Ensemble Data Assimilation and Uncertainty Quantification

Jeffrey Anderson, Alicia Karspeck, Tim Hoar,

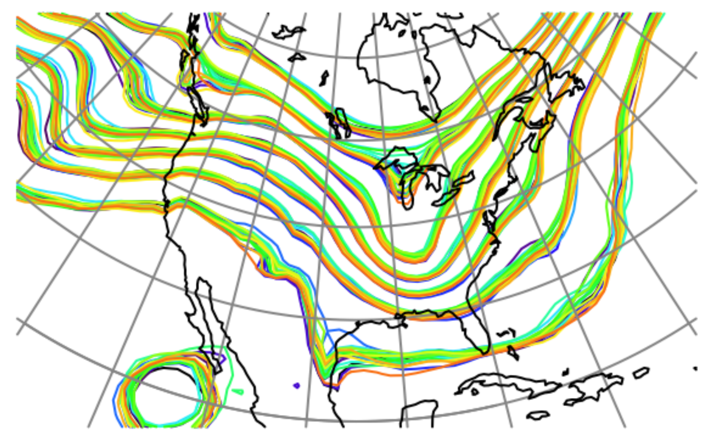
Nancy Collins, Kevin Raeder, Steve Yeager

National Center for Atmospheric Research

What is Data Assimilation?

Observations combined with a Model forecast…

+



…to produce an analysis
(best possible estimate).

What is Ensemble Data Assimilation?

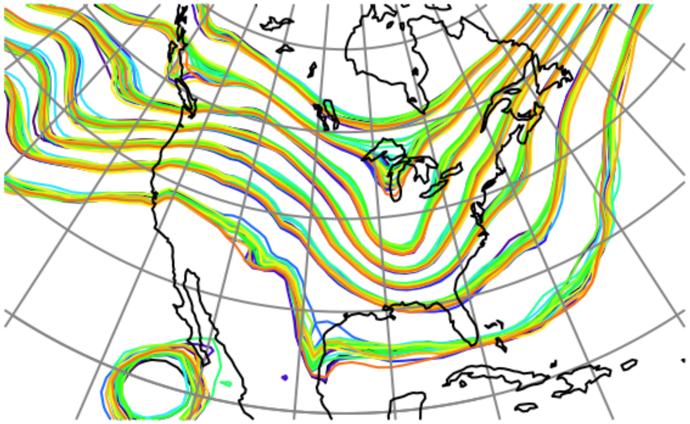
Use an ensemble (set) of model forecasts.

Use sample statistics to get covariance between state and observations.

Often assume that ensemble members are random draw.

Ensemble methods I use are optimal solution when:

1. Model is linear,
2. Observation errors are unbiased gaussian,
3. Relation between model and obs is linear,
4. Ensemble is large enough.

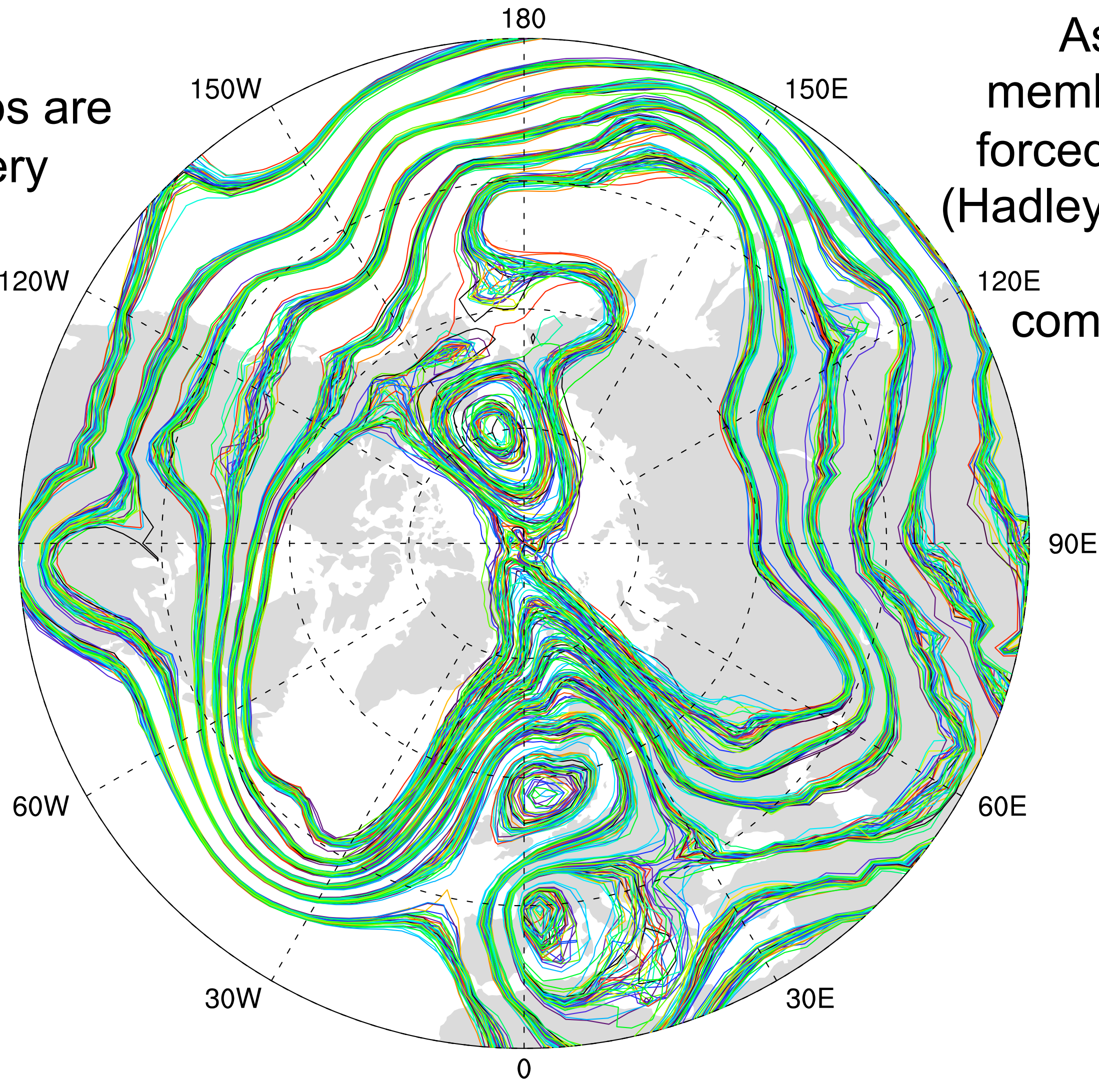


Atmospheric Ensemble Reanalysis, 1998-2010

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

Ensemble Filter for Large Geophysical Models

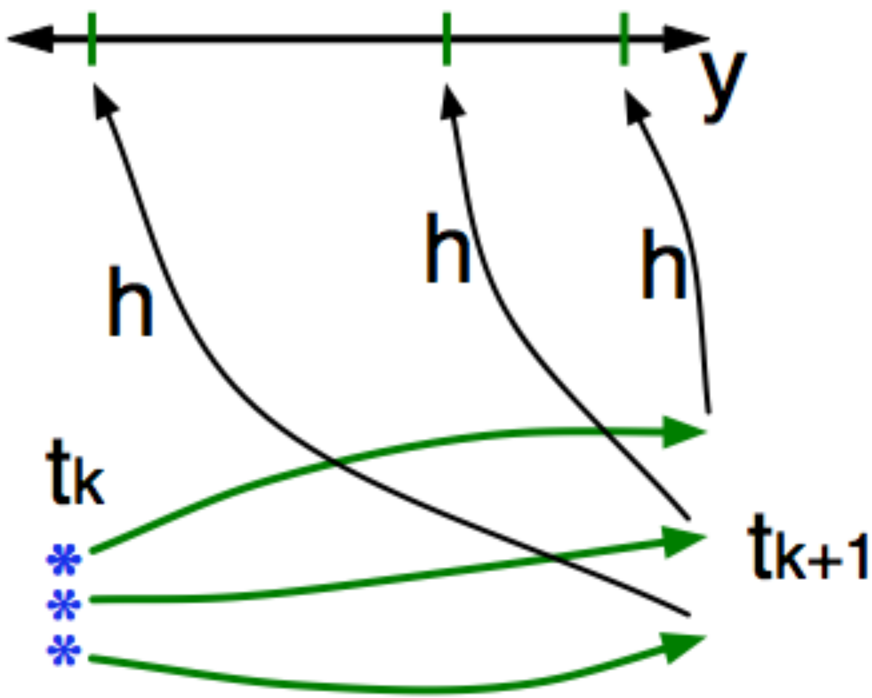
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Ensemble Filter for Large Geophysical Models

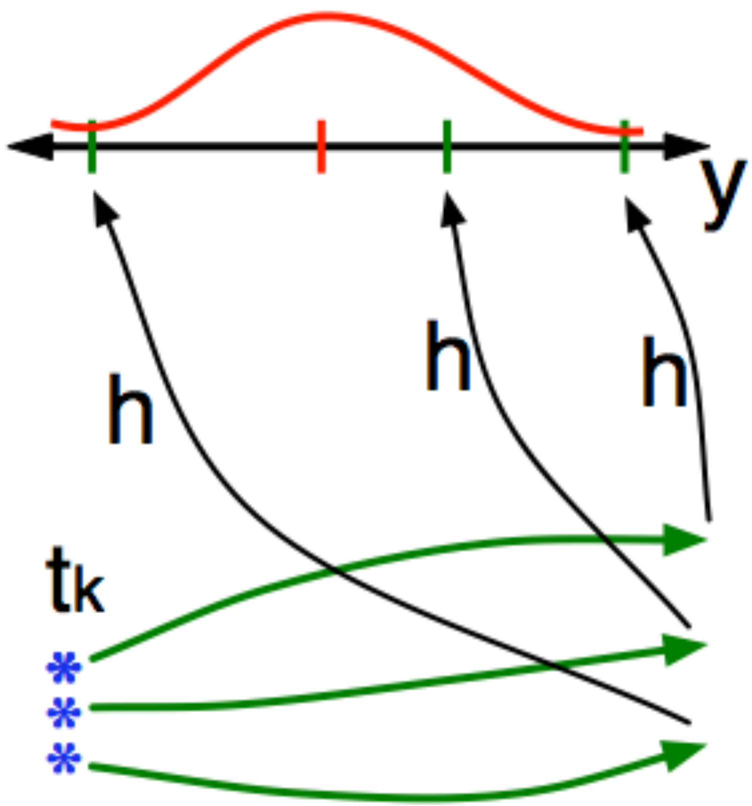
2. Get prior ensemble sample of observation, y = h(x), by applying forward operator **h** to each ensemble member.



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

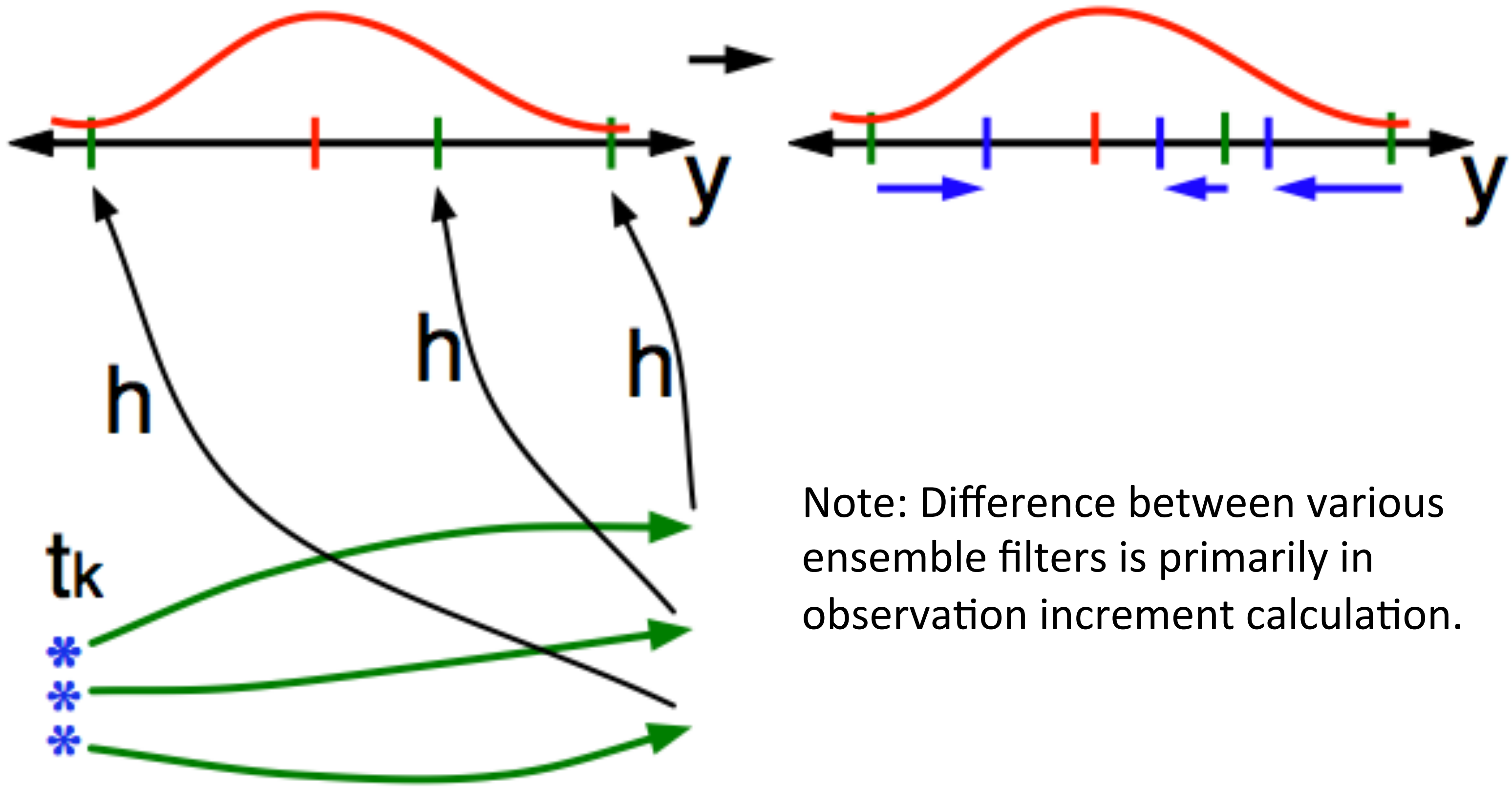
Ensemble Filter for Large Geophysical Models

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



Ensemble Filter for Large Geophysical Models

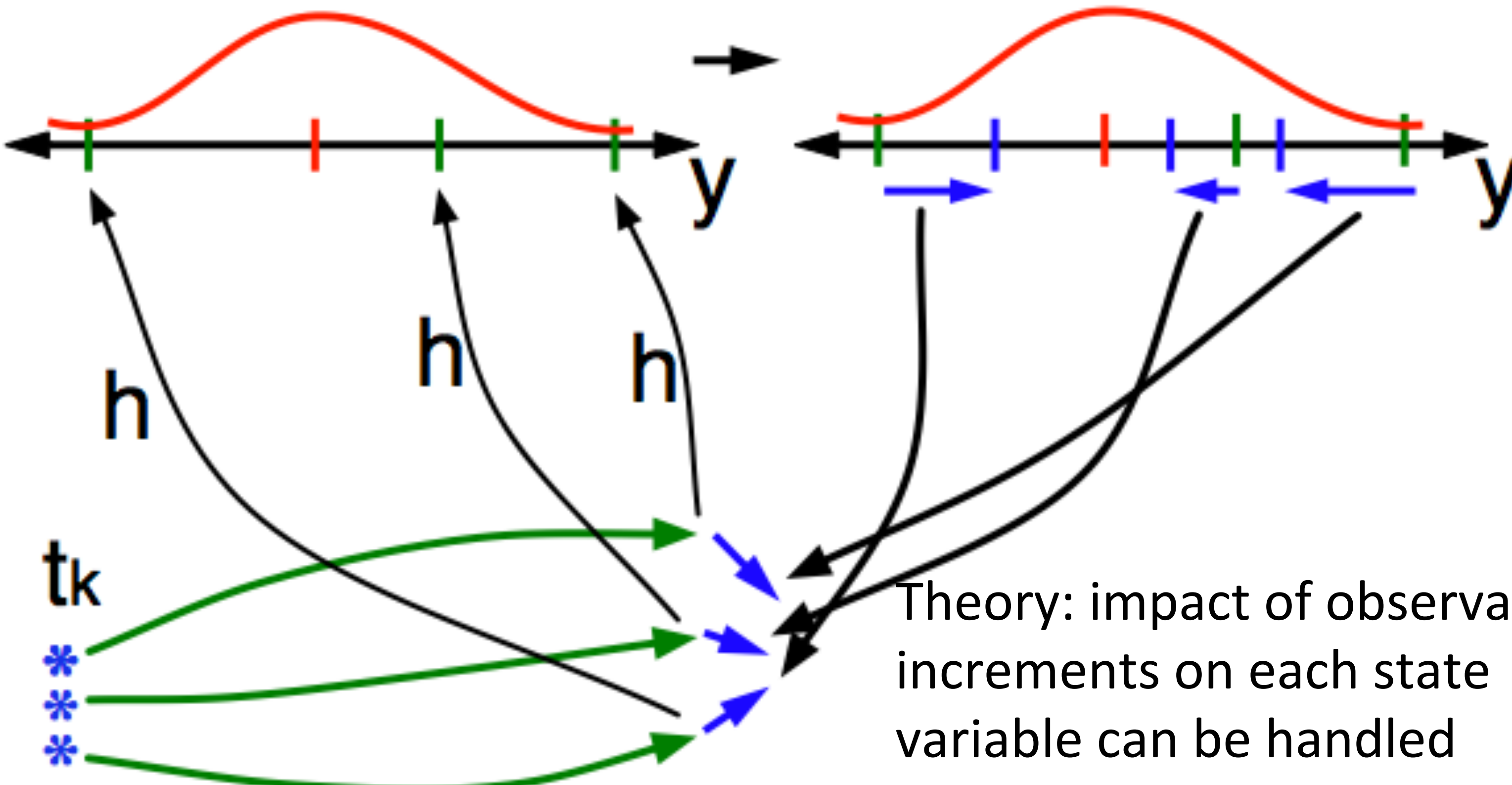
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



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

Ensemble Filter for Large Geophysical Models

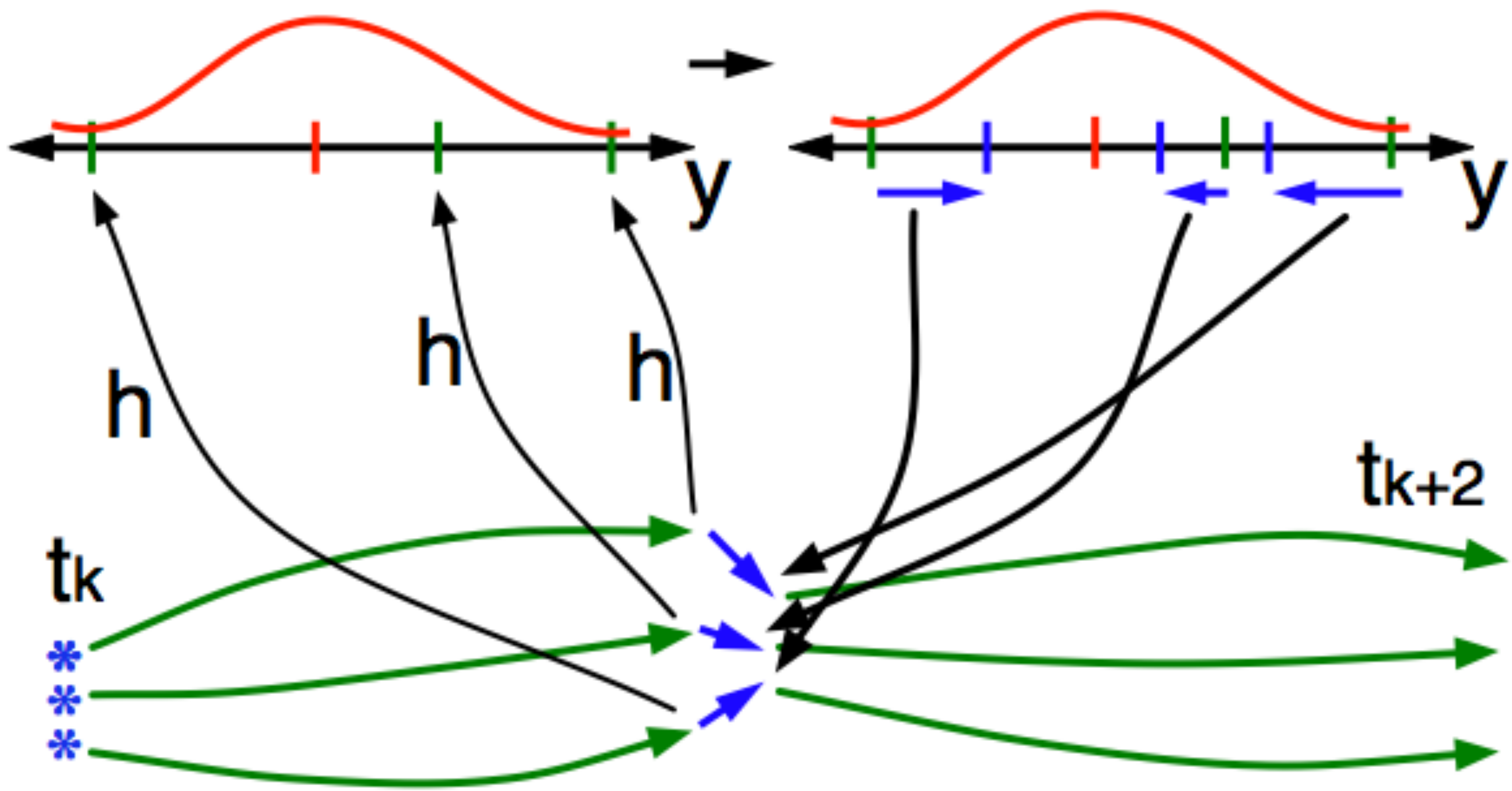
5. Use ensemble samples of **y** and each state variable to linearly regress observation increments onto state variable increments.



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

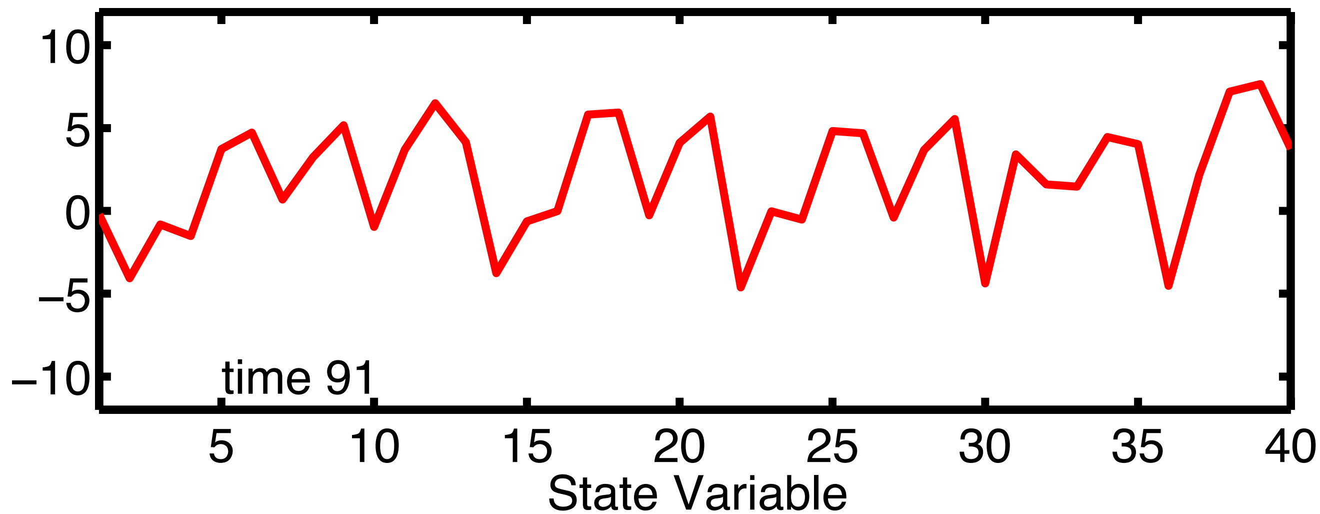
Ensemble Filter for Large Geophysical Models

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



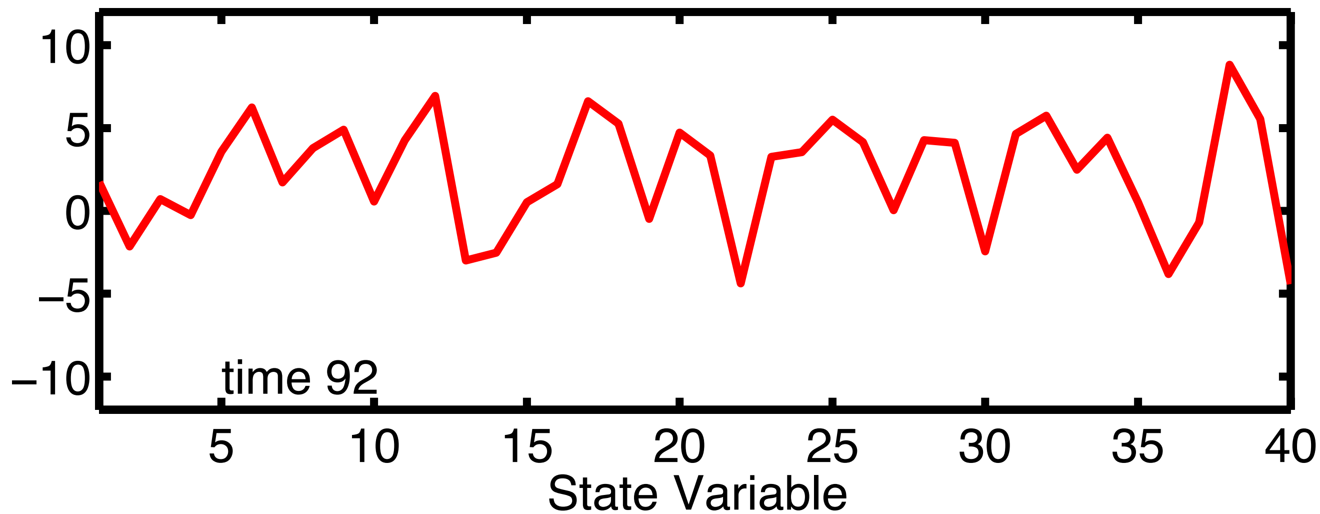
Ensemble Filter for Lorenz-96 40-Variable Model

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.



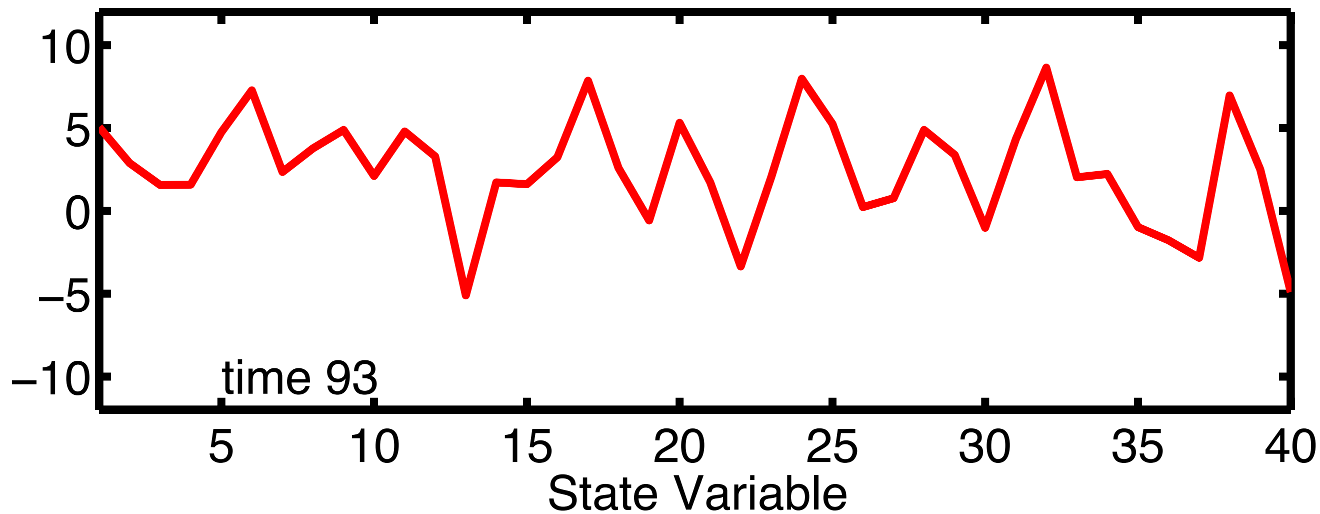
Ensemble Filter for Lorenz-96 40-Variable Model

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.

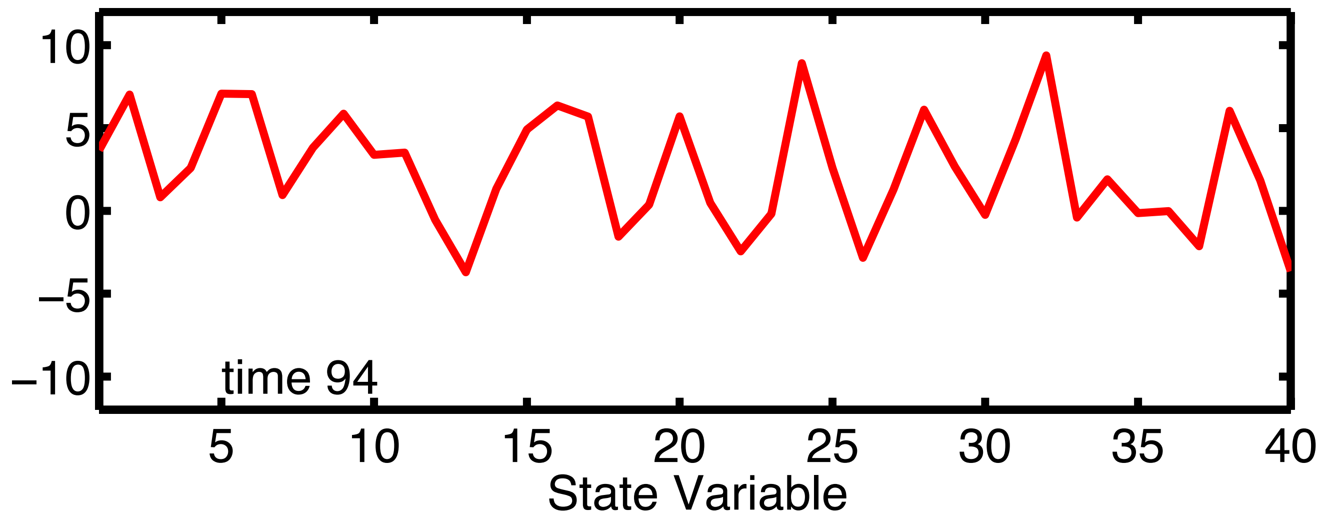


Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.

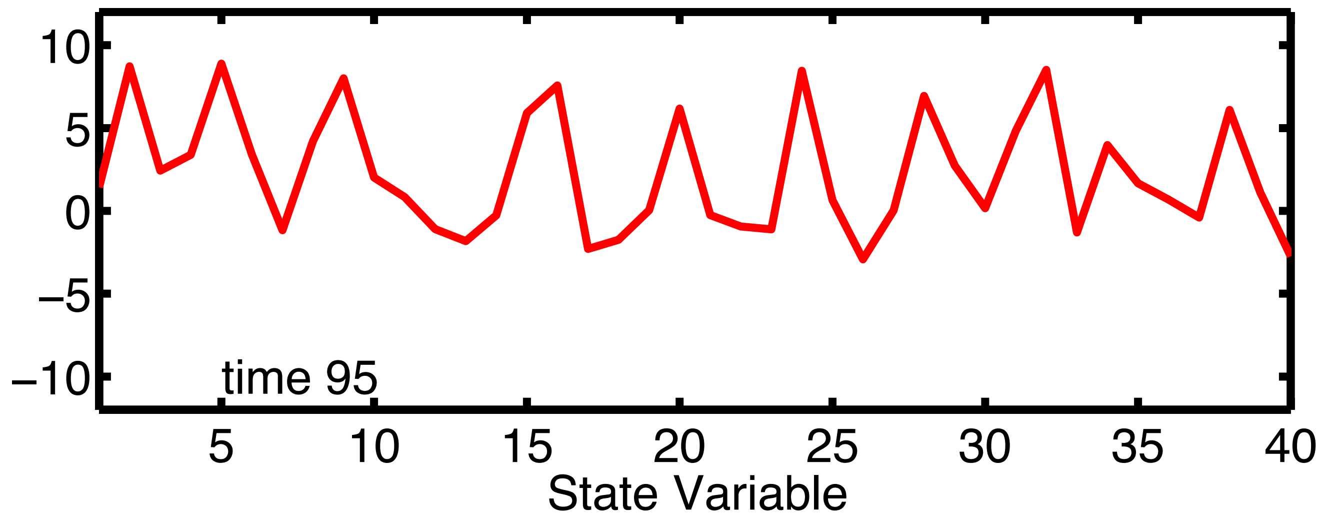
$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$.

Acts 'something' like synoptic weather around a latitude band.



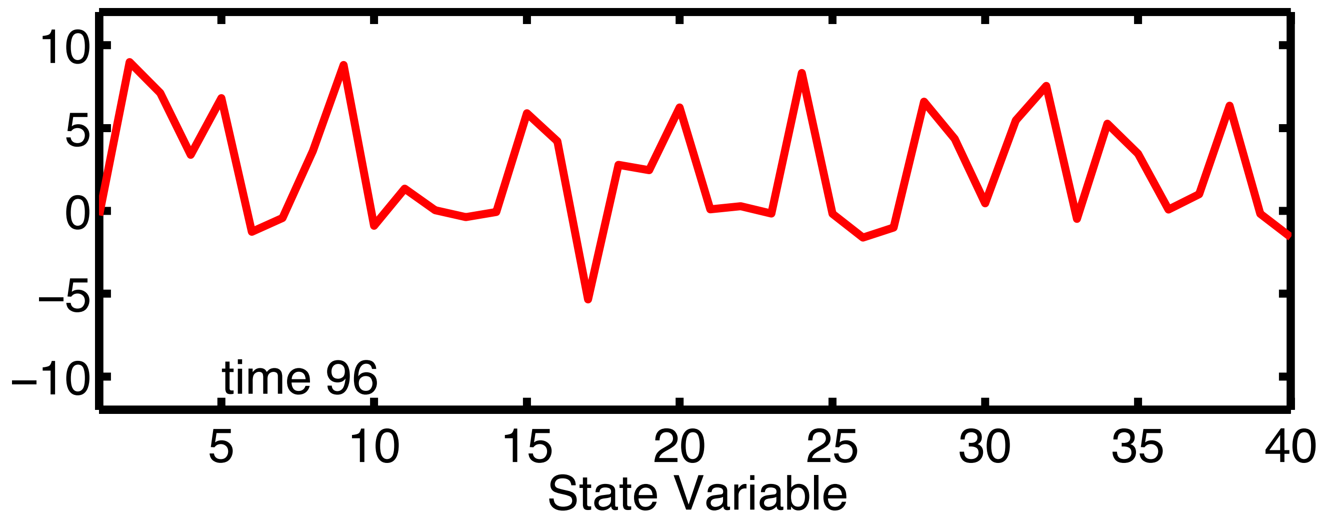
Ensemble Filter for Lorenz-96 40-Variable Model

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.



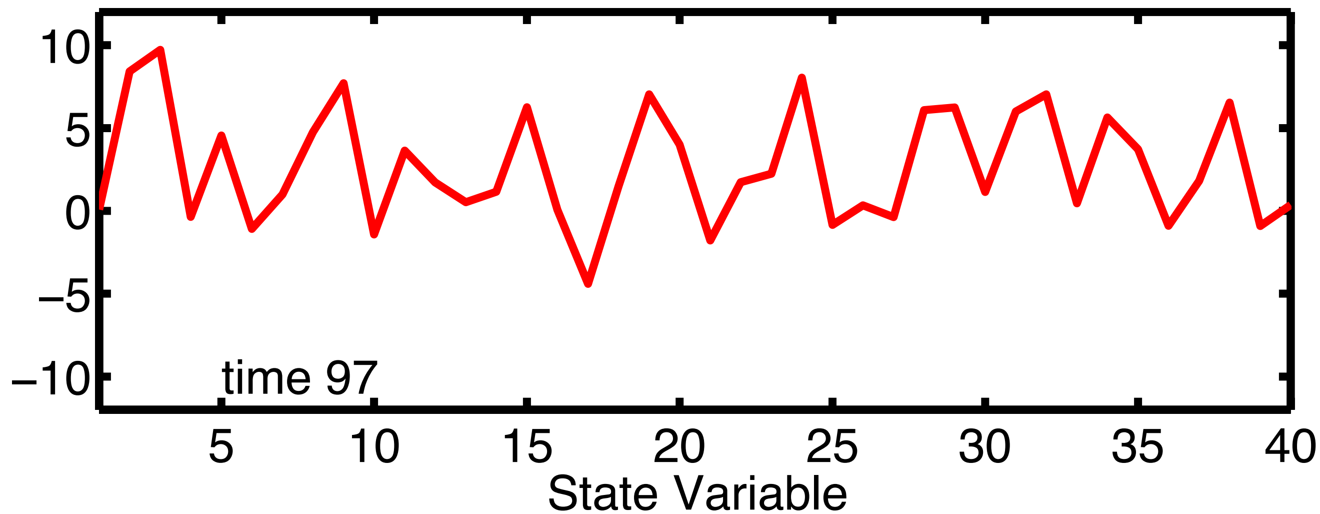
Ensemble Filter for Lorenz-96 40-Variable Model

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.

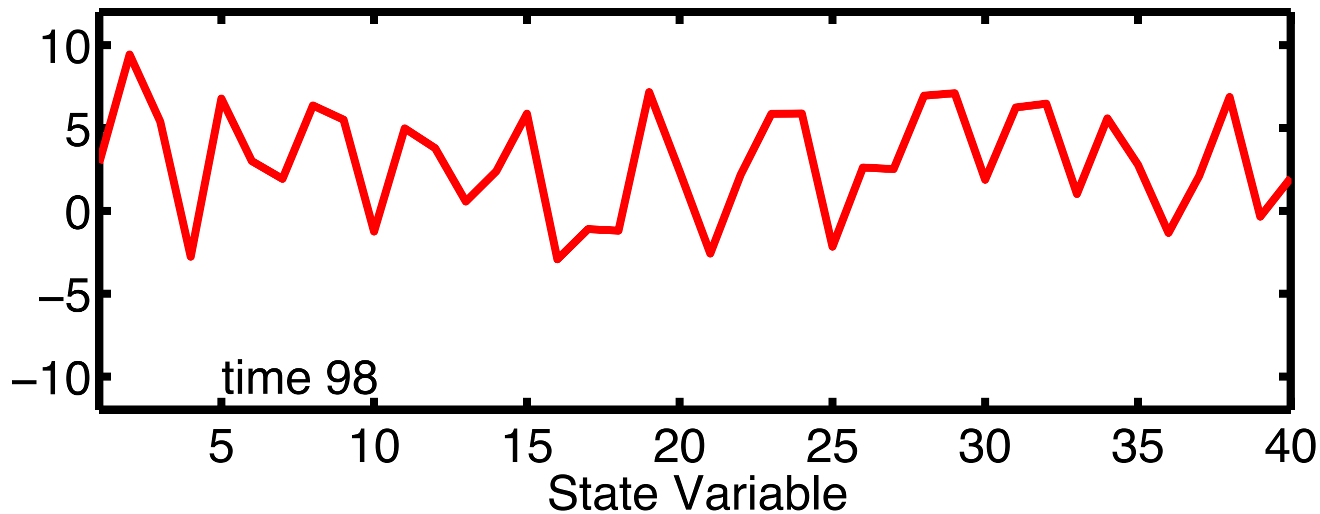


Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: X1, X2,..., X40.

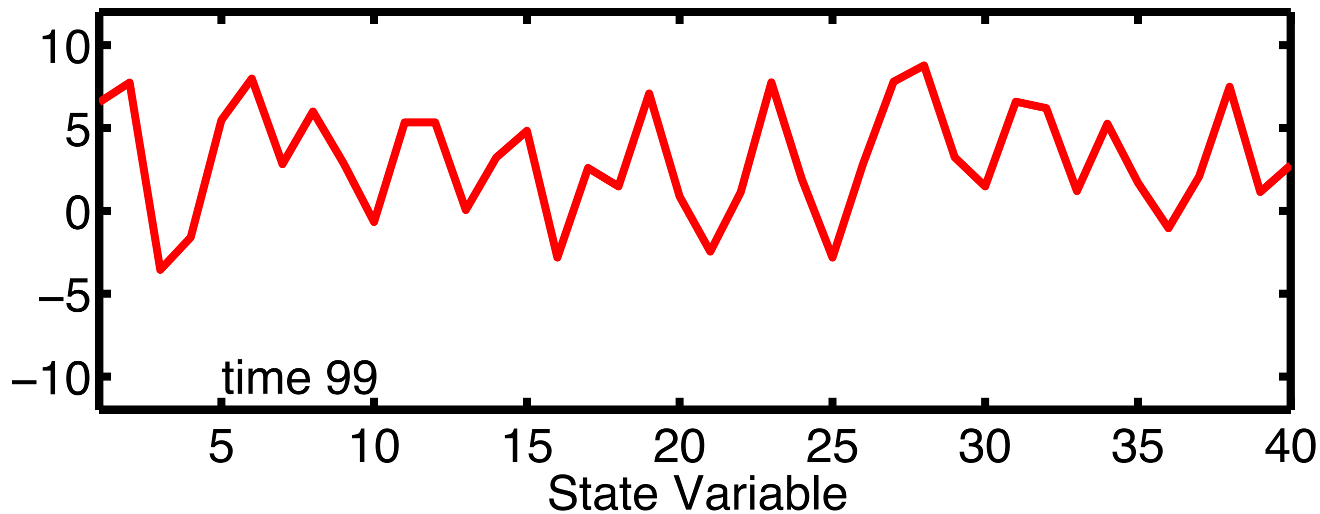
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.

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.

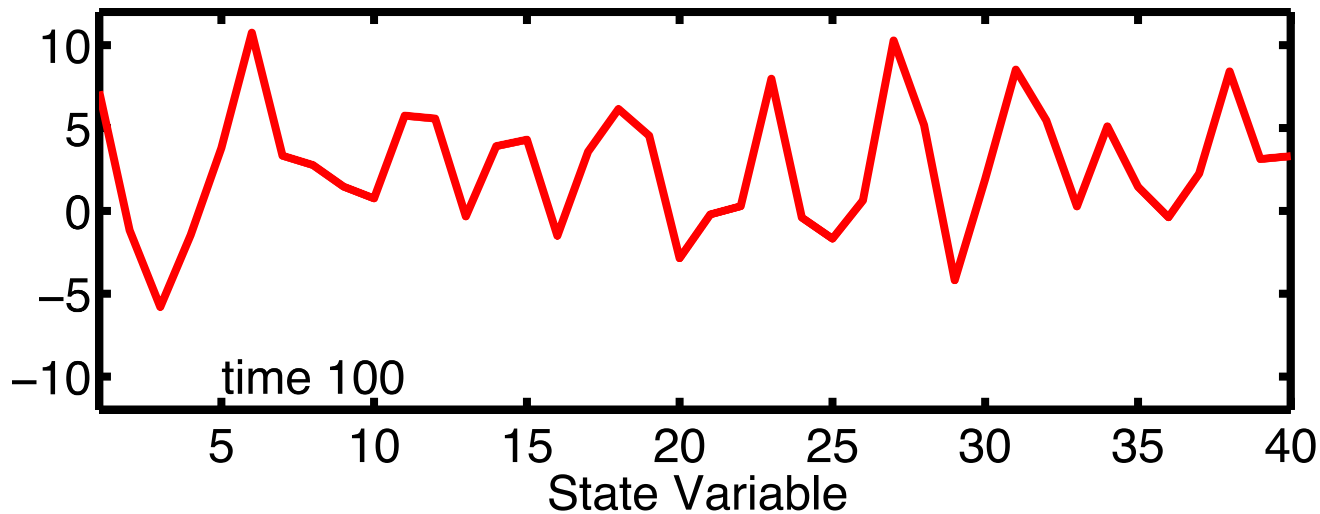


Ensemble Filter for Lorenz-96 40-Variable Model

40 state variables: $X_1, X_2, \ldots, X_{40}$.

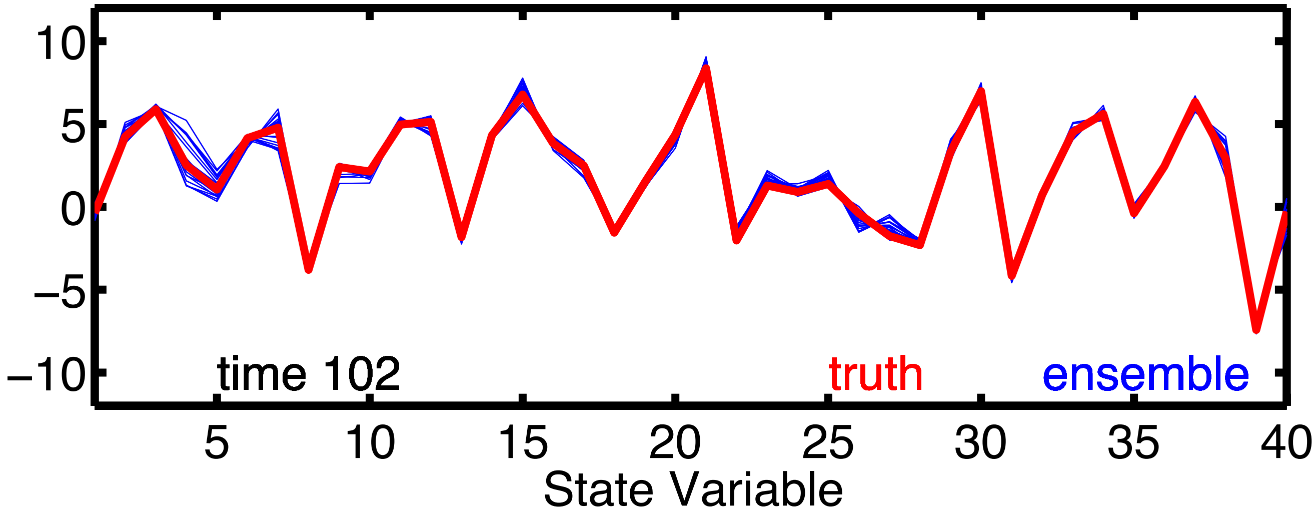
$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$.

Acts ‘something’ like synoptic weather around a latitude band.



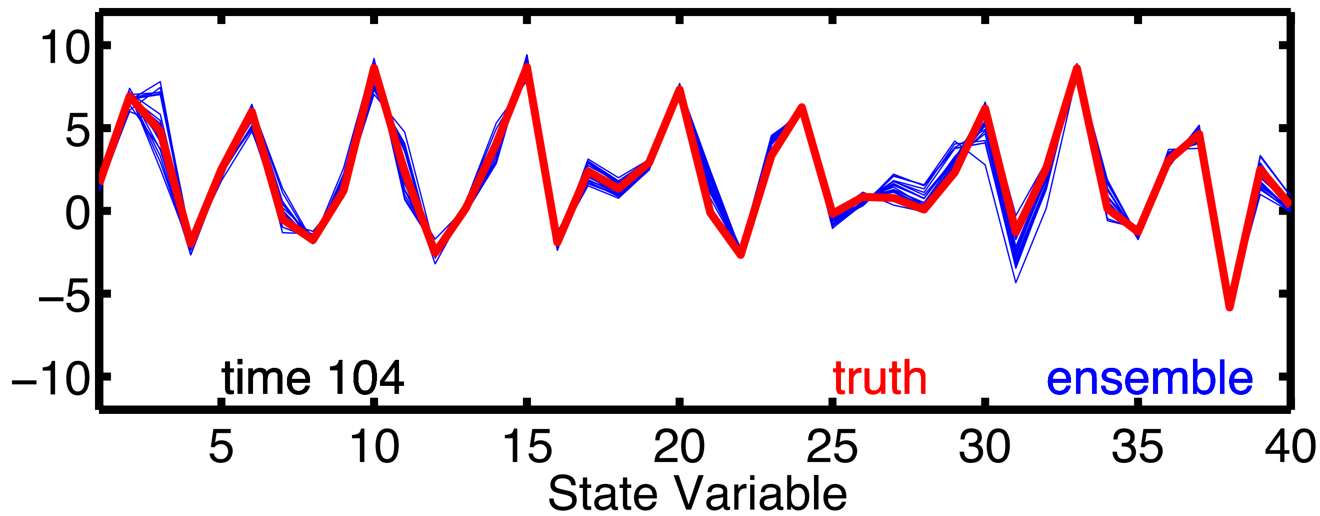
Lorenz-96 is sensitive to small perturbations

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'..



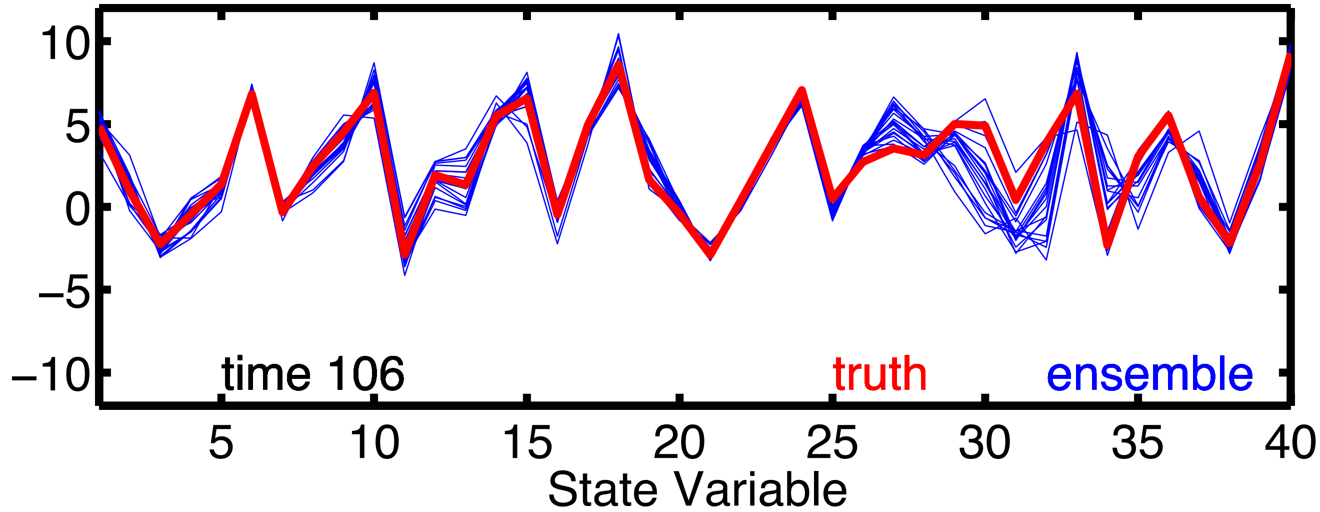
Lorenz-96 is sensitive to small perturbations

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'..



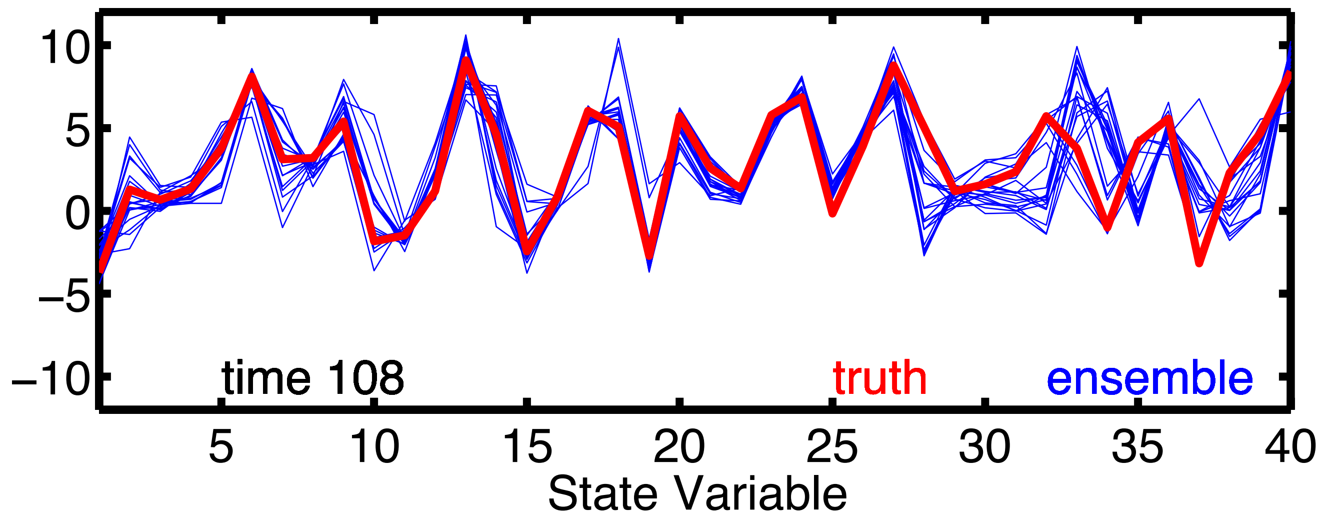
Lorenz-96 is sensitive to small perturbations

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’..



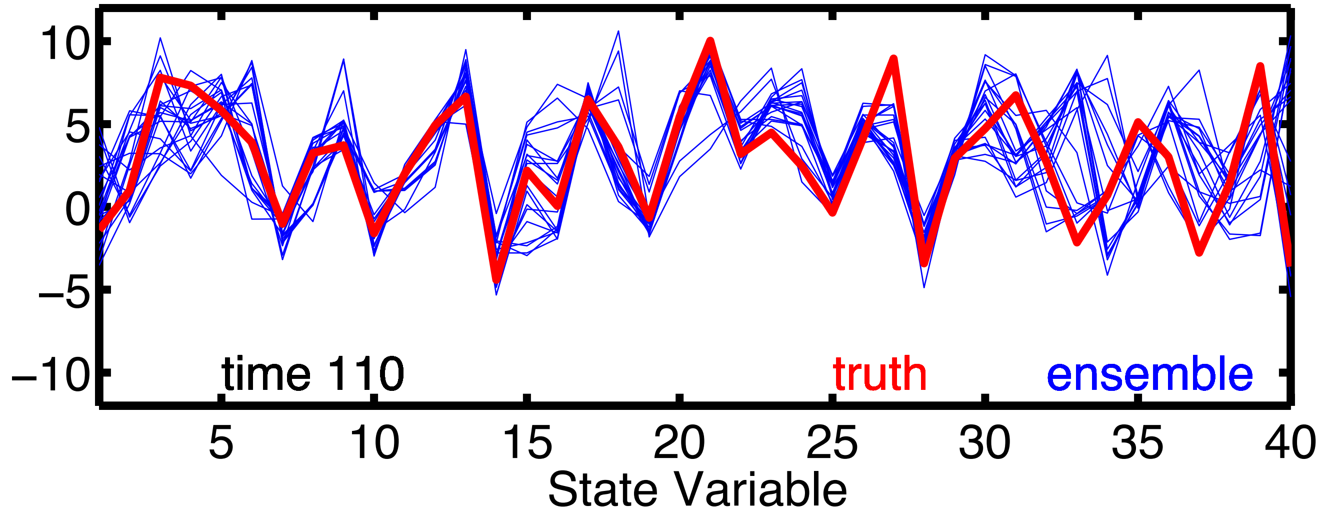
Lorenz-96 is sensitive to small perturbations

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’..



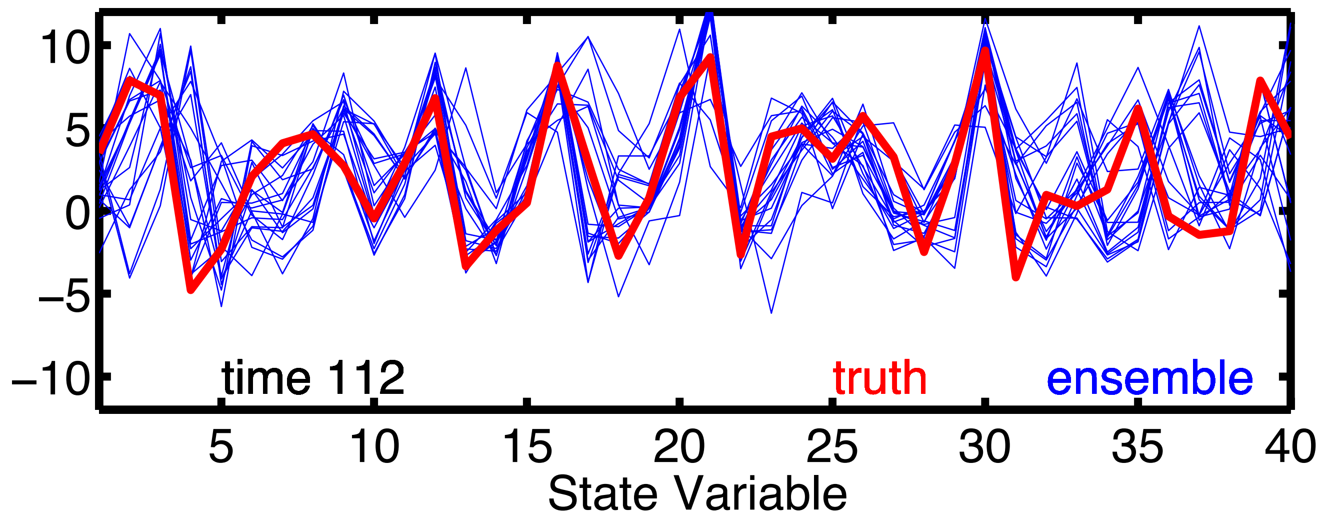
Lorenz-96 is sensitive to small perturbations

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'..



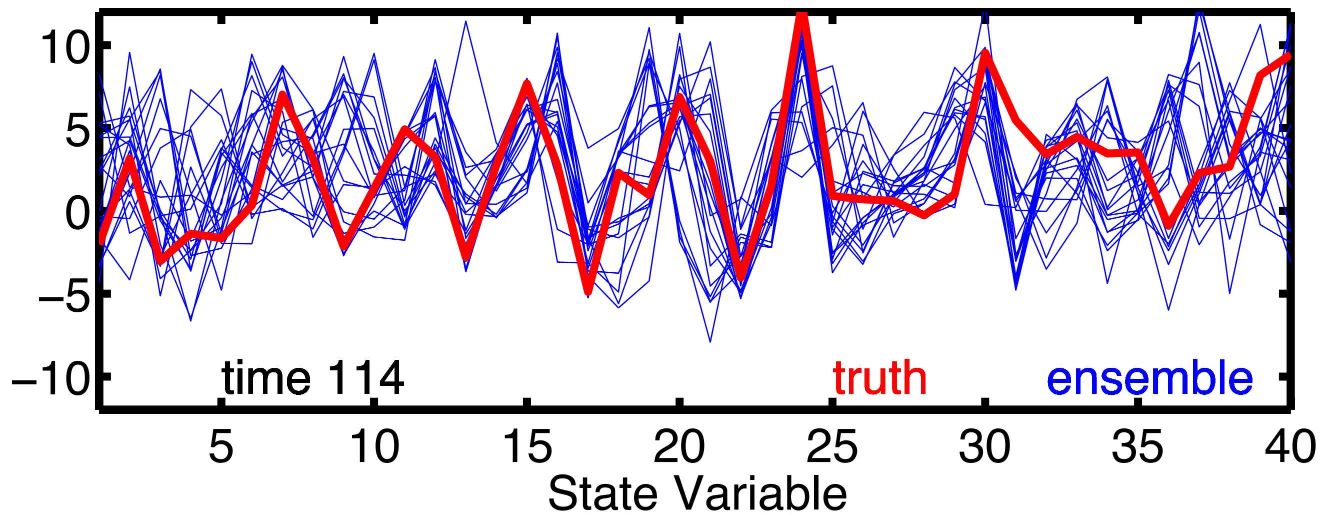
Lorenz-96 is sensitive to small perturbations

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'..



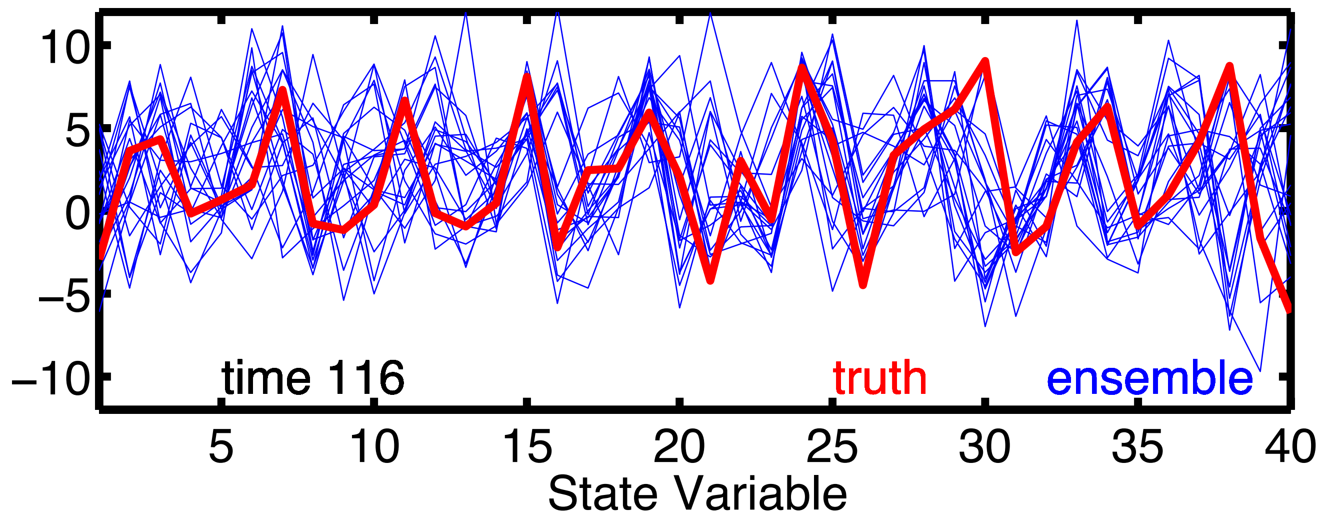
Lorenz-96 is sensitive to small perturbations

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'..



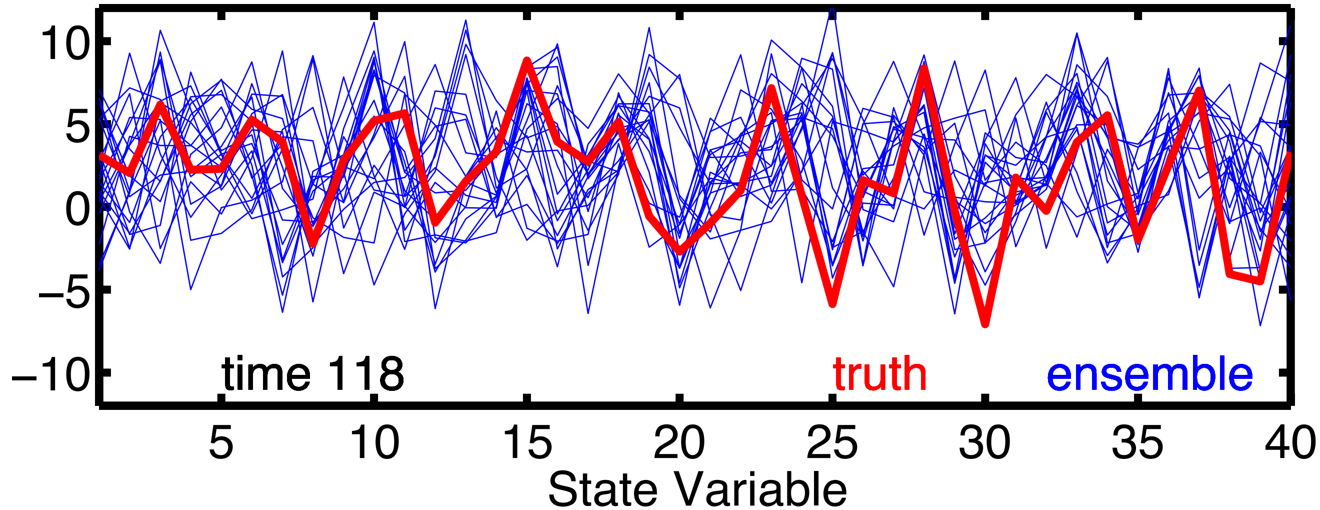
Lorenz-96 is sensitive to small perturbations

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’..



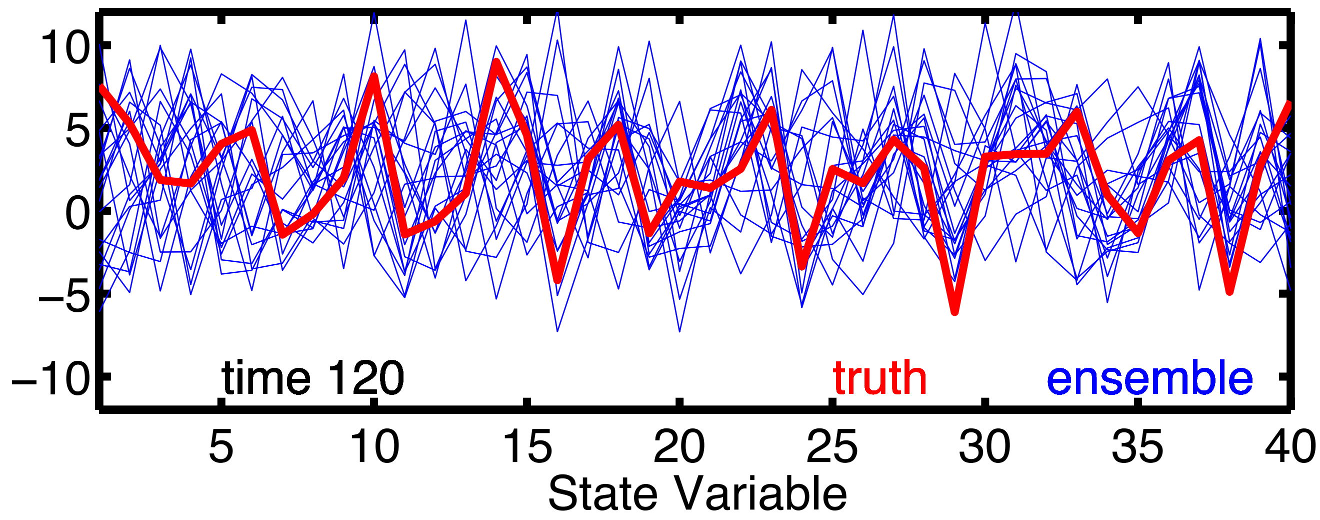
Lorenz-96 is sensitive to small perturbations

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'..



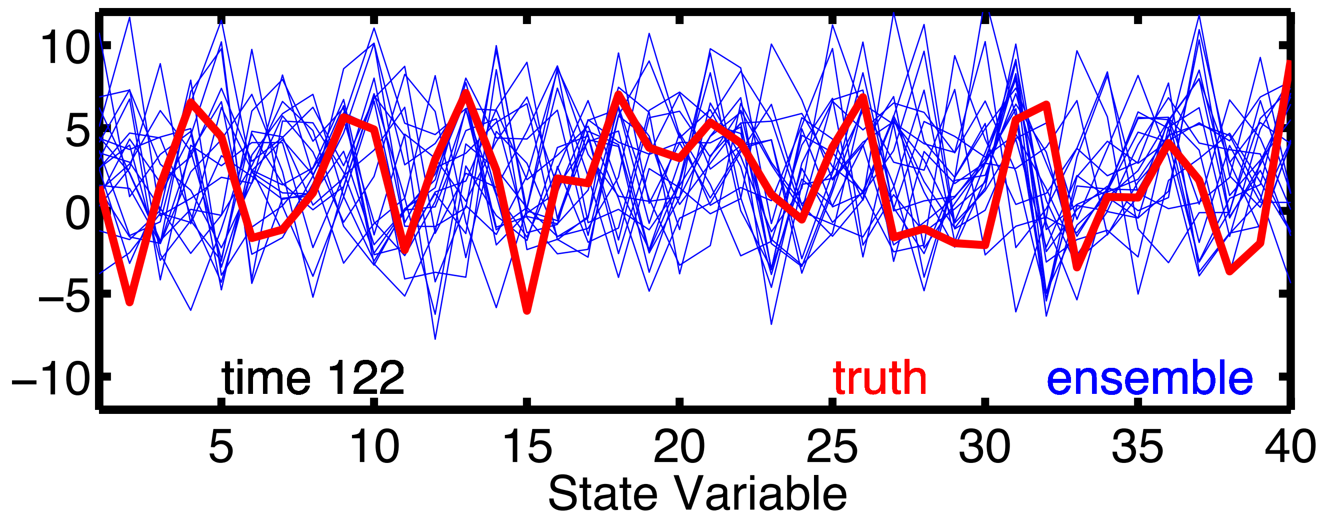
Lorenz-96 is sensitive to small perturbations

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'..



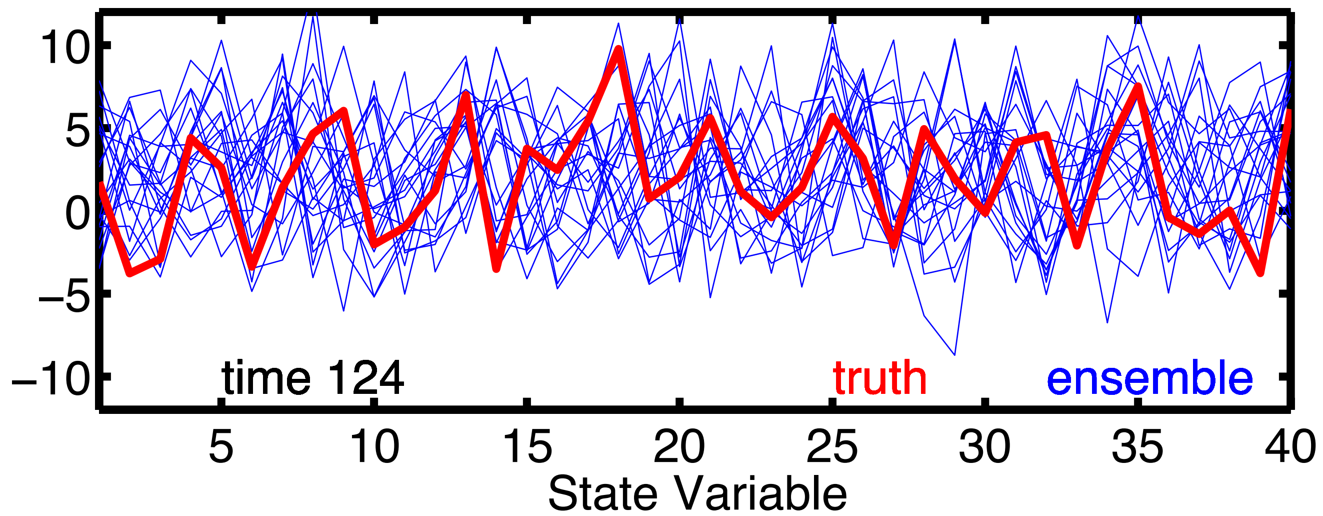
Lorenz-96 is sensitive to small perturbations

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'..



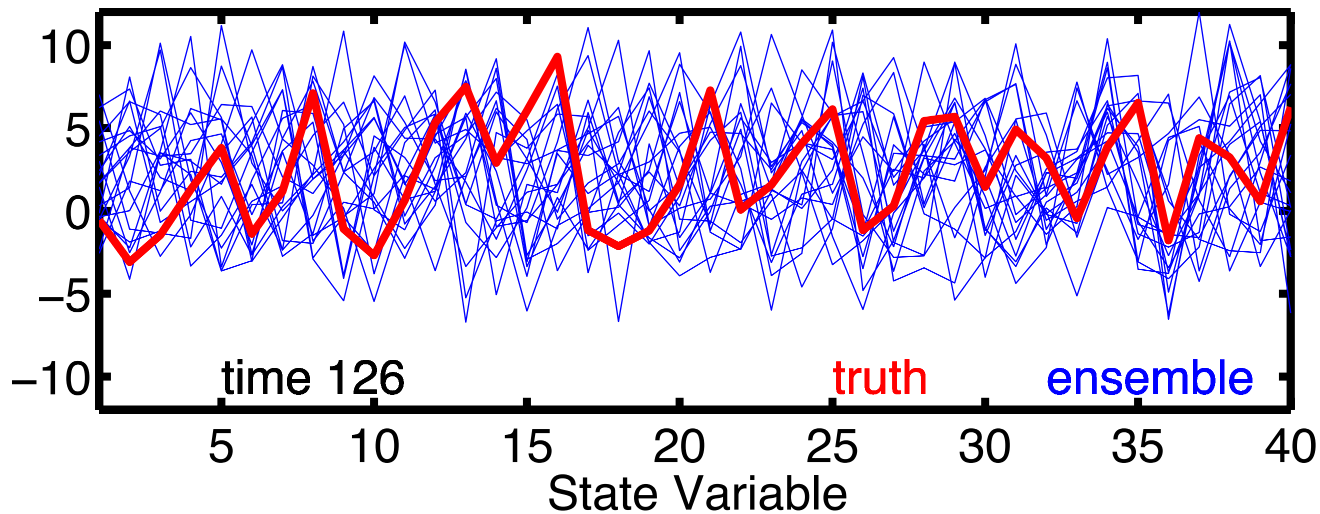
Lorenz-96 is sensitive to small perturbations

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'..



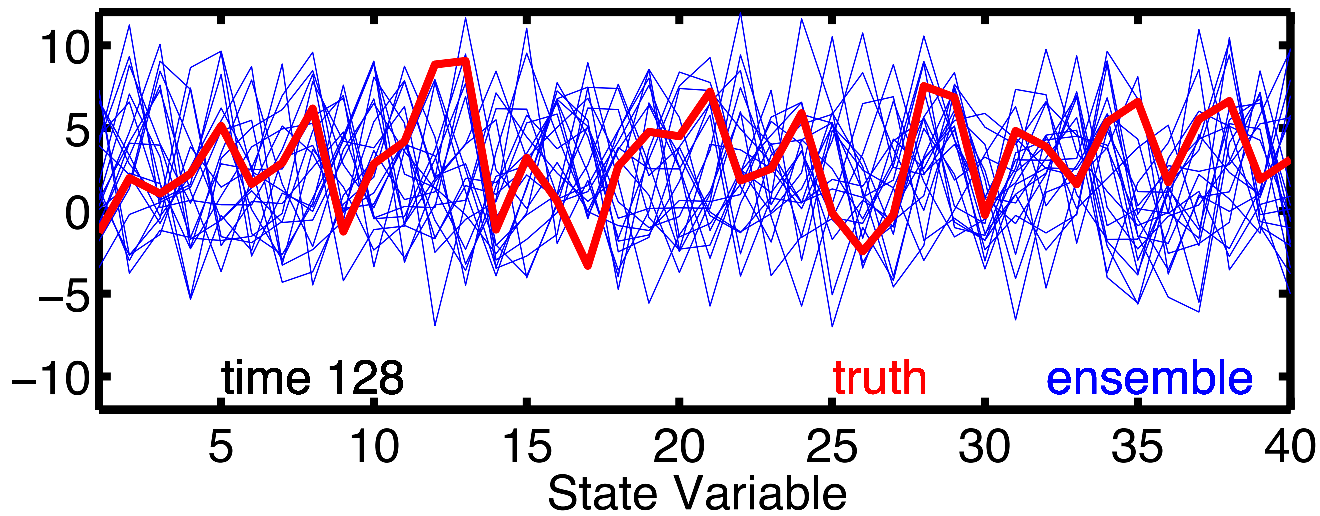
Lorenz-96 is sensitive to small perturbations

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’..



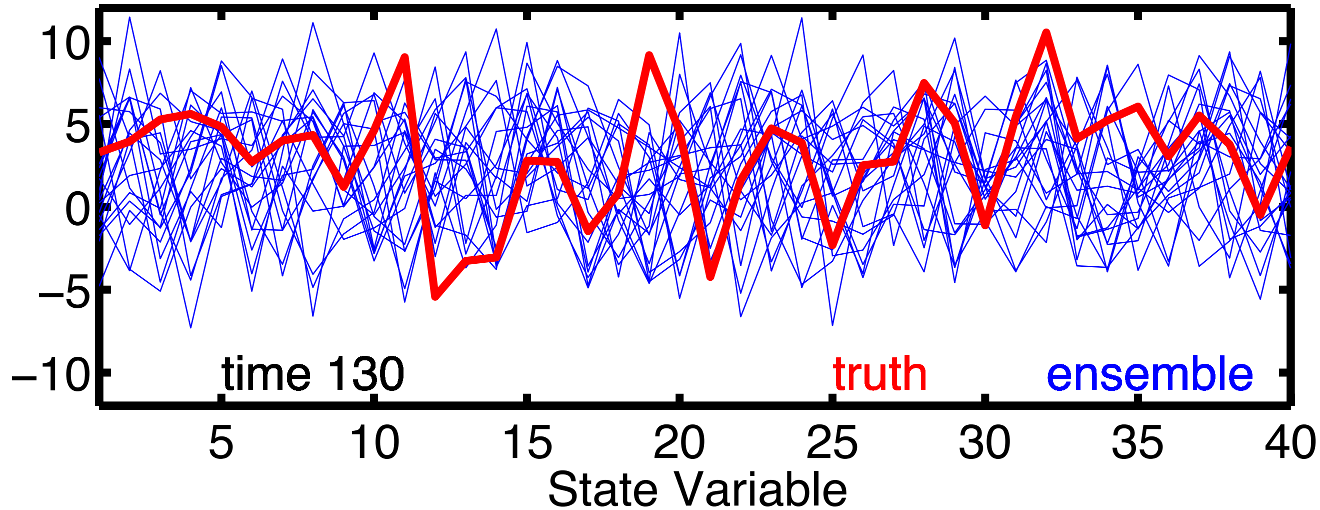
Lorenz-96 is sensitive to small perturbations

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'..



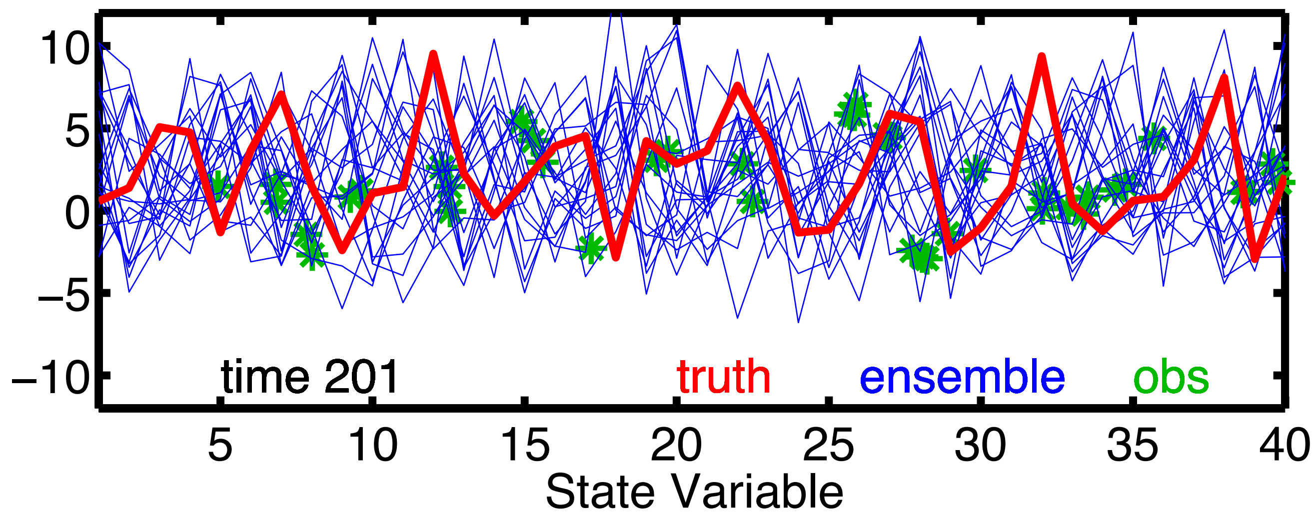
Lorenz-96 is sensitive to small perturbations

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’..



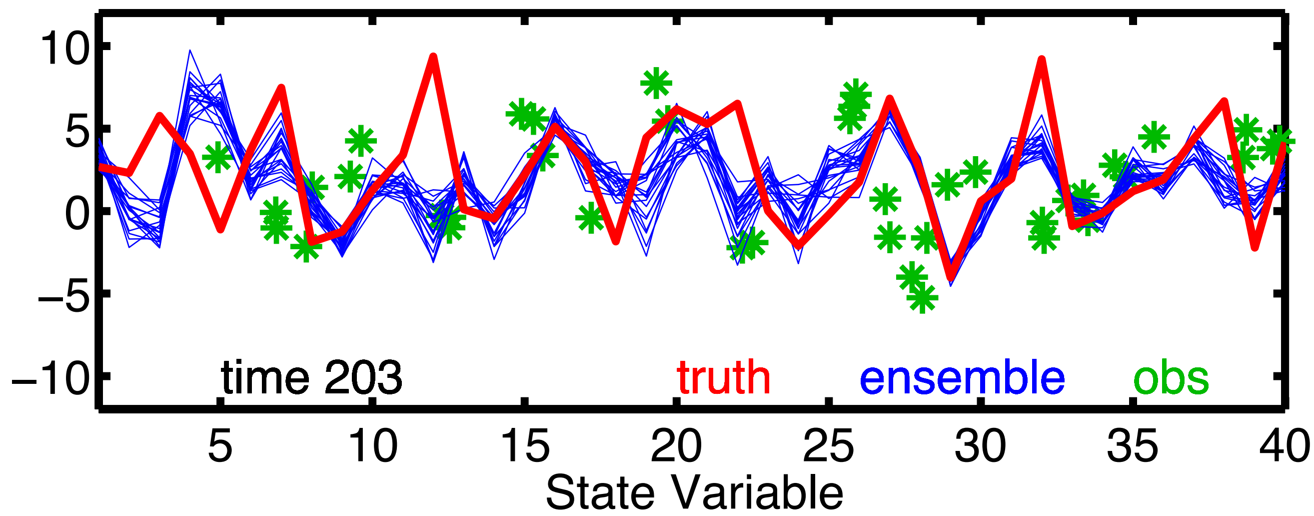
Assimilate ‘observations’ from 40 random locations each step.

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.



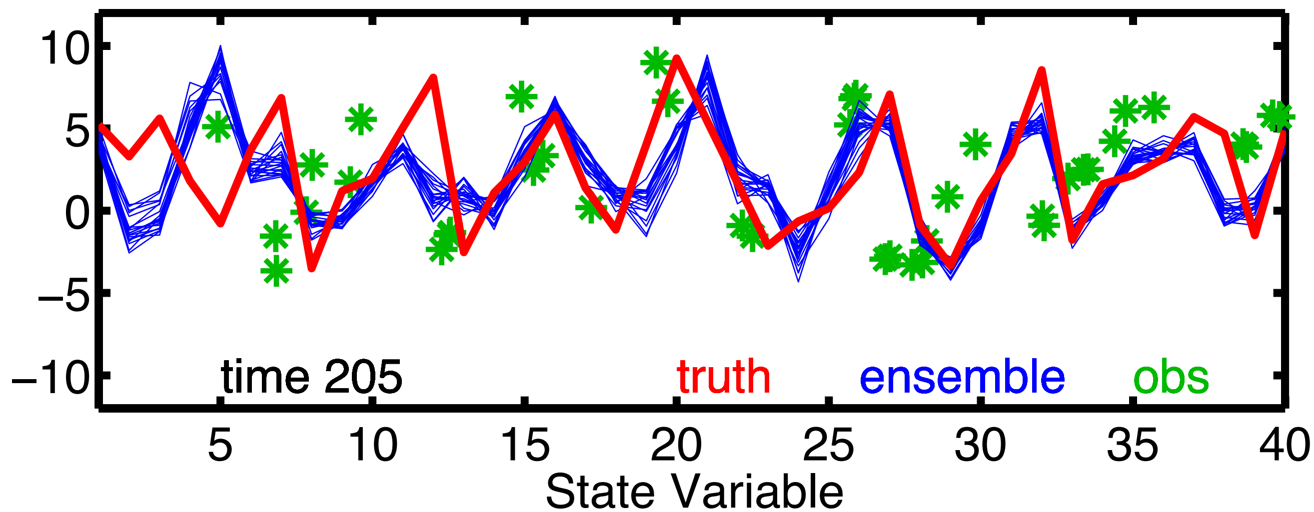
Assimilate ‘observations’ from 40 random locations each step.

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.



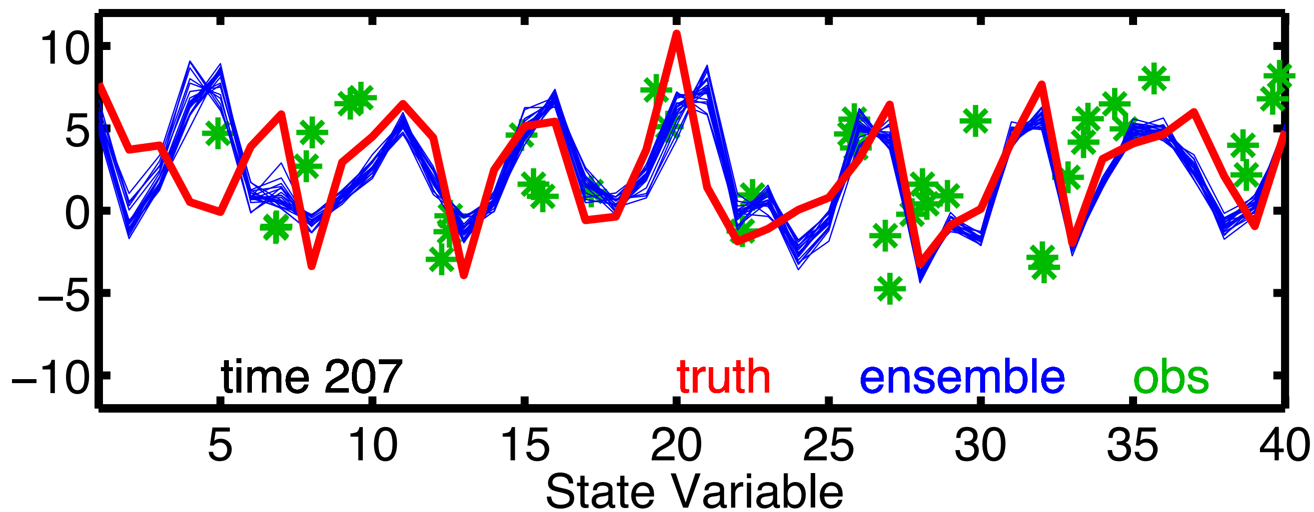
Assimilate 'observations' from 40 random locations each step.

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.



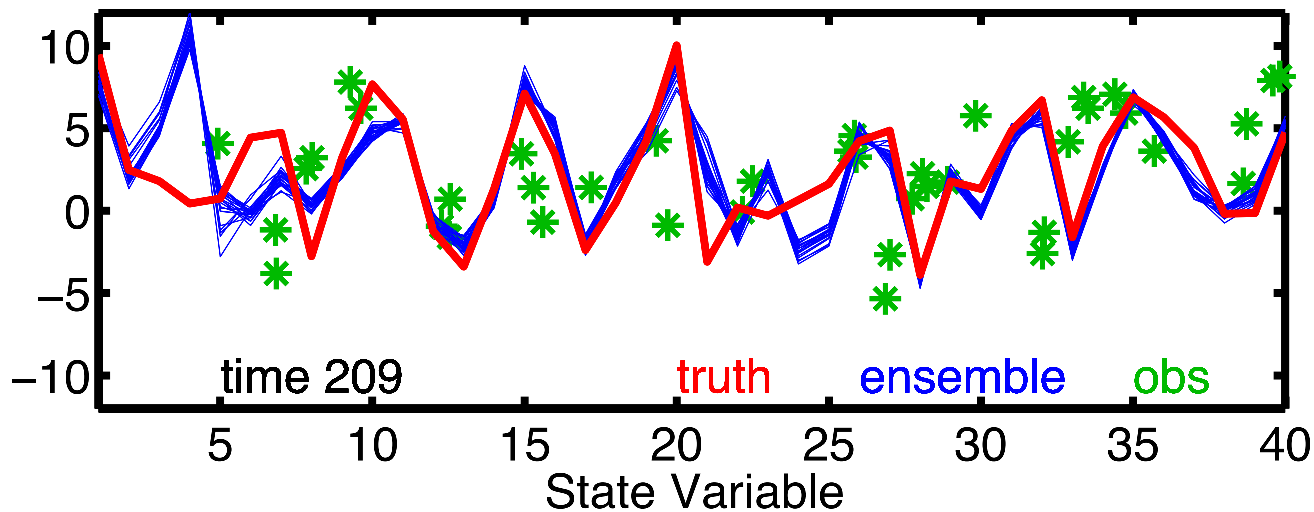
Assimilate 'observations' from 40 random locations each step.

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.



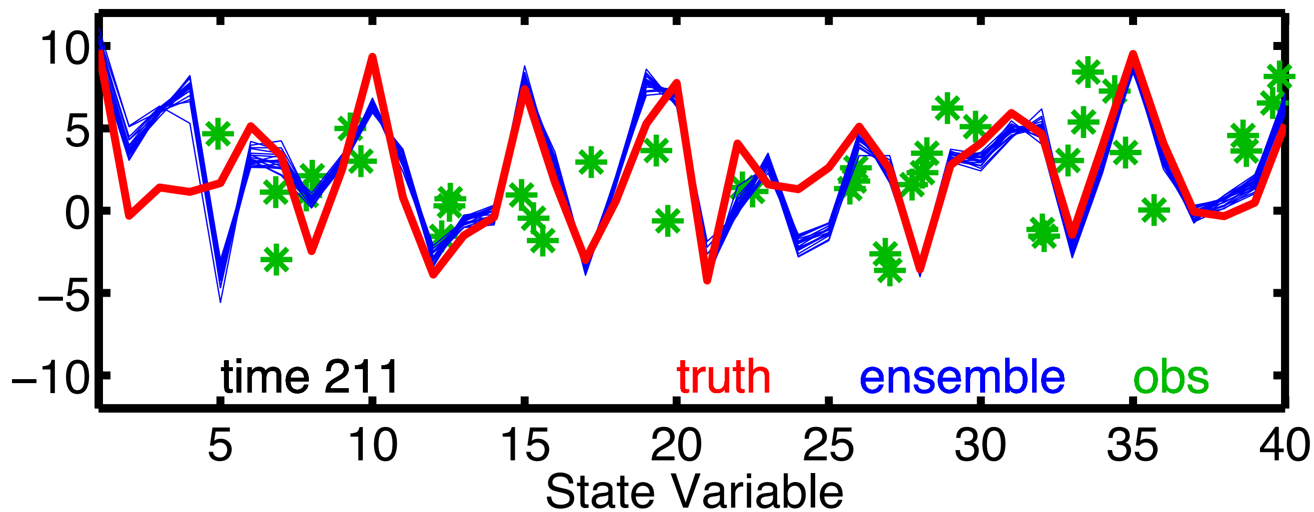
Assimilate ‘observations’ from 40 random locations each step.

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.



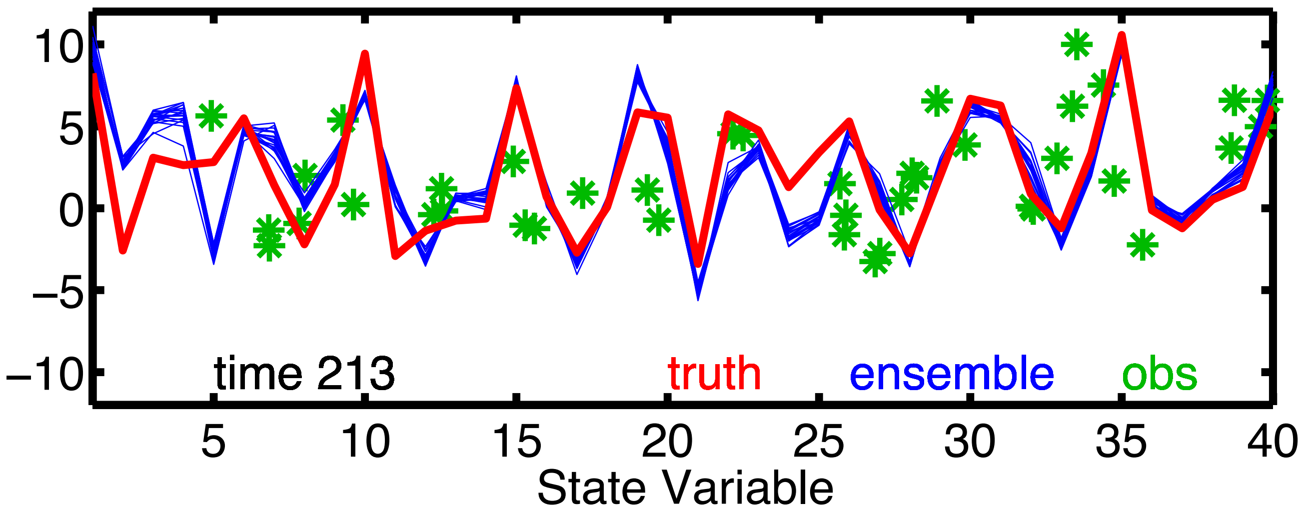
Assimilate 'observations' from 40 random locations each step.

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.



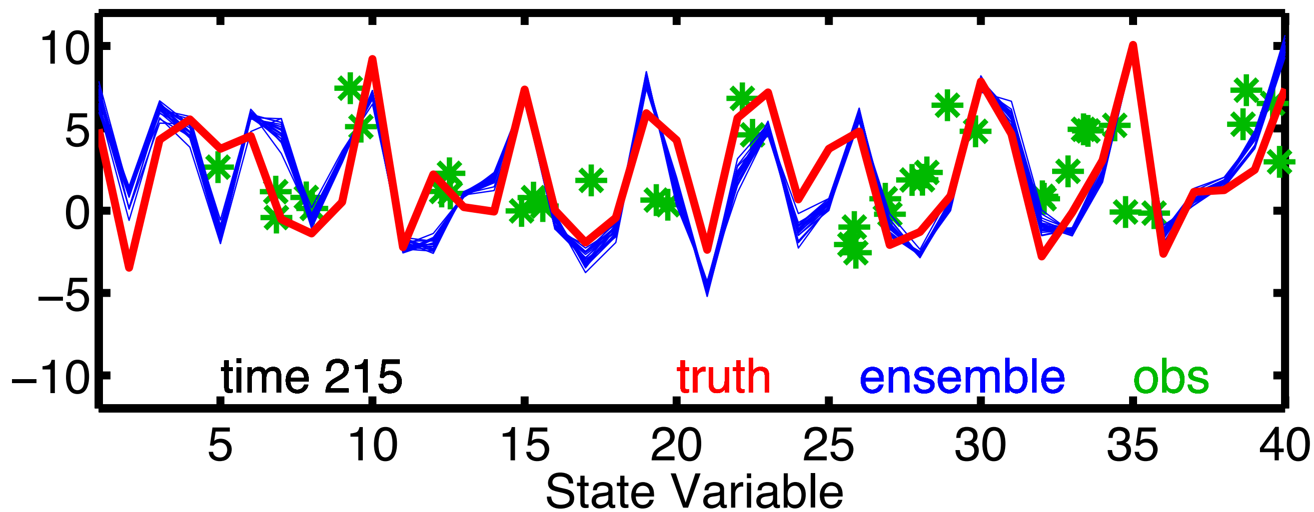
Assimilate 'observations' from 40 random locations each step.

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.



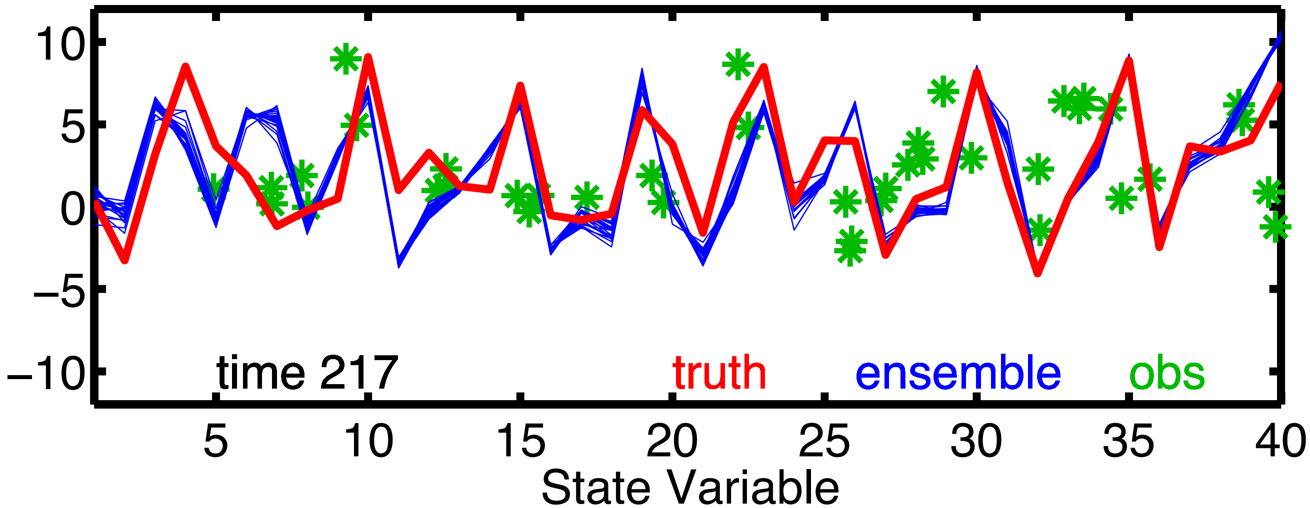
Assimilate 'observations' from 40 random locations each step.

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.



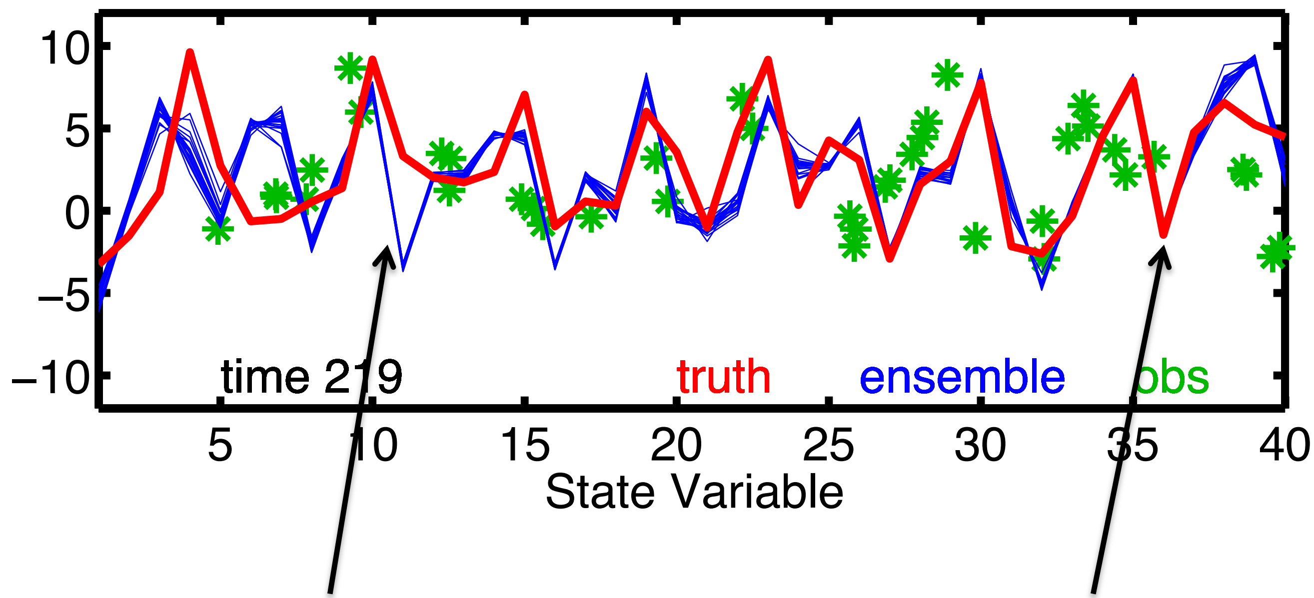
Assimilate 'observations' from 40 random locations each step.

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.

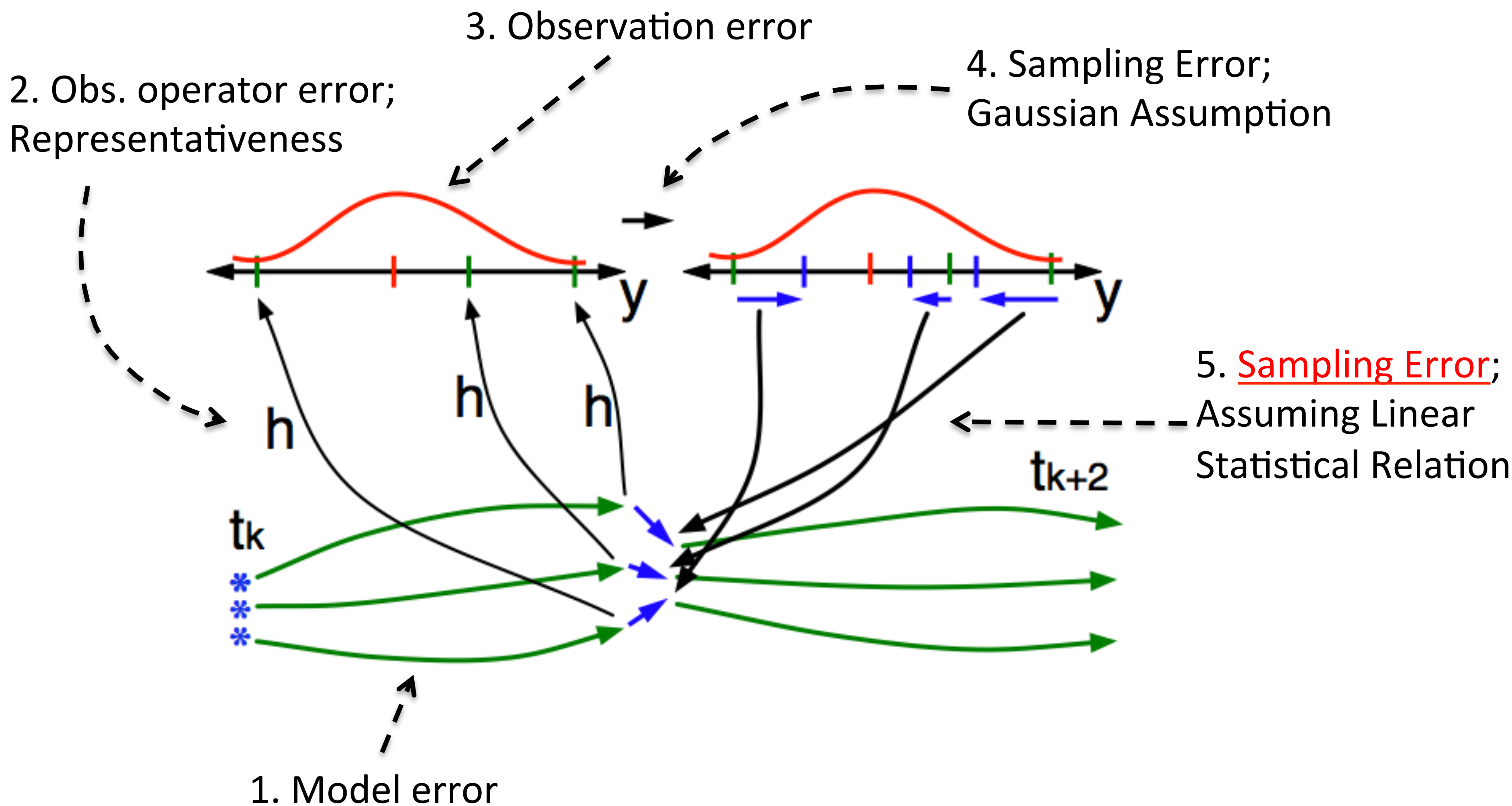
This isn't working very well.
Ensemble spread is reduced, but...,
Ensemble is inconsistent with truth most places.



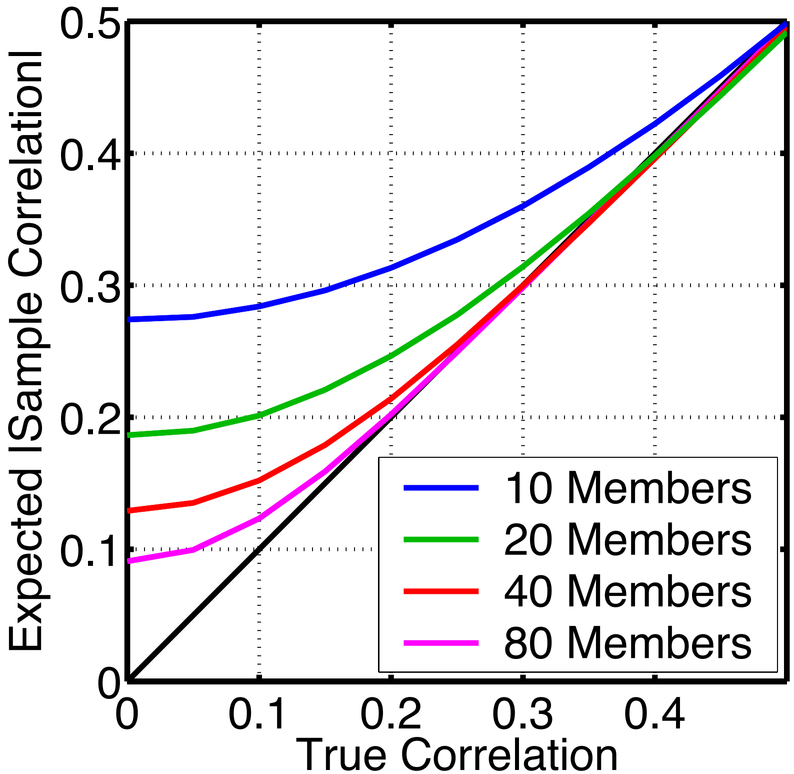
Confident and wrong.

Confident and right.

Some Error Sources in Ensemble Filters



Observations impact unrelated state variables through sampling error.



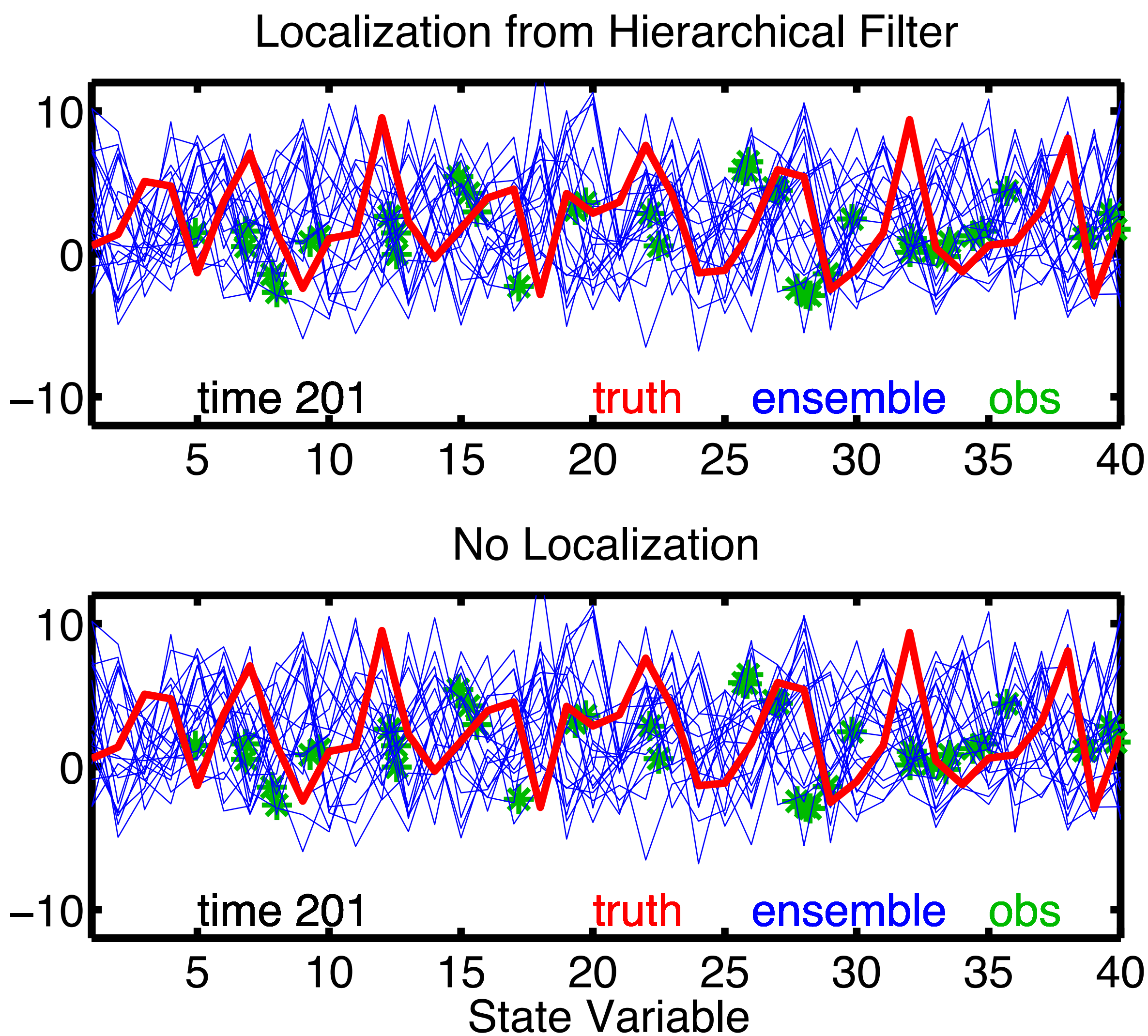
Plot shows expected absolute value of sample correlation vs. true correlation.

Unrelated obs. reduce spread, increase error.

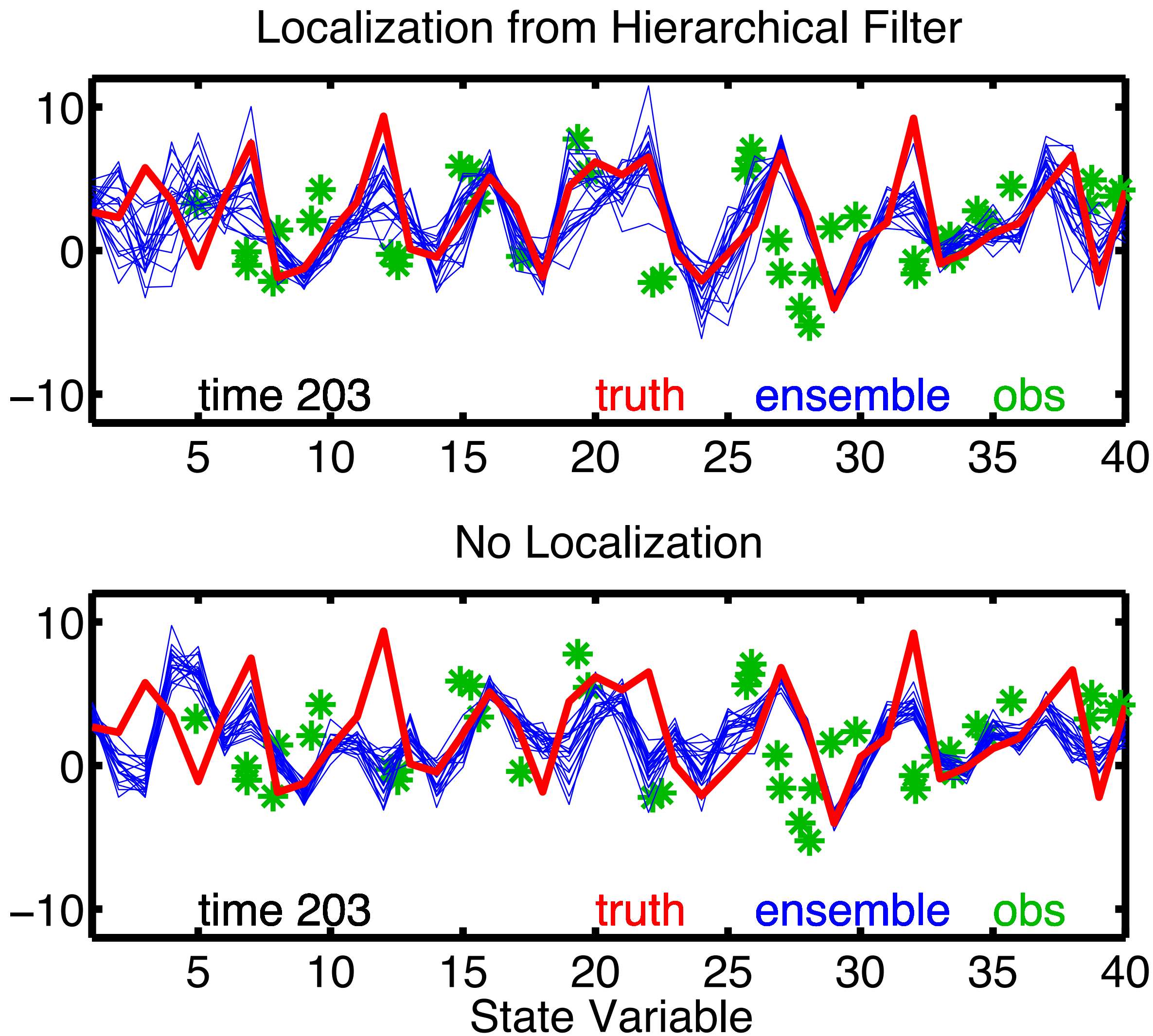
Attack with localization.

Don't let obs. Impact unrelated state.

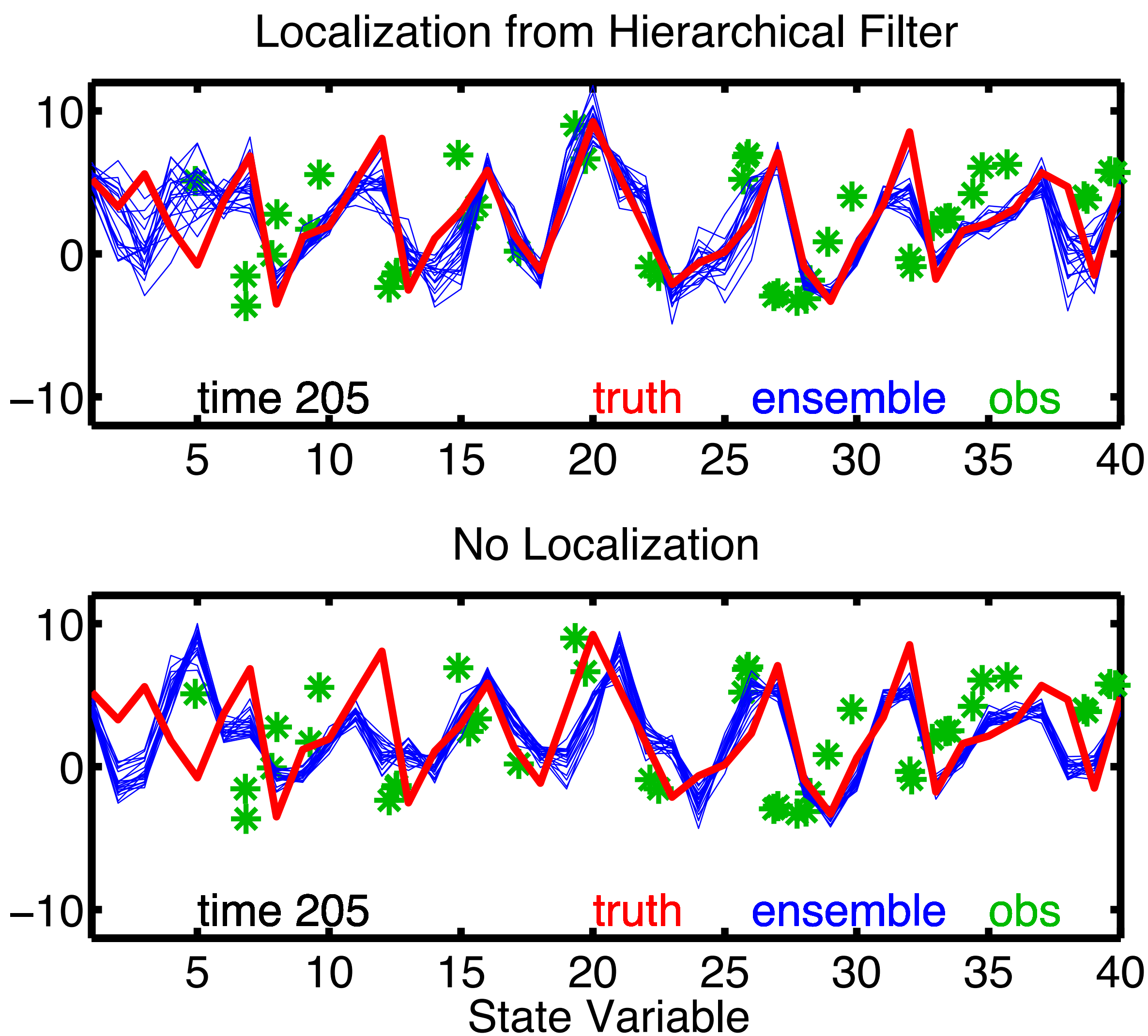
Lorenz-96 Assimilation with localization of observation impact.



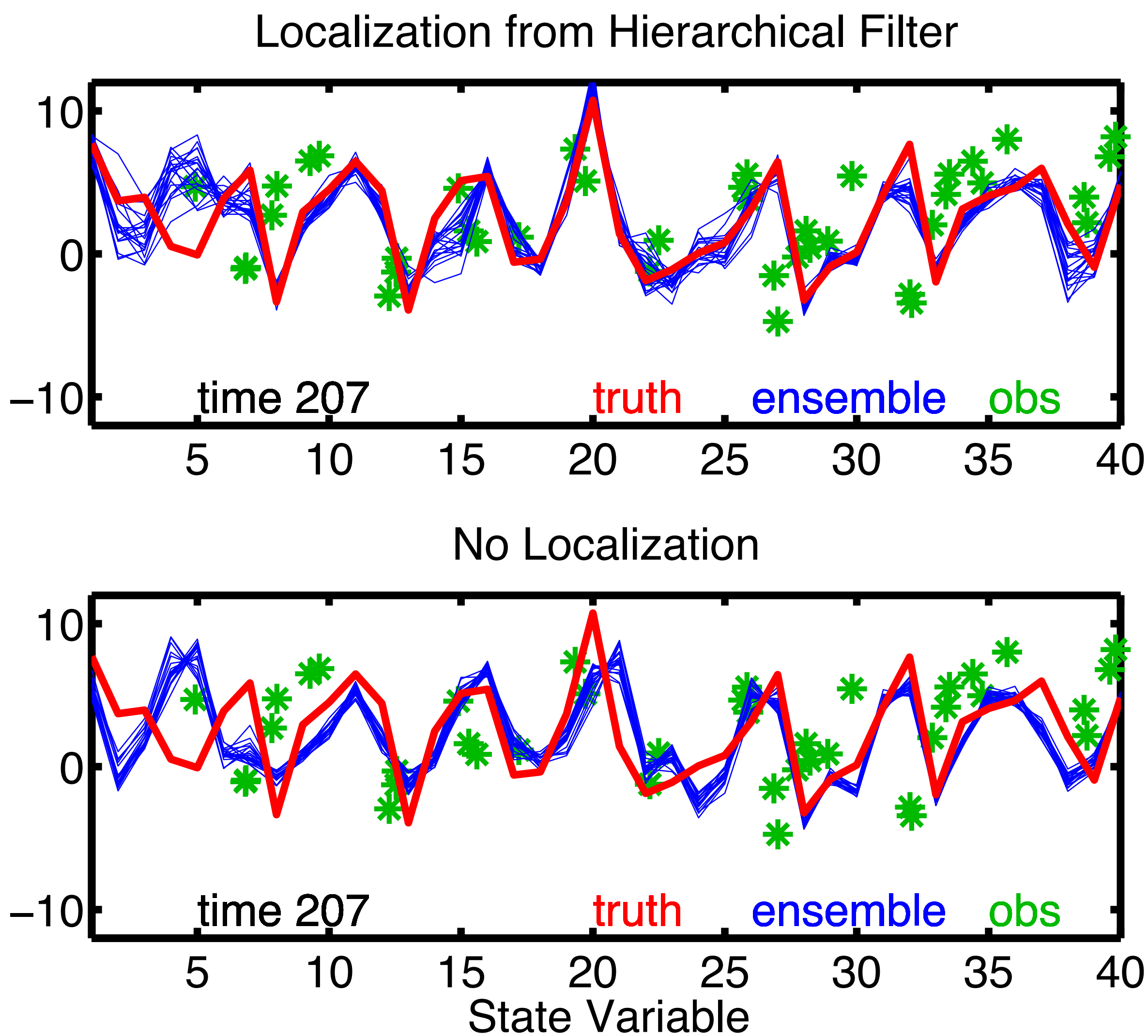
Lorenz-96 Assimilation with localization of observation impact.



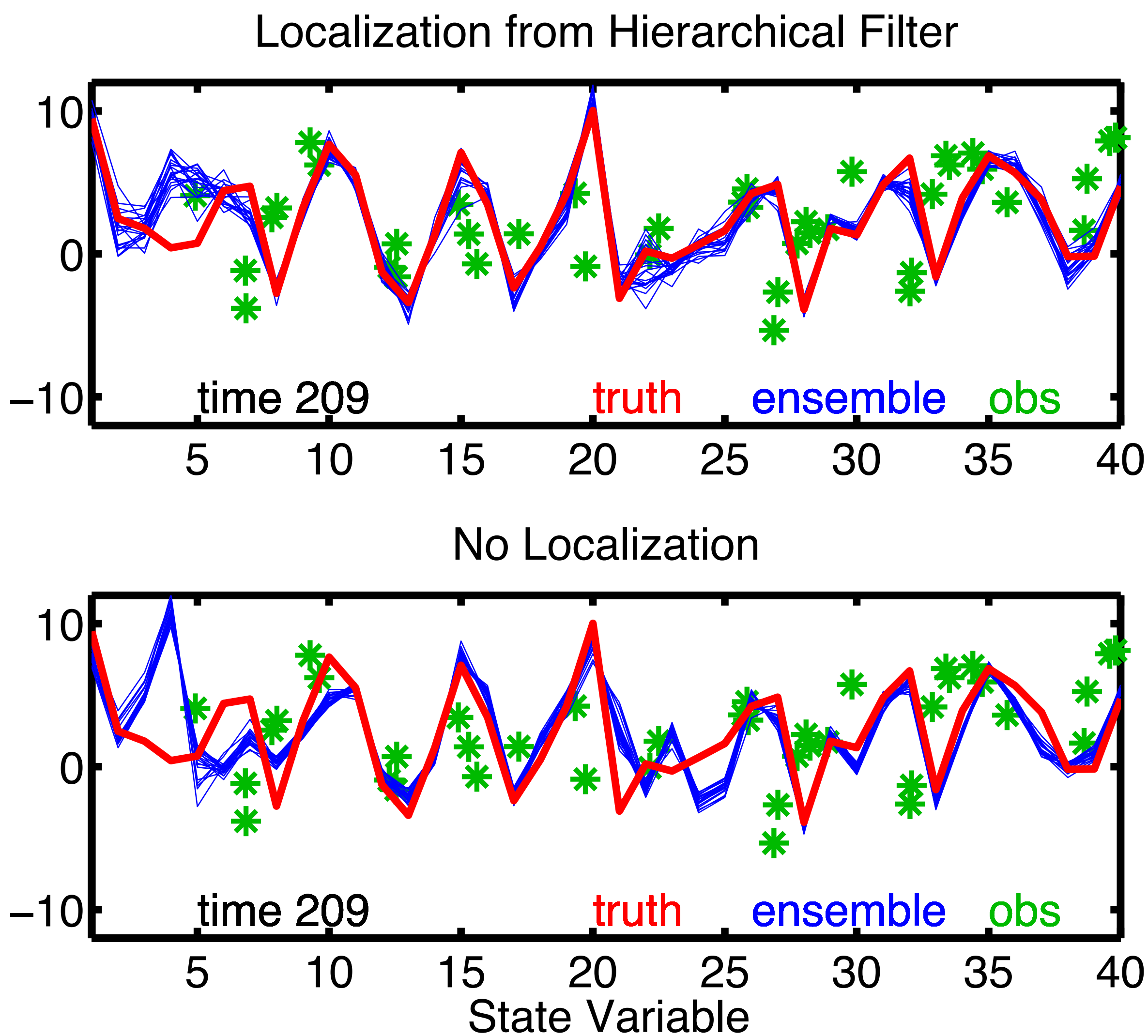
Lorenz-96 Assimilation with localization of observation impact.



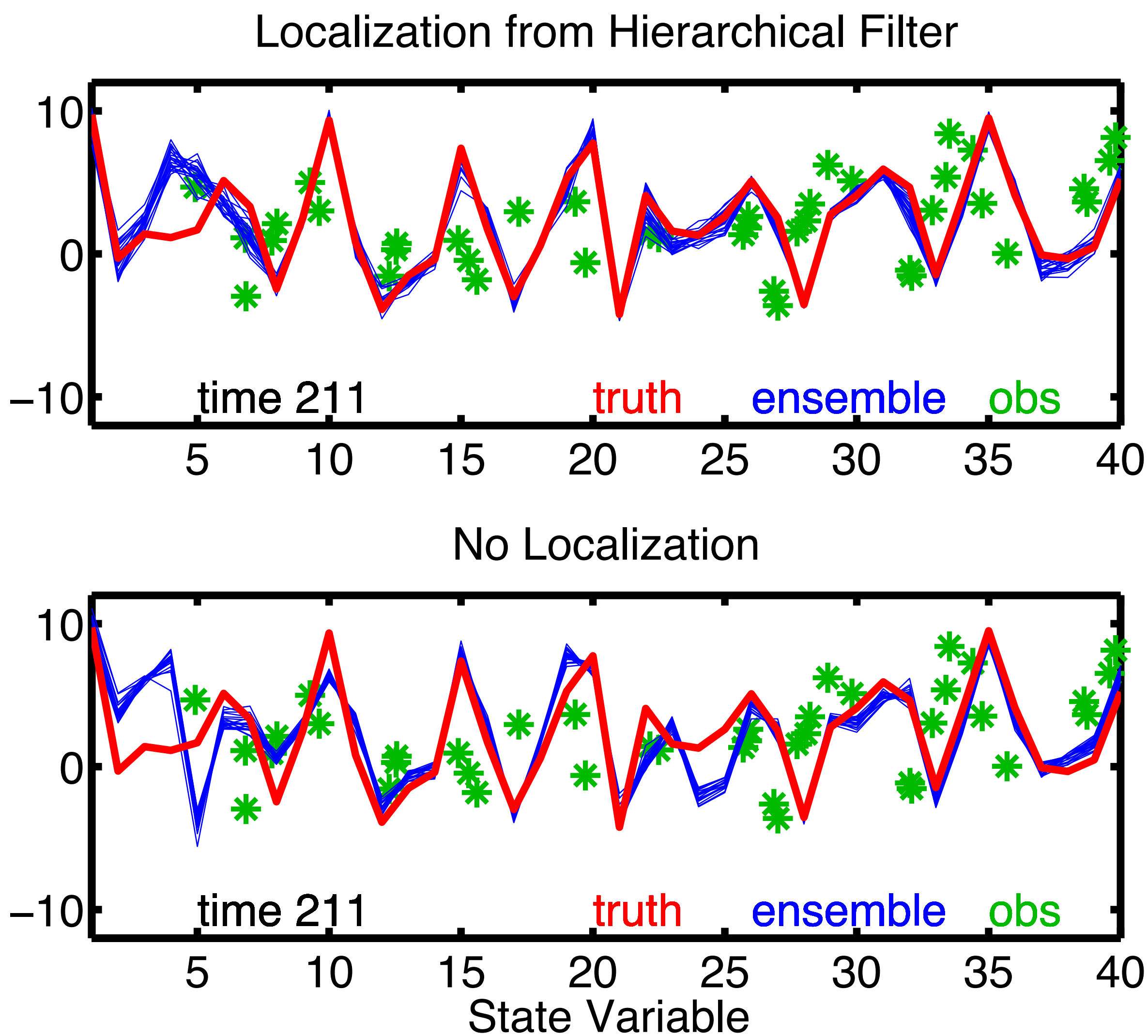
Lorenz-96 Assimilation with localization of observation impact.



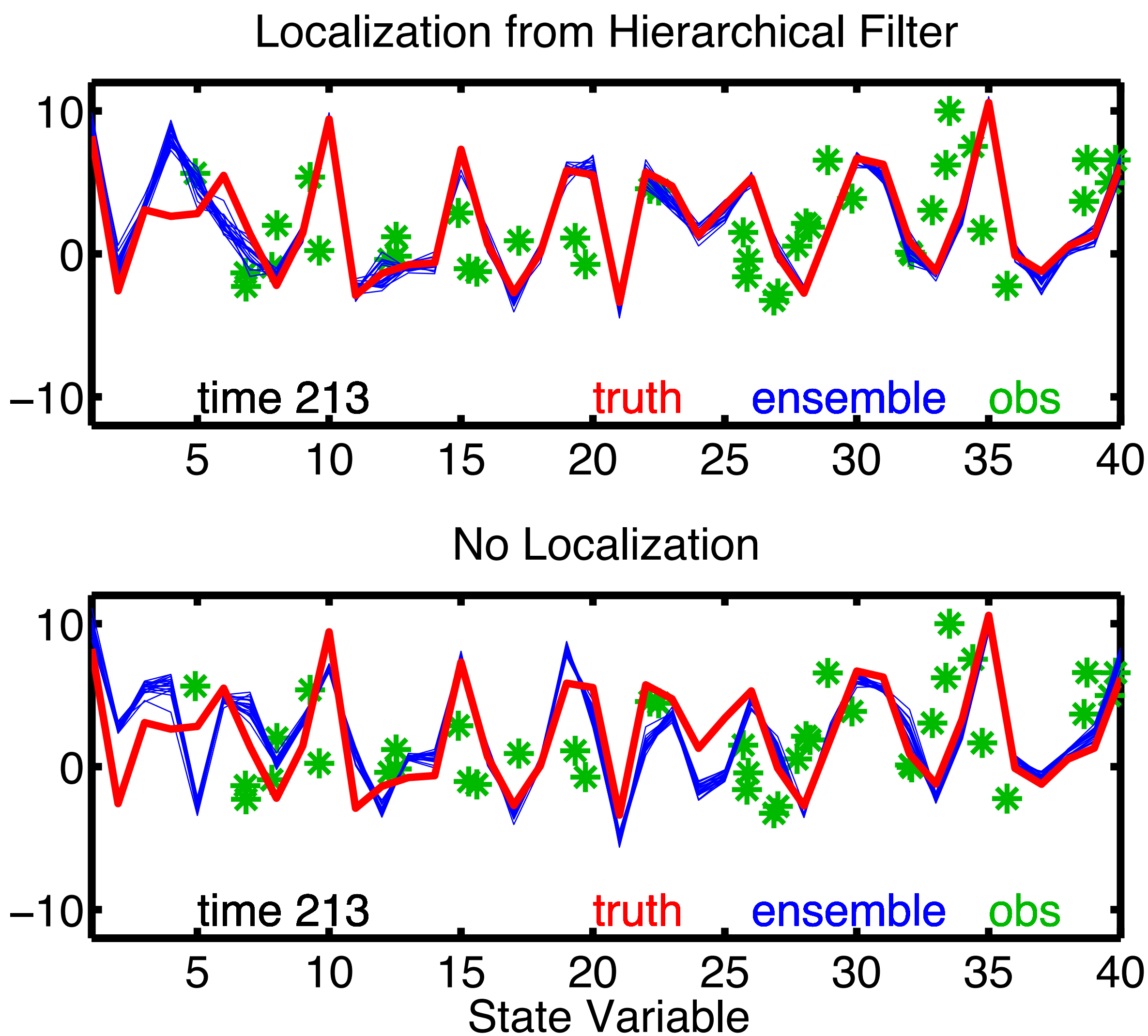
Lorenz-96 Assimilation with localization of observation impact.



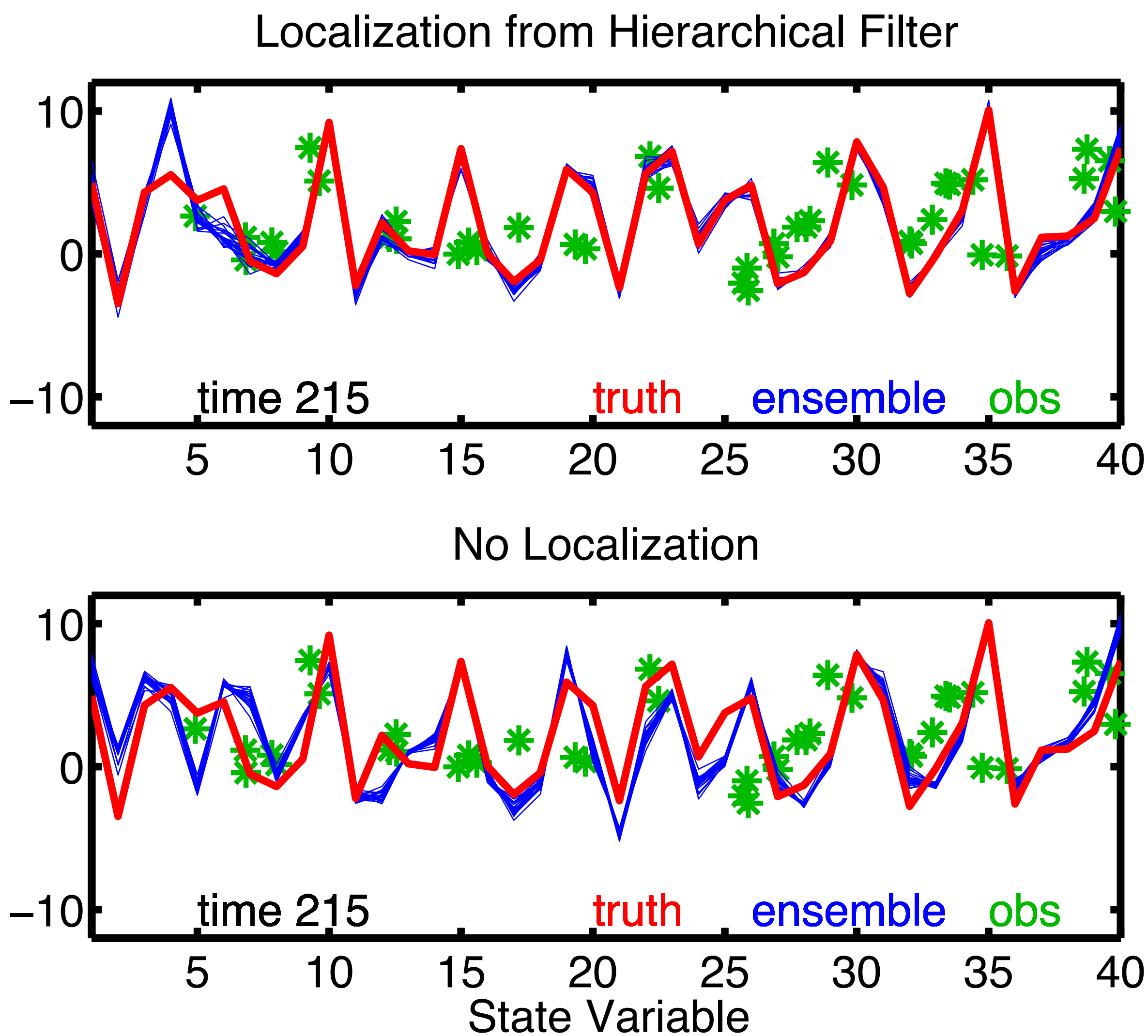
Lorenz-96 Assimilation with localization of observation impact.



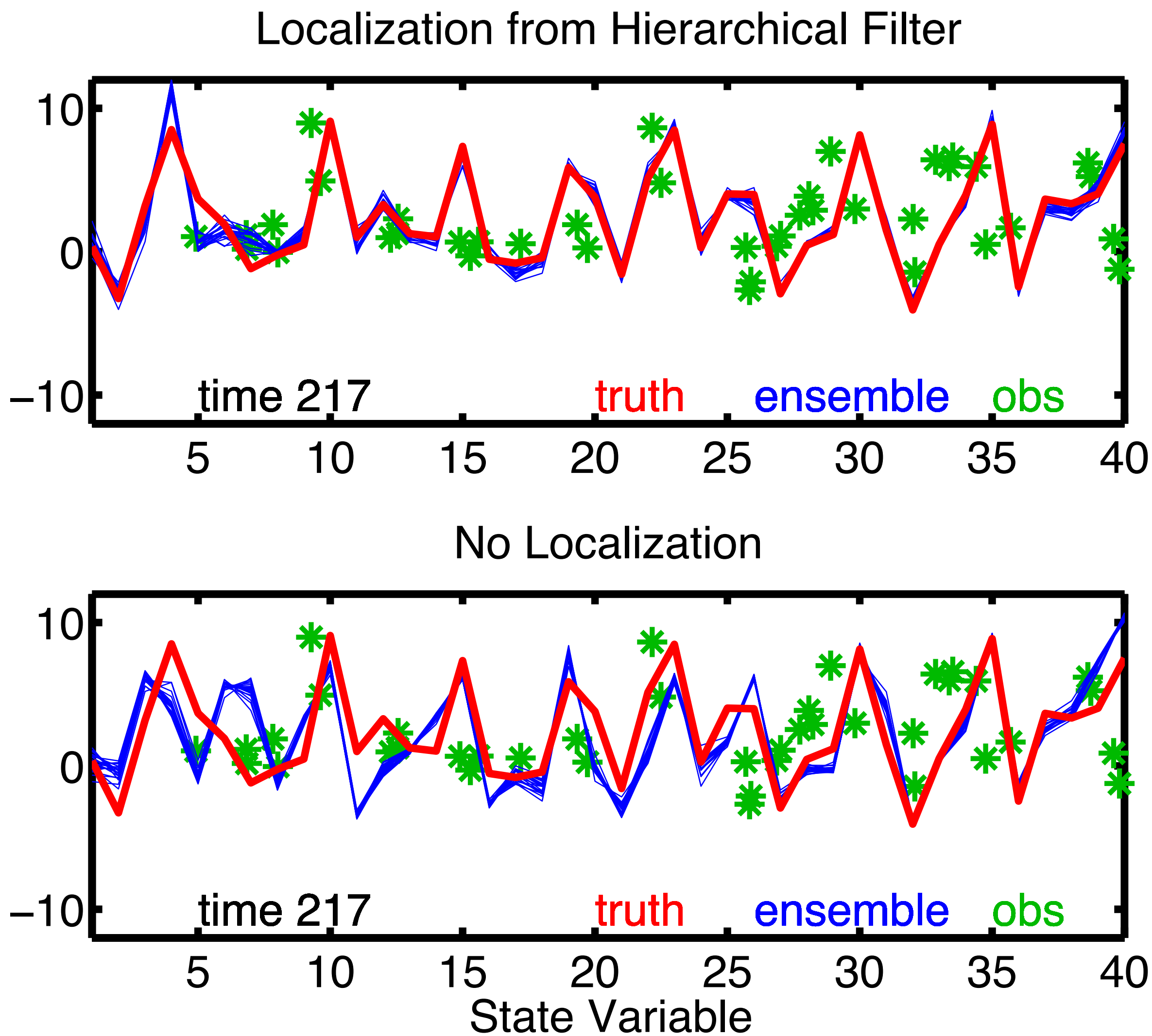
Lorenz-96 Assimilation with localization of observation impact.



Lorenz-96 Assimilation with localization of observation impact.

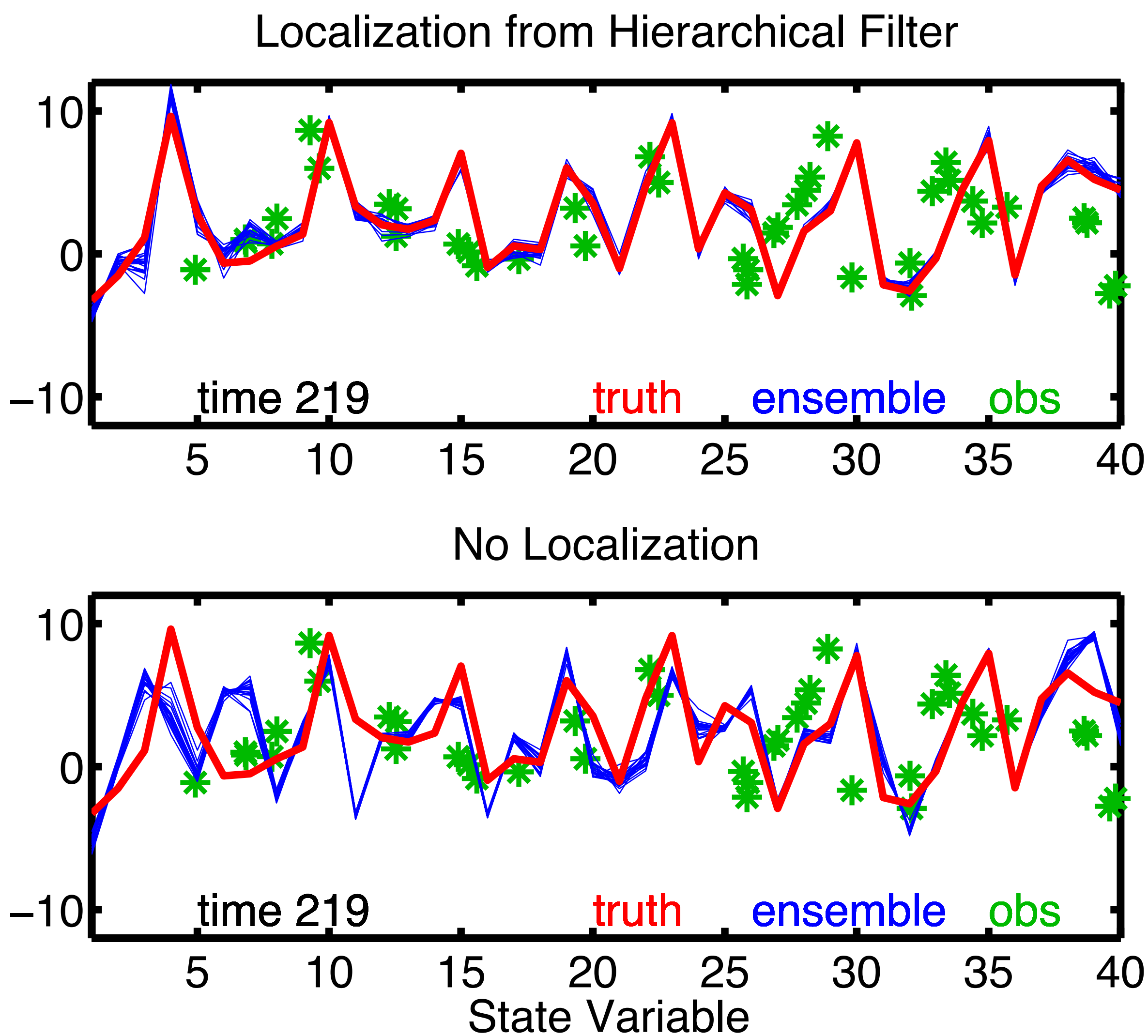


Lorenz-96 Assimilation with localization of observation impact.

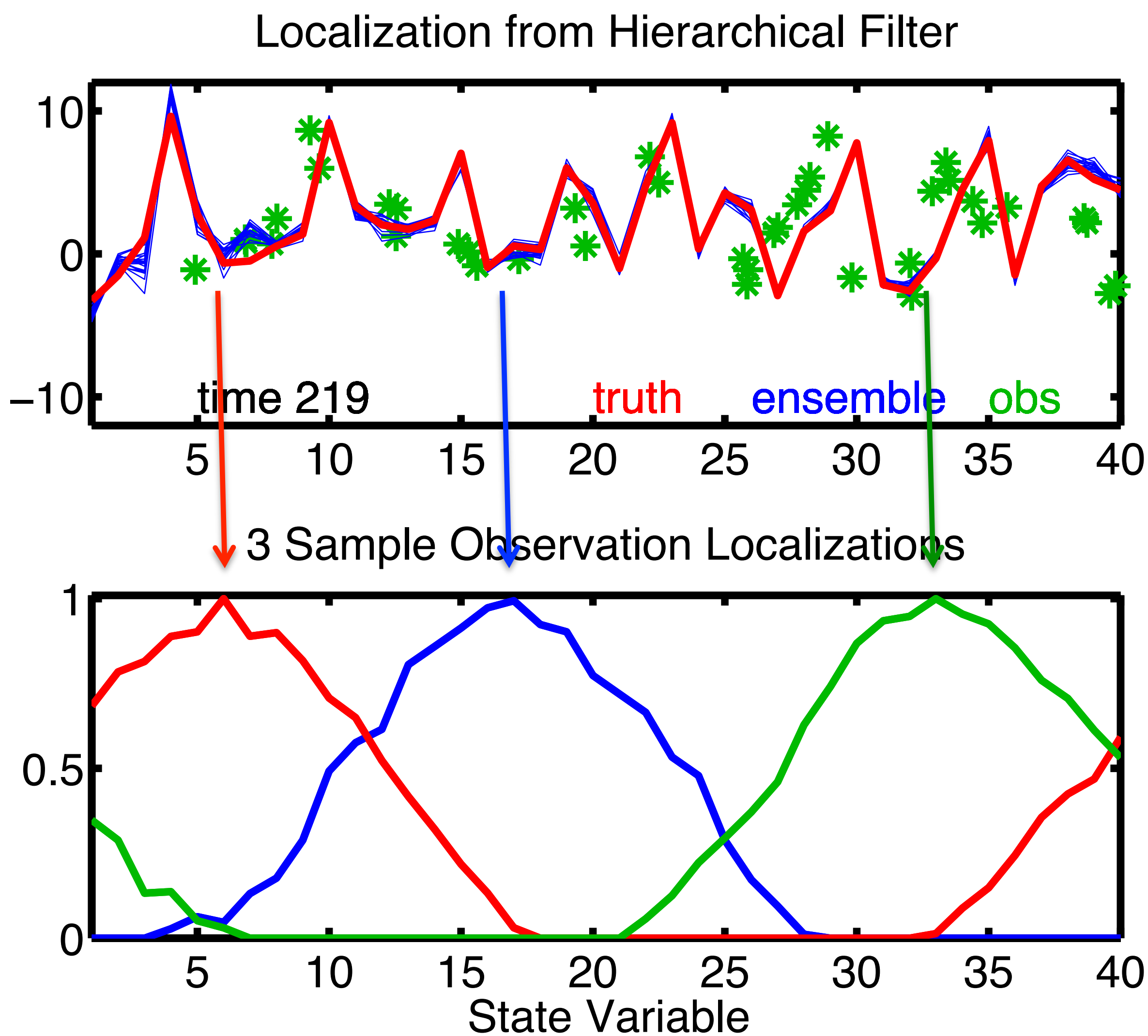


Lorenz-96 Assimilation with localization of observation impact.

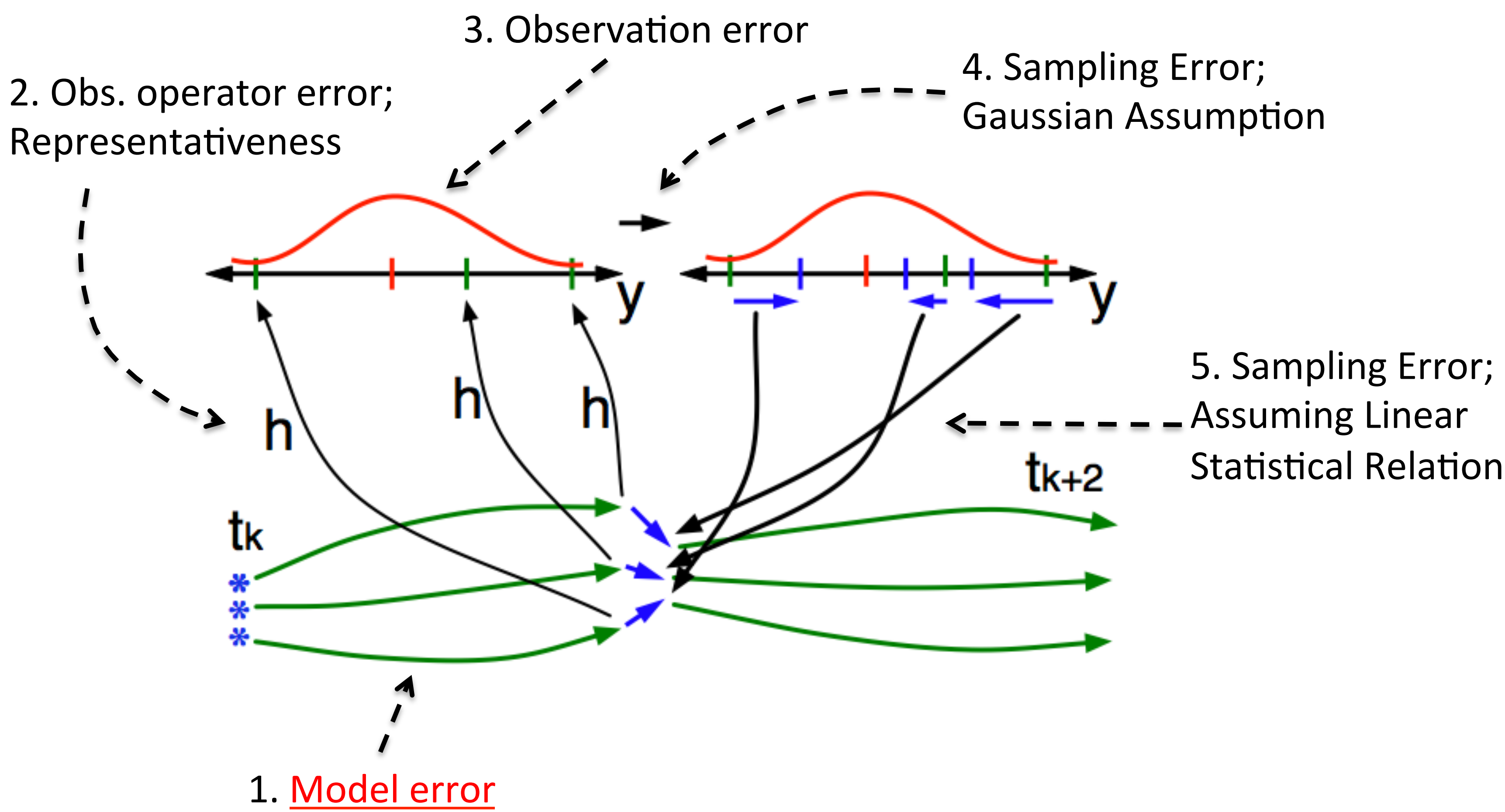
This ensemble is more consistent with truth.



Localization computed by empirical offline computation.

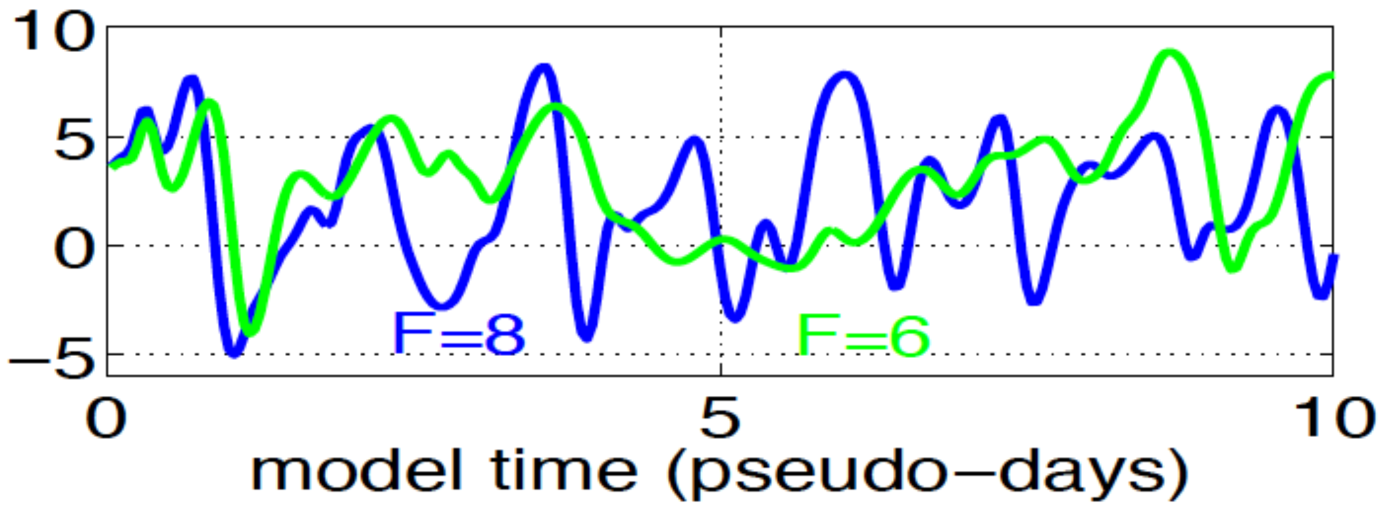


Some Error Sources in Ensemble Filters



Assimilating in the presence of simulated model error.

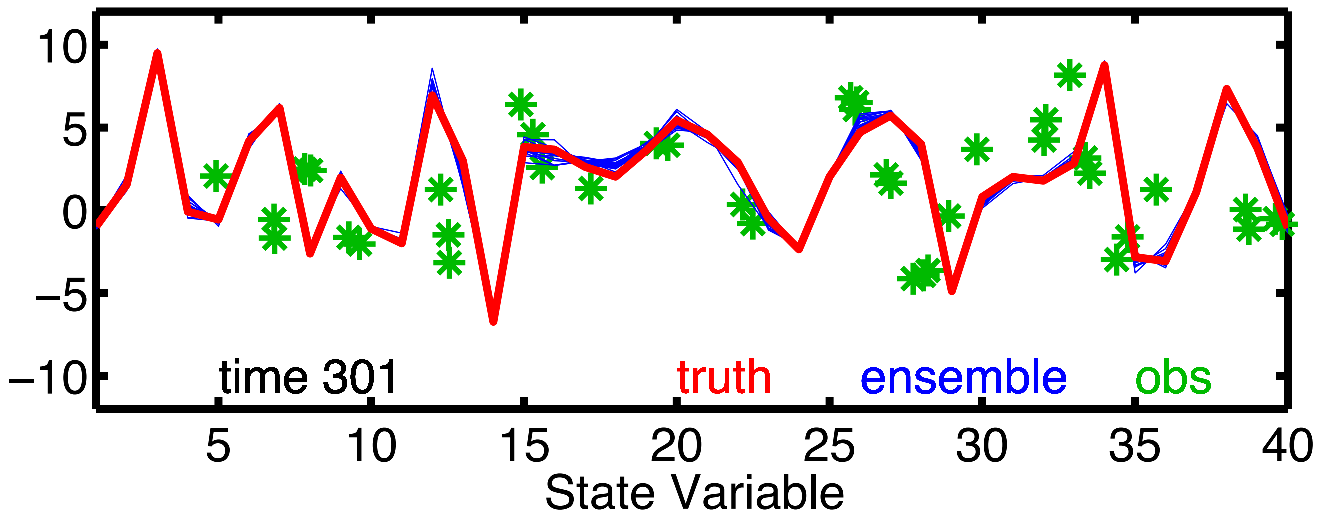
dXi / dt = (Xi+1 - Xi-2)Xi-1 - Xi + F.
For truth, use F = 8.
In assimilating model, use F = 6.



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

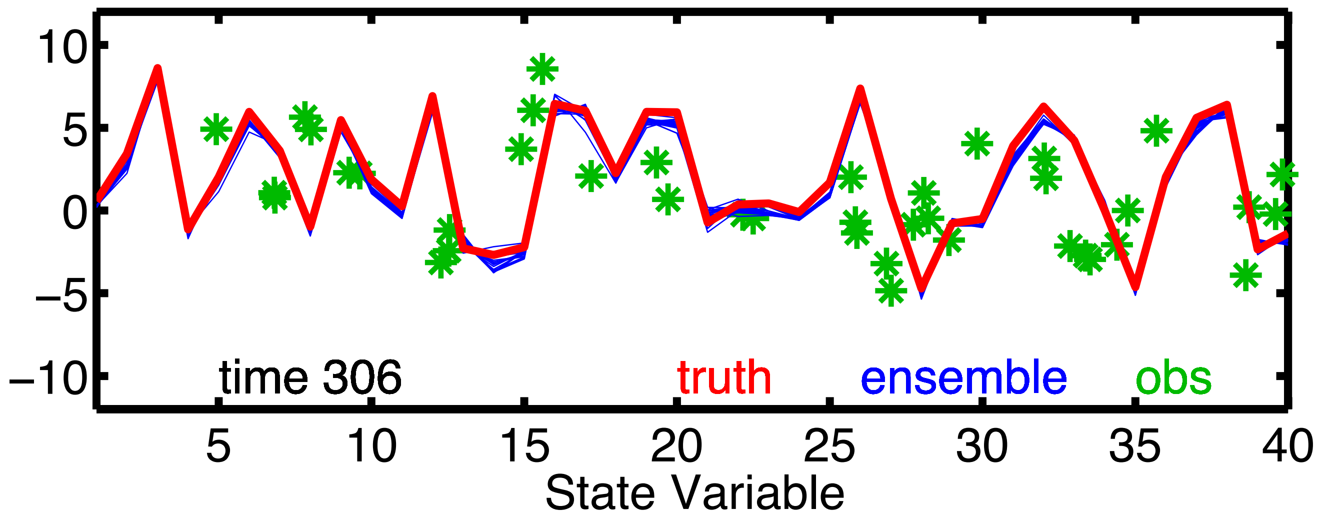
Assimilating in the presence of simulated model error.

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



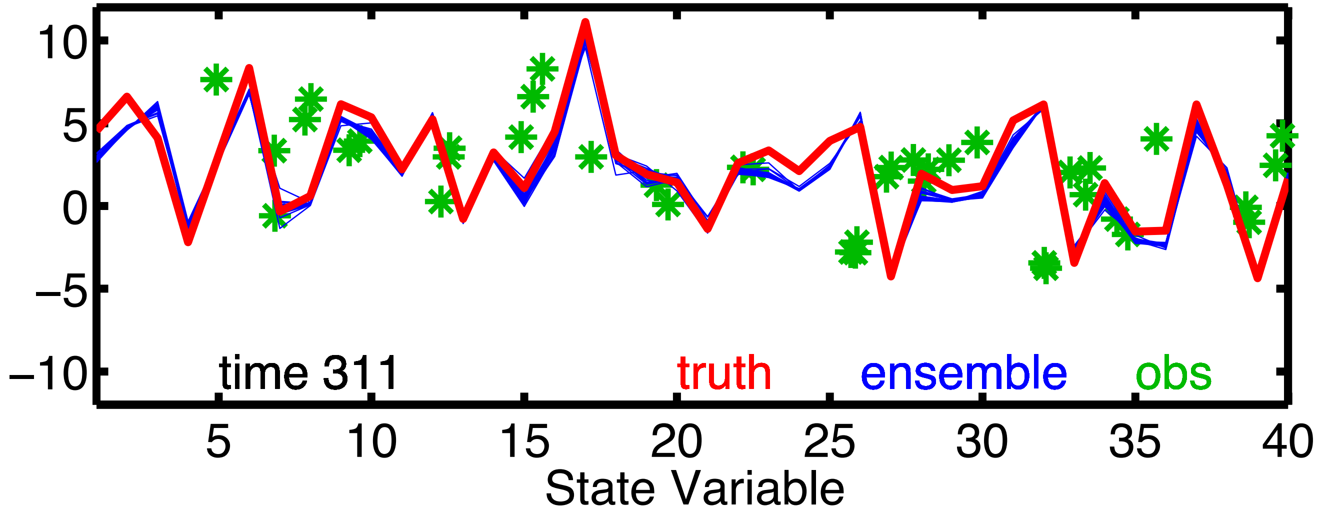
Assimilating in the presence of simulated model error.

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



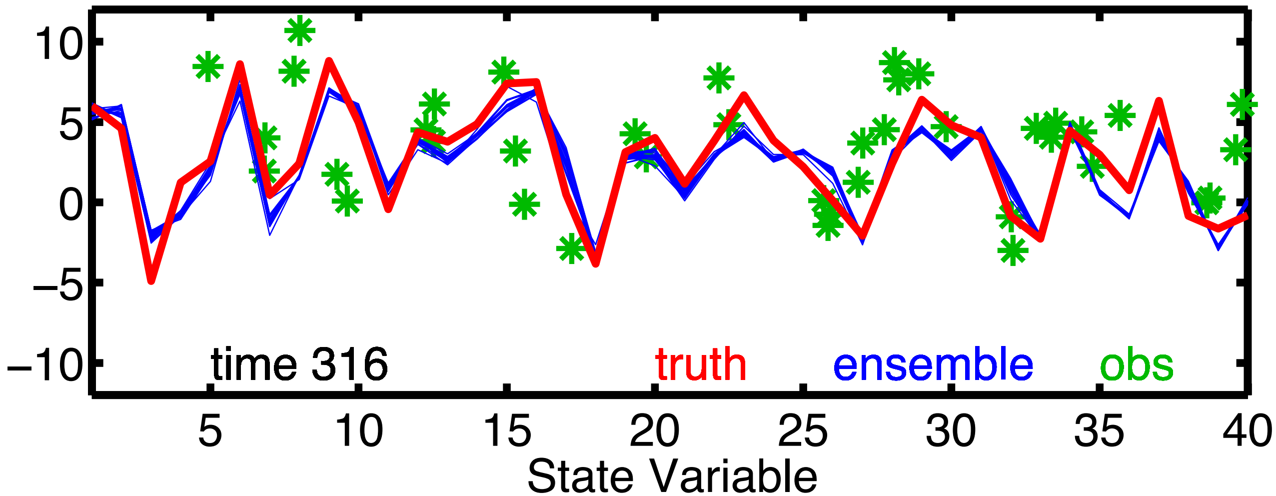
Assimilating in the presence of simulated model error.

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



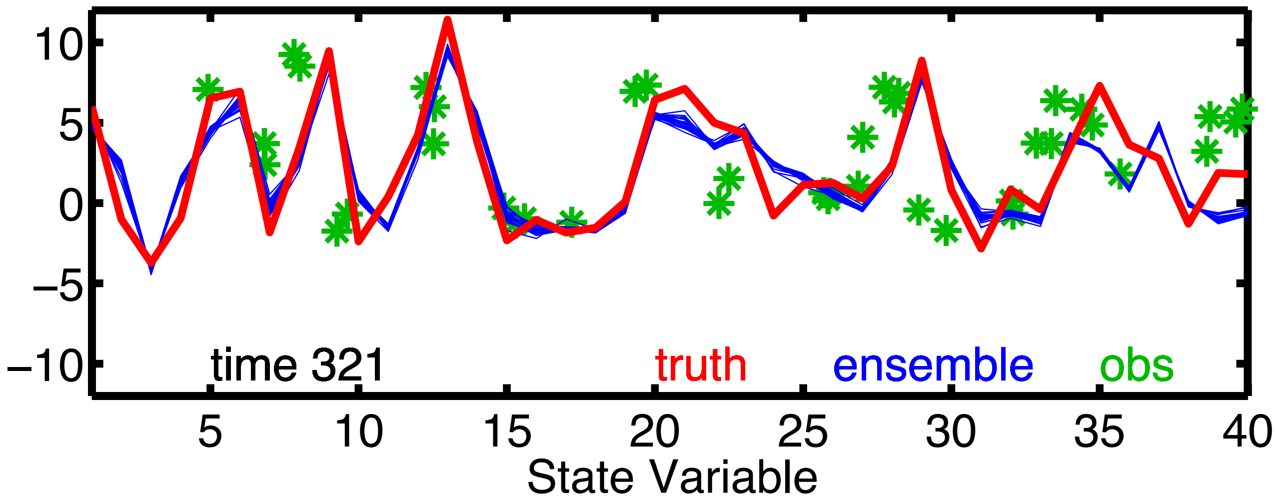
Assimilating in the presence of simulated model error.

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



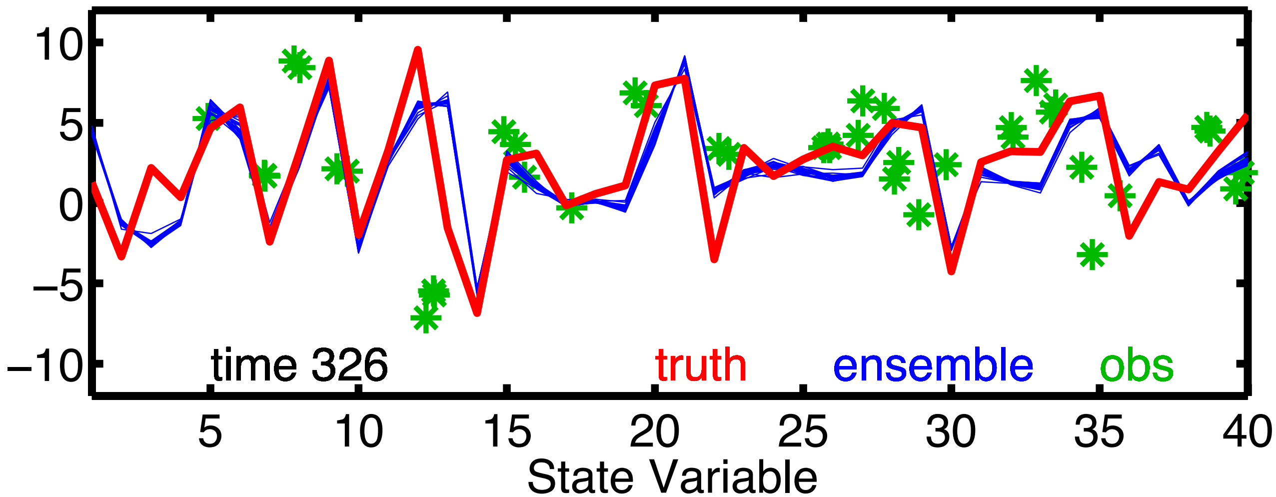
Assimilating in the presence of simulated model error.

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



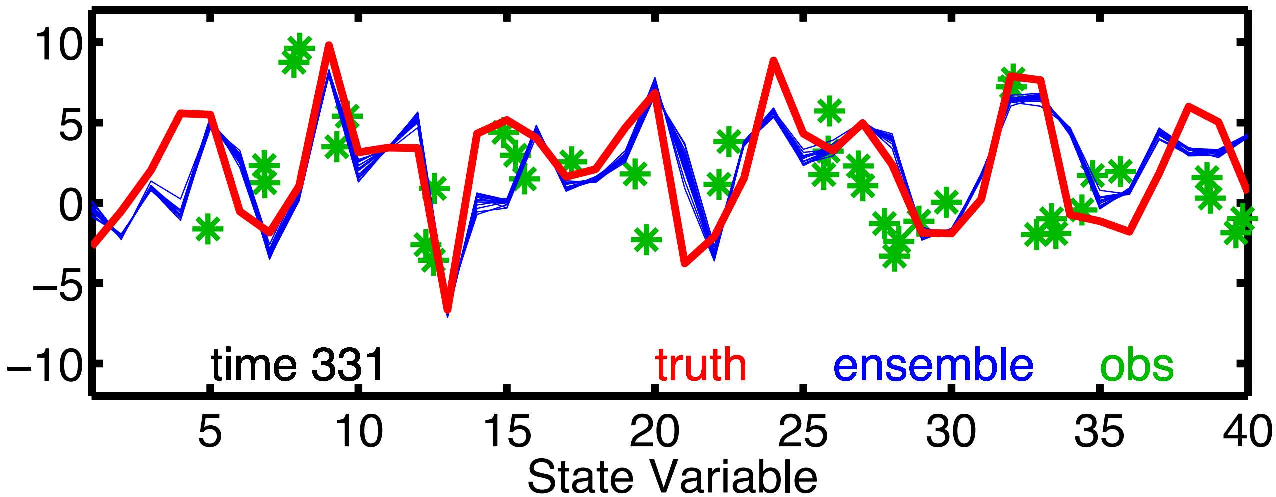
Assimilating in the presence of simulated model error.

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



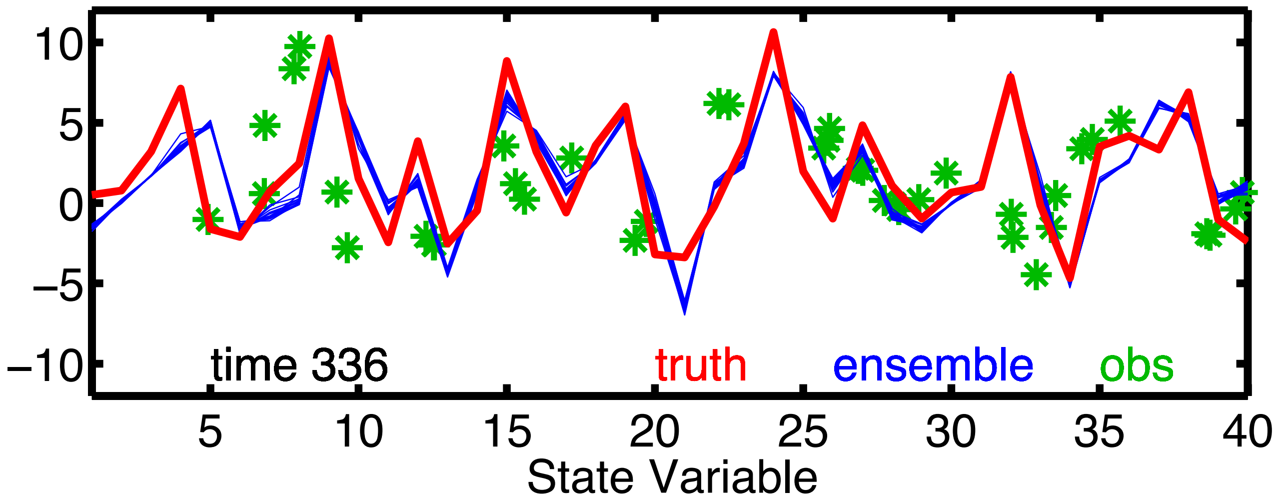
Assimilating in the presence of simulated model error.

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



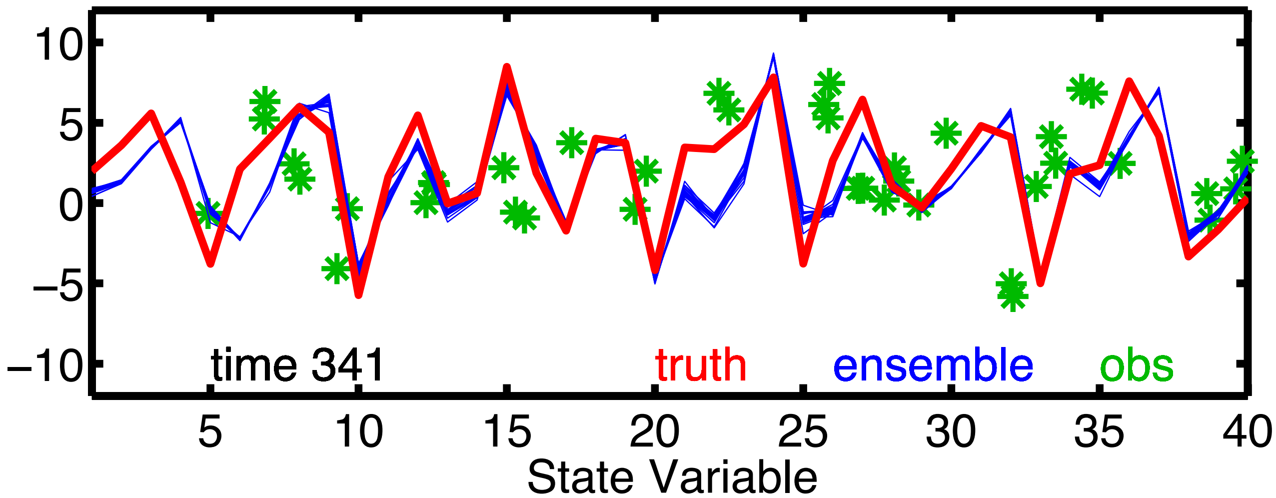
Assimilating in the presence of simulated model error.

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



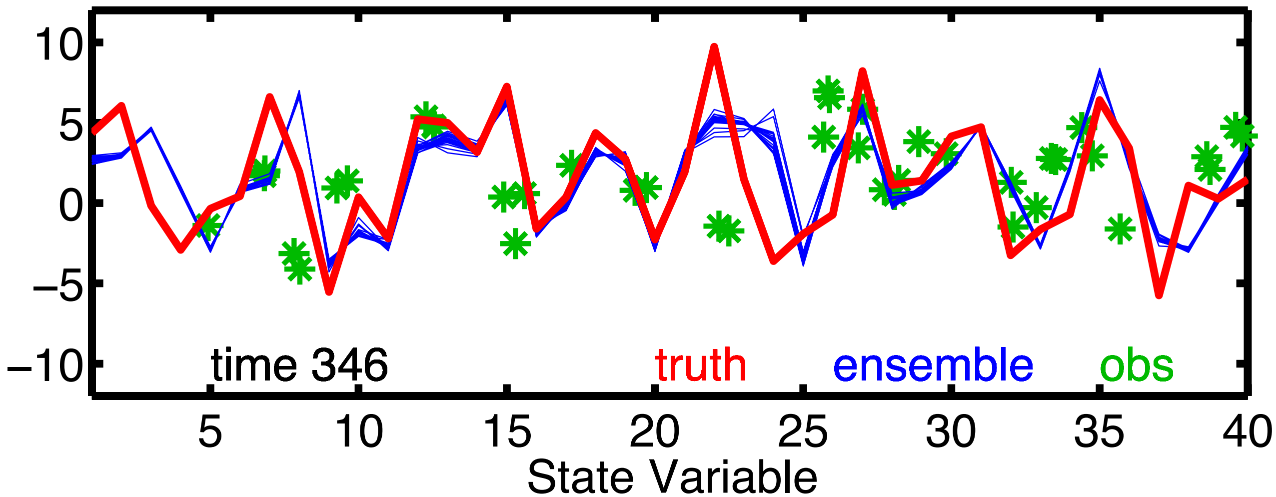
Assimilating in the presence of simulated model error.

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.

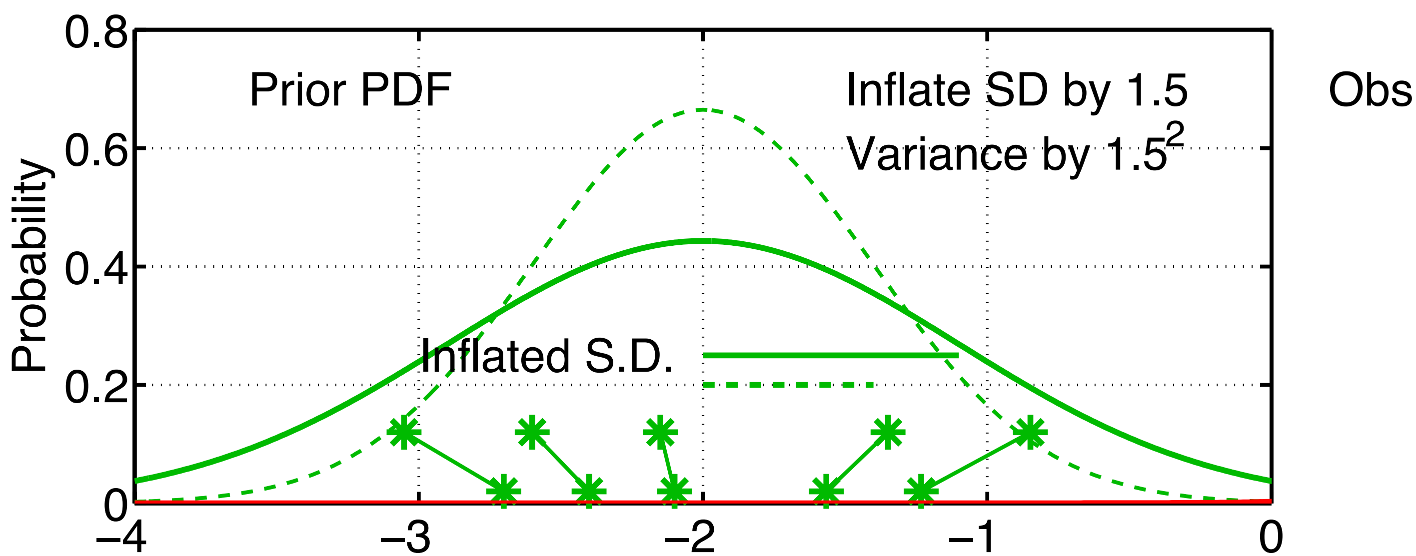
$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F$.
For truth, use F = 8.
In assimilating model, use F = 6.



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

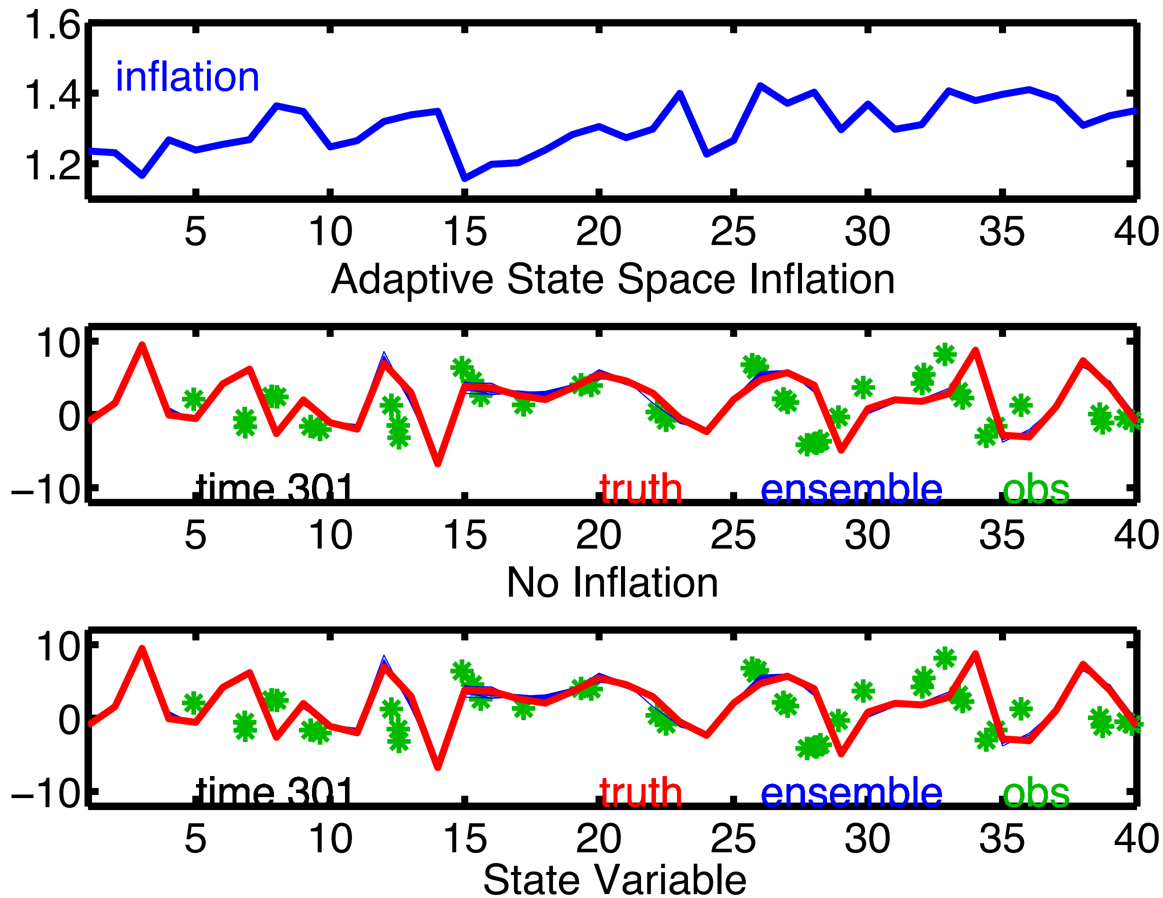
Reduce confidence in model prior to deal with model error

Use inflation.
Simply increase prior ensemble variance for each state variable.
Adaptive algorithms use observations to guide this.



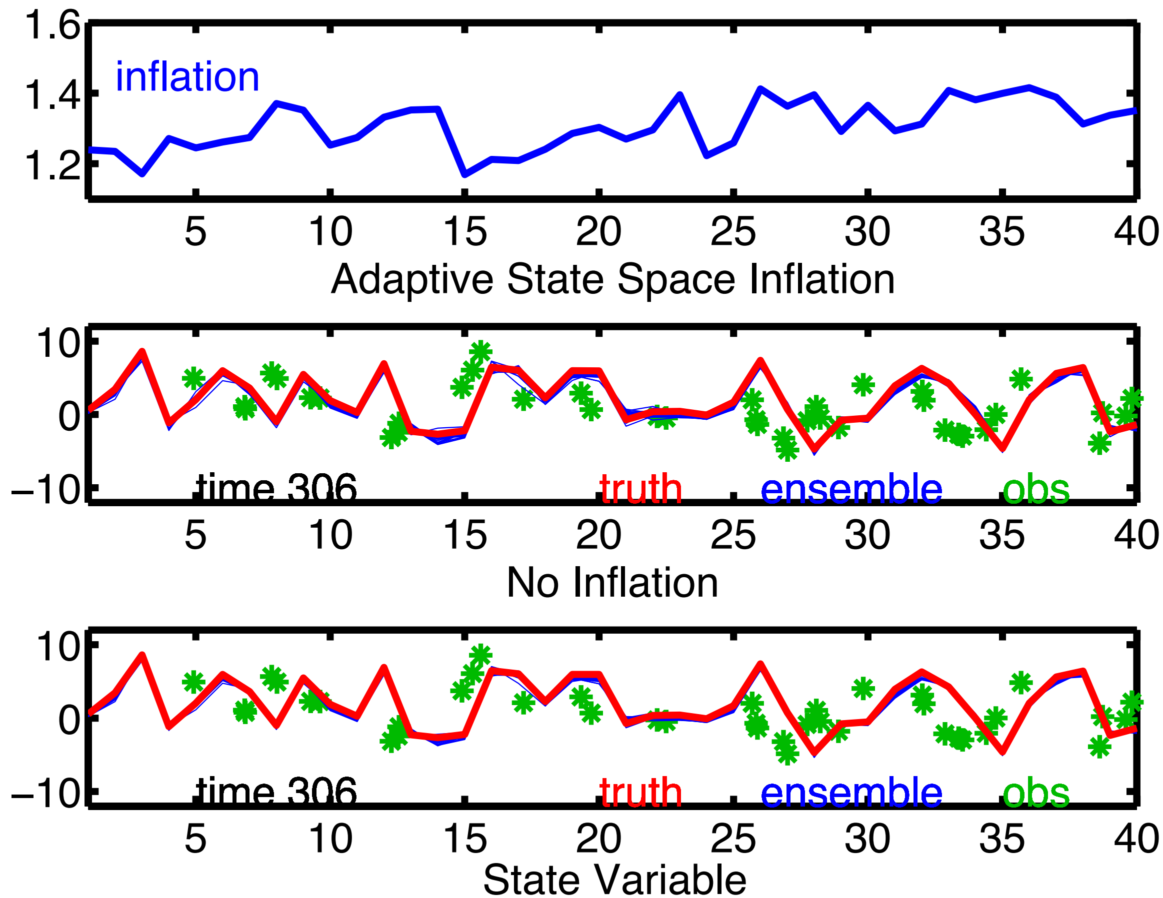
Assimilating with Inflation in presence of model error.

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



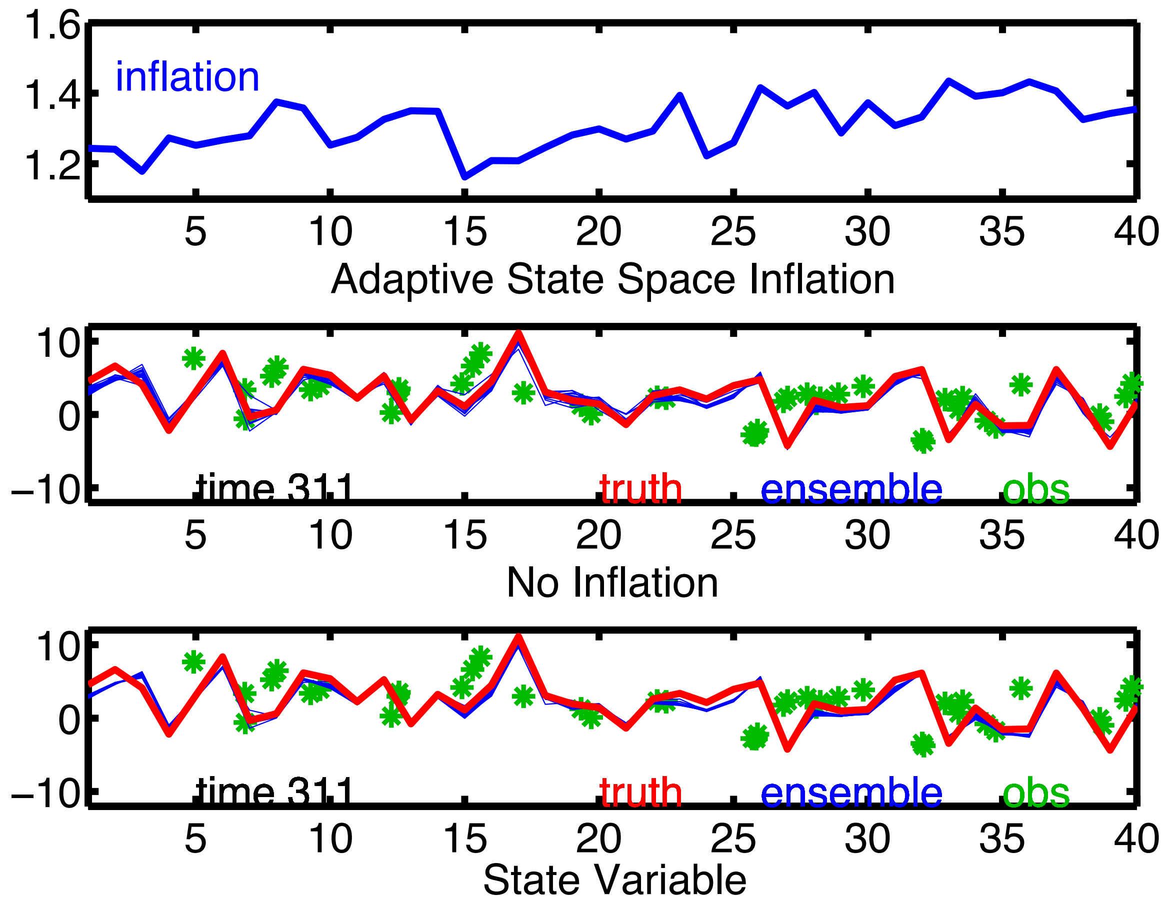
Assimilating with Inflation in presence of model error.

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



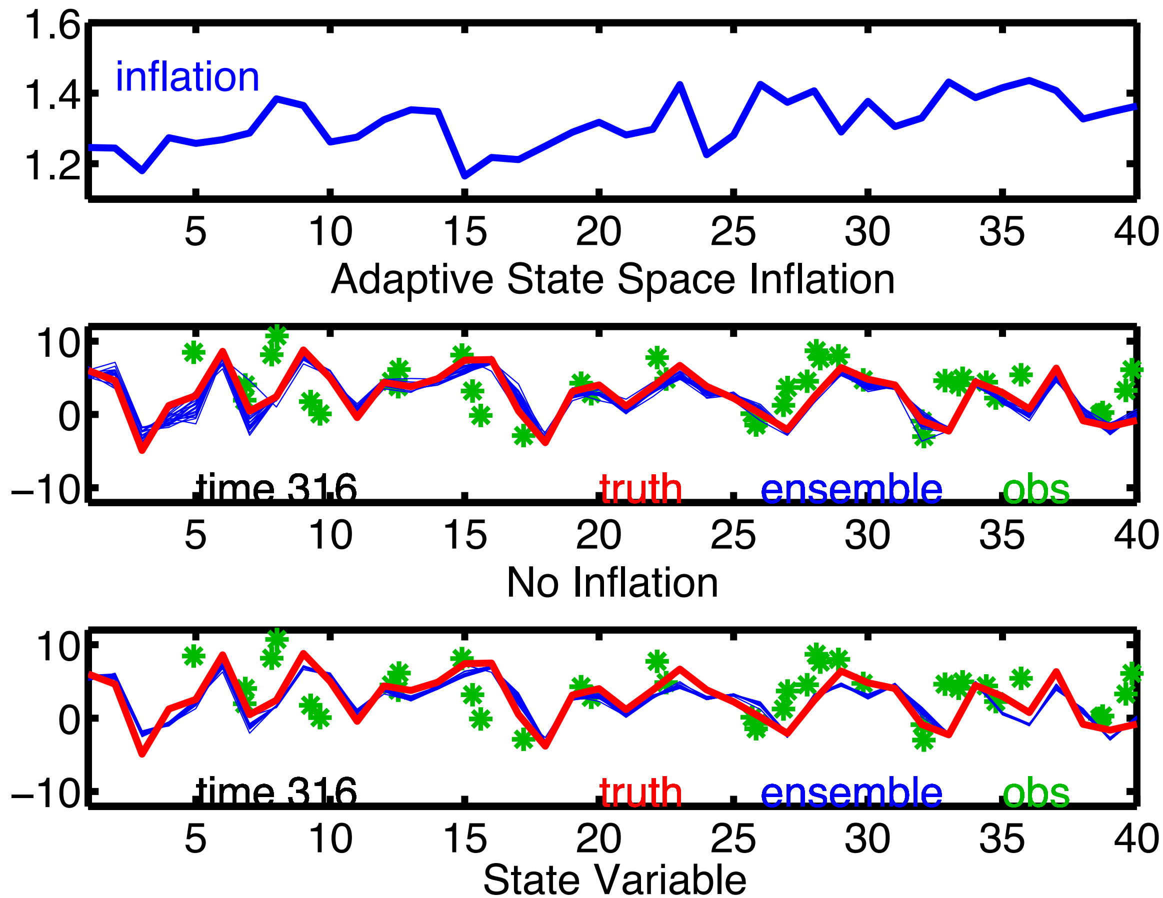
Assimilating with Inflation in presence of model error.

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



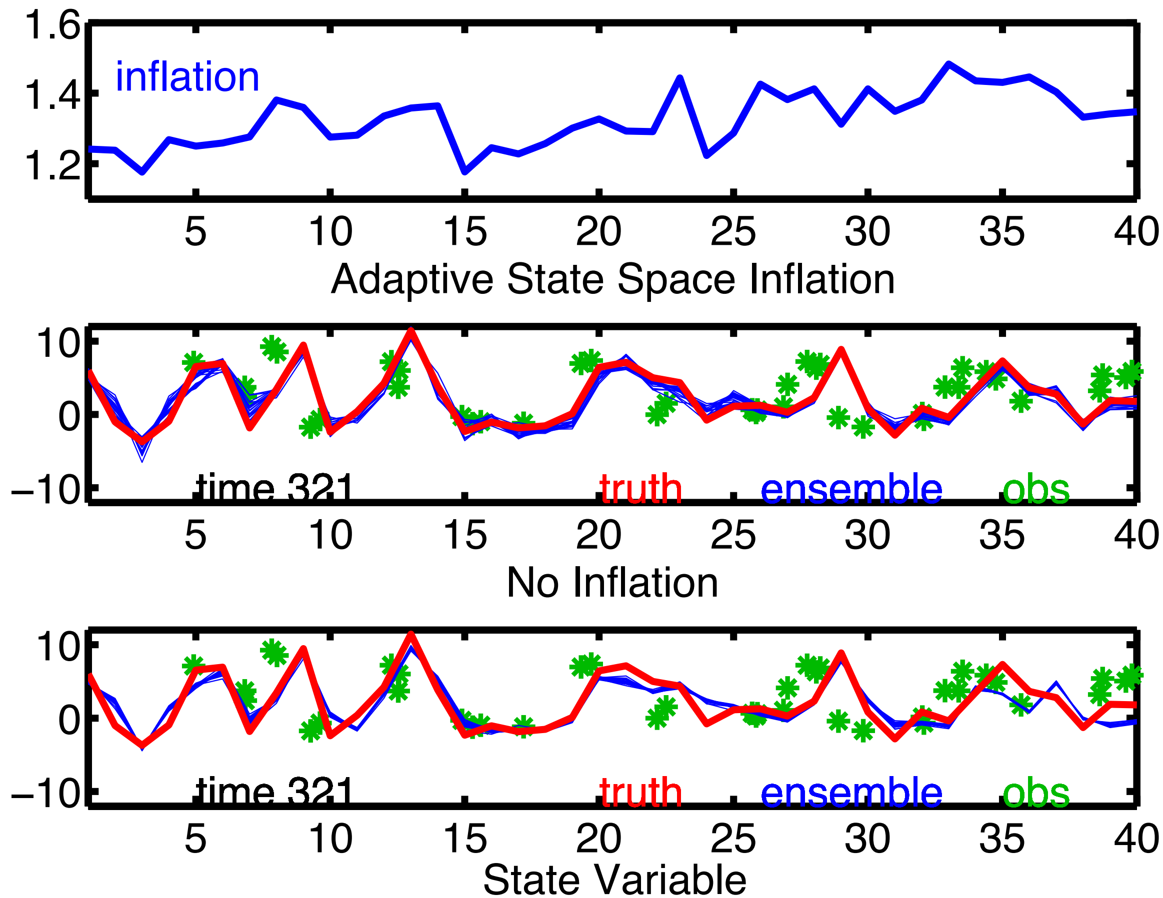
Assimilating with Inflation in presence of model error.

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



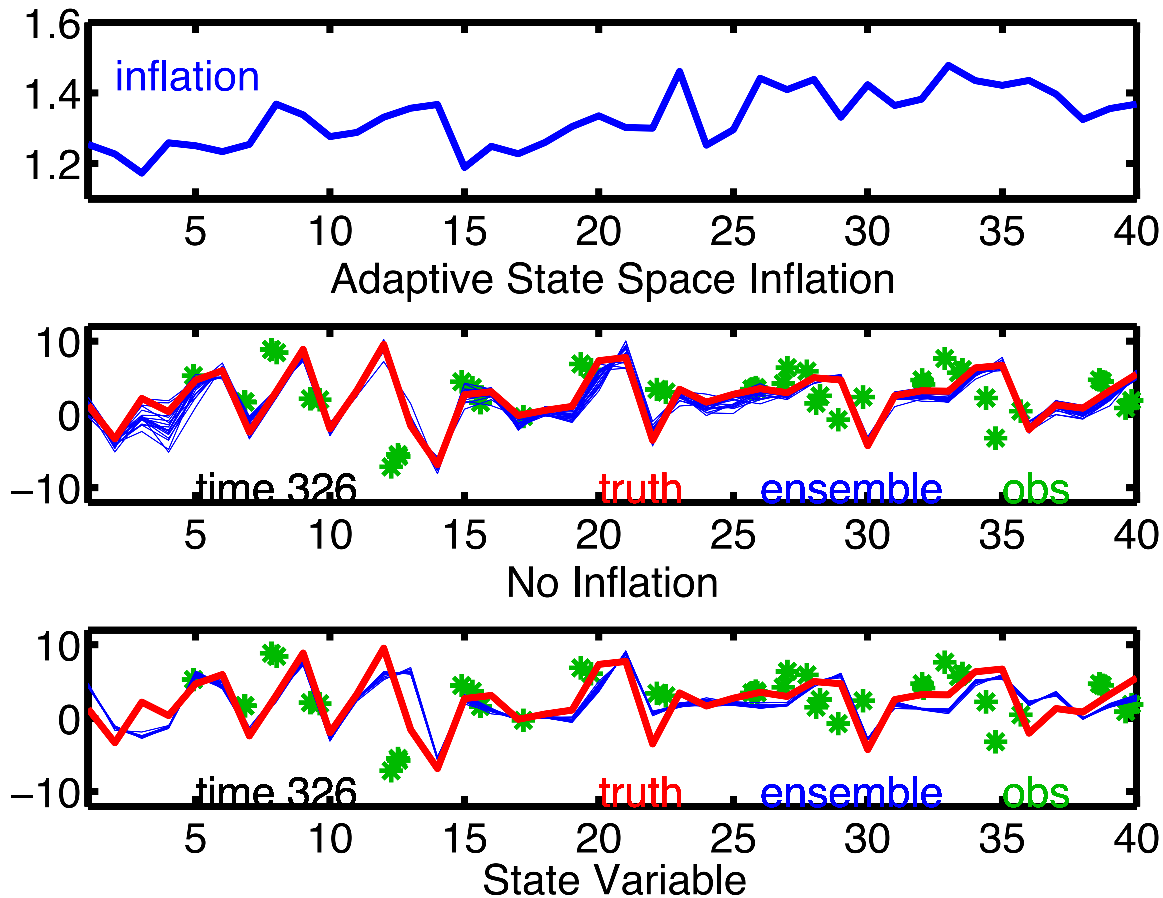
Assimilating with Inflation in presence of model error.

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



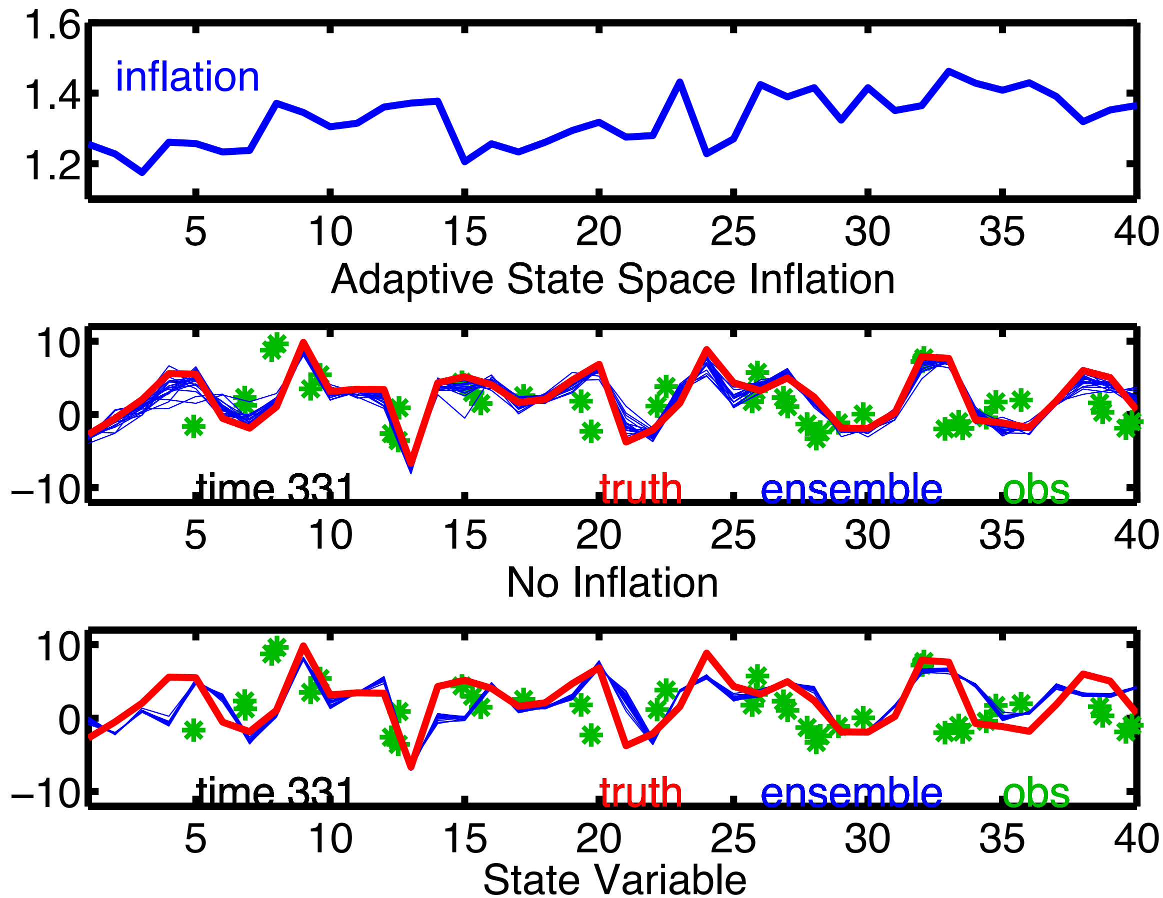
Assimilating with Inflation in presence of model error.

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



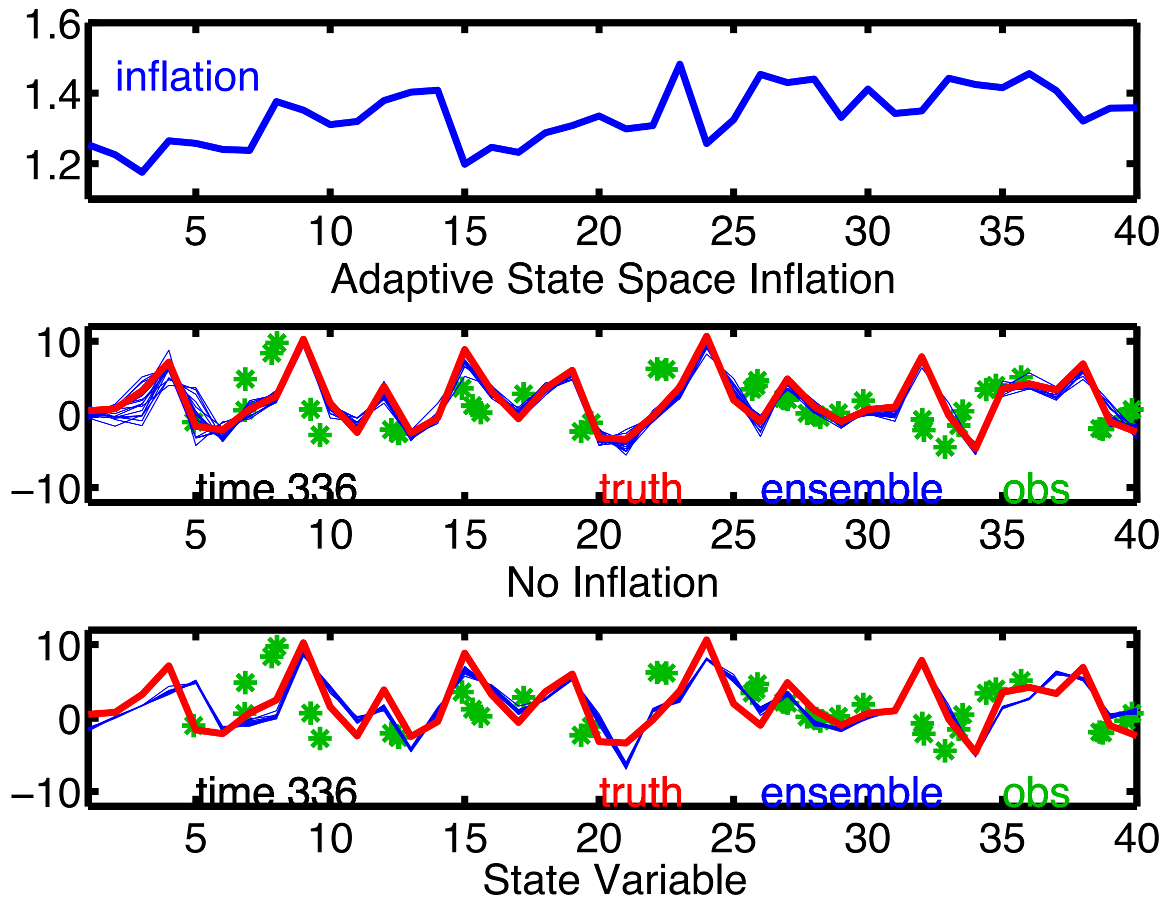
Assimilating with Inflation in presence of model error.

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



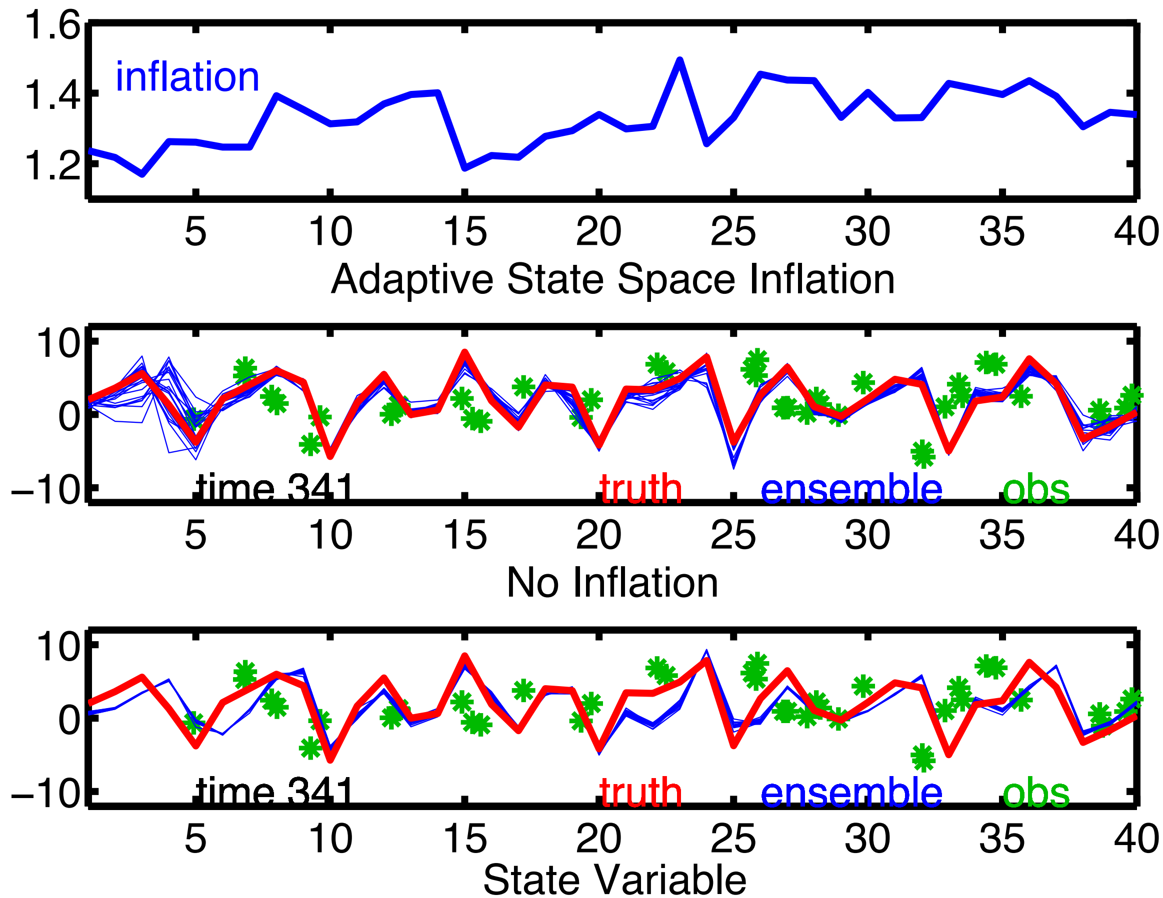
Assimilating with Inflation in presence of model error.

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



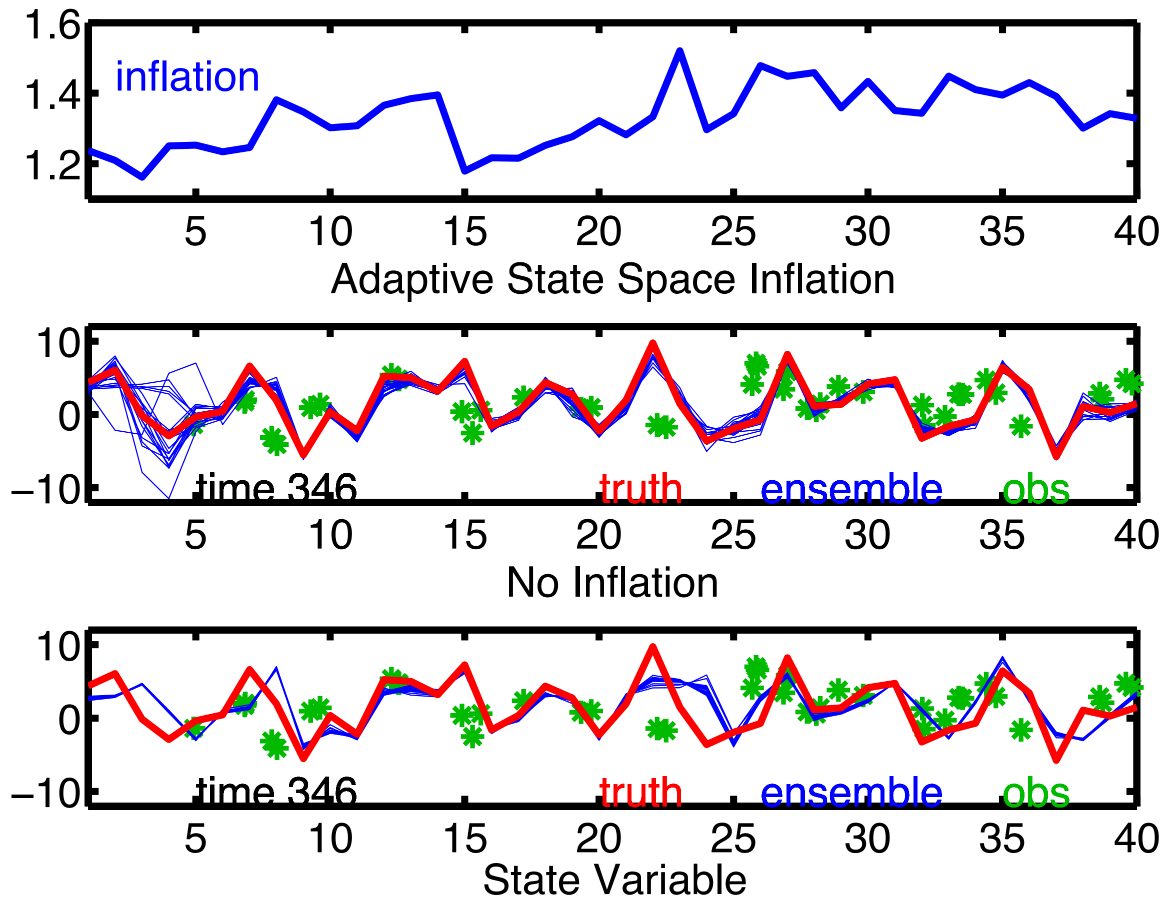
Assimilating with Inflation in presence of model error.

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.



Uncertainty Quantification from an Ensemble Kalman Filter

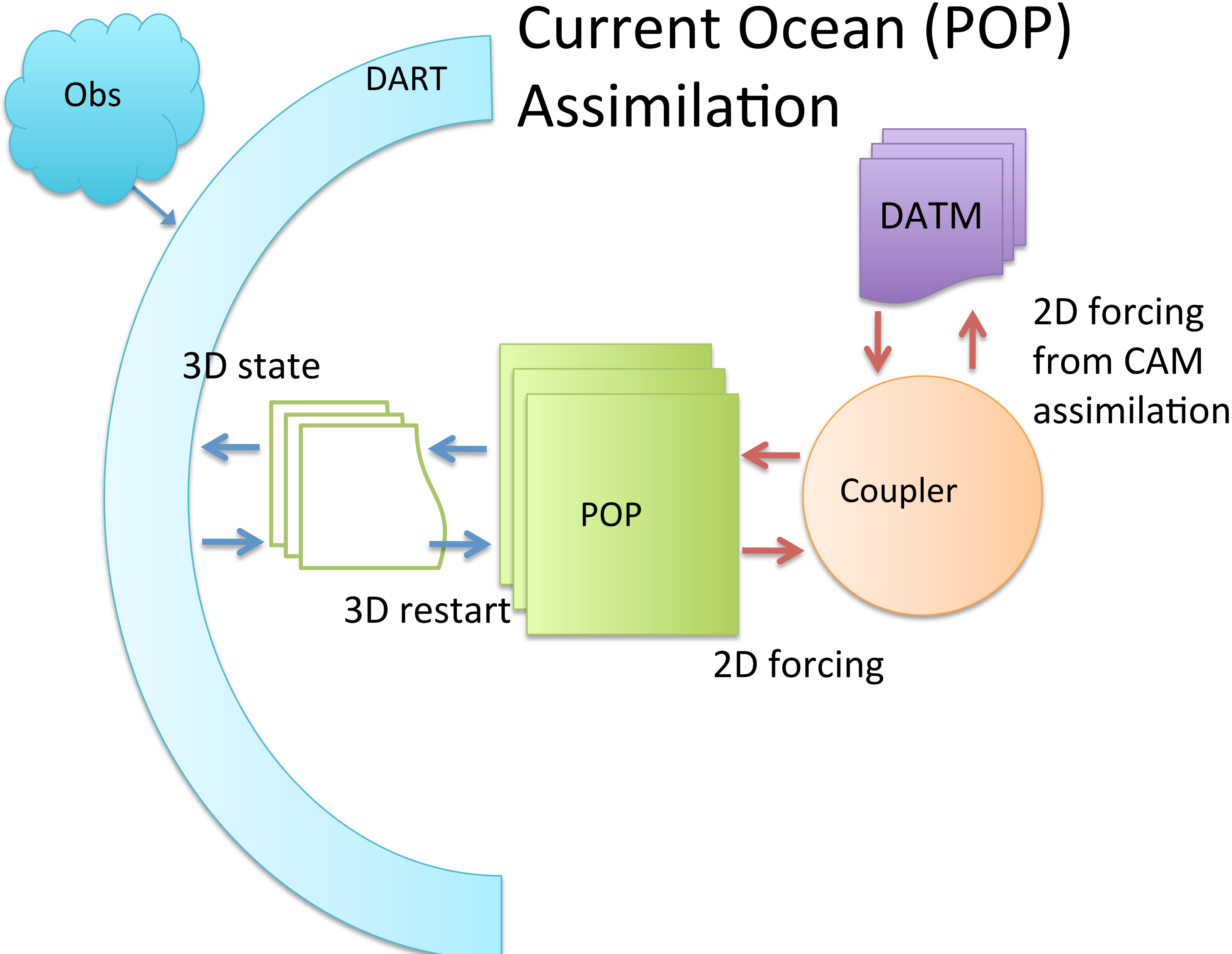
- (Ensemble) KF optimal for linear model, gaussian likelihood, perfect model.
- In KF, only mean and covariance have meaning.

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

- What do they mean when there are all sorts of error? Even less clear.

- Must Calibrate and Validate results.

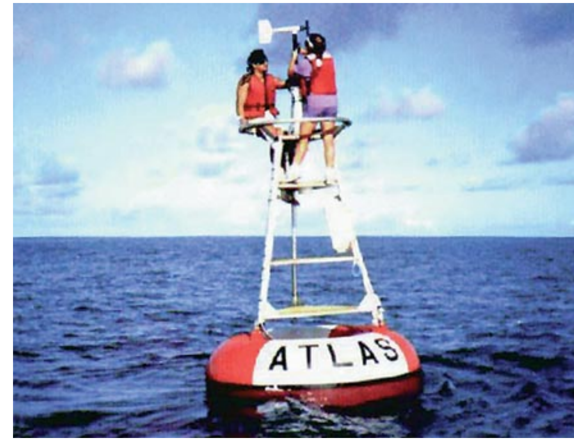
Current Ocean (POP) Assimilation



World Ocean Database T,S observation counts

These counts are for 1998 & 1999 and are representative.

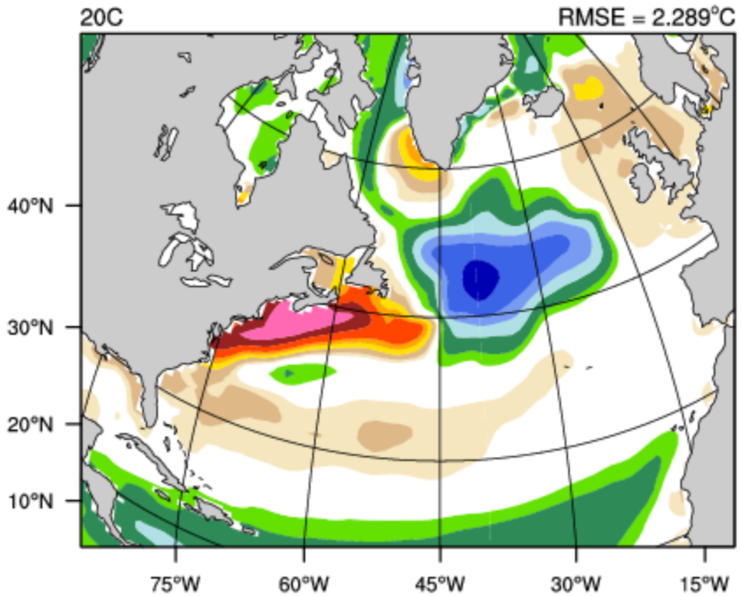
FLOAT_SALINITY	68200
FLOAT_TEMPERATURE	395032
DRIFTER_TEMPERATURE	33963
MOORING_SALINITY	27476
MOORING_TEMPERATURE	623967
BOTTLE_SALINITY	79855
BOTTLE_TEMPERATURE	81488
CTD_SALINITY	328812
CTD_TEMPERATURE	368715
STD_SALINITY	674
STD_TEMPERATURE	677
XCTD_SALINITY	3328
XCTD_TEMPERATURE	5790
MBT_TEMPERATURE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

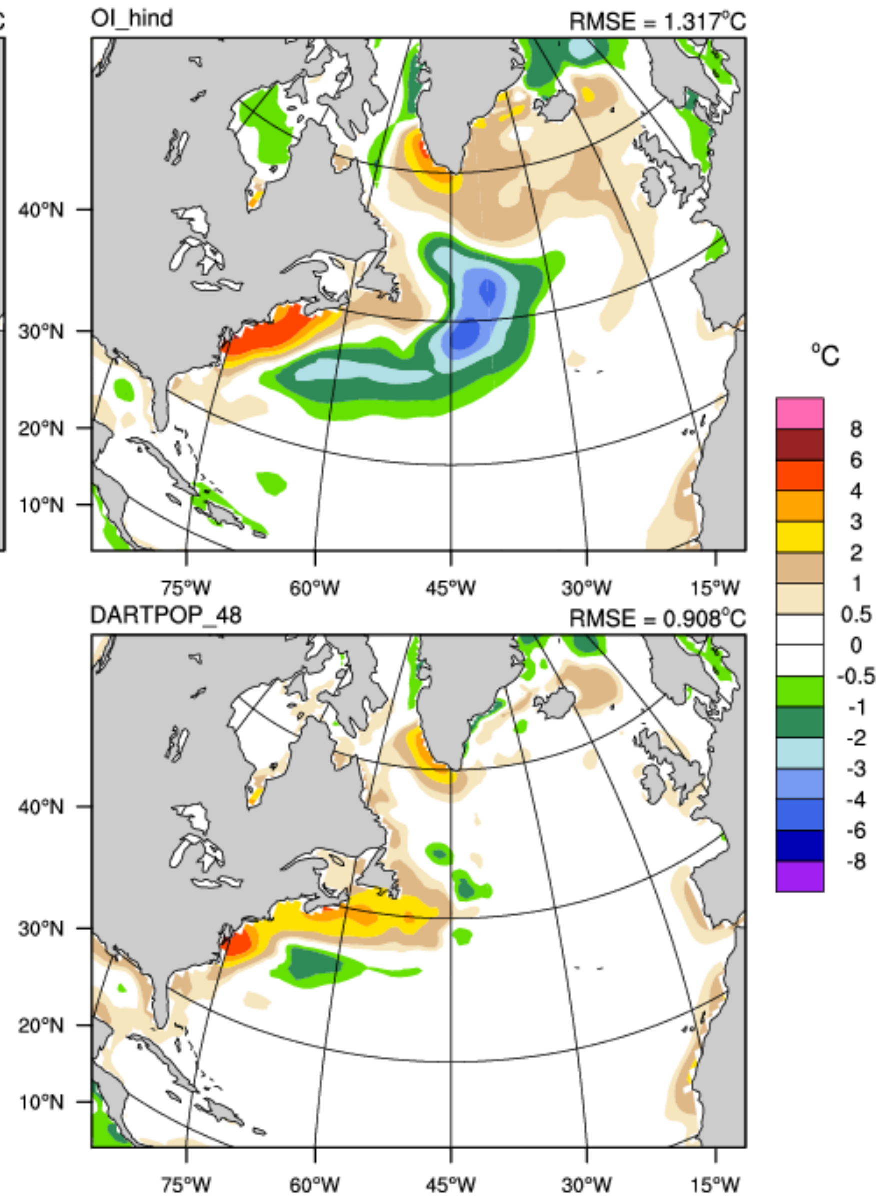


- temperature observation error standard deviation == 0.5 K.
- salinity observation error standard deviation == 0.5 msu.

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

Coupled Free Run





POP forced by observed atmosphere (hindcast)

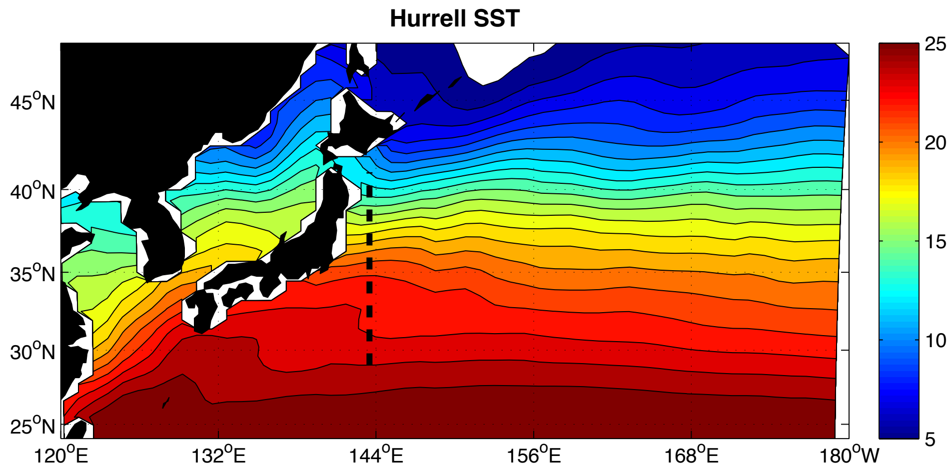
Ensemble Assimilation

48 POP oceans

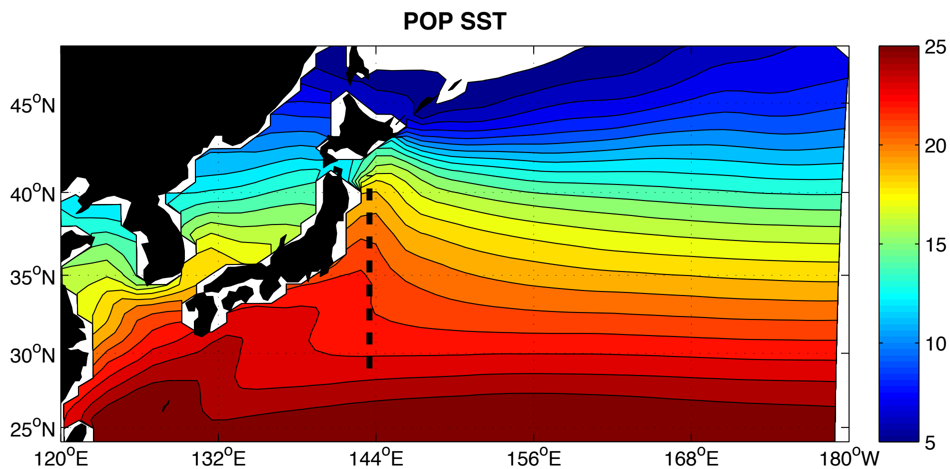
Forced by 48 CAM reanalyses

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

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



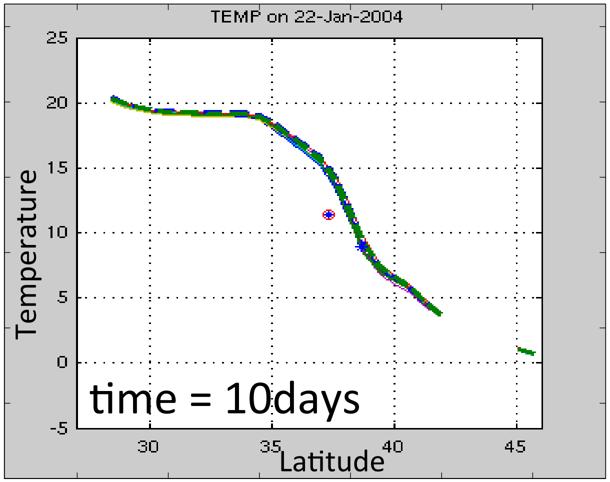
Regarded to be accurate.

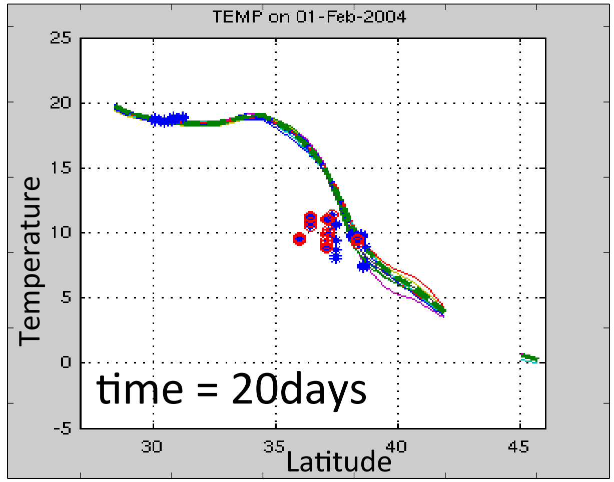


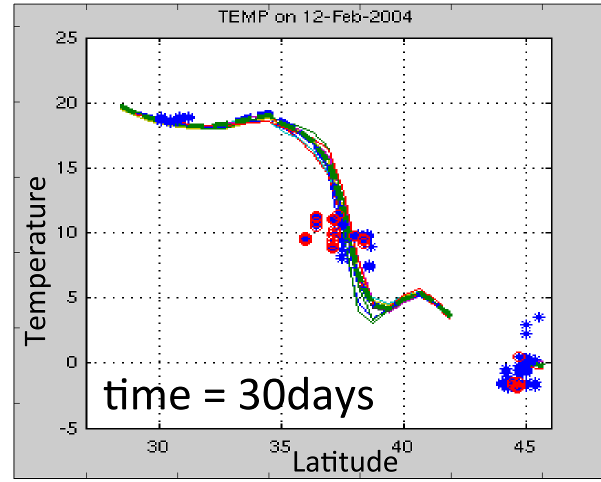
Free run of POP, the warm water is too far North.

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.

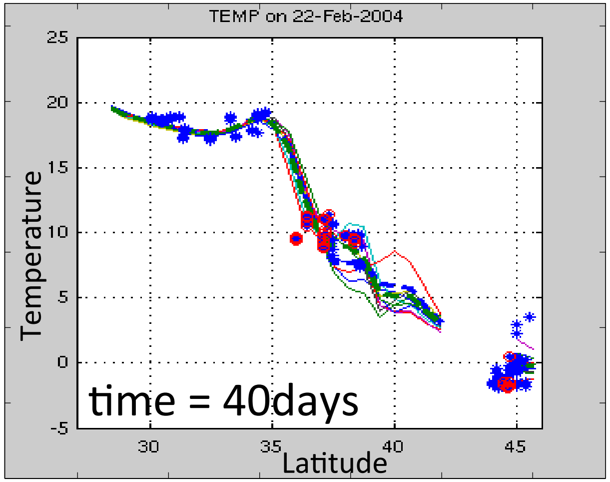
Challenges in correcting position of Kuroshio.

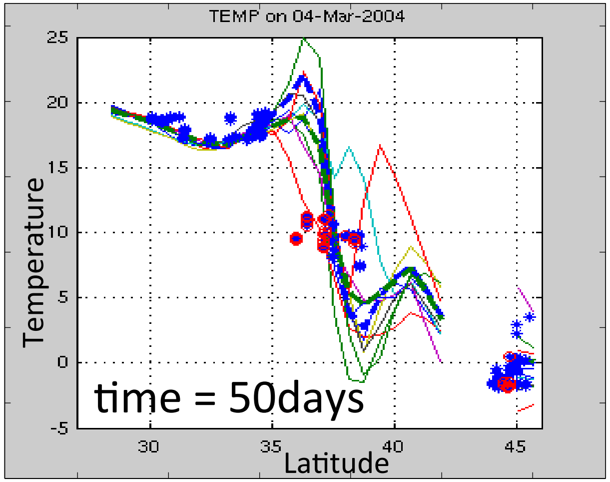
60-day assimilation starting from model climatology on 1 January 2004.

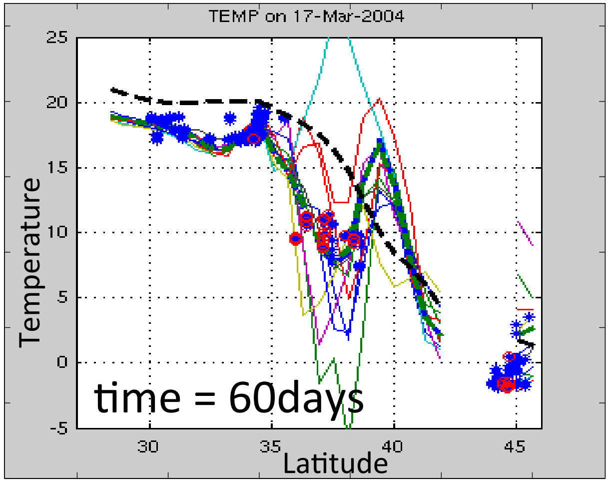
Green dashed line is posterior at previous time,
Blue dashed line is prior at current time,
Ensembles are thin lines.

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.



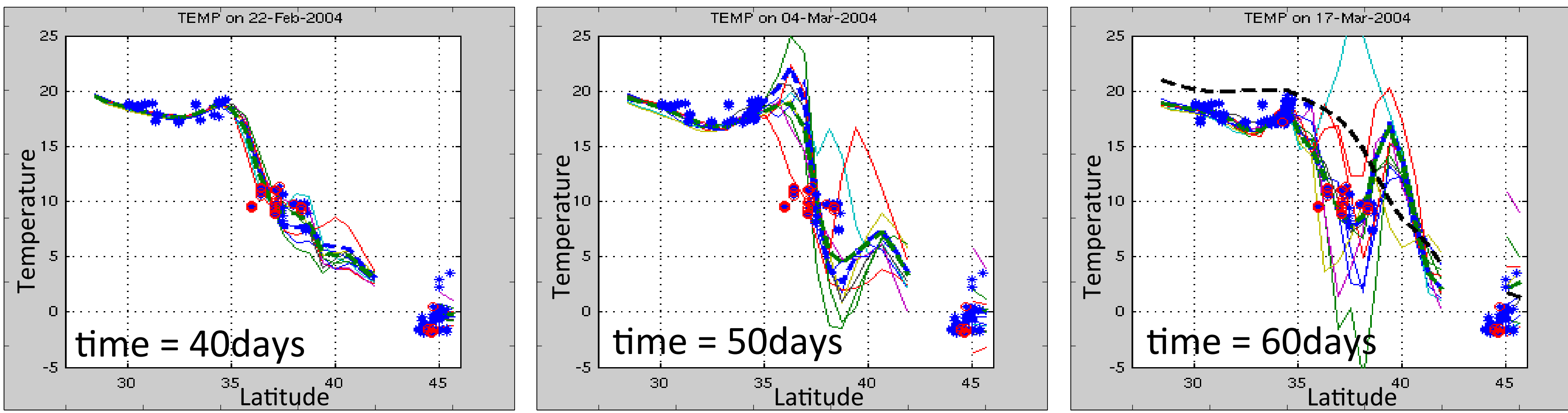




Challenges in correcting position of Kuroshio.

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

- Assimilation cannot force model to fit observations.
- Use of adaptive inflation leads to eventual model failure.
- Reduced adaptive inflation can lead to compromise between observations and model.



Ensemble DA and UQ for Global Ocean Models

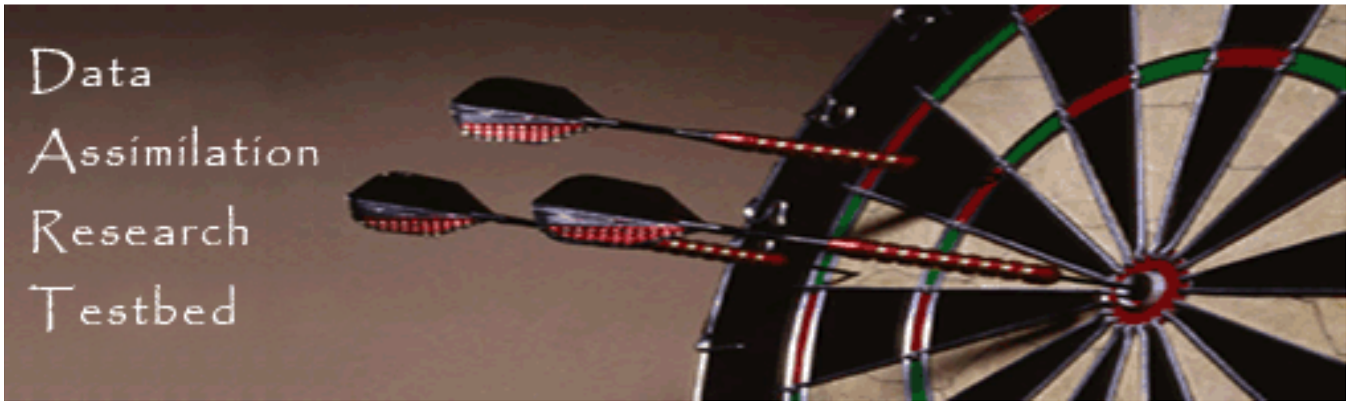
- Very certain that model predictions are different from observations.
- Very certain that small correlations have large errors.
- Moderately confident that large correlations are ‘realistic’.
- Very uncertain about state in sparse/unobserved regions.

- 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.

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



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