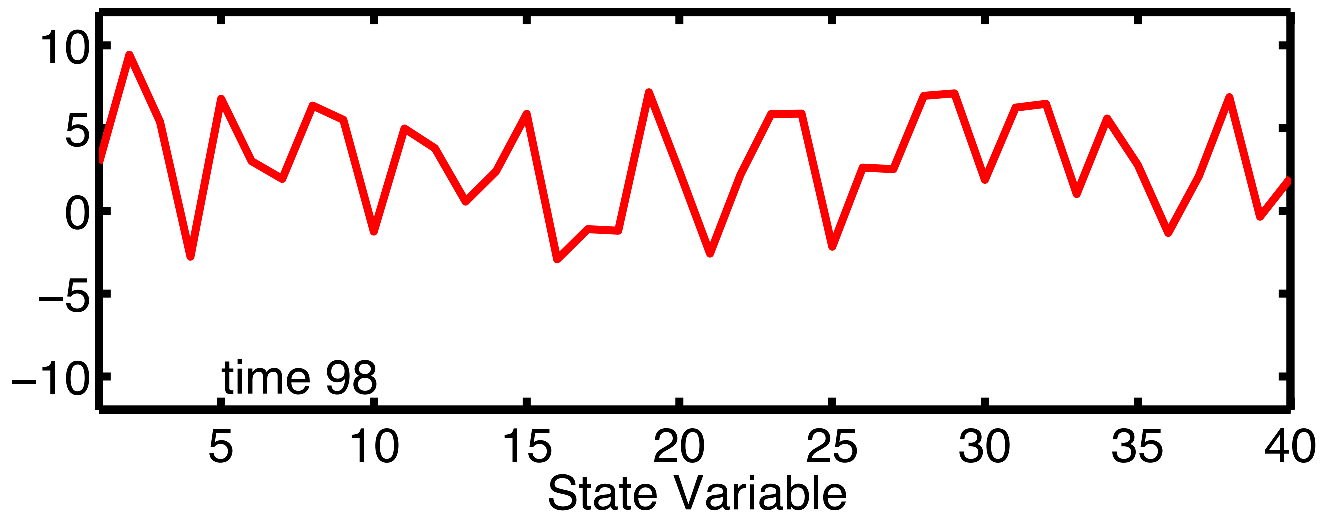
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



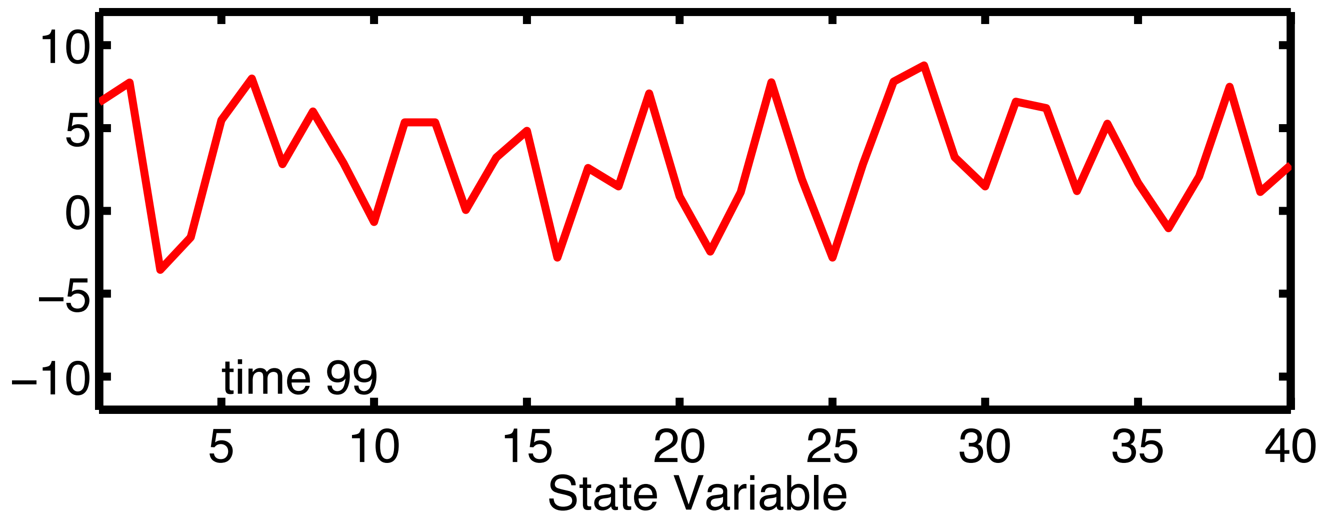
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts ‘something’ like synoptic weather around a latitude band.



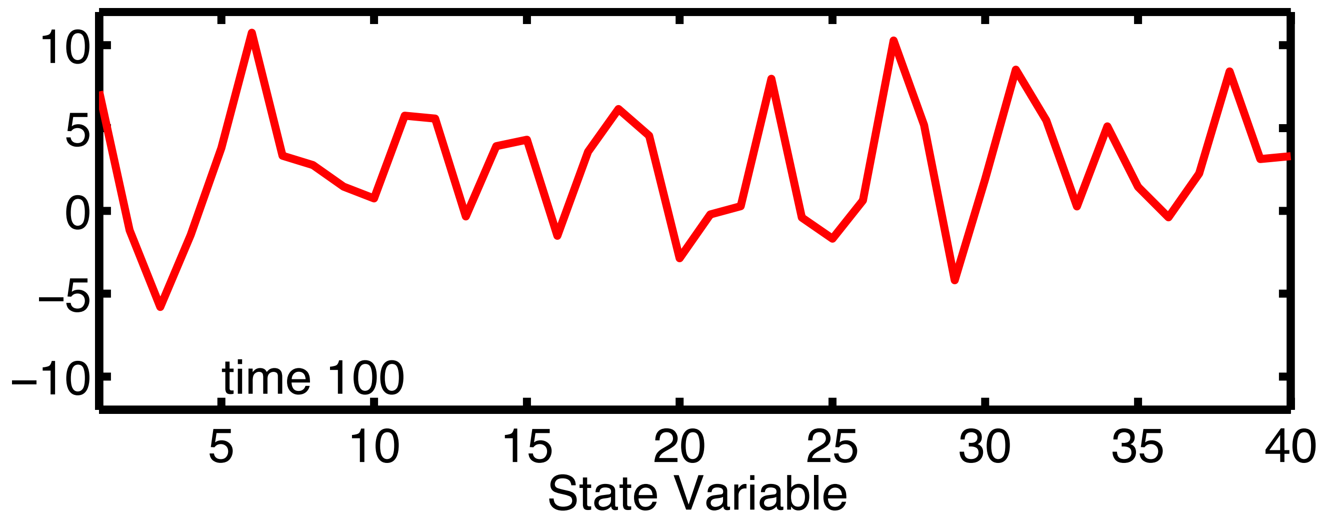
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



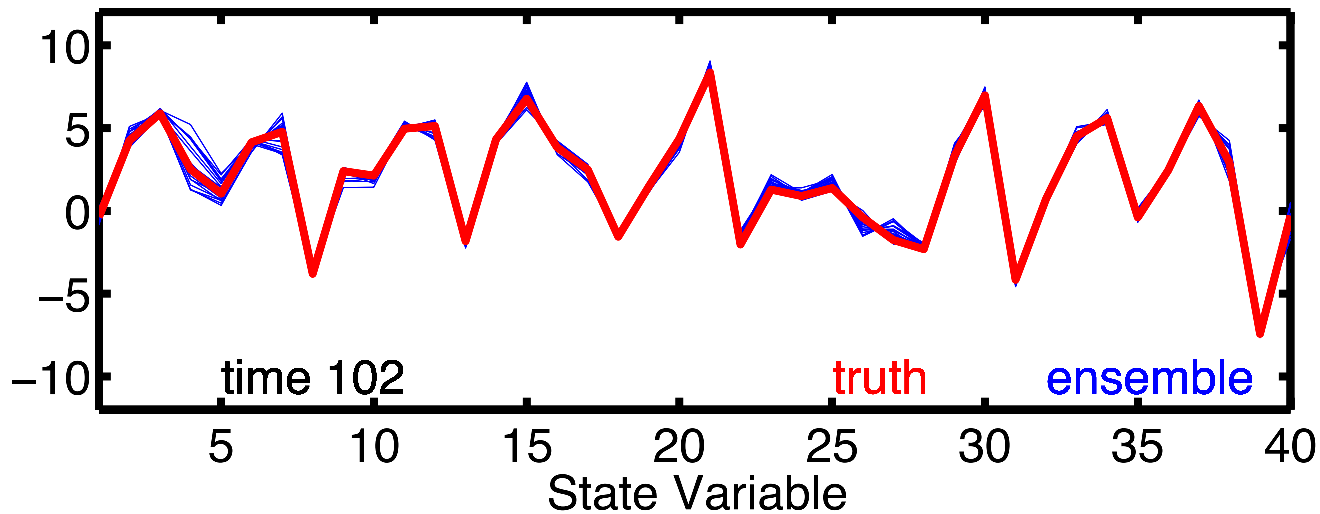
Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
Acts 'something' like synoptic weather around a latitude band.



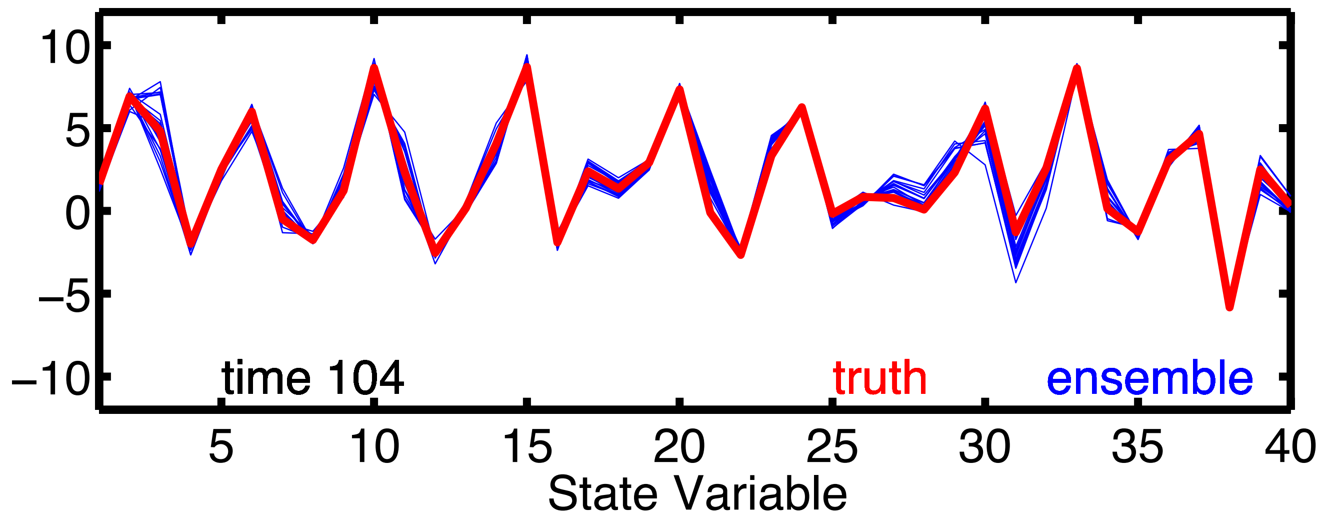
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



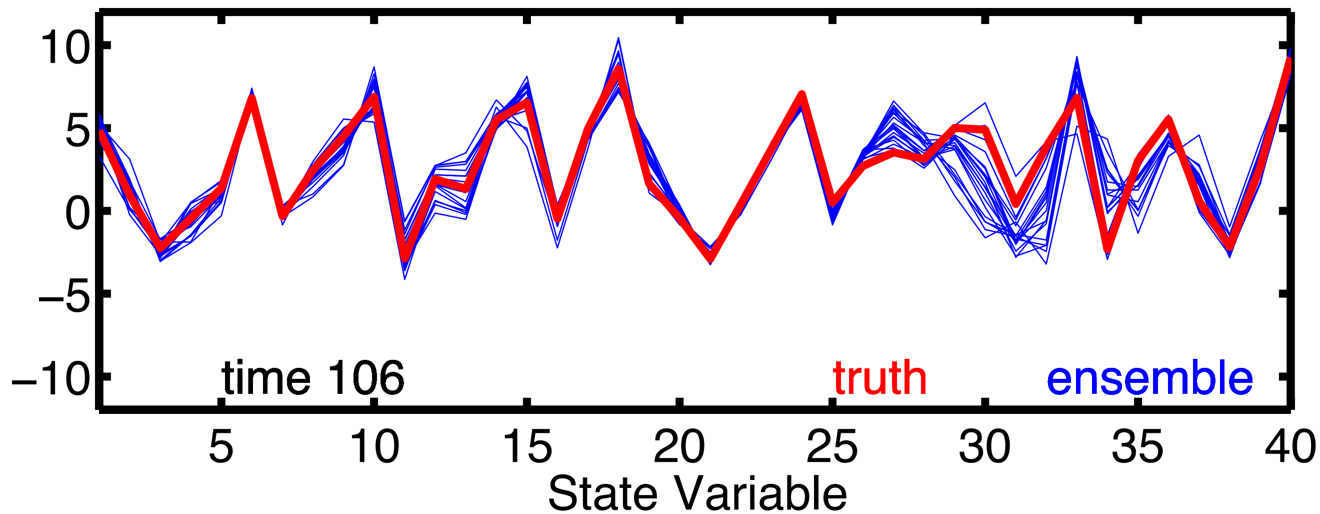
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



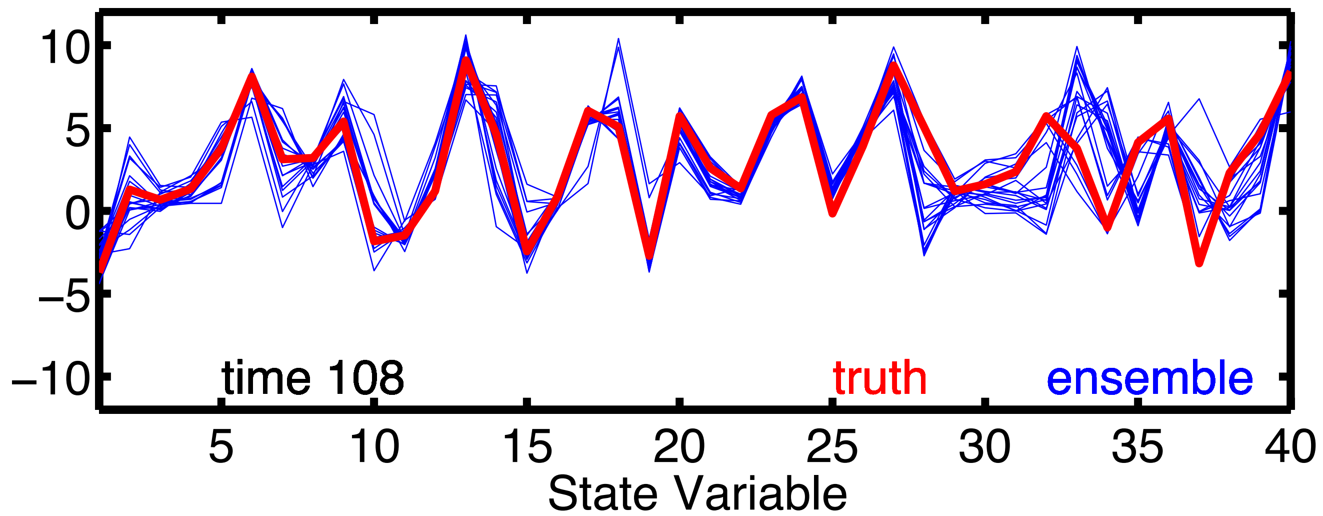
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



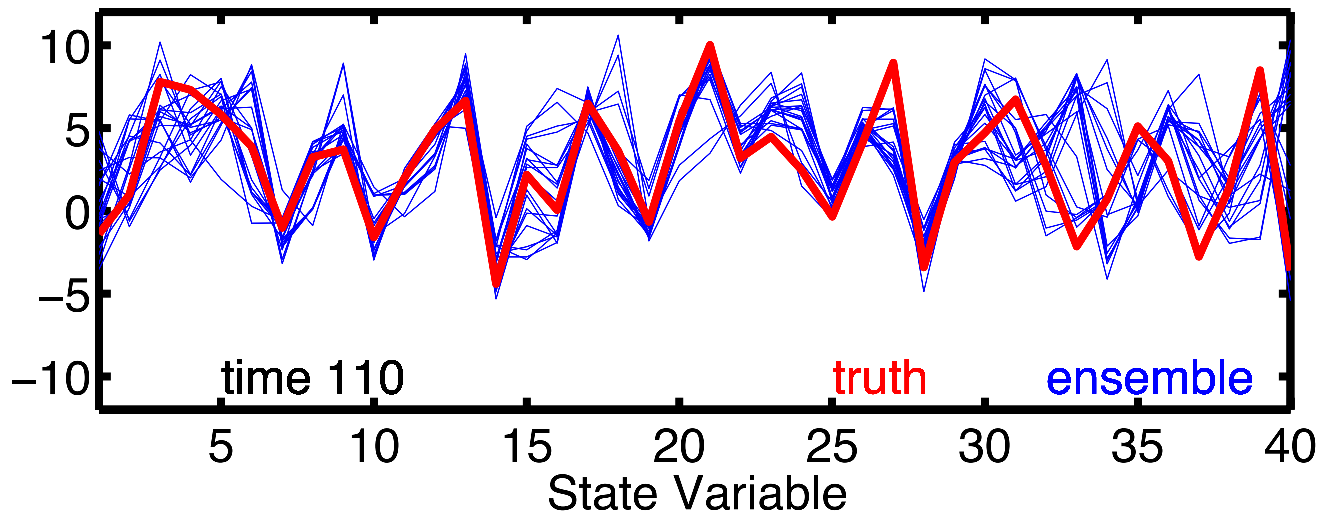
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



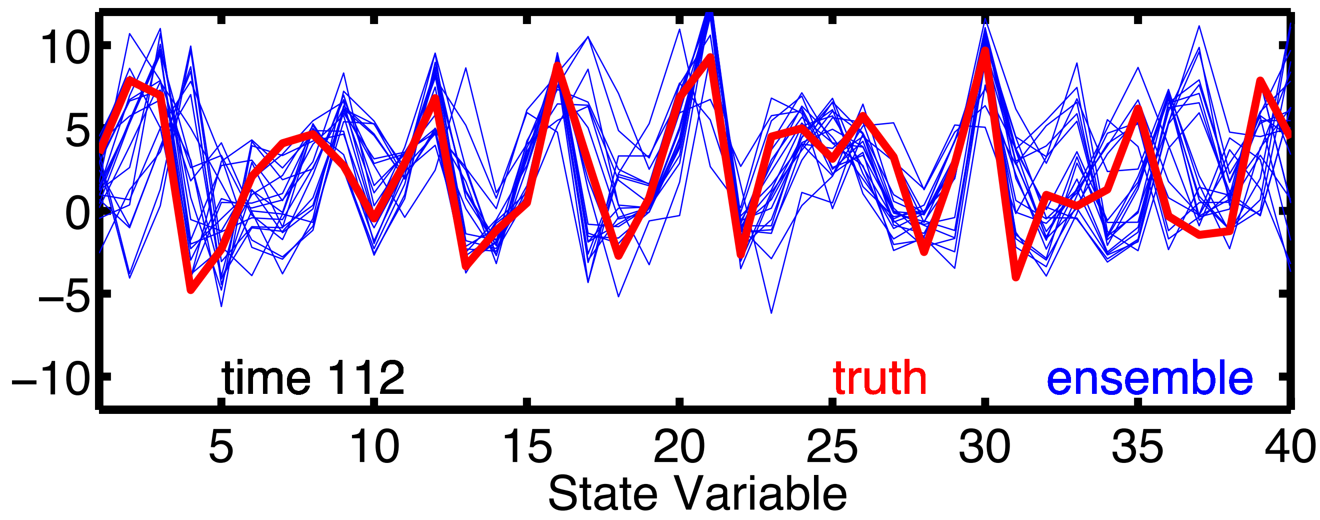
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



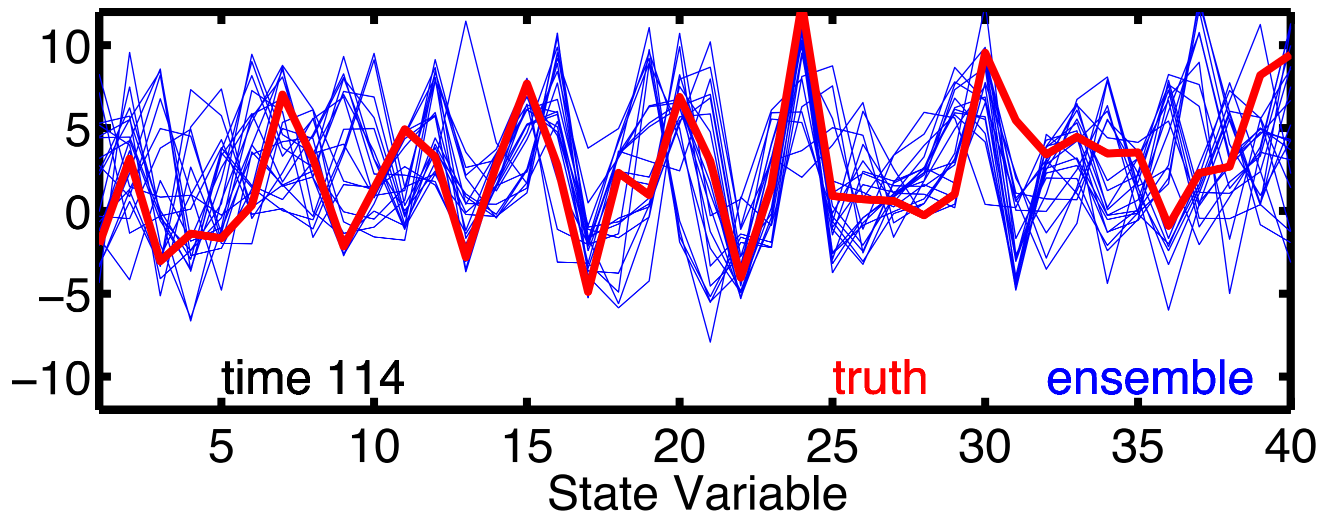
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



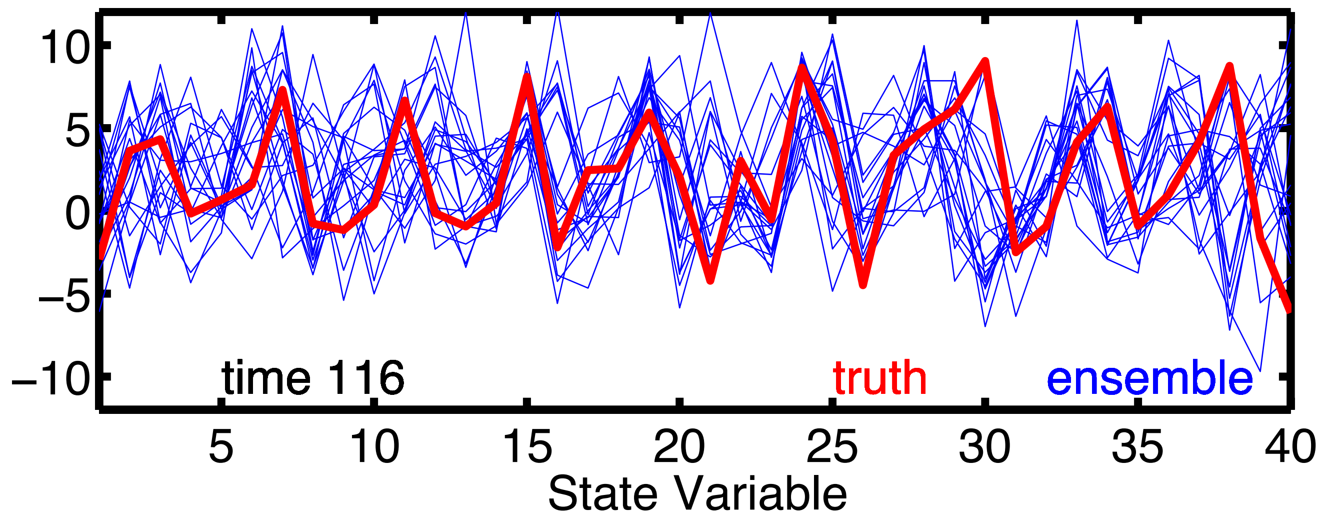
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



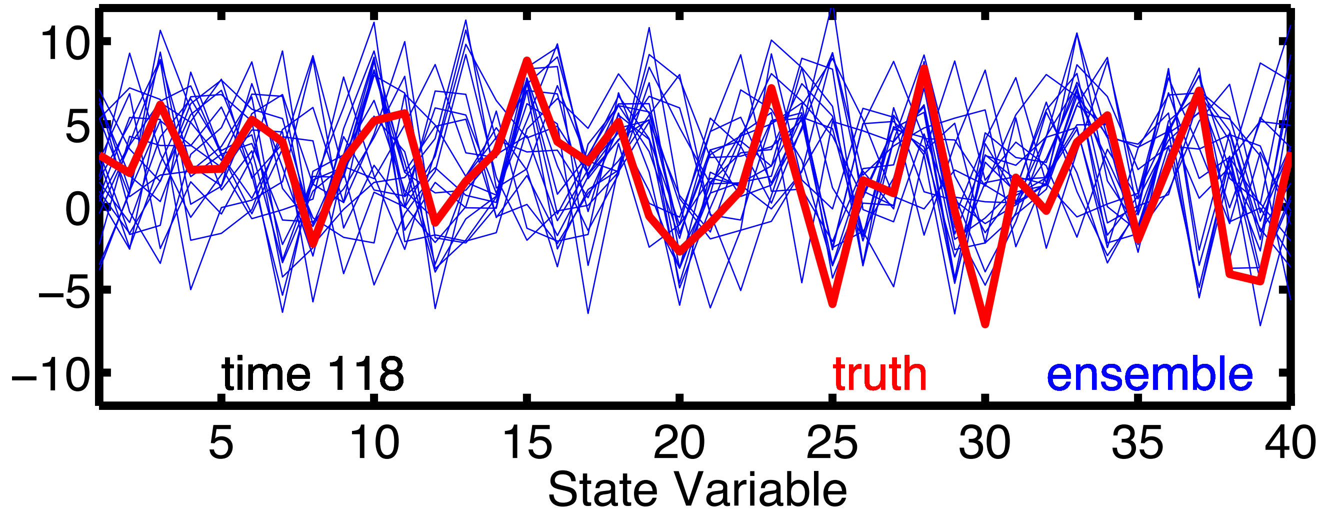
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



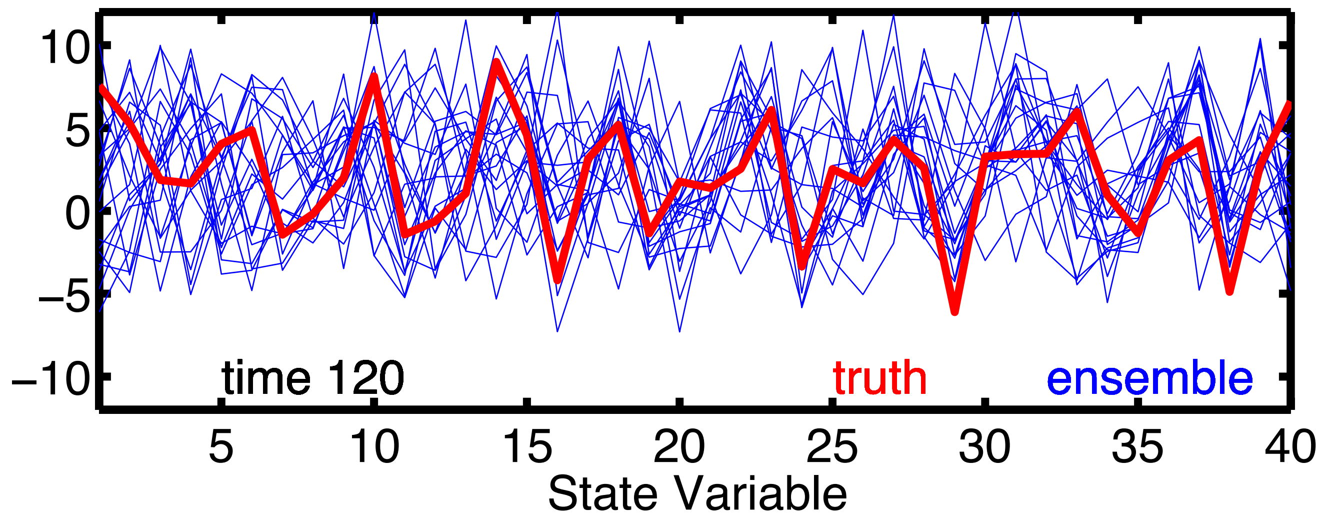
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



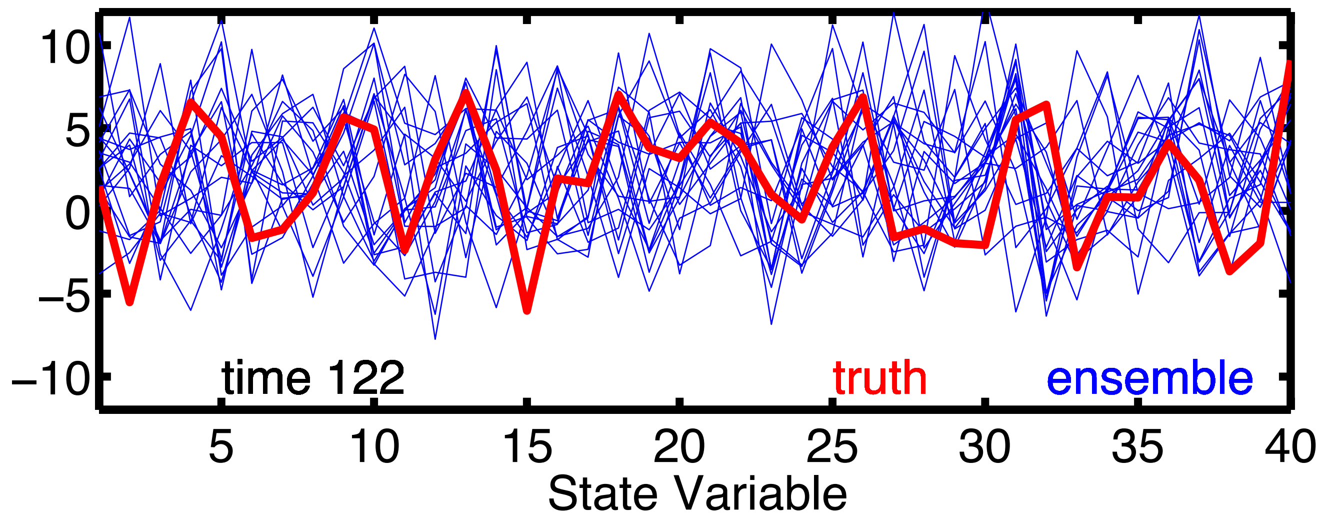
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



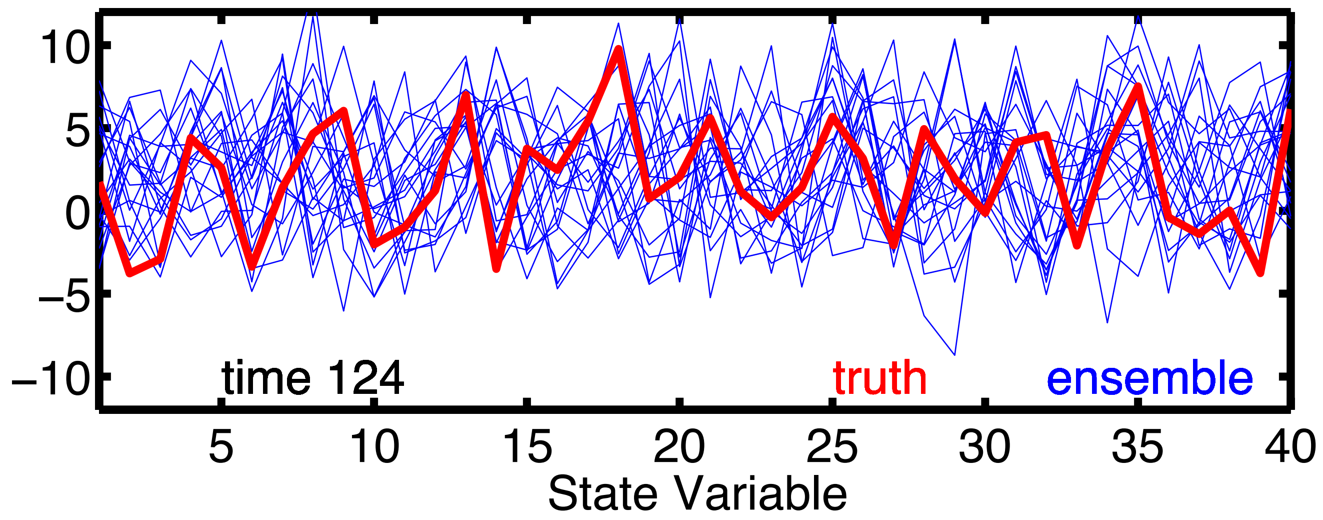
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



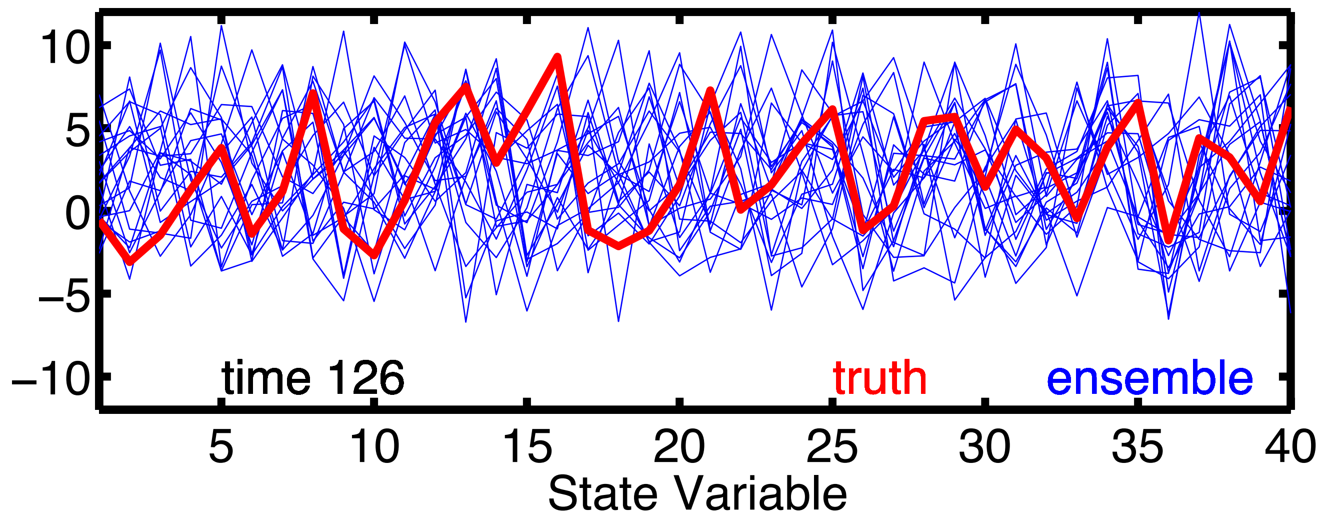
Lorenz-96 is sensitive to small perturbations

Introduce 20 'ensemble' state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as 'truth'..



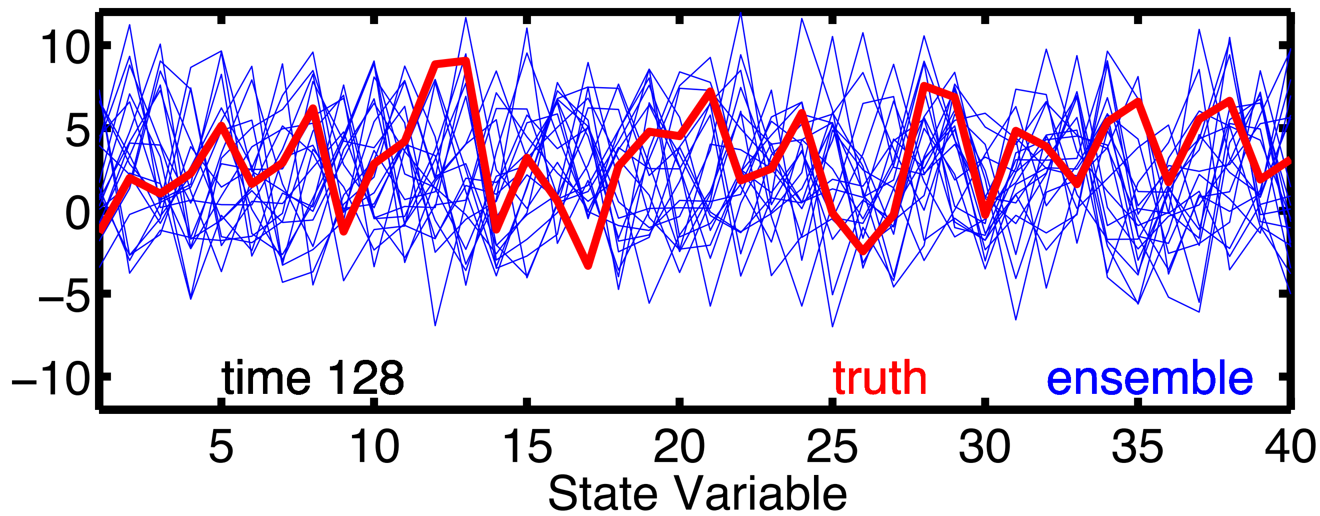
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



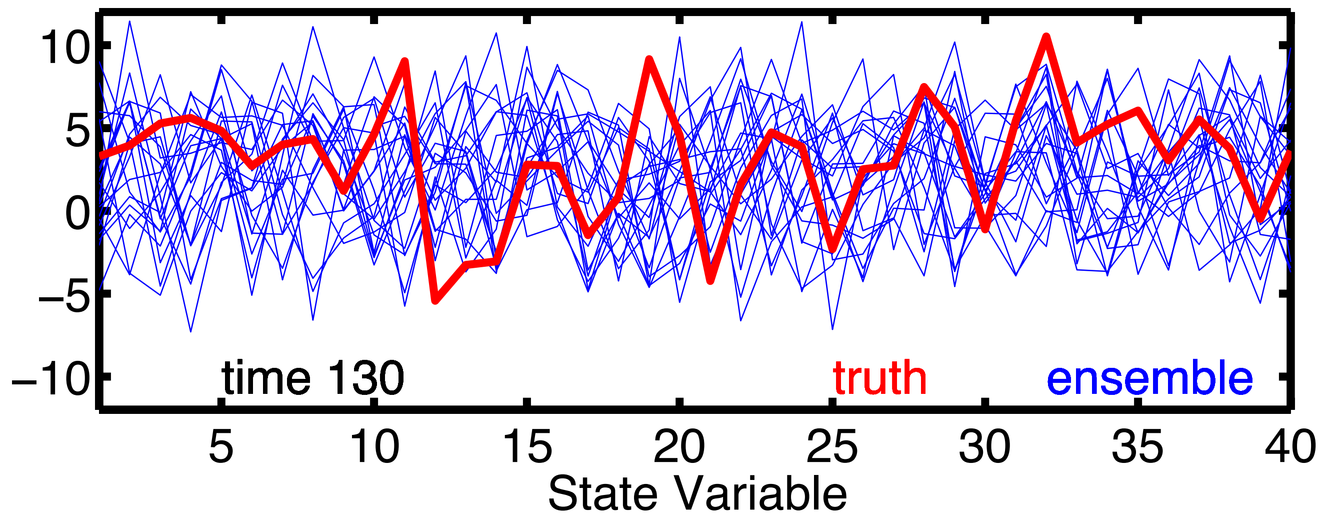
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



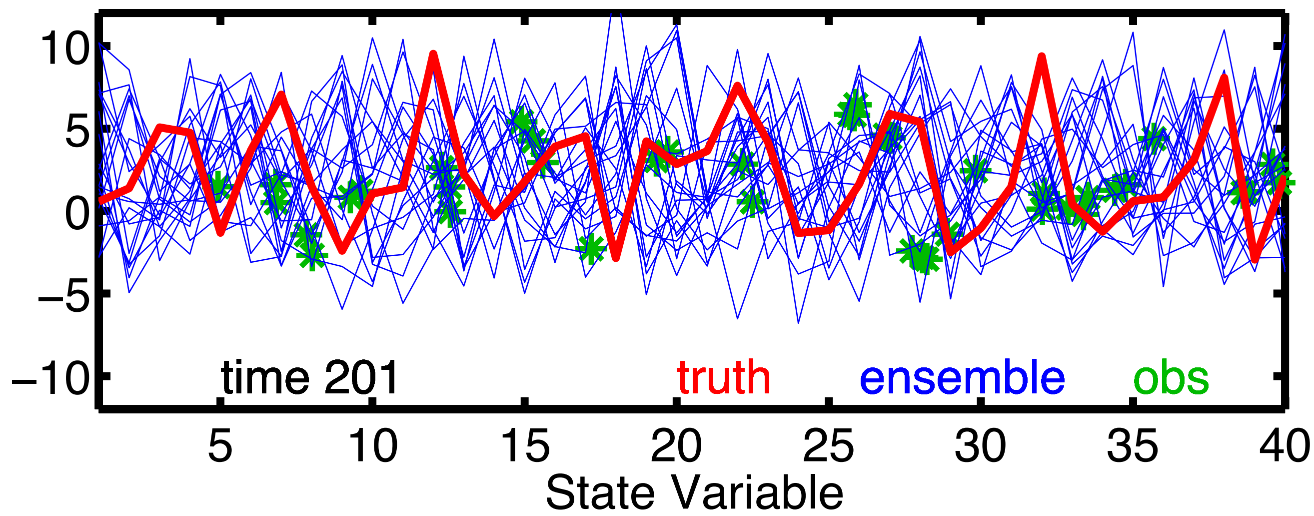
Lorenz-96 is sensitive to small perturbations

Introduce 20 ‘ensemble’ state estimates.
Each is slightly perturbed for each of the 40-variables at time 100.
Refer to unperturbed control integration as ‘truth’..



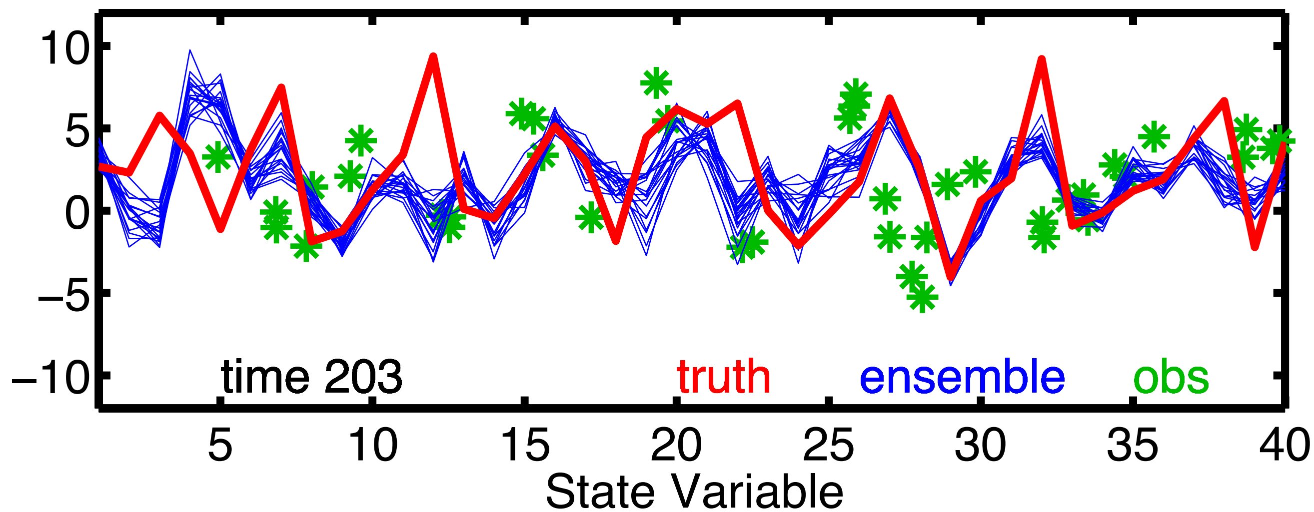
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



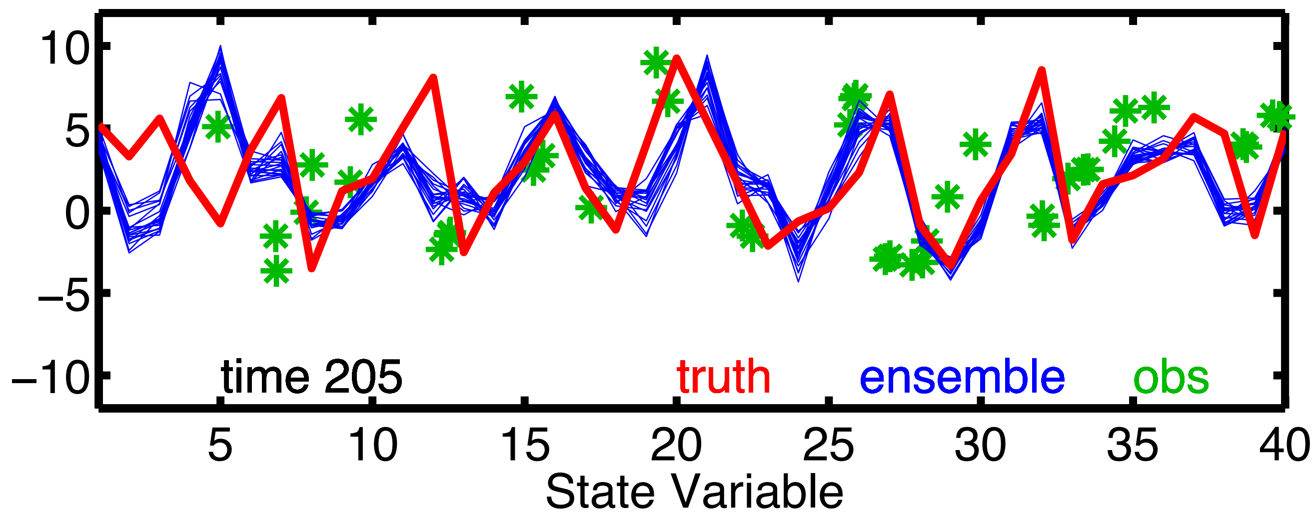
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



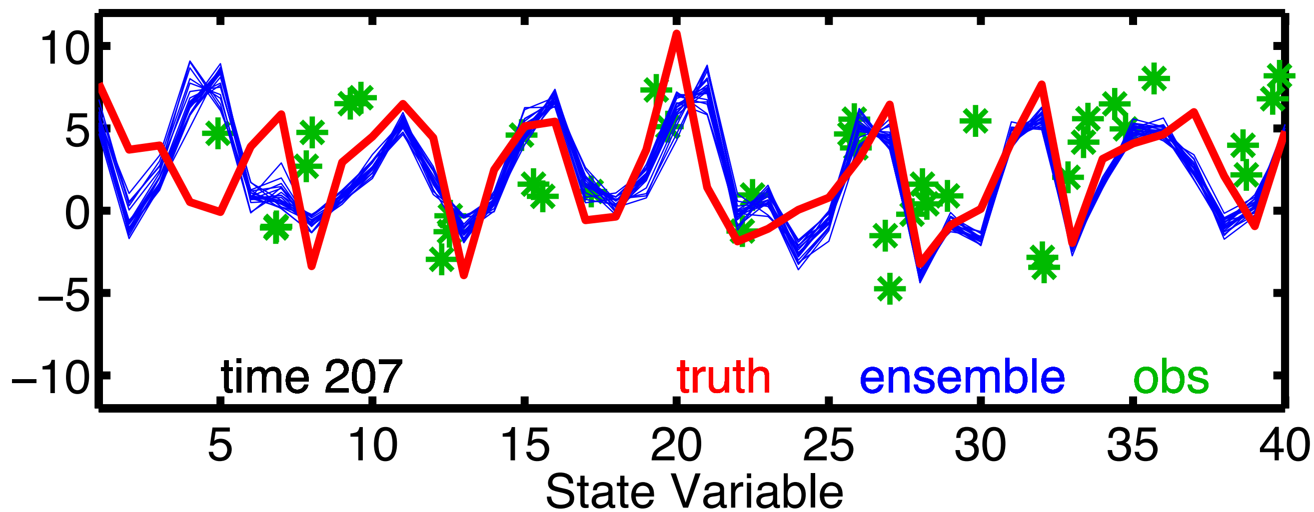
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



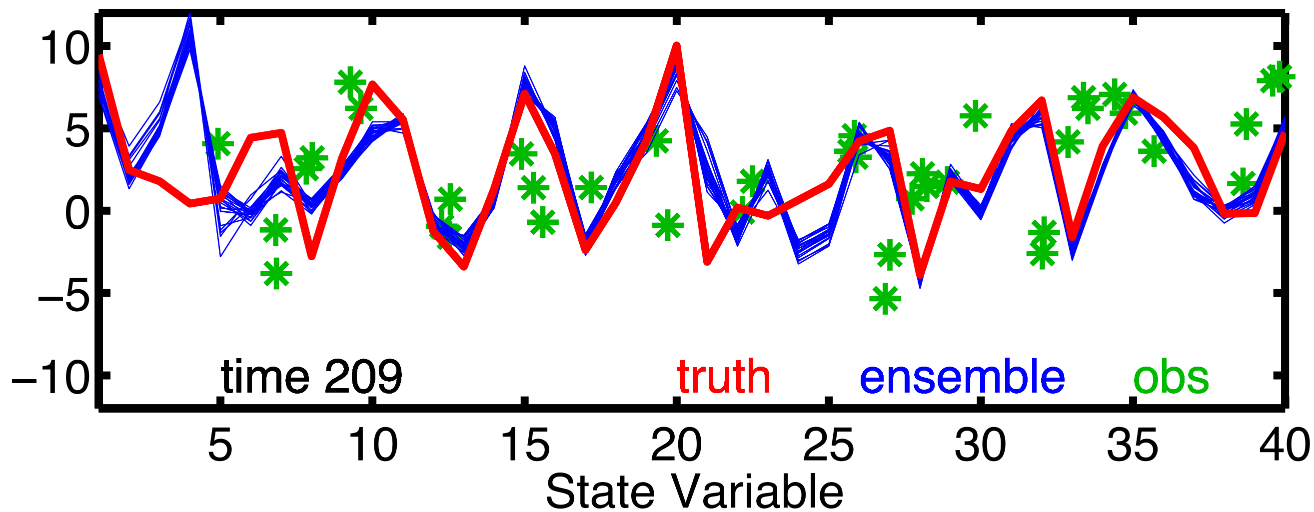
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



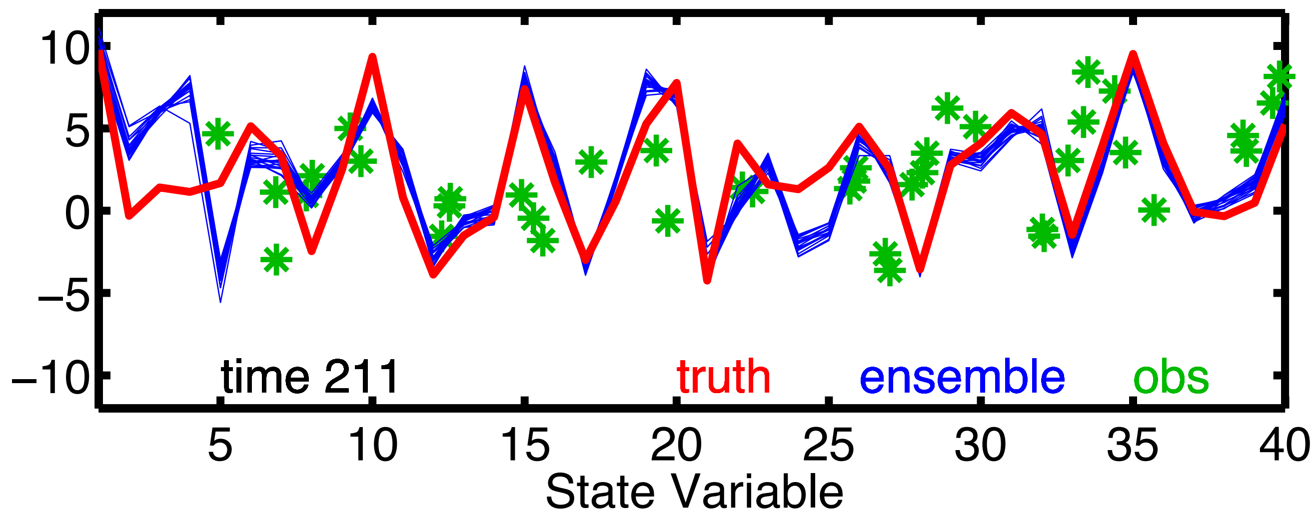
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



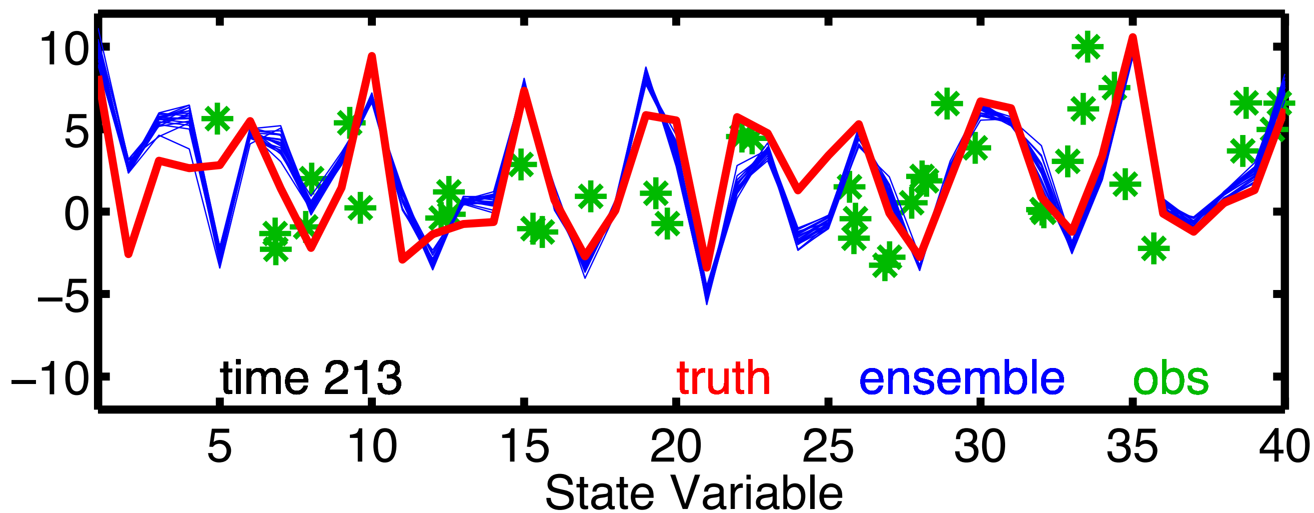
Assimilate ‘observations’ from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from ‘climatological’ 20-member ensemble.



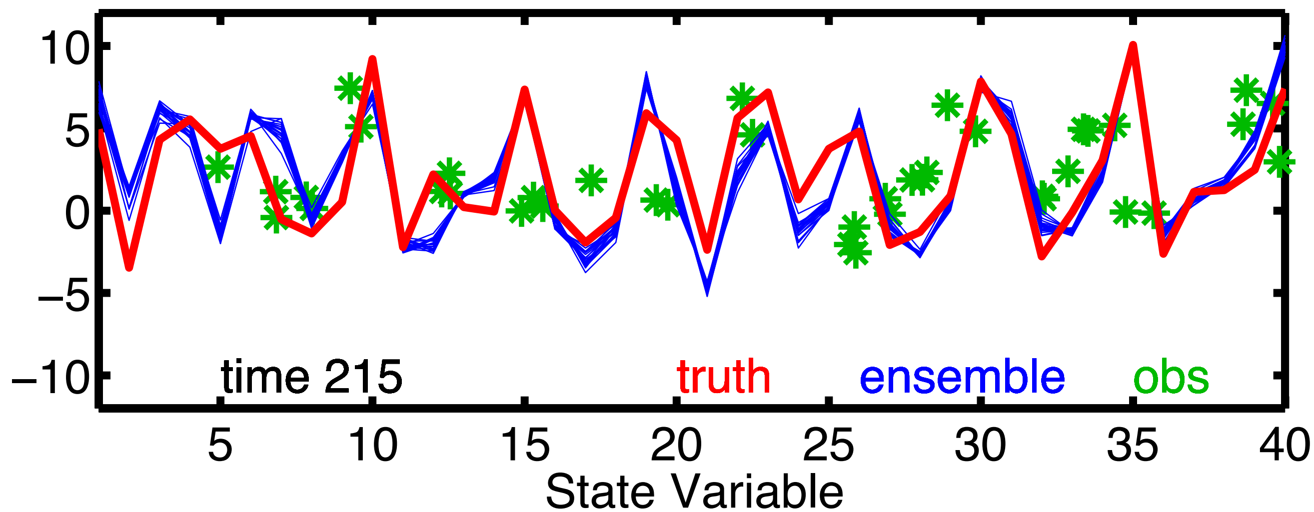
Assimilate ‘observations’ from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from ‘climatological’ 20-member ensemble.



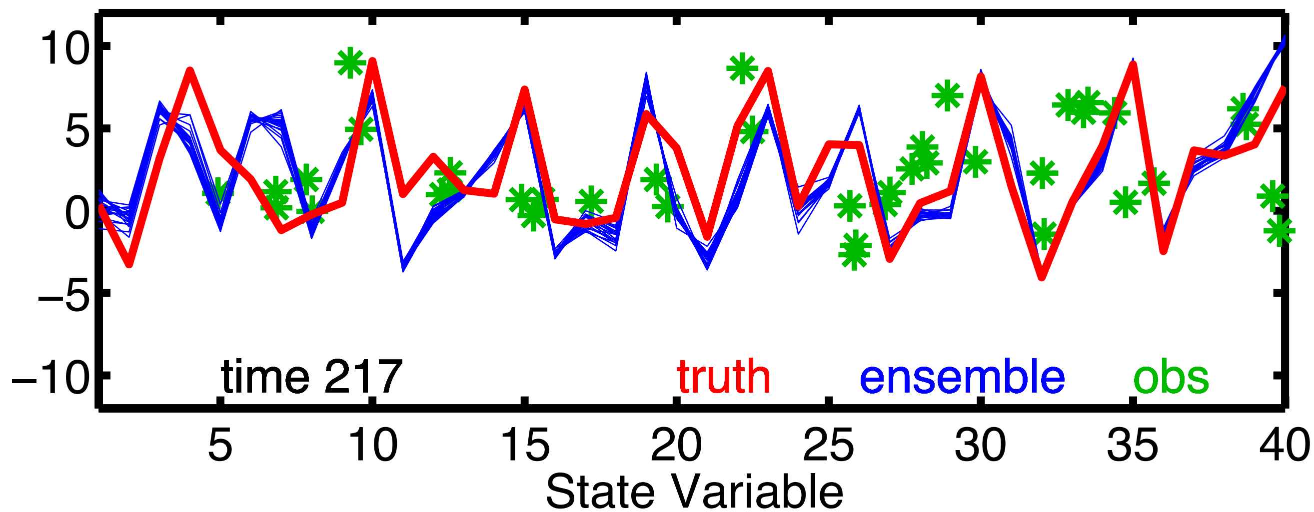
Assimilate 'observations' from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from 'climatological' 20-member ensemble.



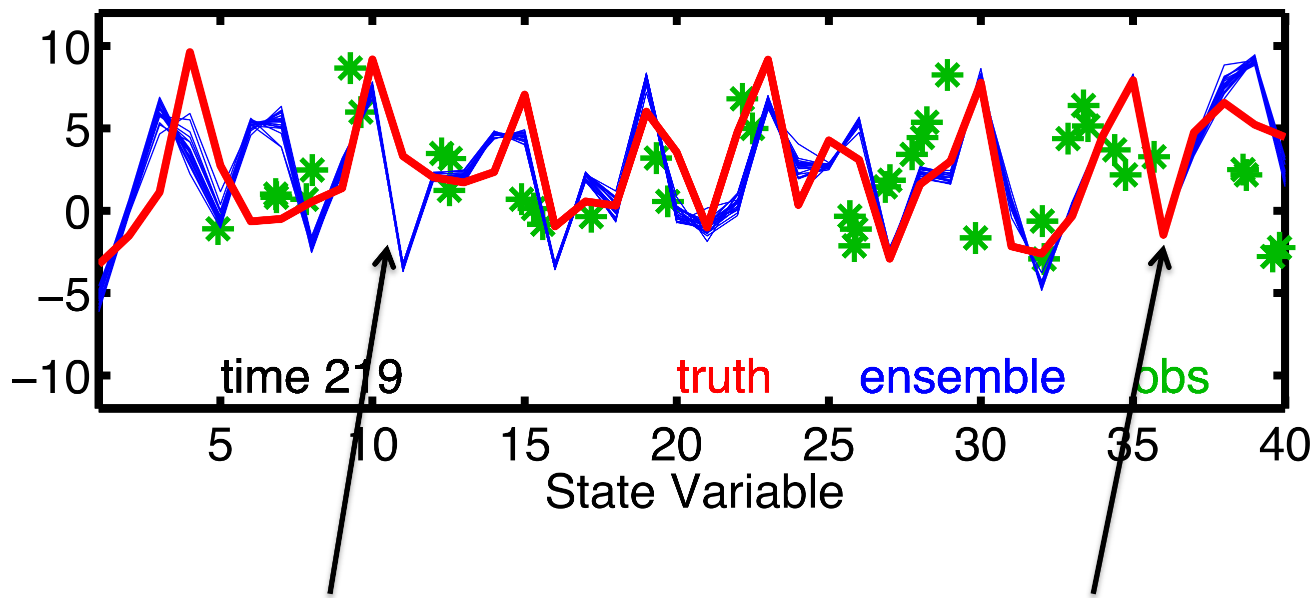
Assimilate ‘observations’ from 40 random locations each step.

Observations generated by interpolating truth to station location.
Simulate observational error: Add random draw from N(0, 16) to each.
Start from ‘climatological’ 20-member ensemble.



Assimilate 'observations' from 40 random locations each step.

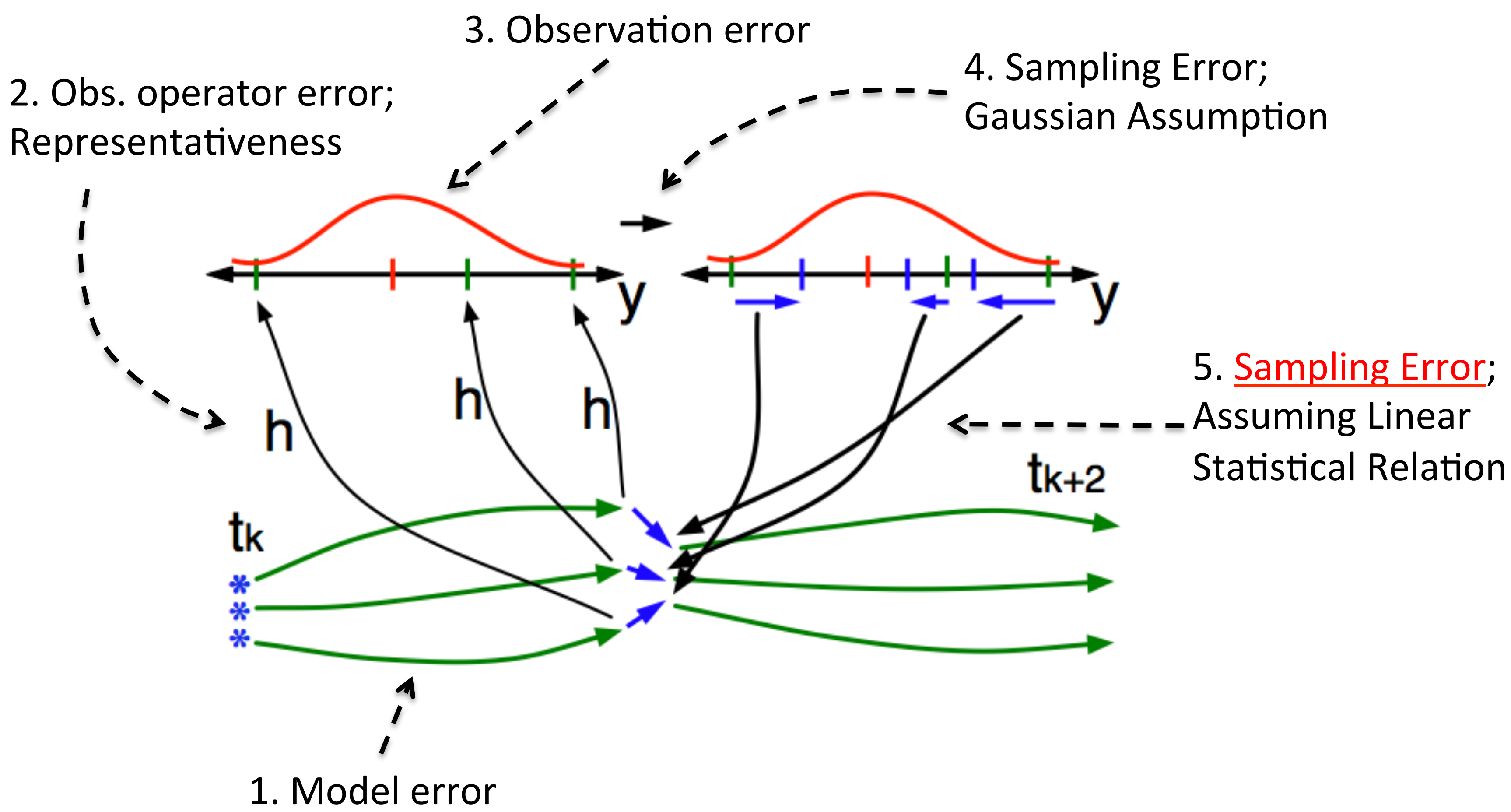
This isn't working very well.
Ensemble spread is reduced, but...,
Ensemble is inconsistent with truth most places.



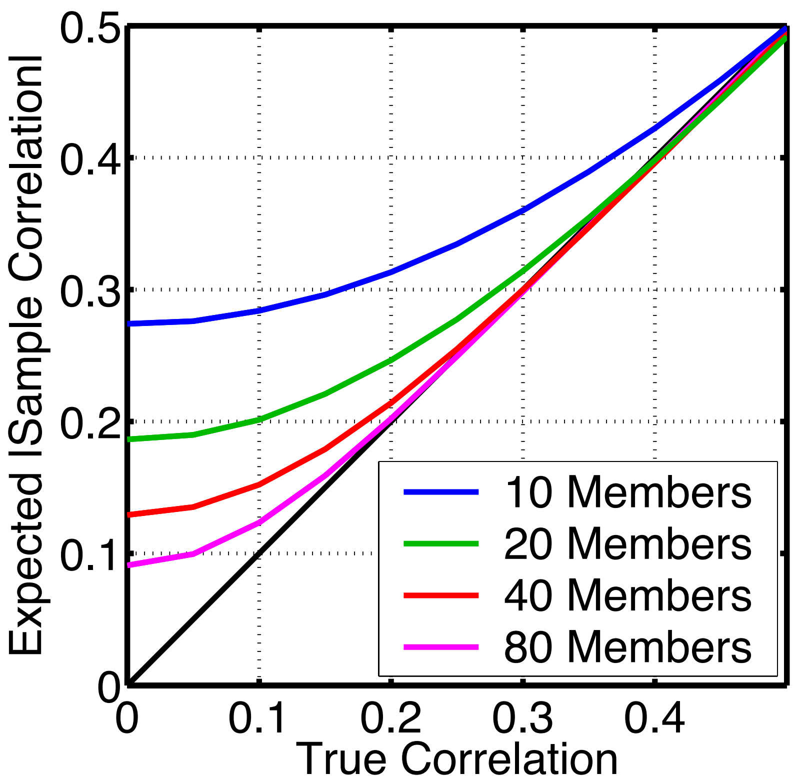
Confident and wrong.

Confident and right.

Some Error Sources in Ensemble Filters



Observations impact unrelated state variables through sampling error.



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

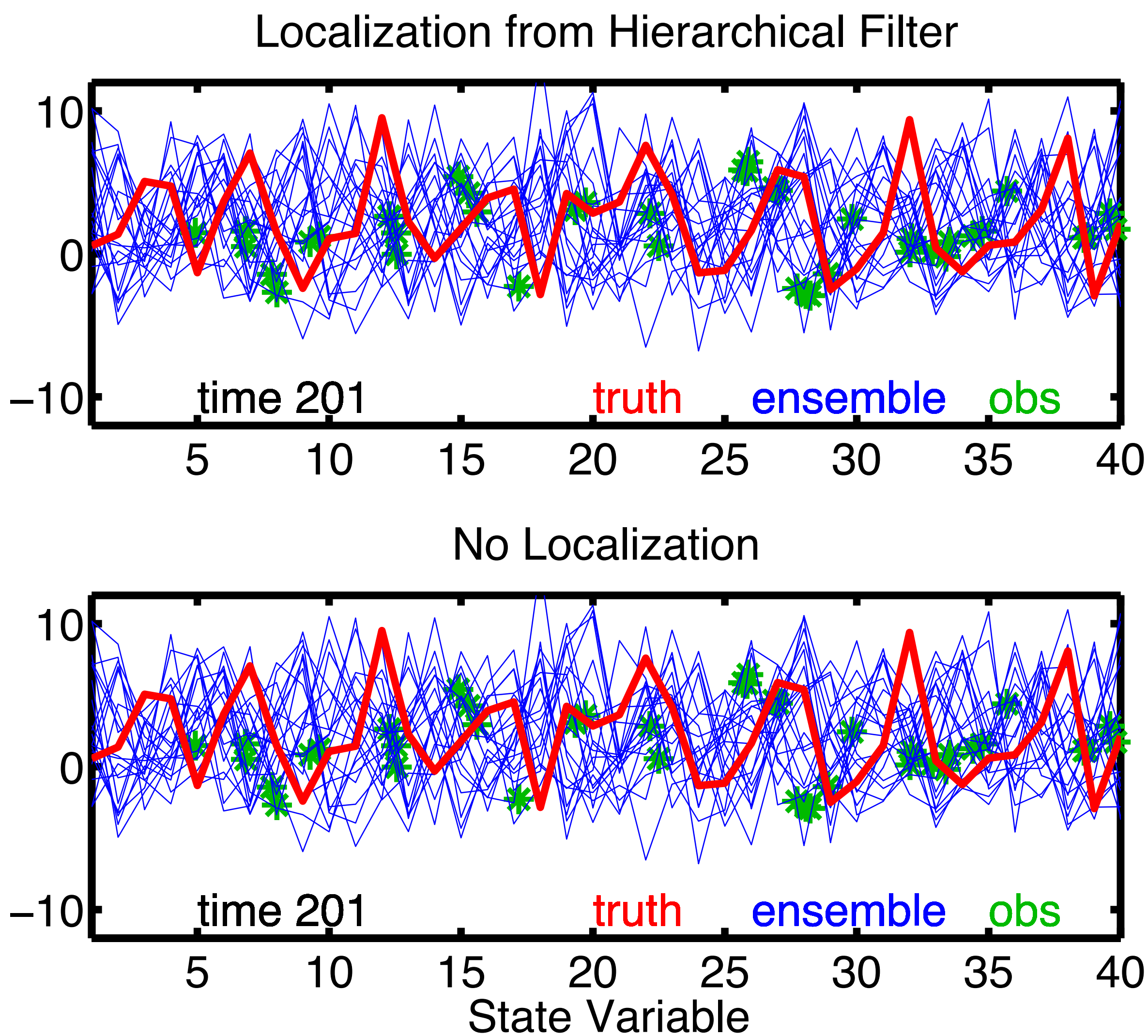
Unrelated obs. reduce spread, increase error.

Attack with localization.

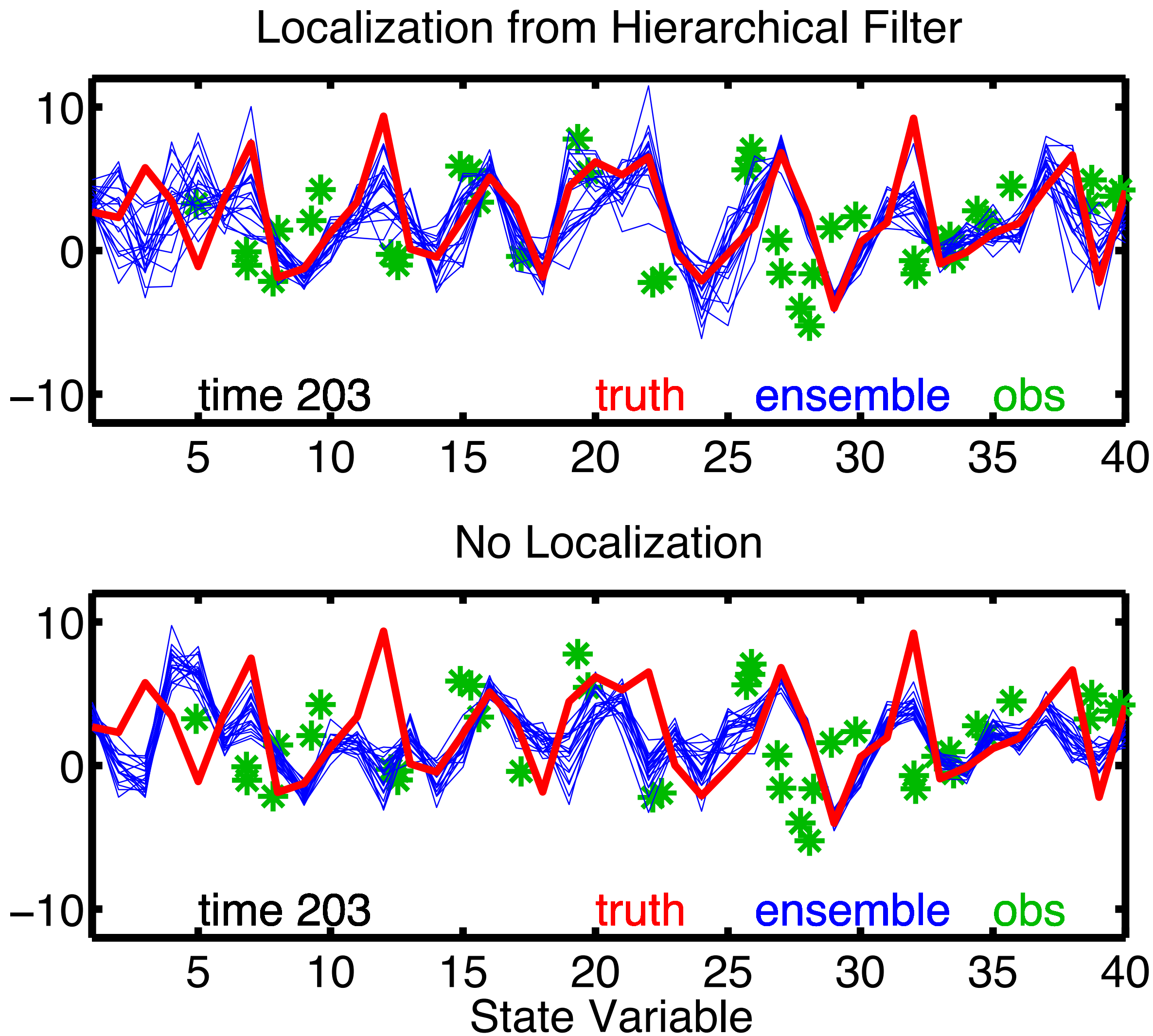
Reduce impact of observation on weakly correlated state variables.

Let weight go to zero for many 'unrelated' variables to save on computing.

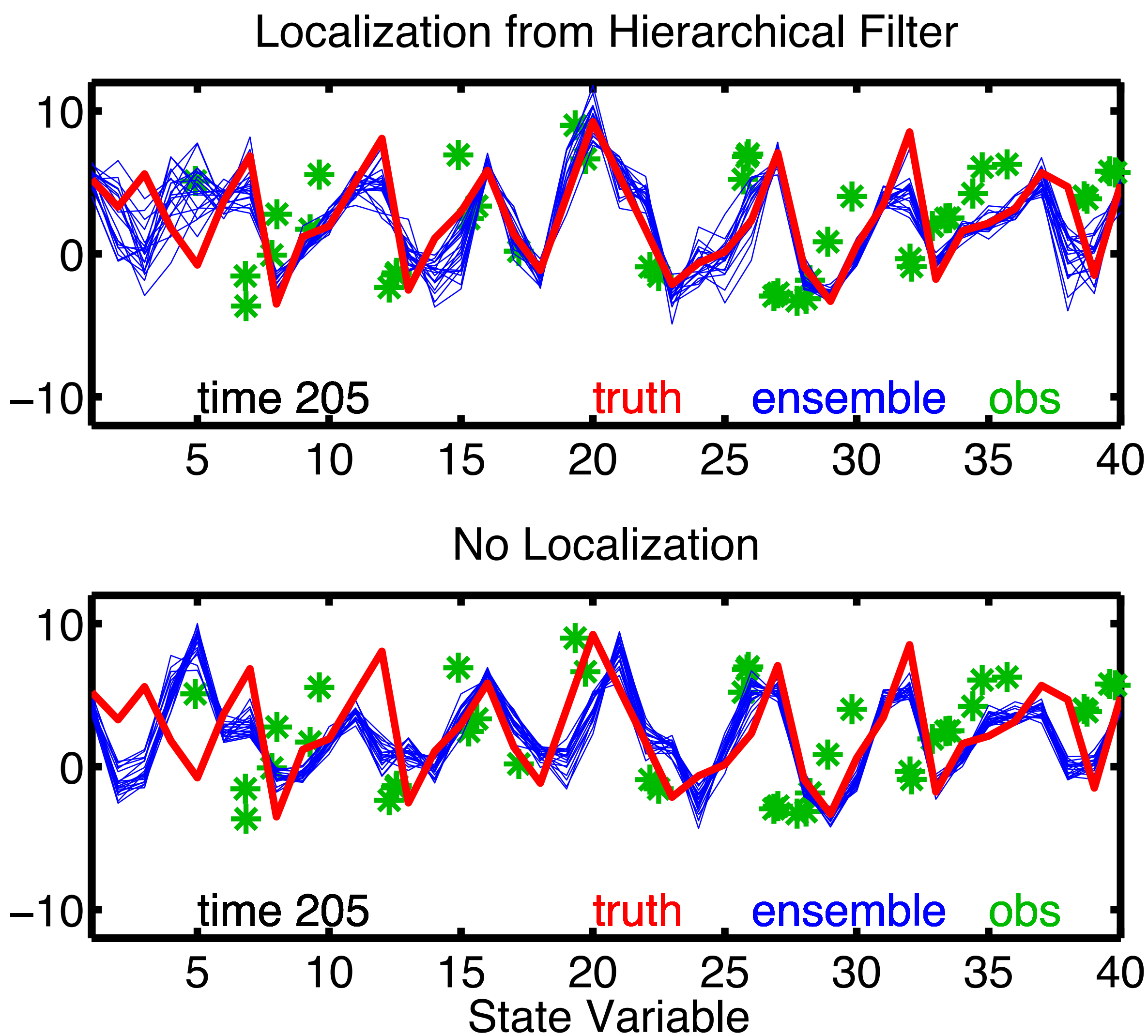
Lorenz-96 Assimilation with localization of observation impact.



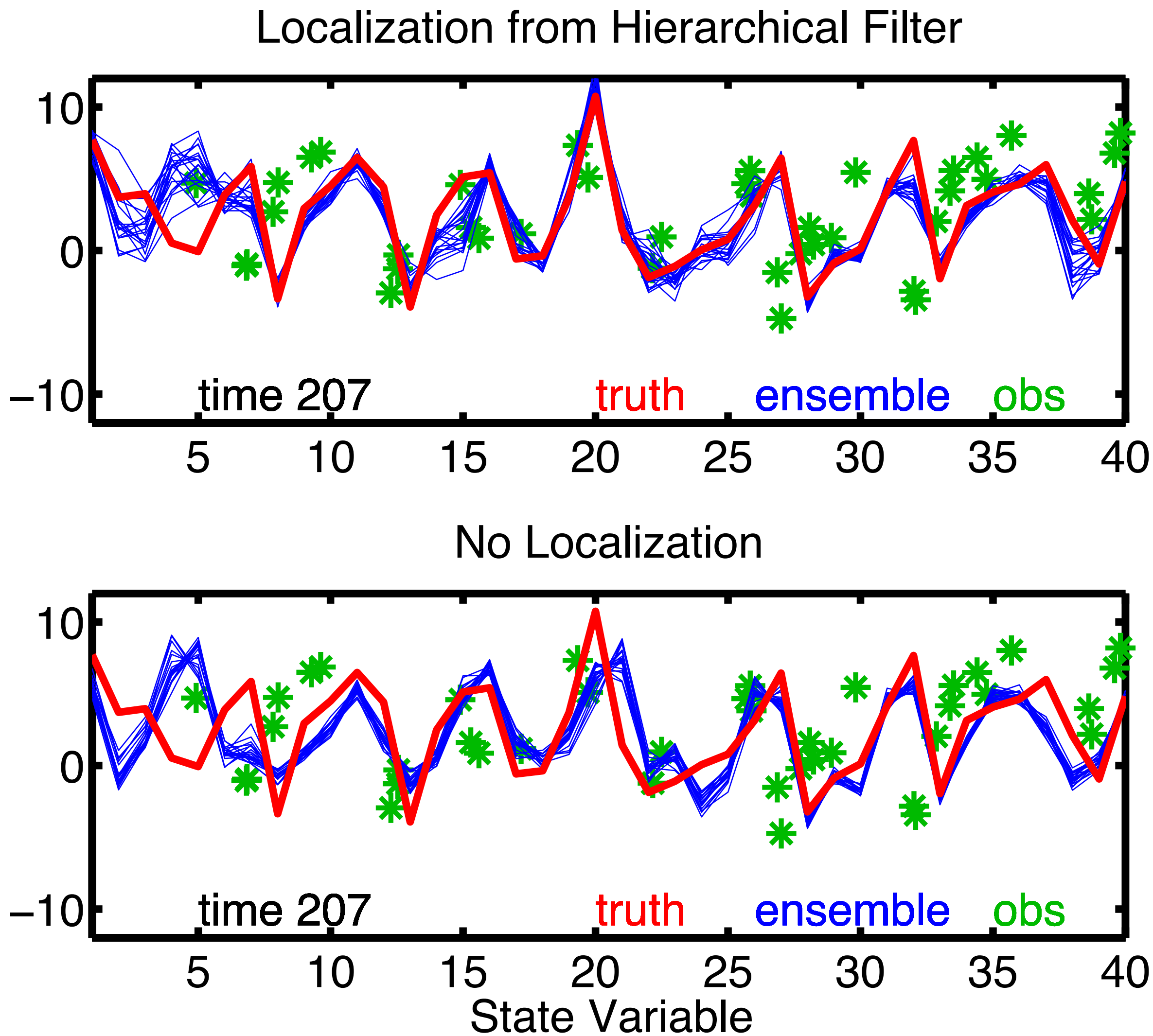
Lorenz-96 Assimilation with localization of observation impact.



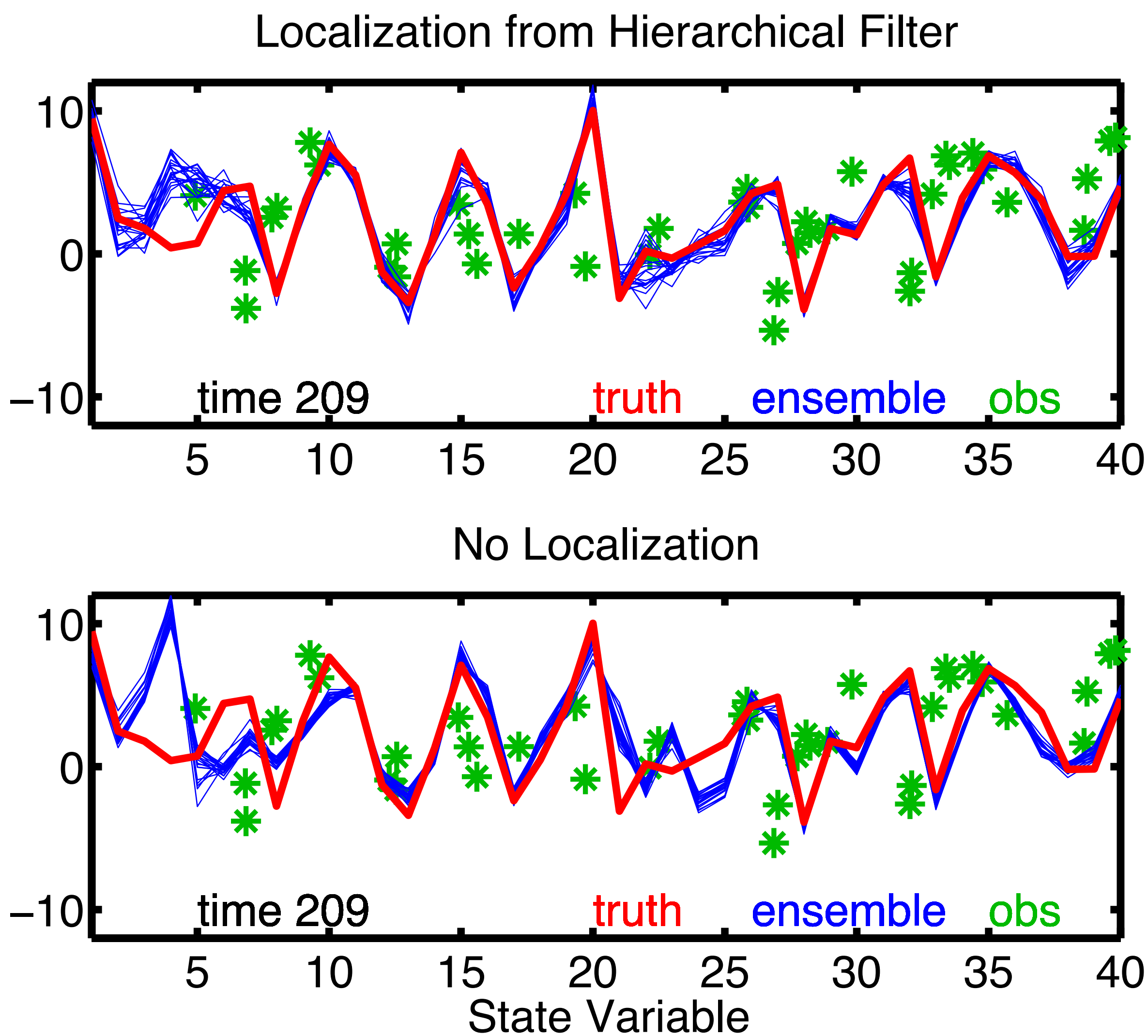
Lorenz-96 Assimilation with localization of observation impact.



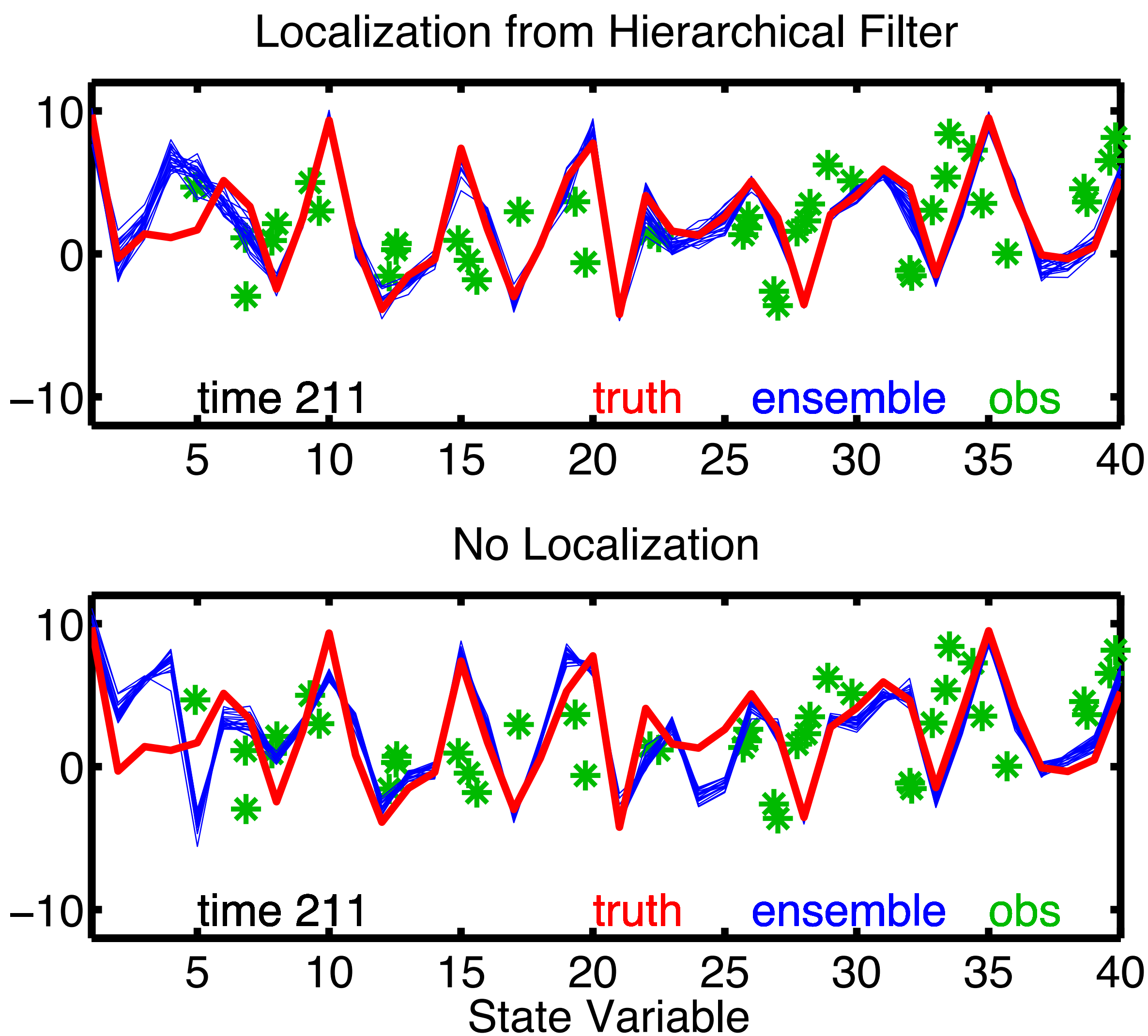
Lorenz-96 Assimilation with localization of observation impact.



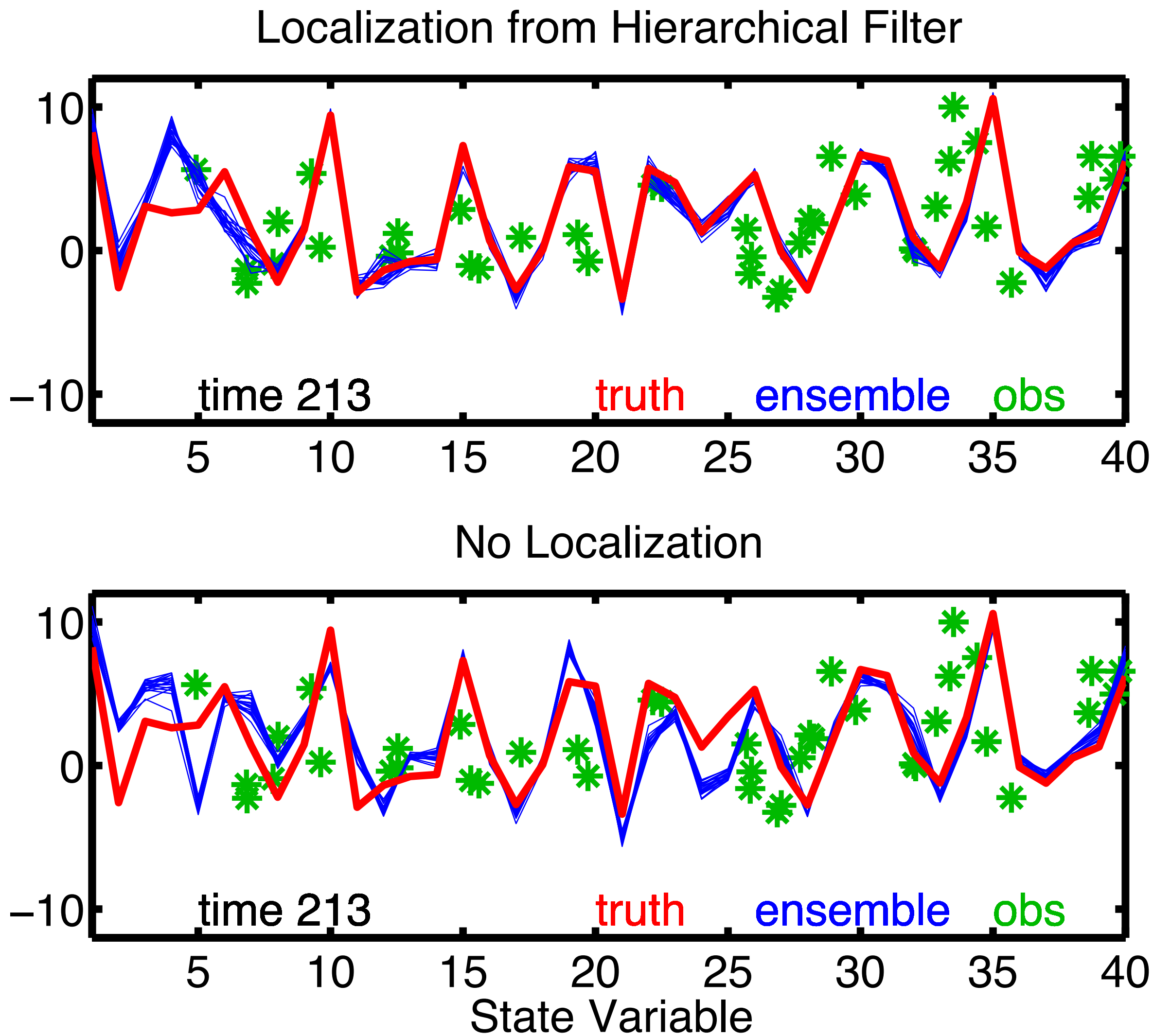
Lorenz-96 Assimilation with localization of observation impact.



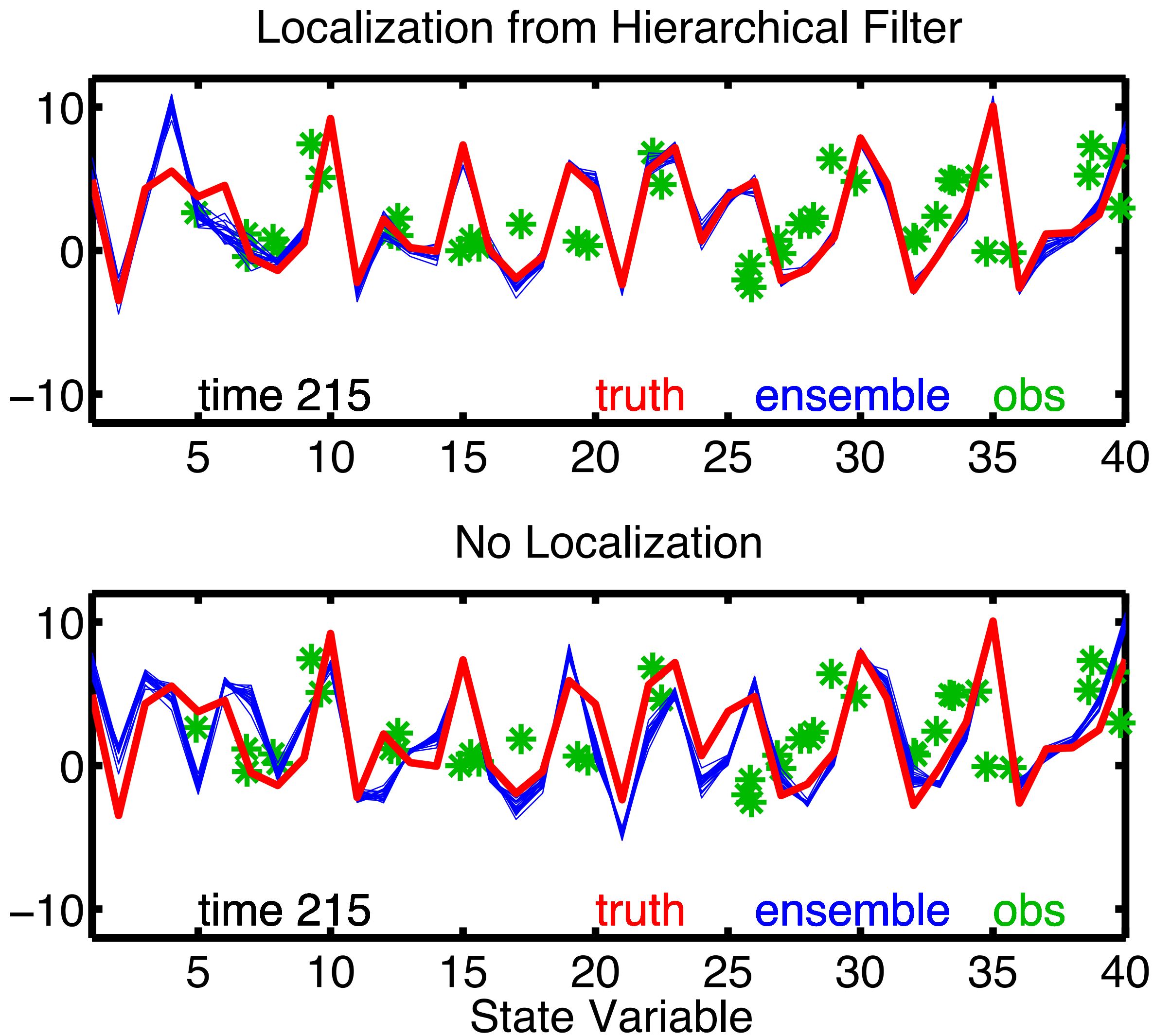
Lorenz-96 Assimilation with localization of observation impact.



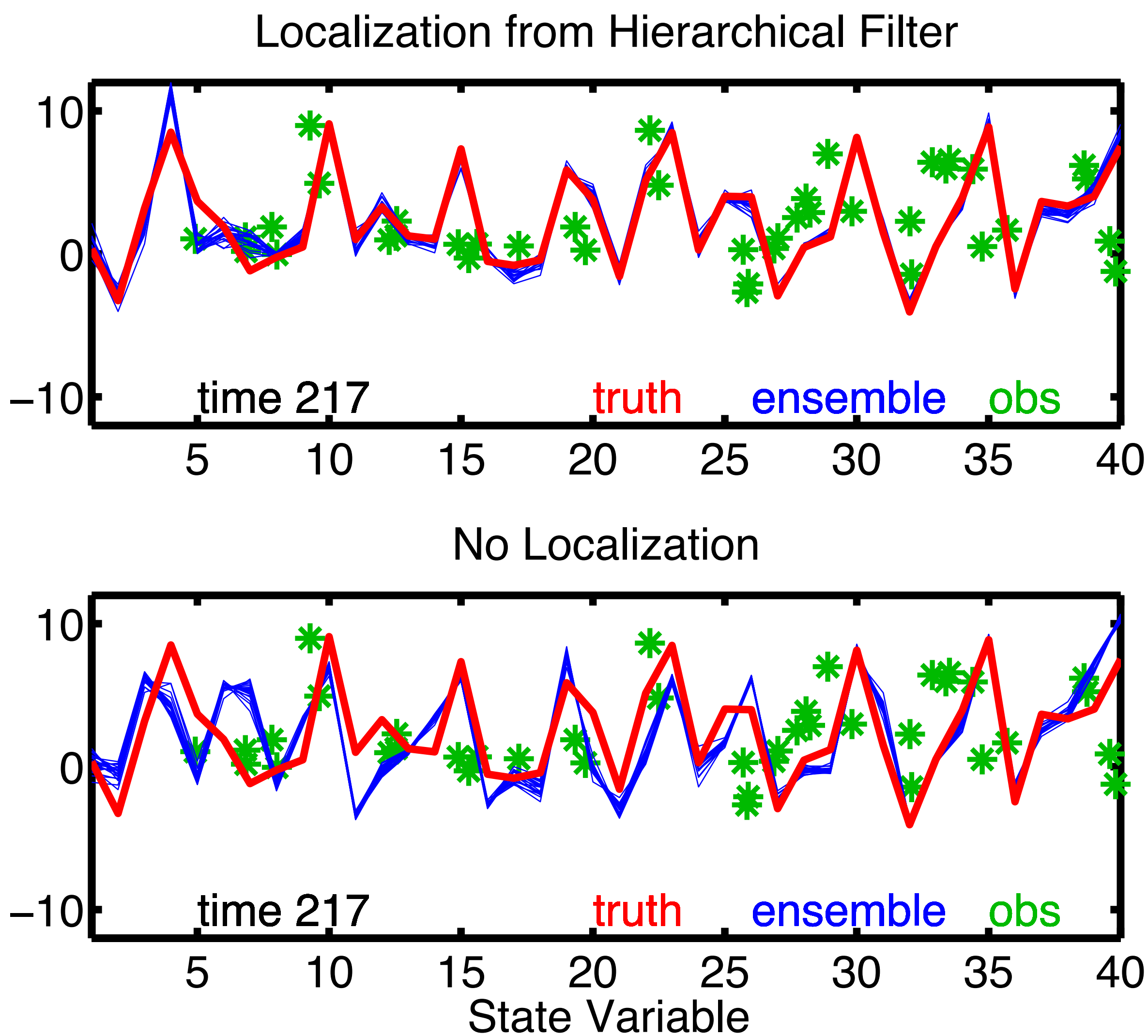
Lorenz-96 Assimilation with localization of observation impact.



Lorenz-96 Assimilation with localization of observation impact.



Lorenz-96 Assimilation with localization of observation impact.



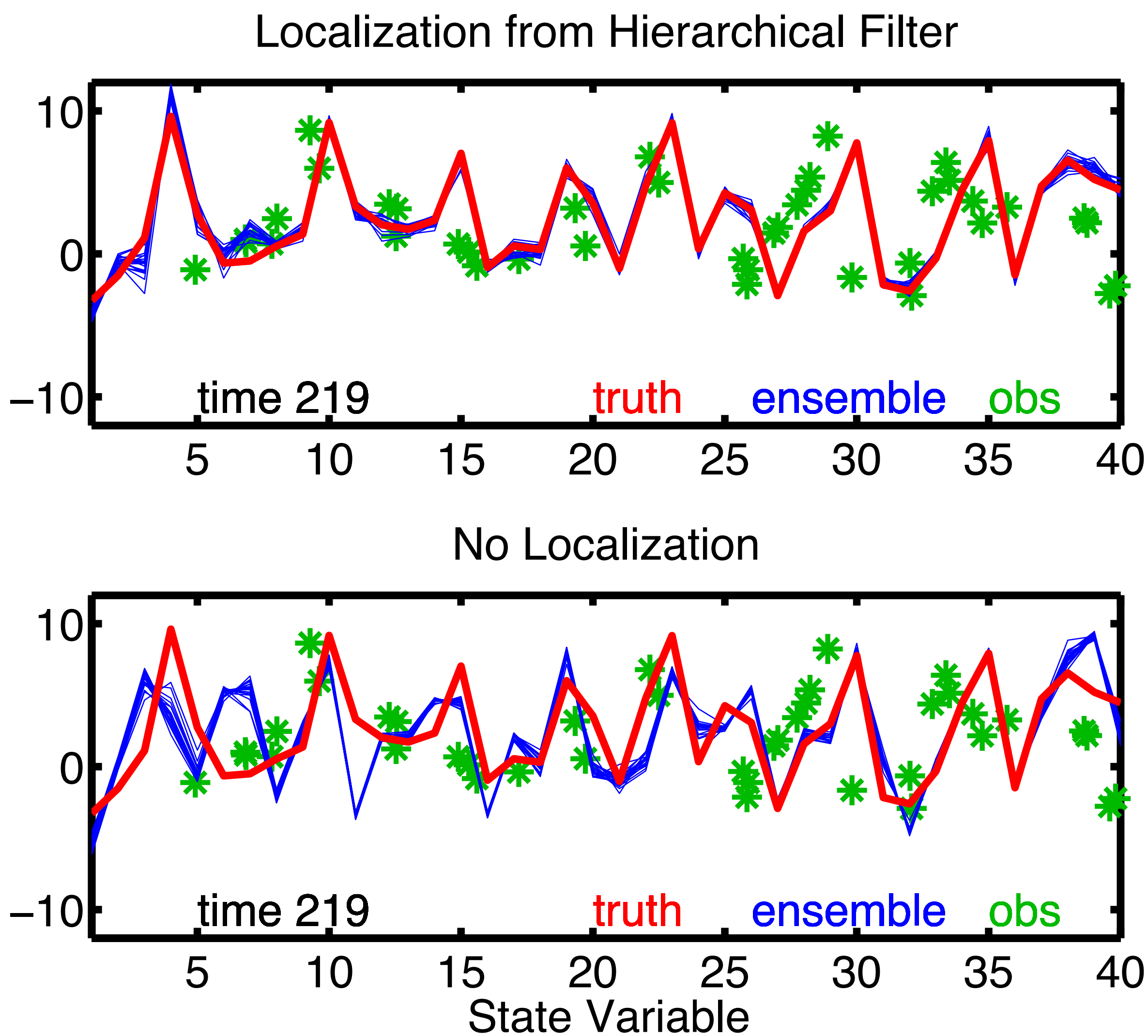
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

NSF

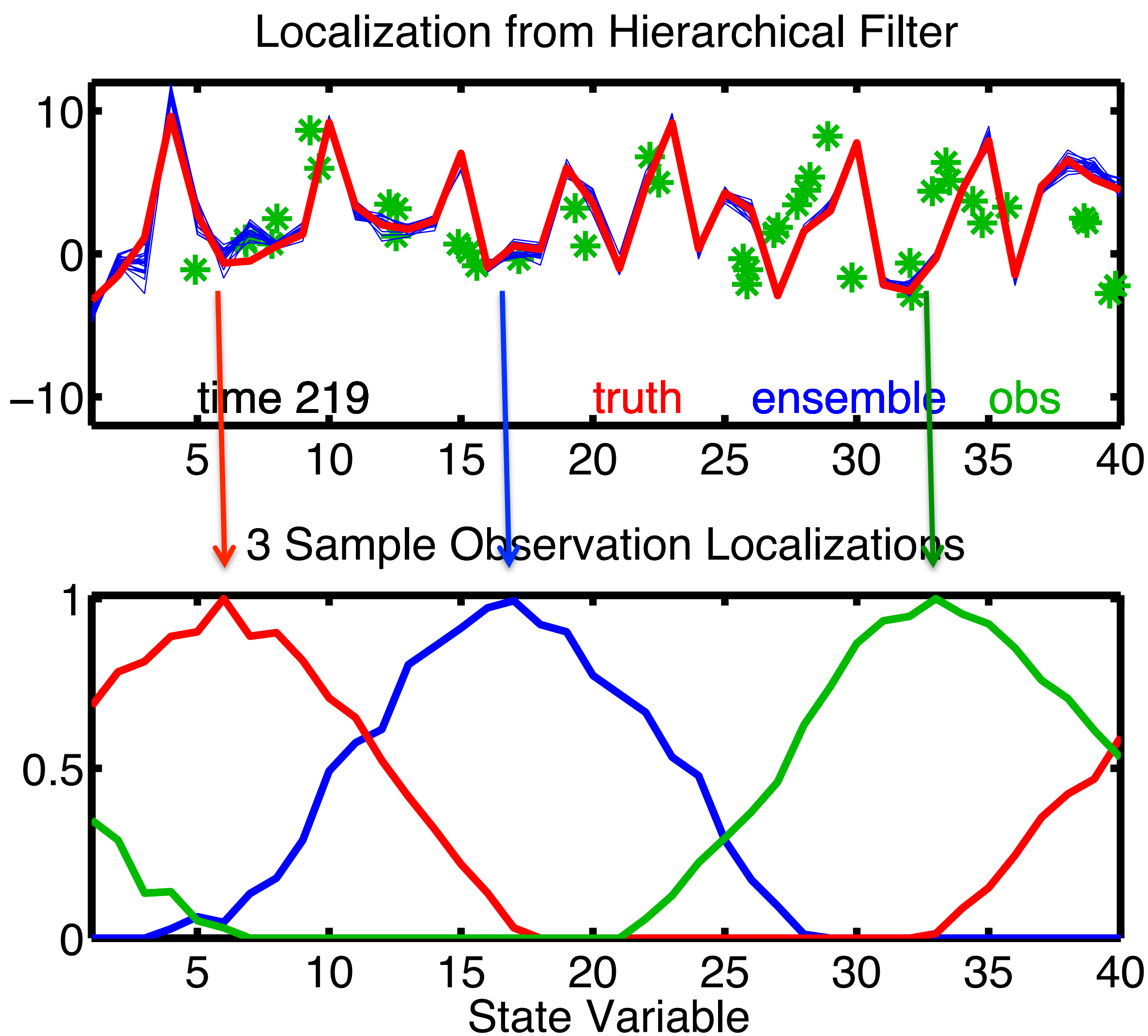
Data Assimilation Research Testbed

Lorenz-96 Assimilation with localization of observation impact.

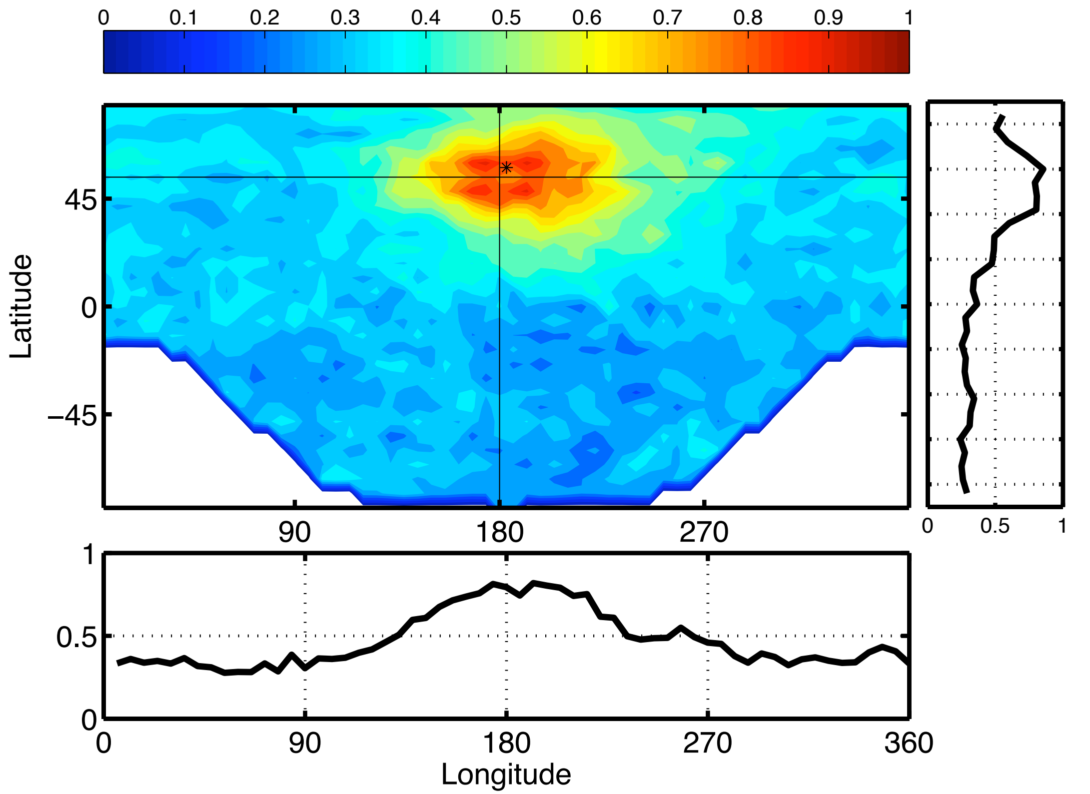
This ensemble is more consistent with truth.



Localization computed by empirical offline computation.

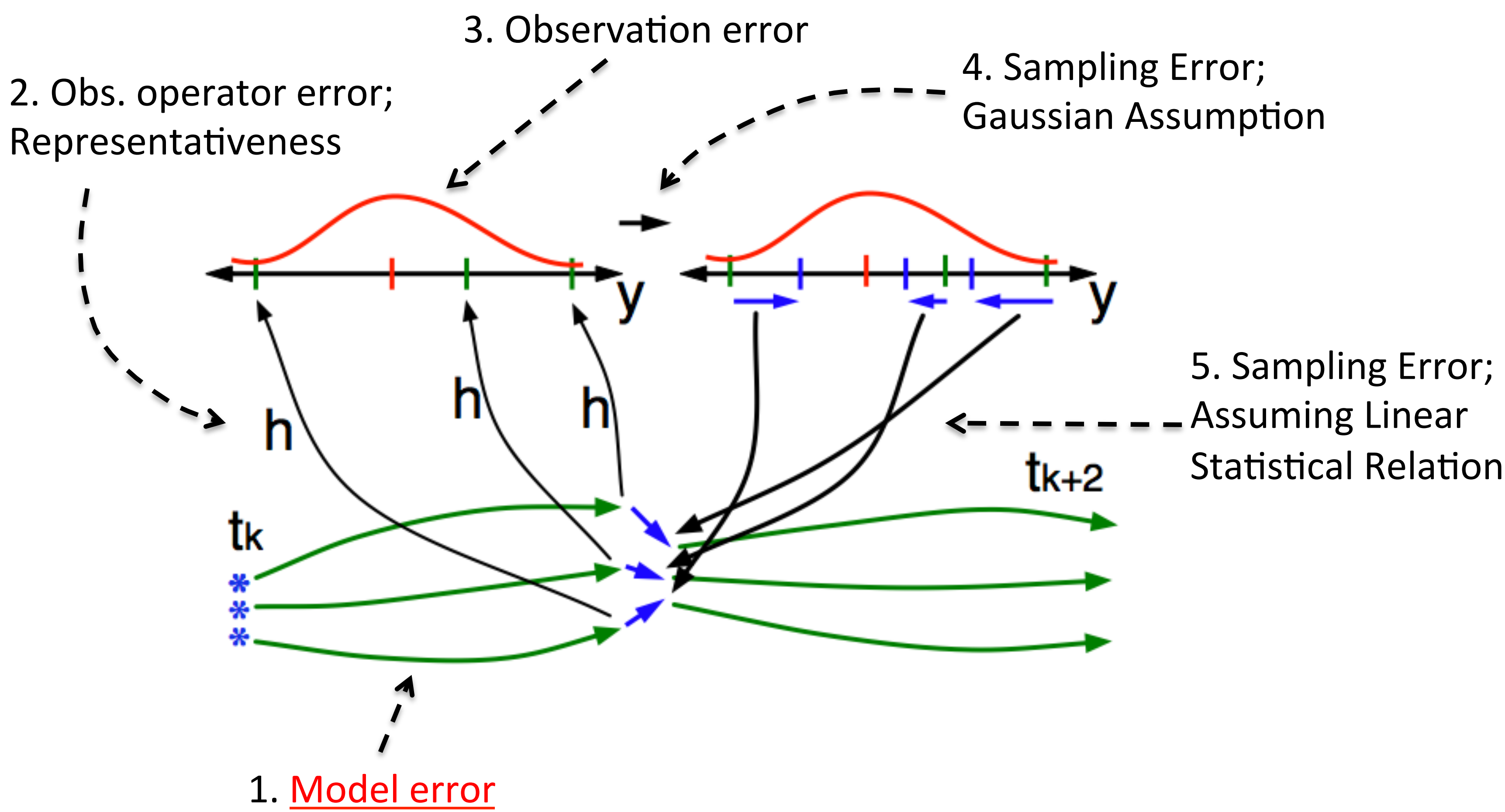


Localization in dry dynamical core



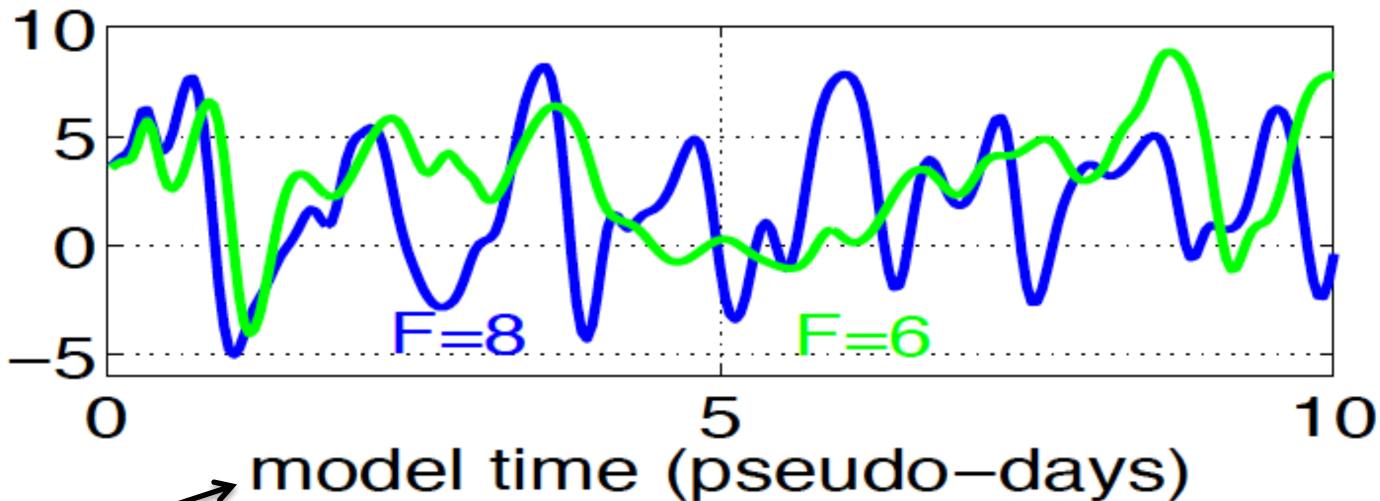
Localization for V ob. on U state variables.
Has statistically significant quadrupole structure in horizontal.
Localization can have lots of structure in realistic models.

Some Error Sources in Ensemble Filters



Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



Note Axis!

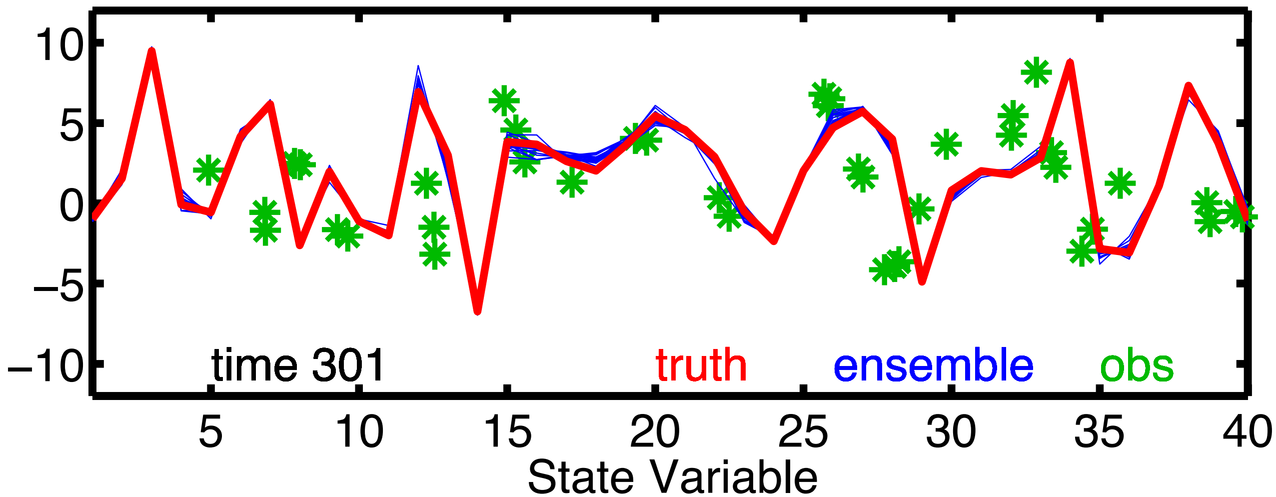
Time evolution for first state variable shown.
Assimilating model quickly diverges from 'true' model.

Assimilating in the presence of simulated model error.

$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F.$

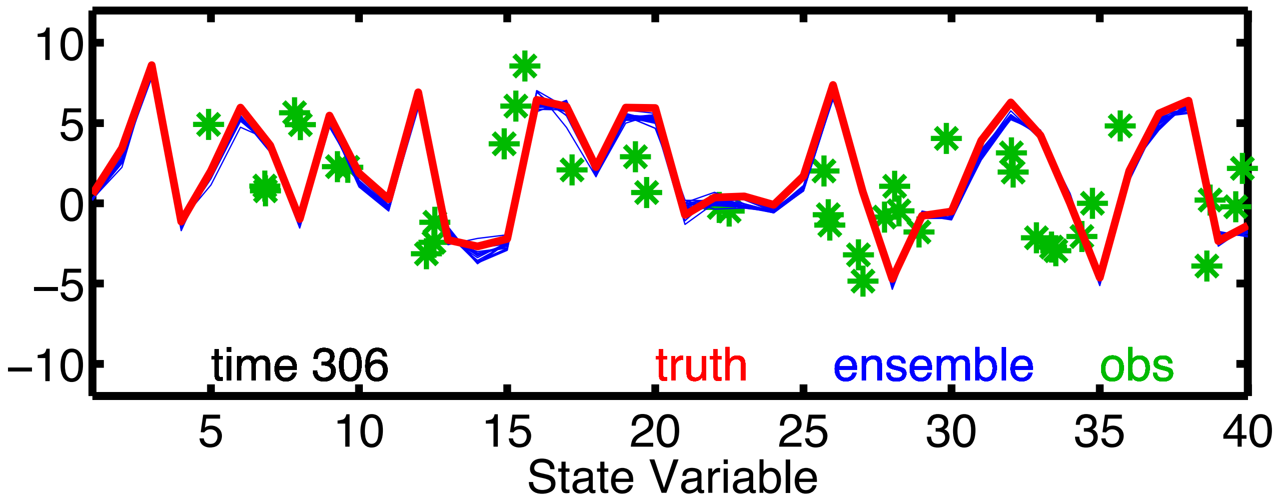
For truth, use F = 8.

In assimilating model, use F = 6.



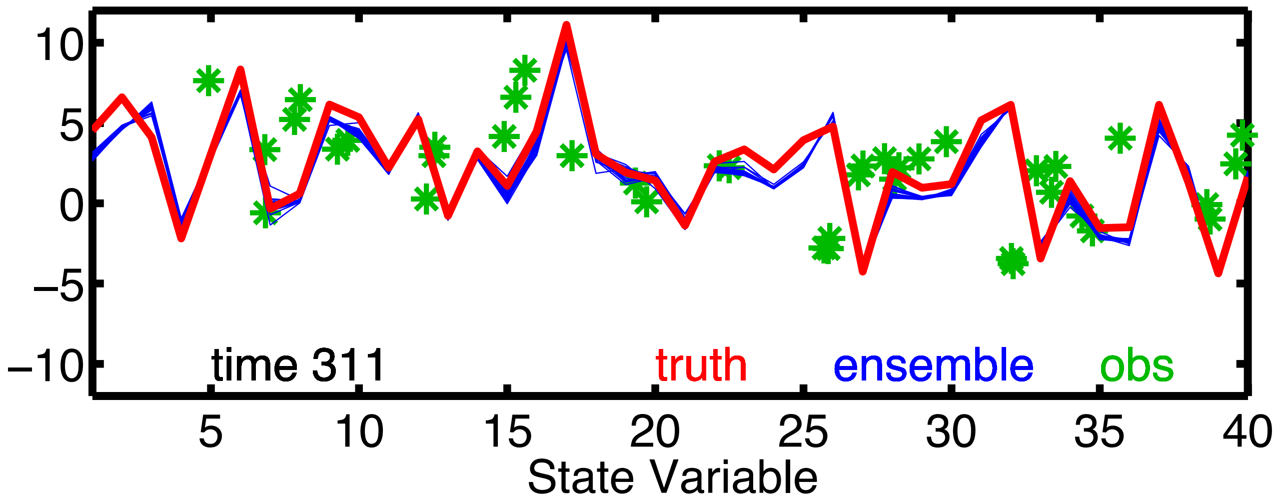
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



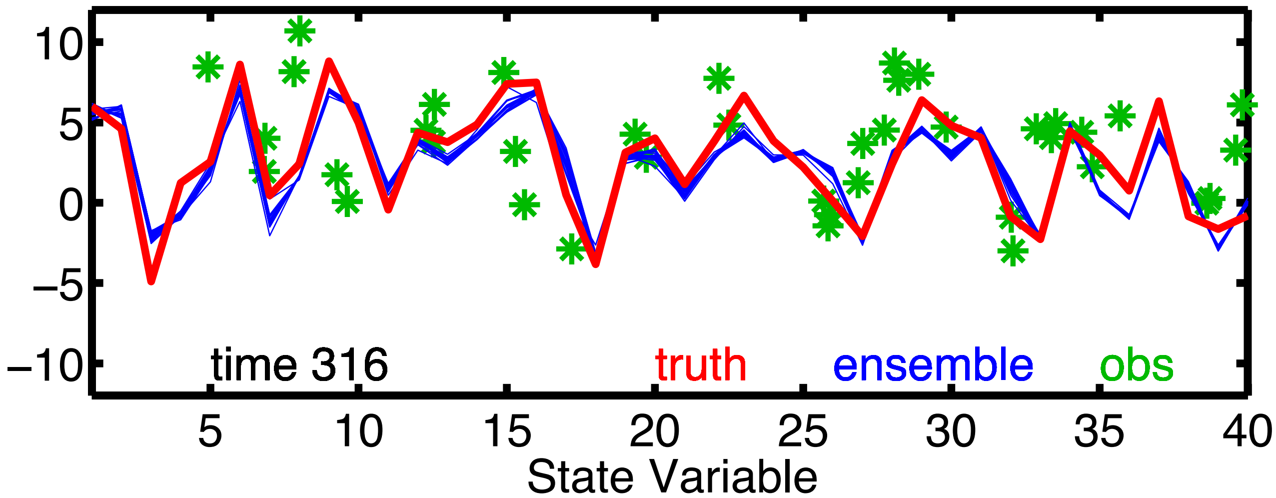
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



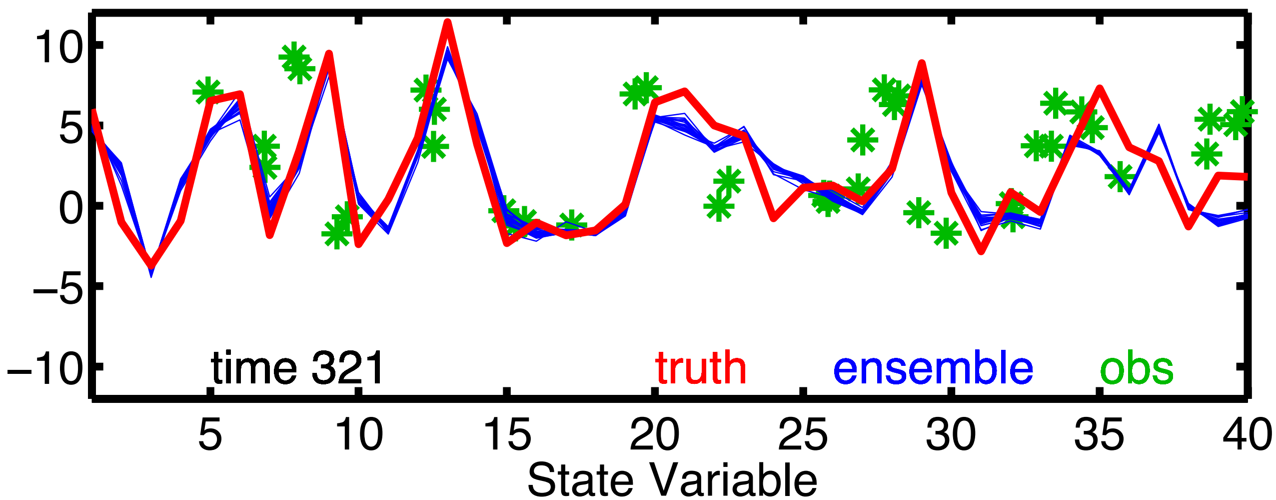
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



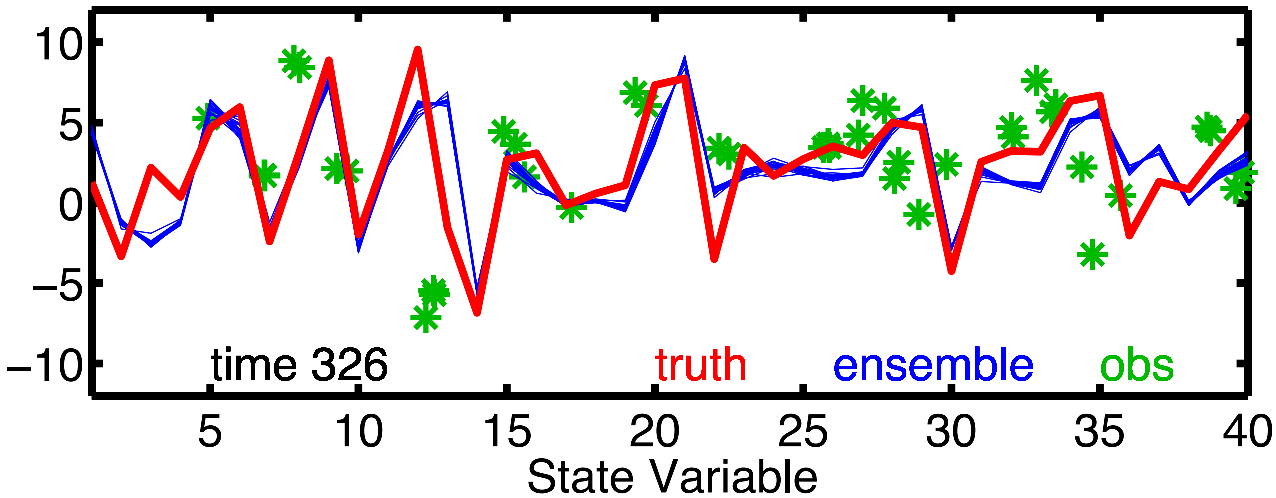
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



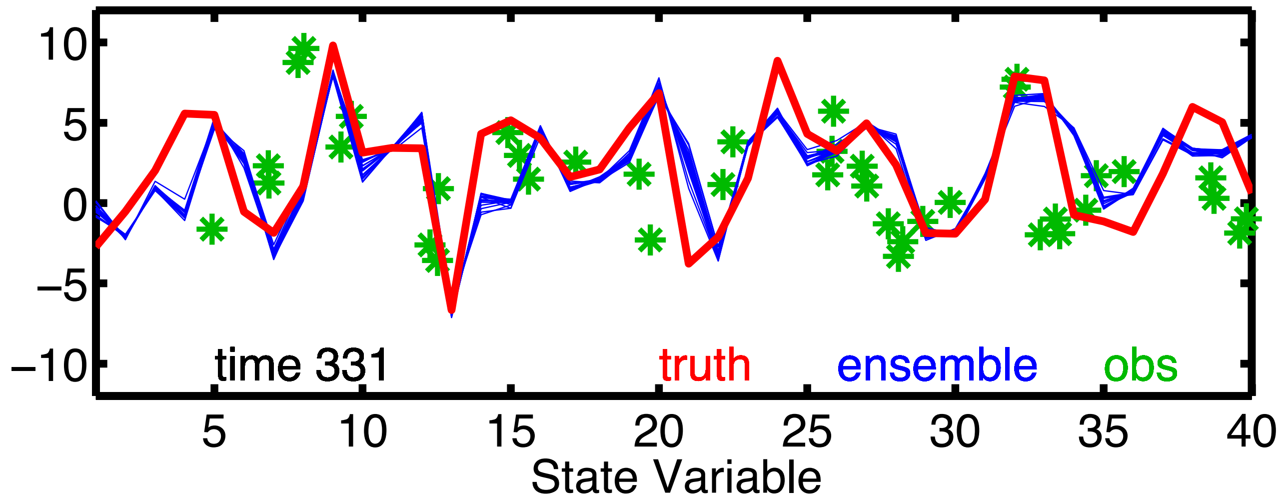
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



Assimilating in the presence of simulated model error.

$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$.
For truth, use F = 8.
In assimilating model, use F = 6.

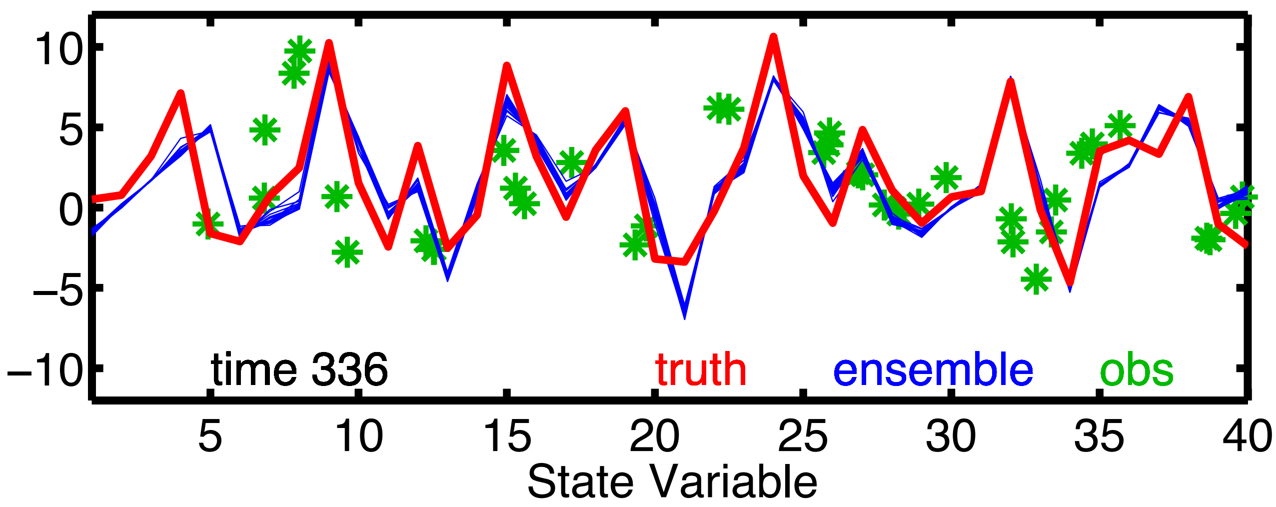


Assimilating in the presence of simulated model error.

$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$.

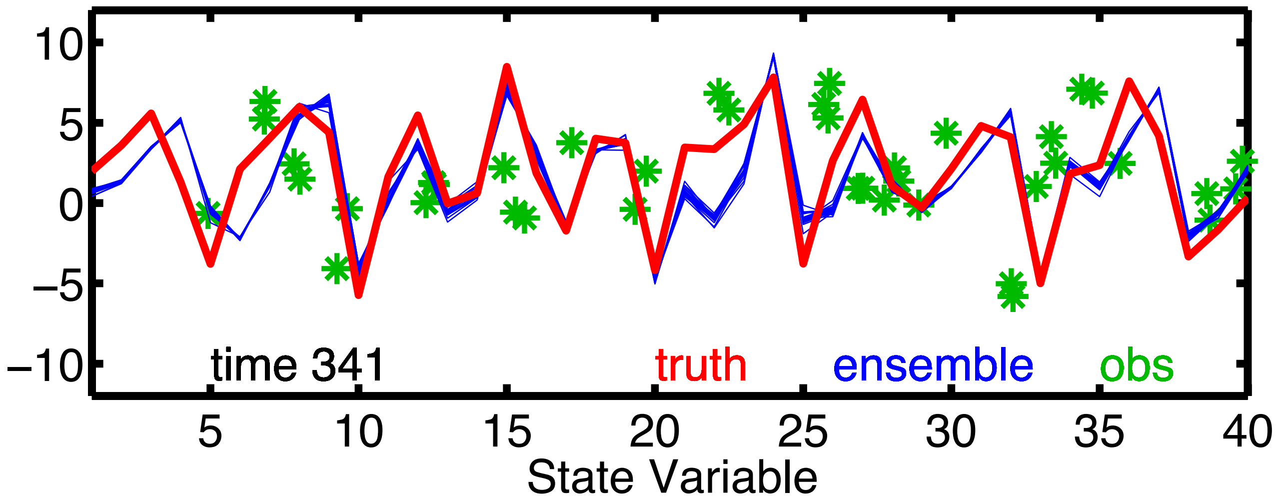
For truth, use F = 8.

In assimilating model, use F = 6.



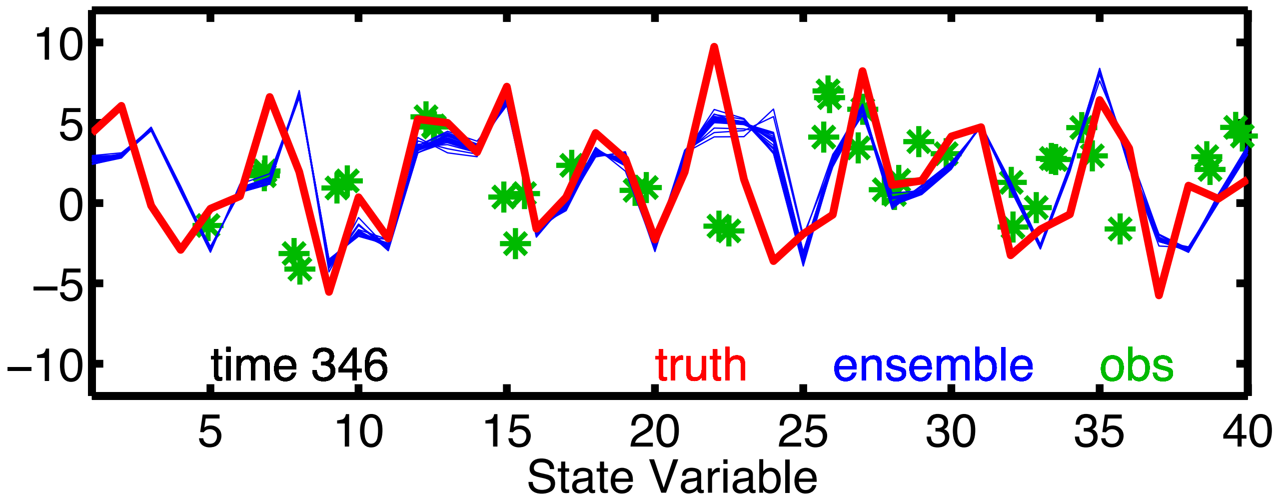
Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



Assimilating in the presence of simulated model error.

dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



This isn't working again!
It will just keep getting worse.

How to deal with model error in assimilation?

- Just fix the model! (May not be such an easy thing to do)

- Model the model error and include this in prior:

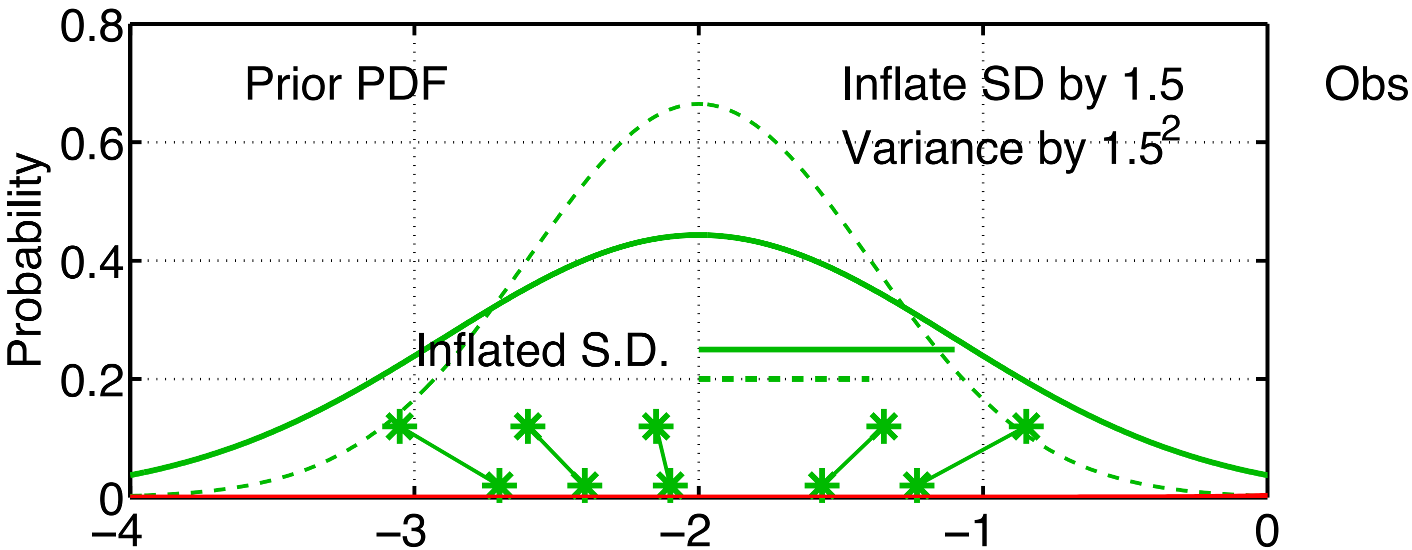
  - If we knew the error accurately could just fix the model.

  - Modeling or parameterizing model uncertainty in prior can help.

  - This looks like a job for stochastic modeling.

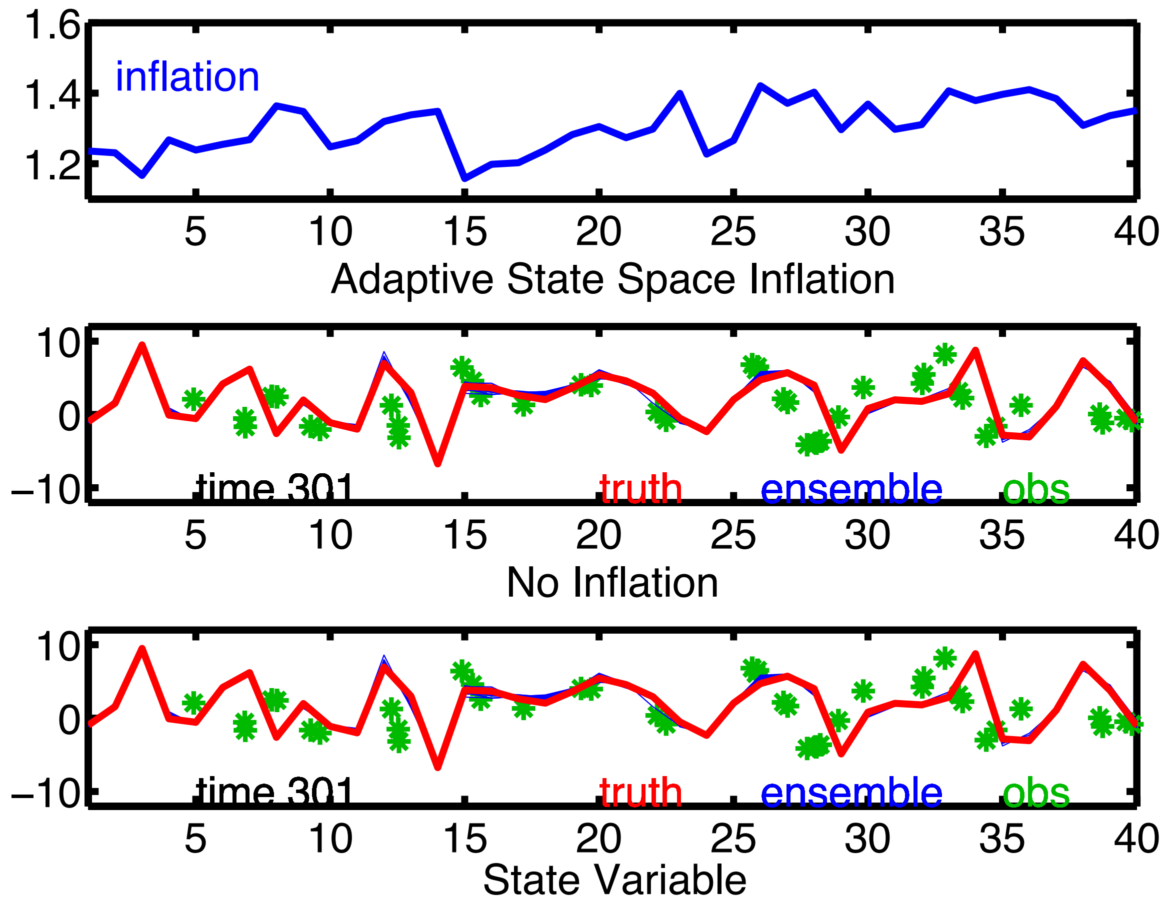
Assimilation fix for model error.

- Use prior variance inflation.
- Simply increase prior ensemble variance of each state variable before computing observation increments.
- Adaptive algorithms use observations to guide this.



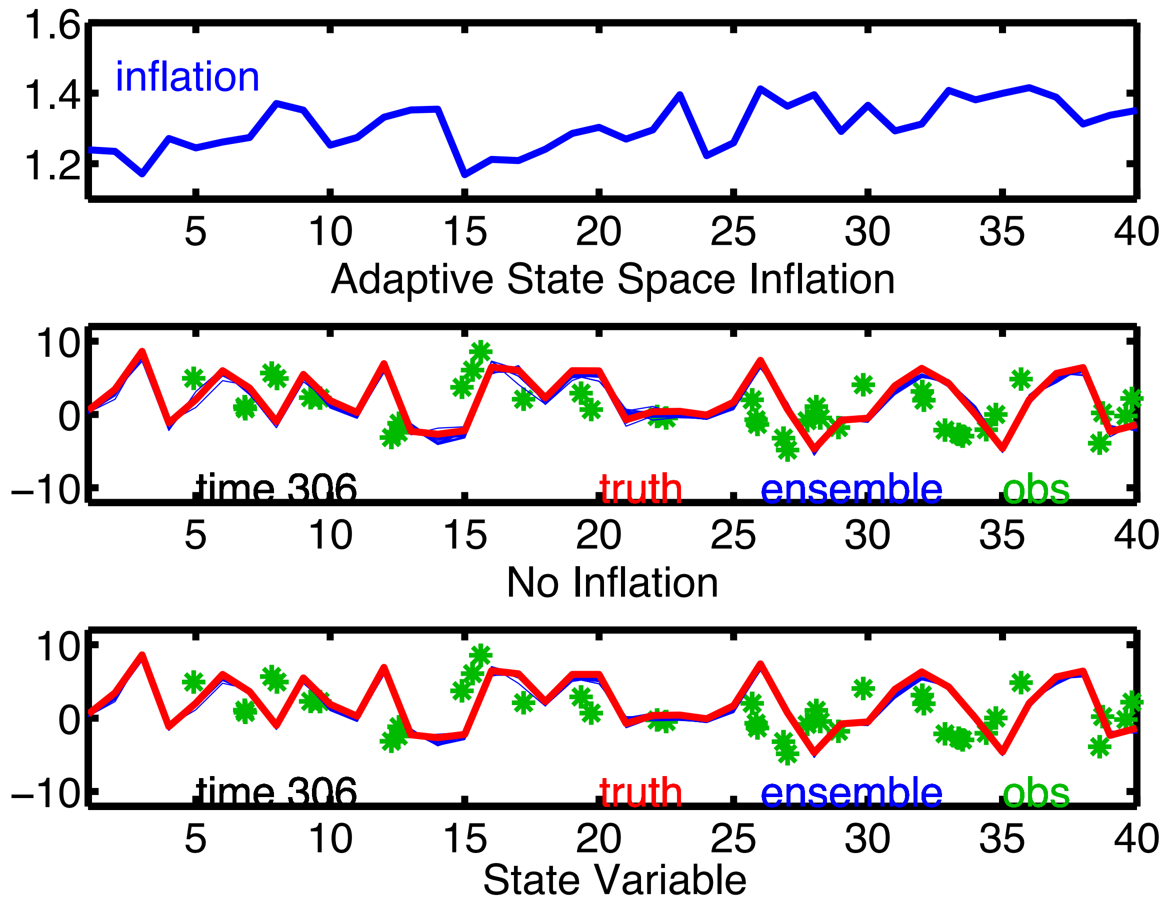
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



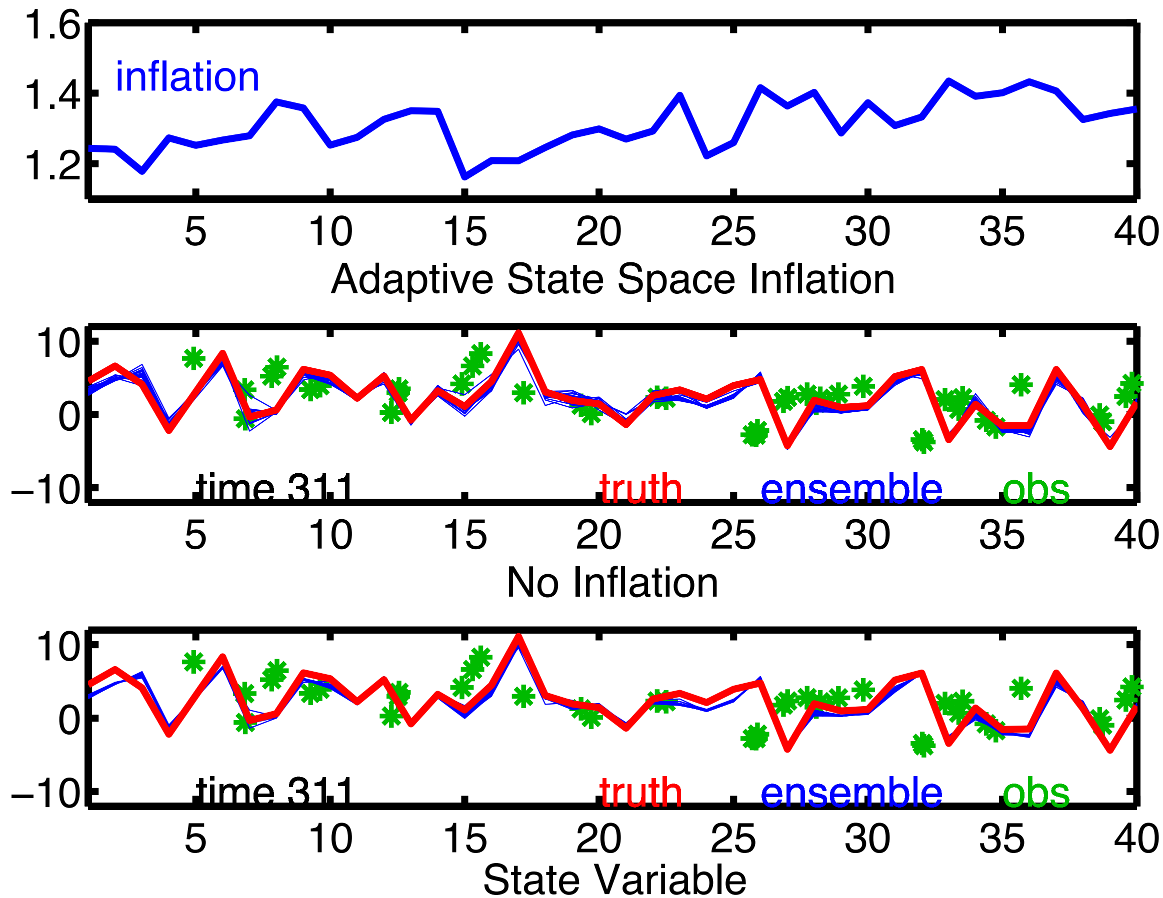
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



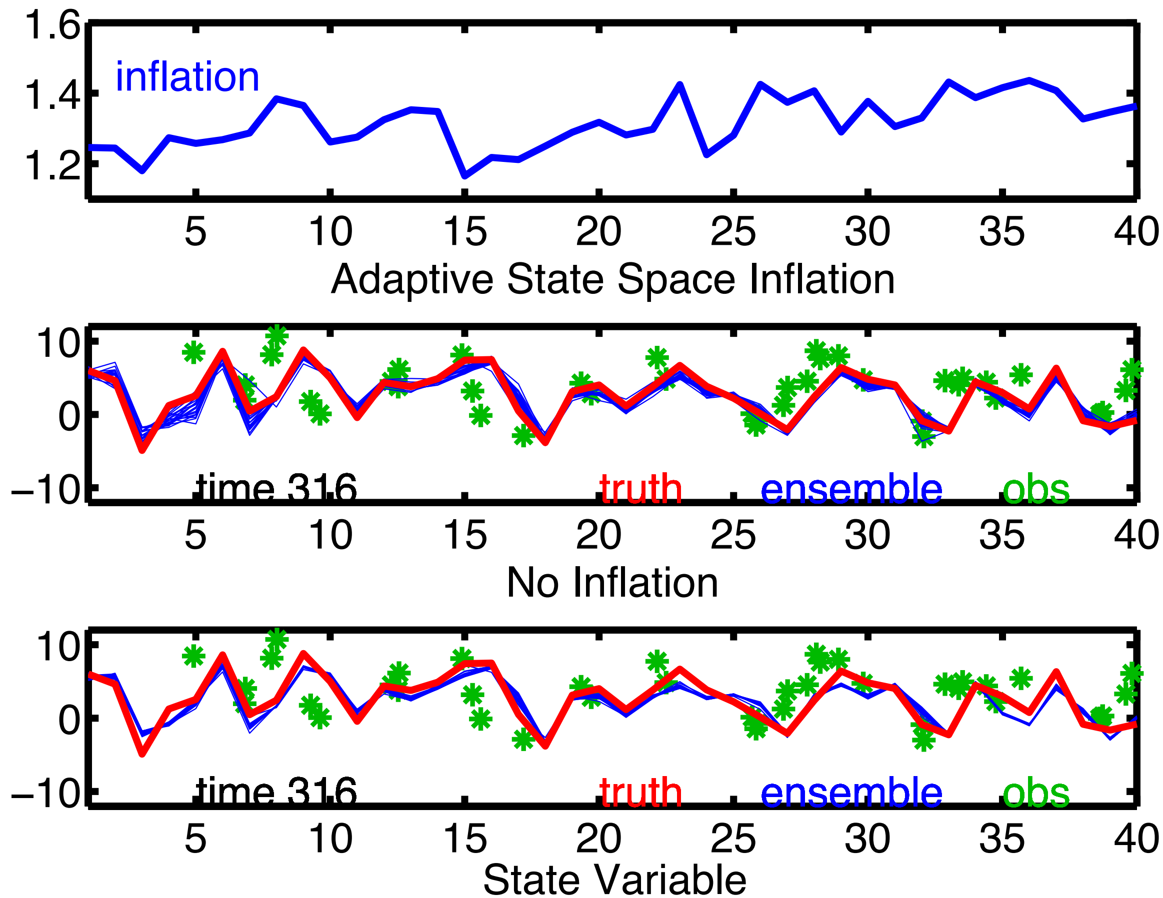
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



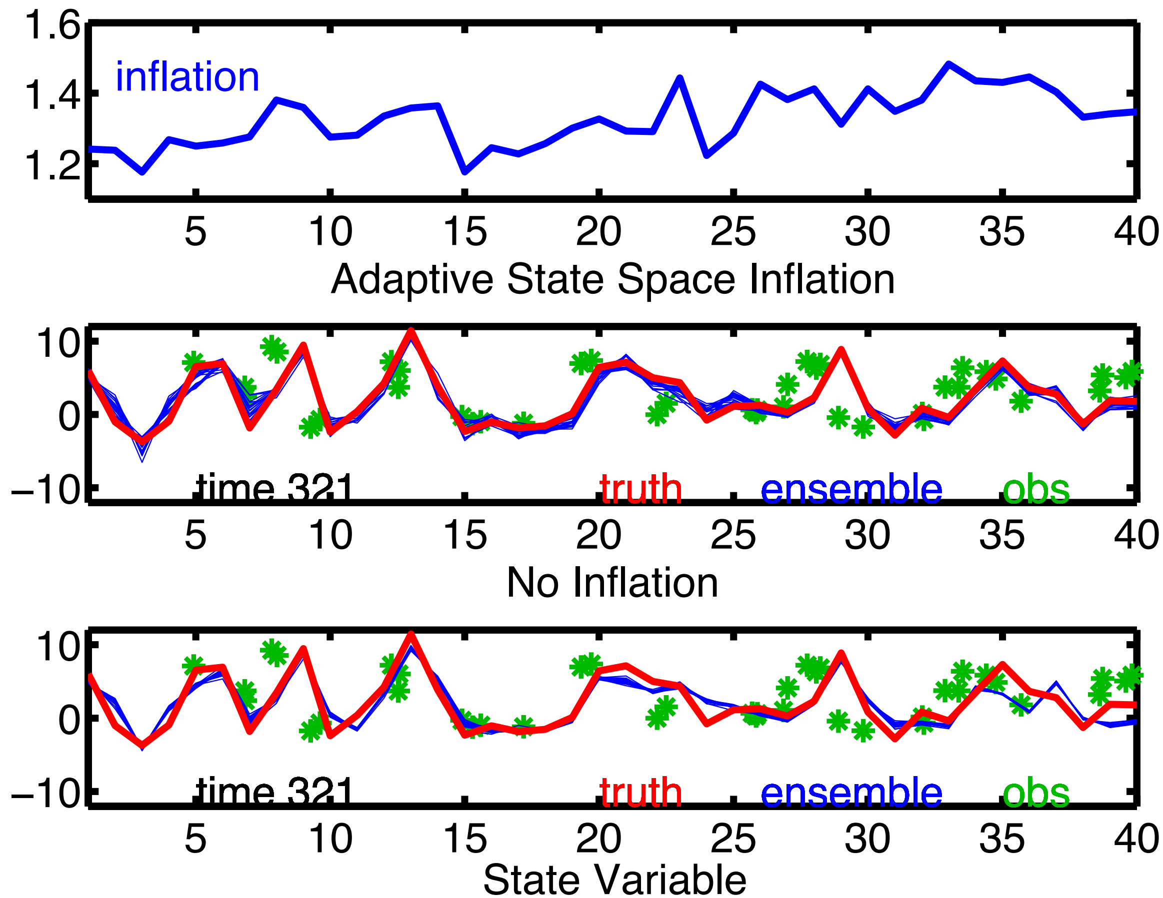
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



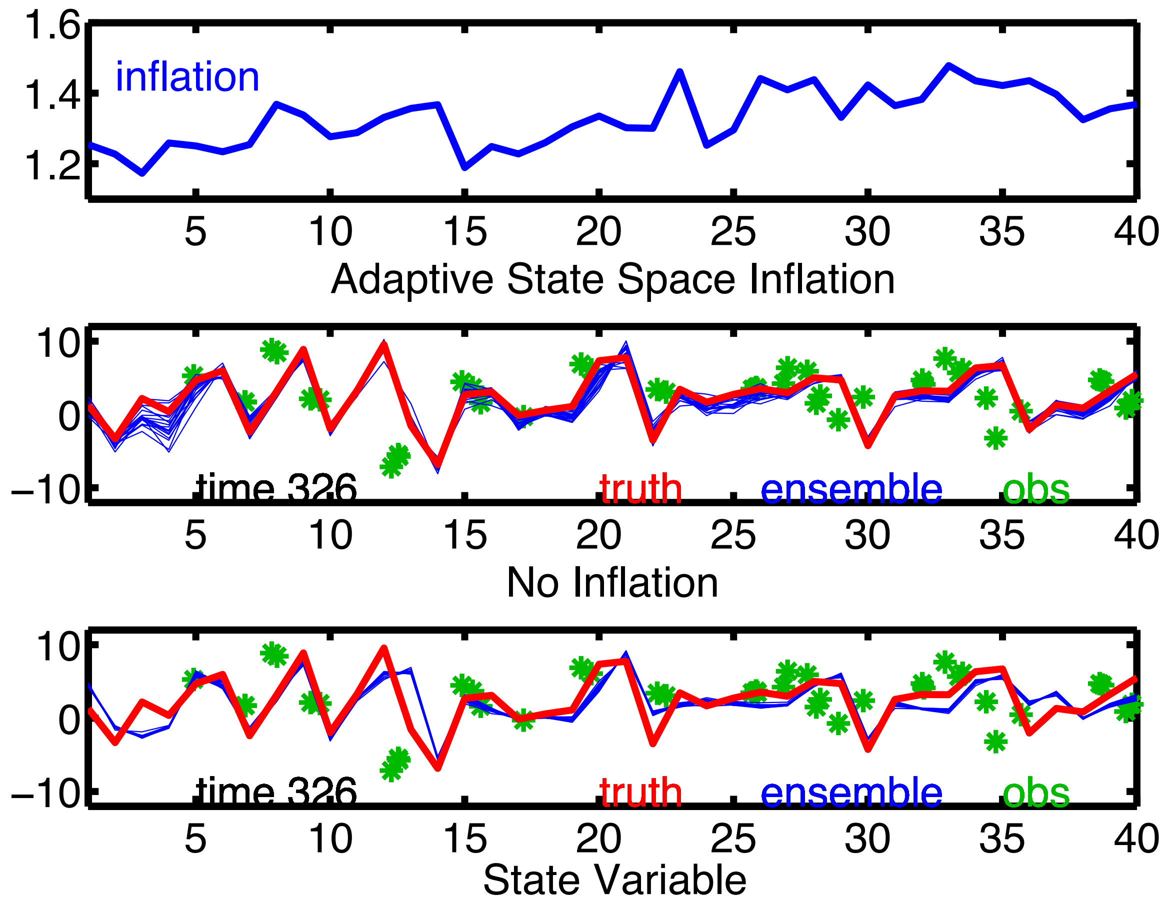
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



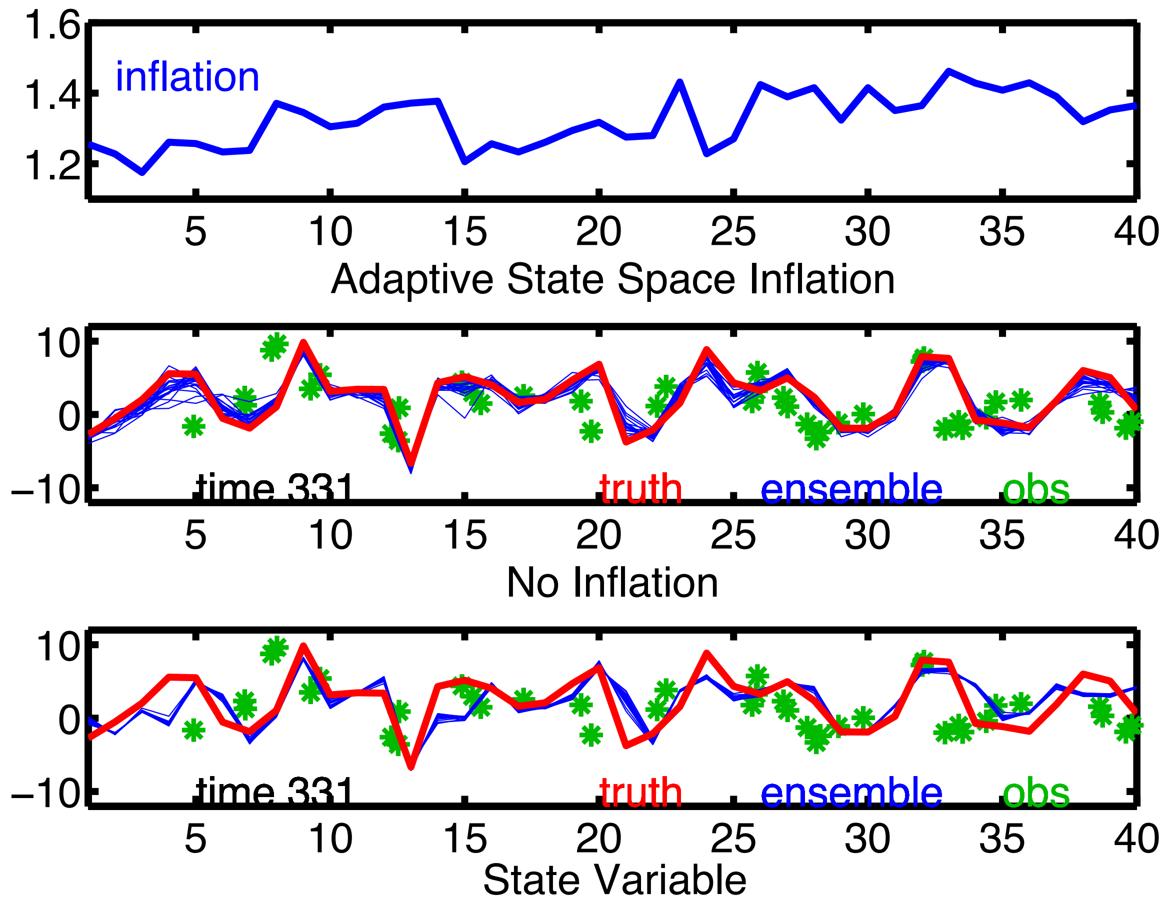
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



Assimilating with Inflation in presence of model error.

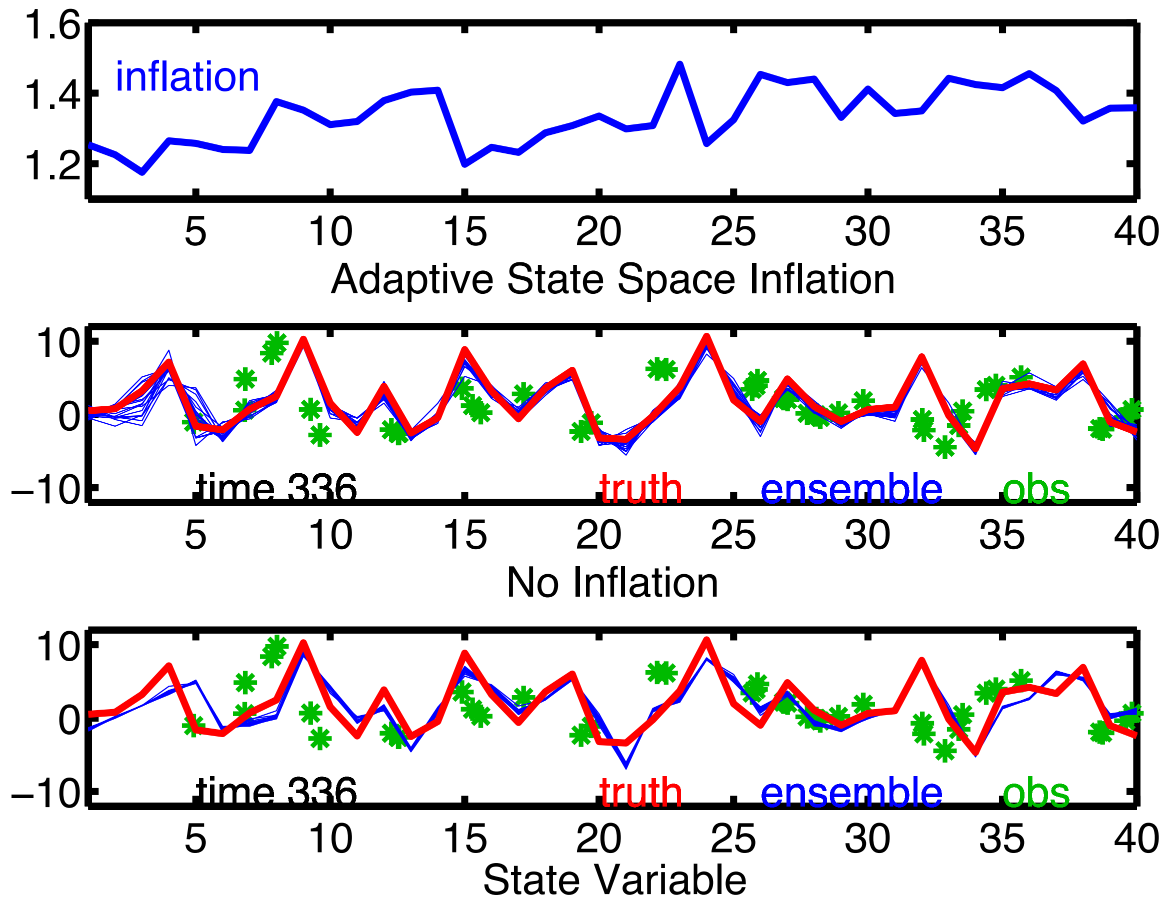
Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.





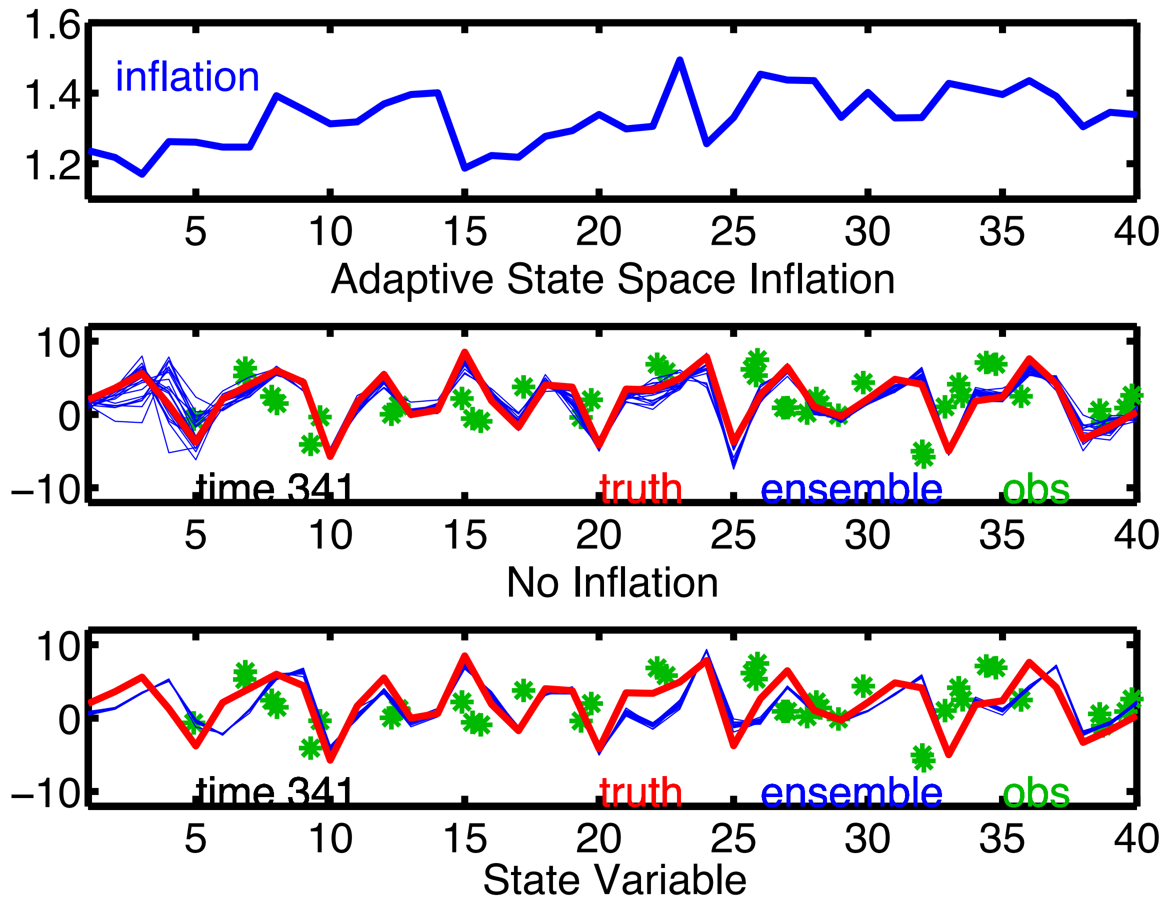
Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.



Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.
Automatically selected by adaptive inflation algorithm.

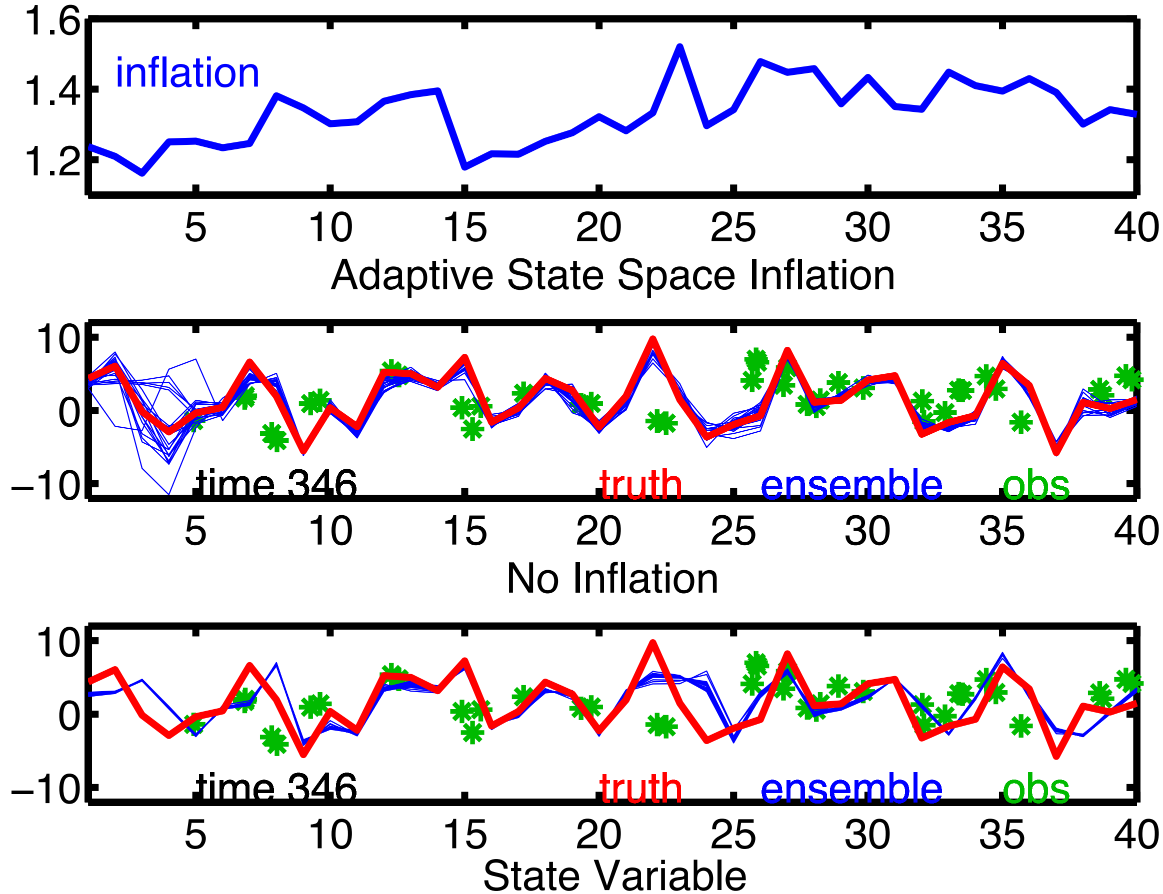


Assimilating with Inflation in presence of model error.

Inflation is a function of state variable and time.

Automatically selected by adaptive inflation algorithm.

It can work, even in presence of severe model error.

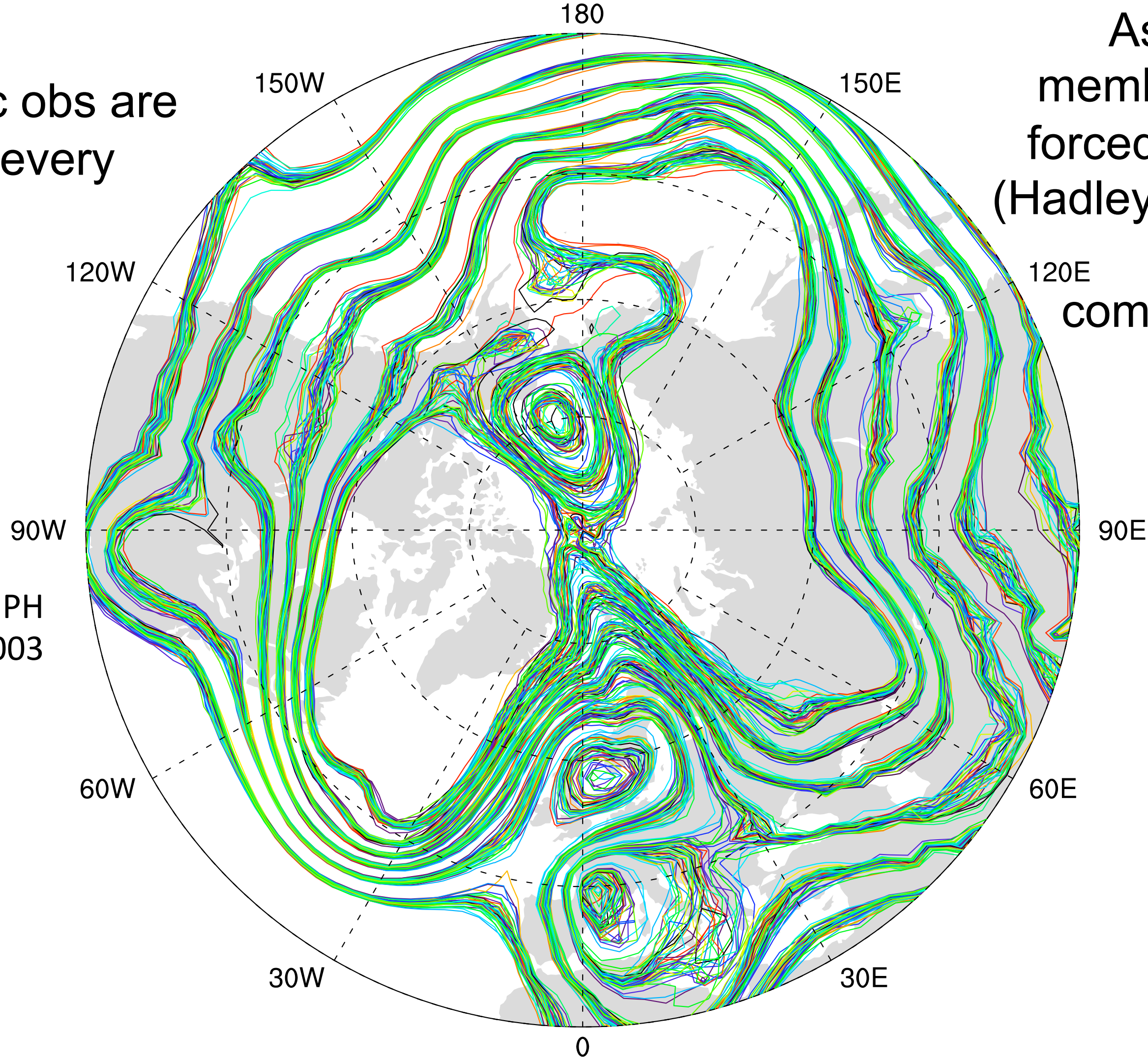


Atmospheric Ensemble Reanalysis, 1998-2010

O(1 million) atmospheric obs are assimilated every day.

Assimilation uses 80 members of 2° FV CAM forced by a single ocean (Hadley+ NCEP-OI2) and produces a very competitive reanalysis.

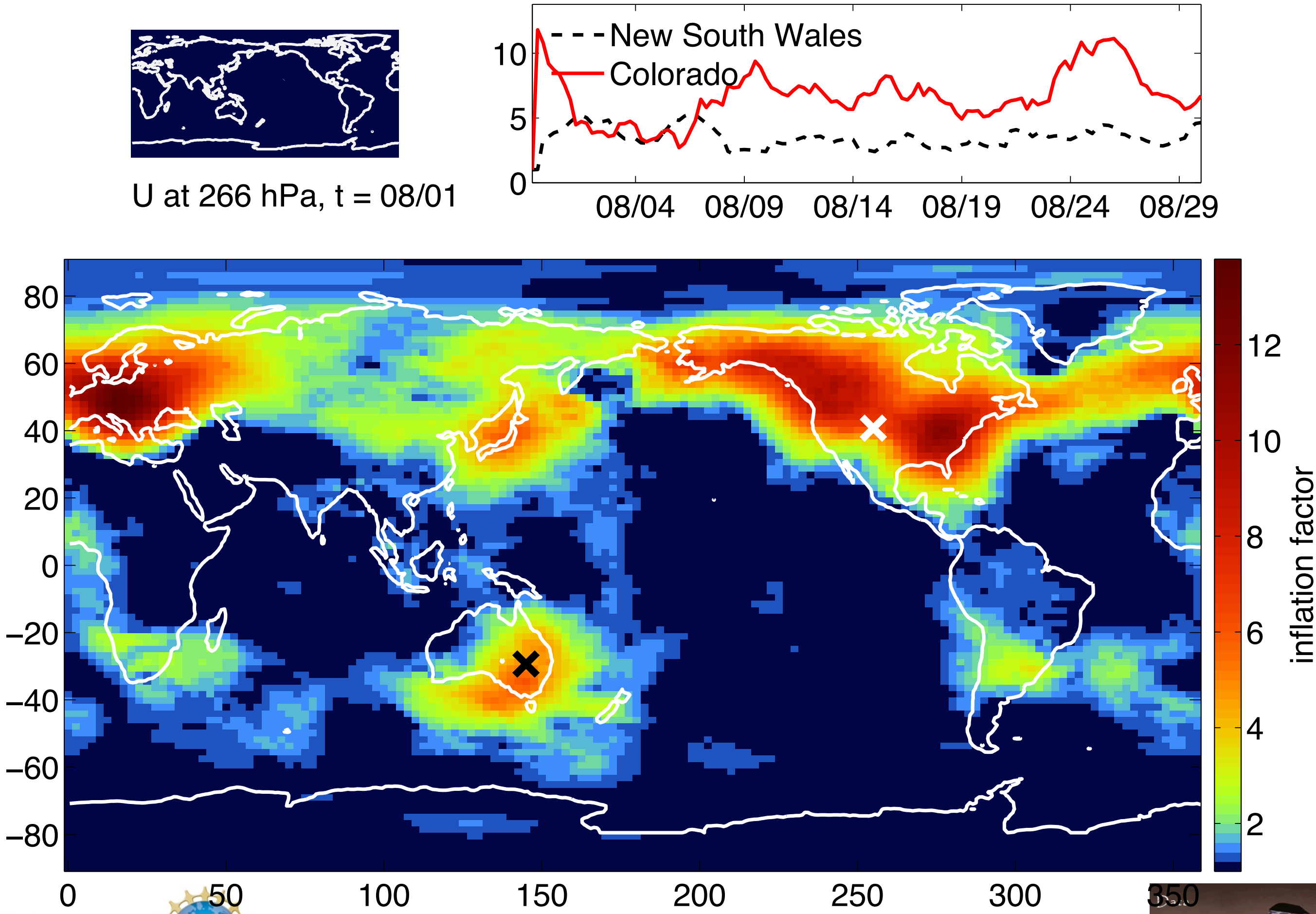
500 hPa GPH
Feb 17 2003



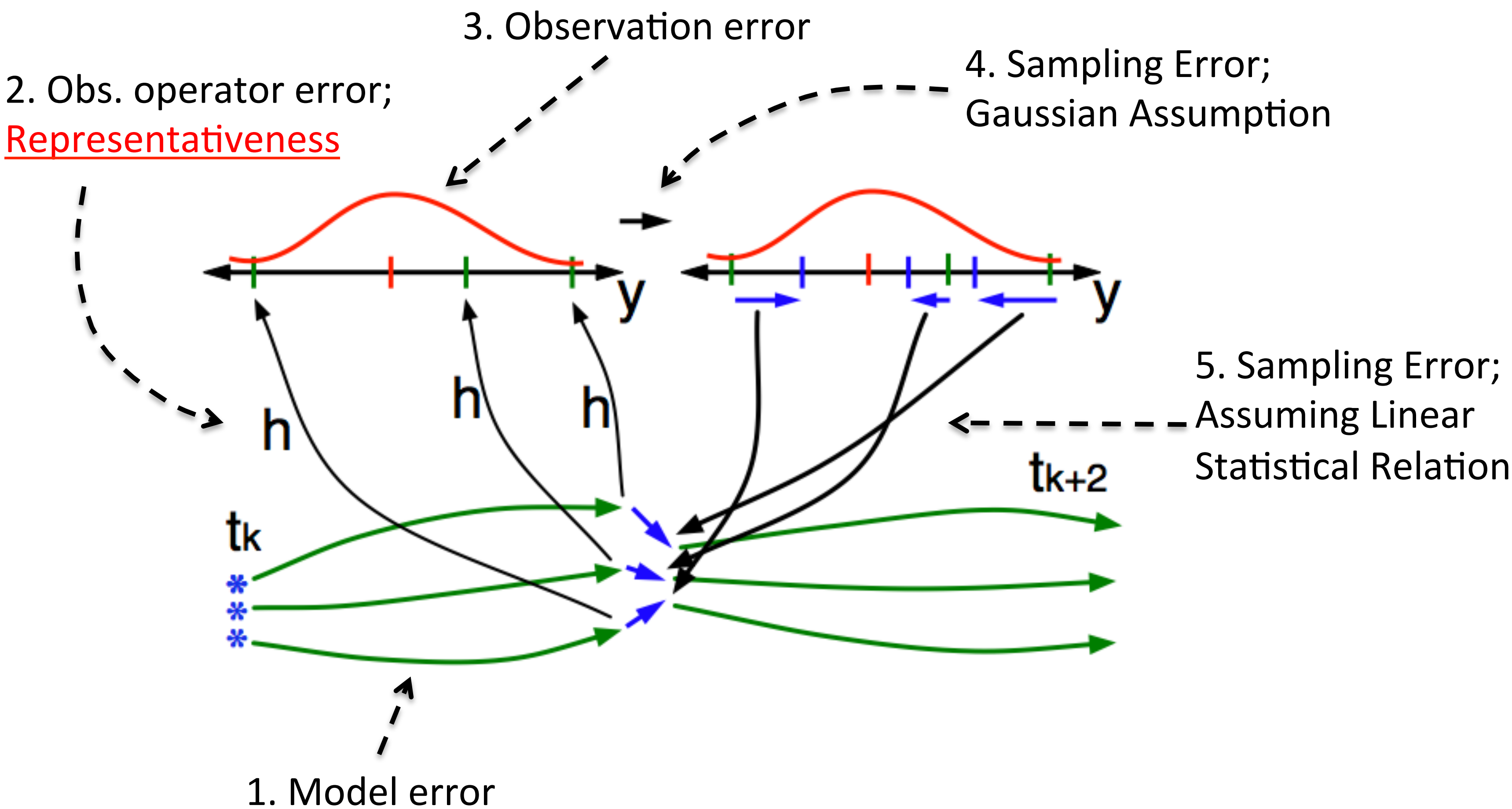
CONTOUR FROM 5200 TO 5700 BY 100

266 hPa U wind inflation

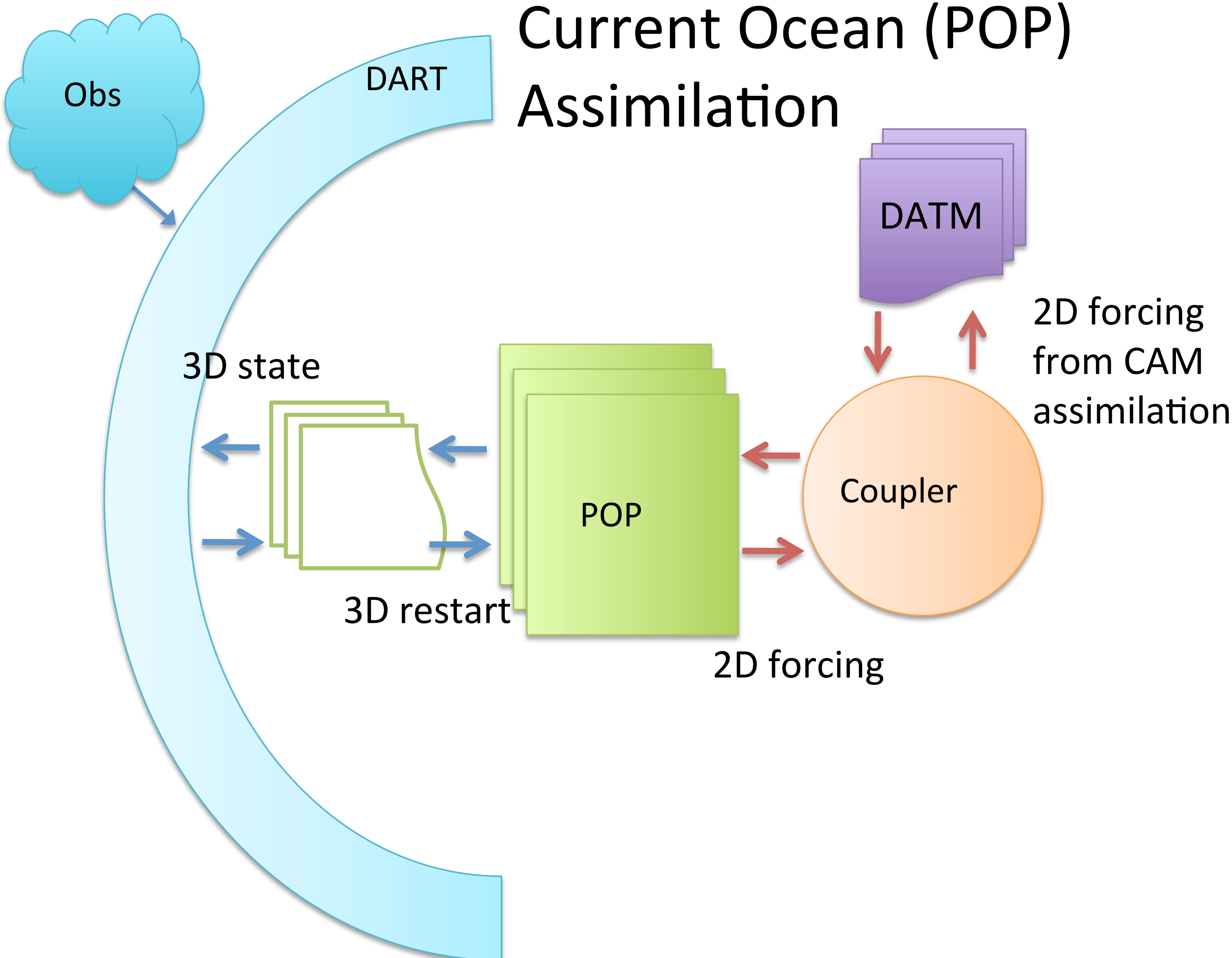
Inflation is large where model bias is detected. Mostly where there are dense observations.



Some Error Sources in Ensemble Filters



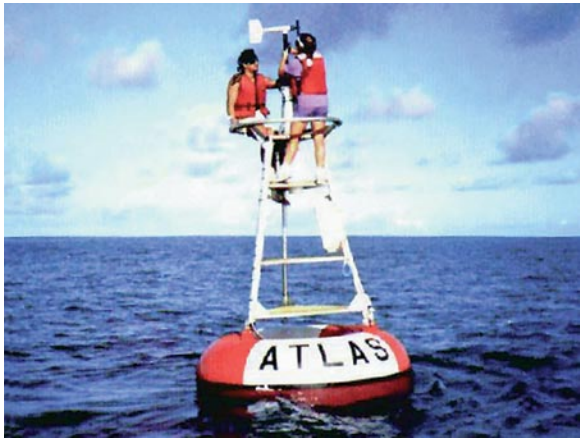
Current Ocean (POP) Assimilation



World Ocean Database T,S observation counts

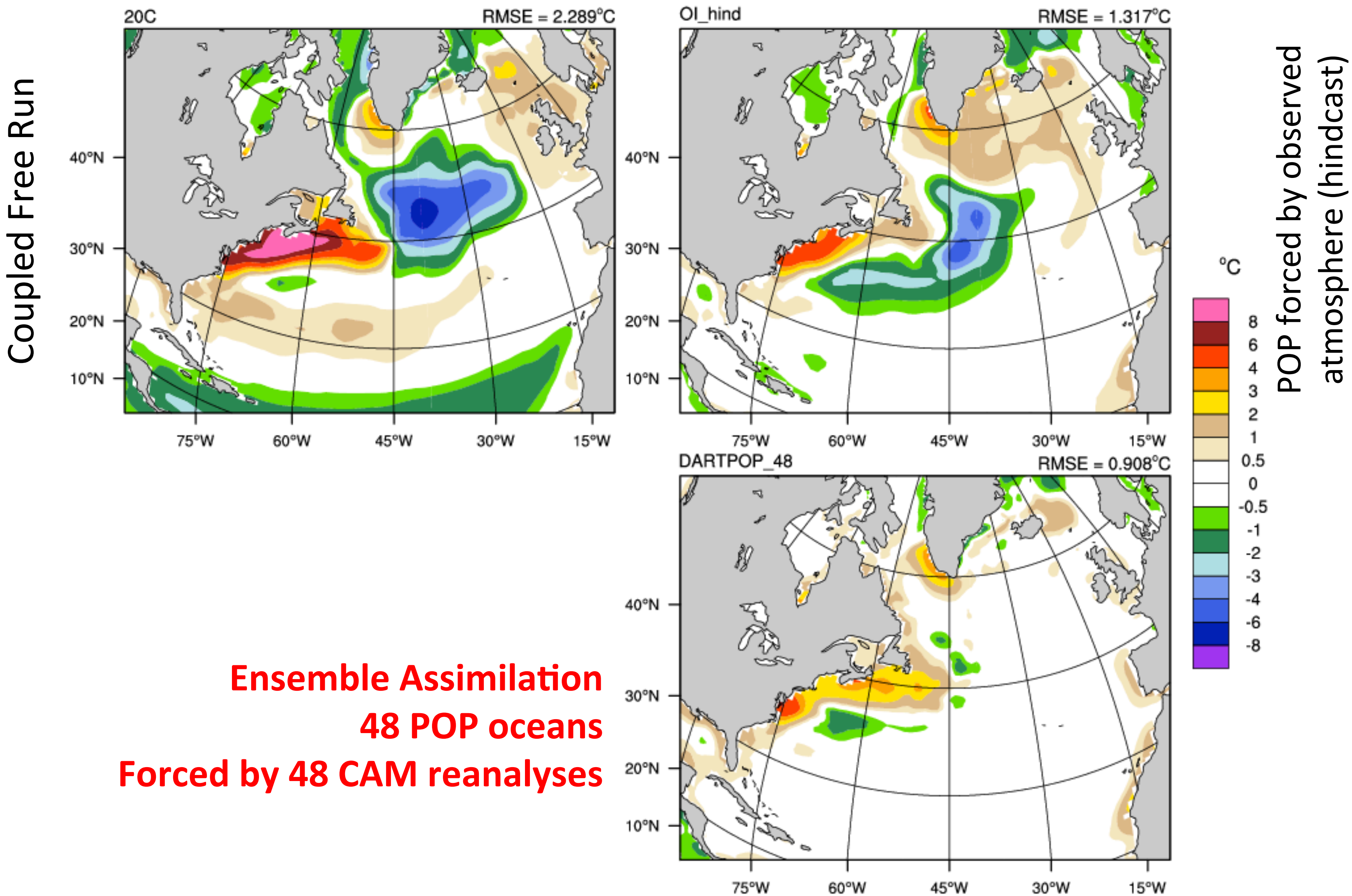
These counts are for 1998 & 1999 and are representative.

FLOAT_SALINITY	68200
FLOAT_TEMPERATURE	395032
DRIFTER_TEMPERATURE	33963
MOORING_SALINITY	27476
MOORING_TEMPERATURE	623967
BOTTLE_SALINITY	79855
BOTTLE_TEMPERATURE	81488
CTD_SALINITY	328812
CTD_TEMPERATURE	368715
STD_SALINITY	674
STD_TEMPERATURE	677
XCTD_SALINITY	3328
XCTD_TEMPERATURE	5790
MBT_TEMPERATURE	58206
XBT_TEMPERATURE	1093330
APB_TEMPERATURE	580111



- temperature observation error standard deviation == 0.5 K.
- salinity observation error standard deviation == 0.5 msu.

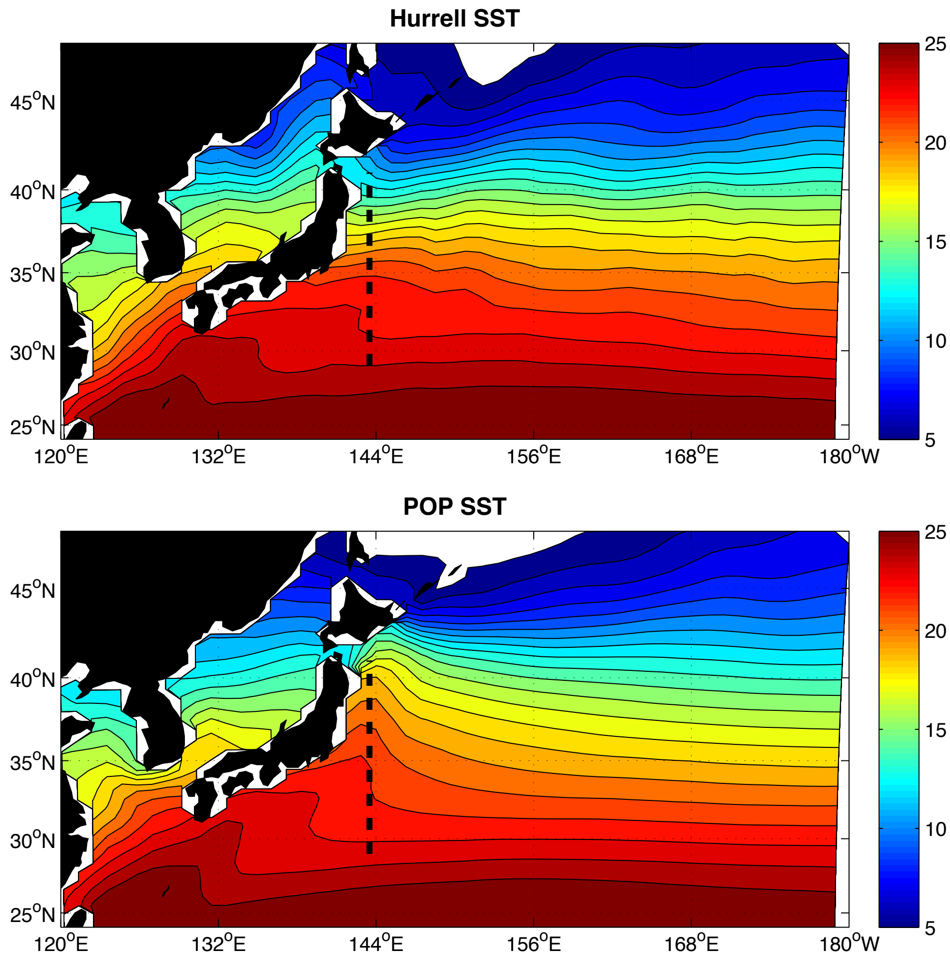
Physical Space: 1998/1999 SST Anomaly from HadOI-SST



Ensemble Assimilation
48 POP oceans
Forced by 48 CAM reanalyses

Challenges where ocean model is unable, or unwilling, to simulate reality.

Example: cross section along Kuroshio; model separates too far north.

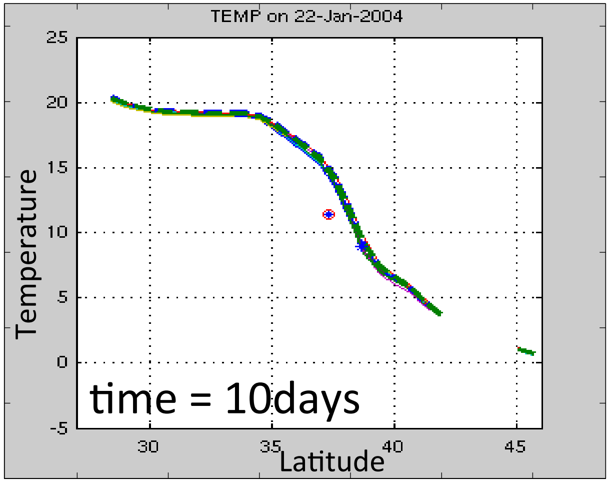


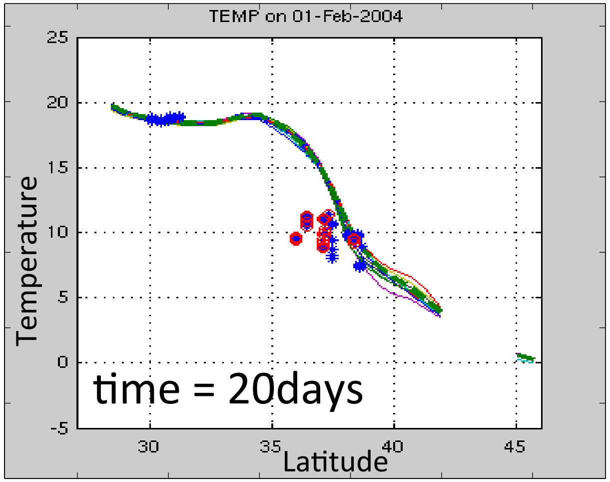
Regarded to be accurate.

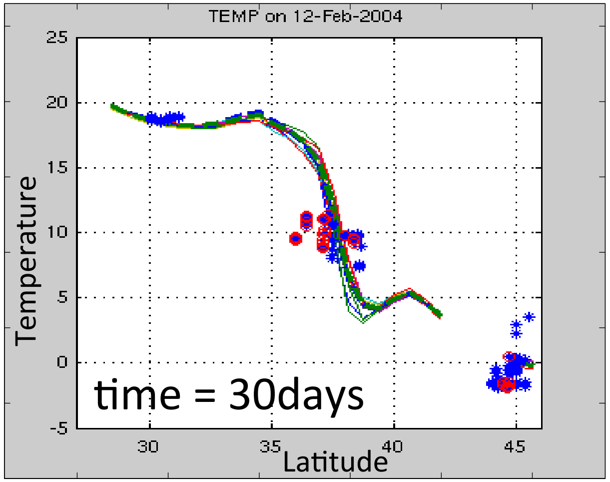
Free run of POP, the warm water is too far North.

Challenges in correcting position of Kuroshio.

60-day assimilation starting from model climatology on 1 January 2004.







Initially warm water goes too far north.

Many observations are rejected (red), but others (blue) move temperature gradient south.

Adaptive inflation increases ensemble spread as assimilation struggles to push model towards obs.

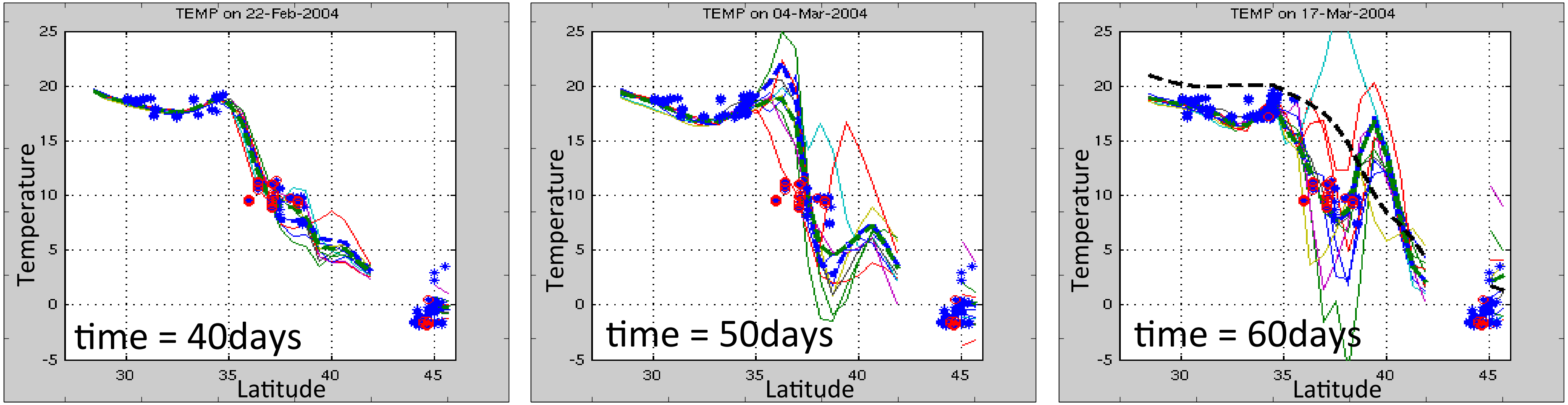
Challenges in correcting position of Kuroshio.

60-day assimilation starting from model climatology on 1 January 2004.

Green dashed line is posterior at previous time,
Blue dashed line is prior at current time,
Ensembles are thin lines.

Observations keep pulling the warm water to the south; Model forecasts continue to quickly move warm water further north. Inflation continues to increase spread.

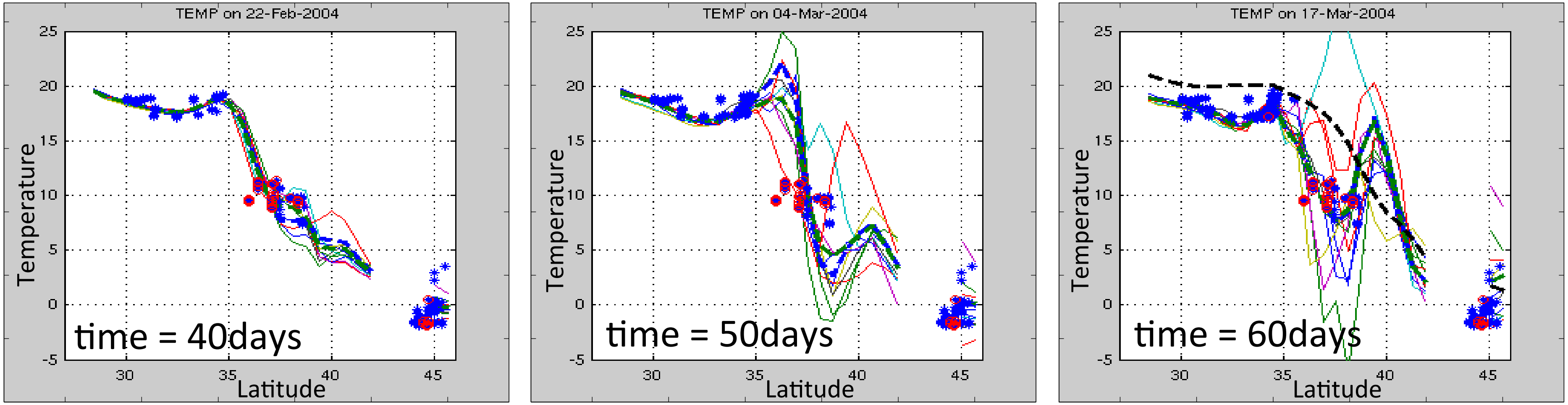
Model forecasts finally fail due to numerical issues. Black dashes show original model state from 10 January.



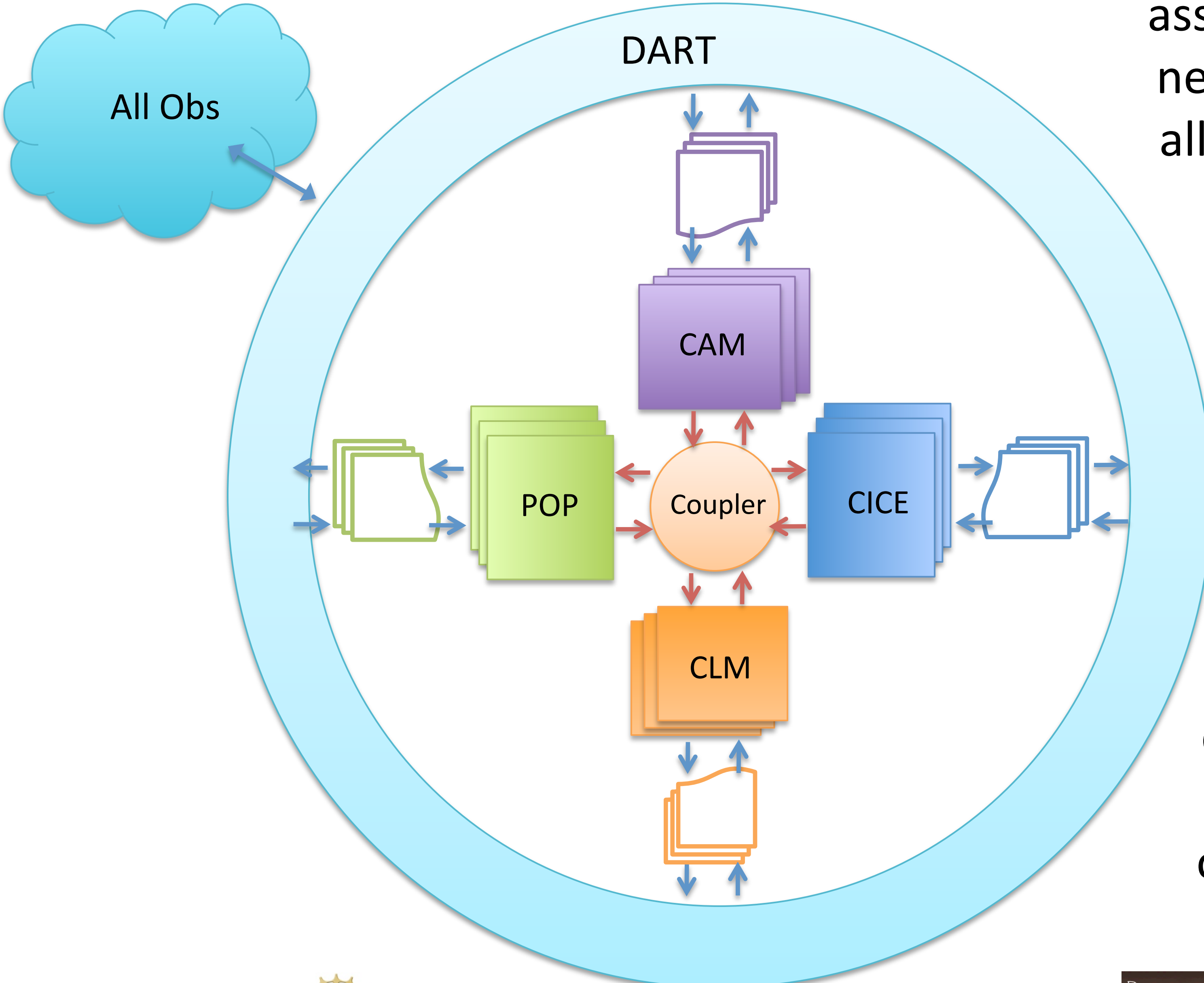
Challenges in correcting position of Kuroshio.

60-day assimilation starting from model climatology on 1 January 2004.

- Assimilation cannot force model to fit observations.
- Model cannot 'represent' the observations.
- Use of adaptive inflation leads to eventual model failure.
- Reduced adaptive inflation can lead to compromise between observations and model.
- Increasing the error associated with forward operator can ameliorate, but what do the answers mean?



Fully coupled assimilation will need data from all components at the same time



Each component corrected by all kinds of observations

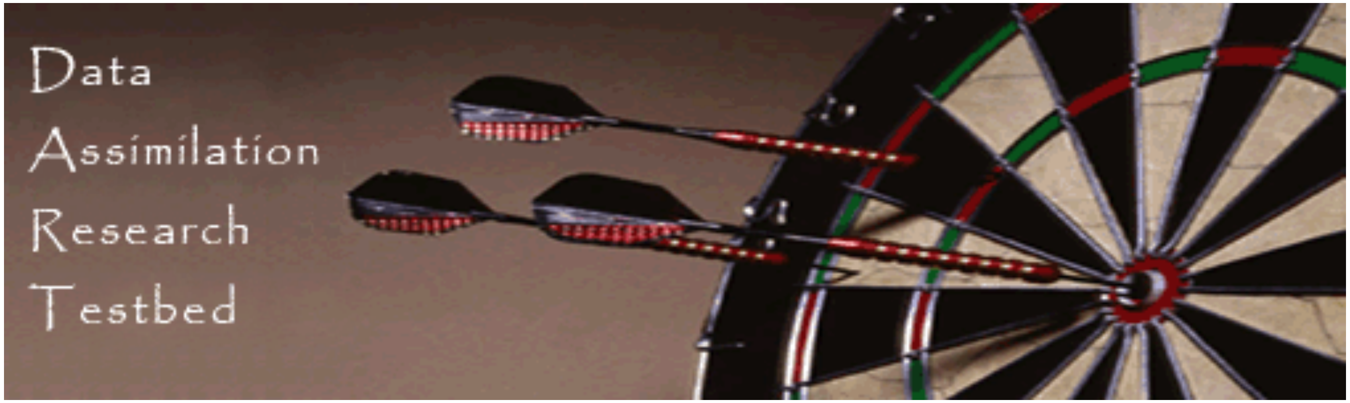
Ensemble Data Assimilation for Large Geophysical Models

- Very certain that model predictions are different from observations.
- Very certain that small correlations have large errors.
- Moderately confident that large correlations are ‘realistic’.
- Very uncertain about state estimates in sparse/unobserved regions.

- Must Calibrate and Validate results (adaptive inflation/localization).
- There may not be enough observations to do this many places.

- First order of business: Improving models. DA can help with this.
- Stochastic models seem like a logical way forward.

Code to implement all of the algorithms discussed are freely available from:



http://www.image.ucar.edu/DAReS/DART/