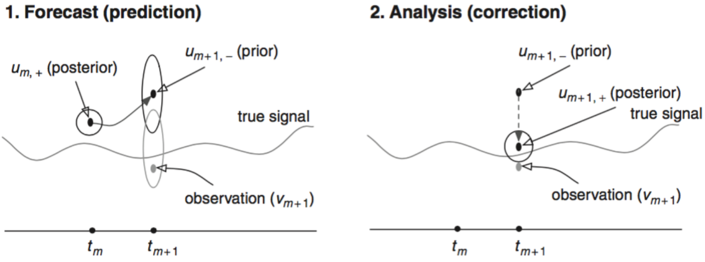
Winter 2019 Math 106
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Issues in High-Dimensional Assimilation

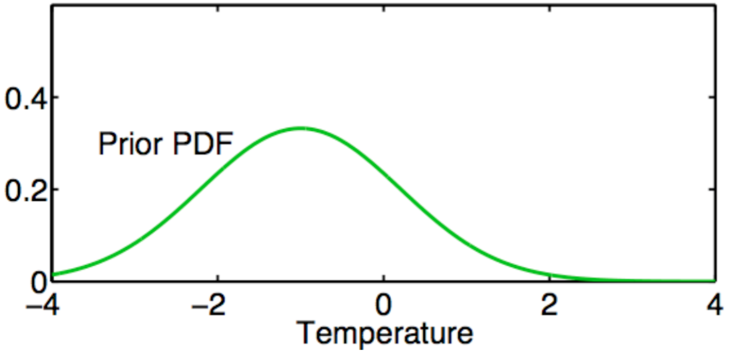
Data Assimilation

Data assimilation (or filtering) provides the best statistical estimate of the true signal from 1) **(mathematical) model prediction** and 2) **(noisy) observation**

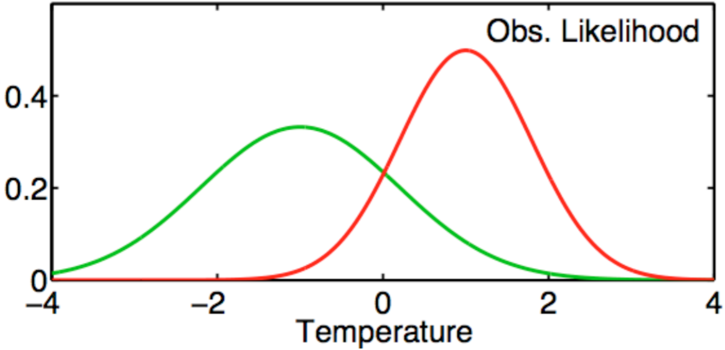


Bayesian Update

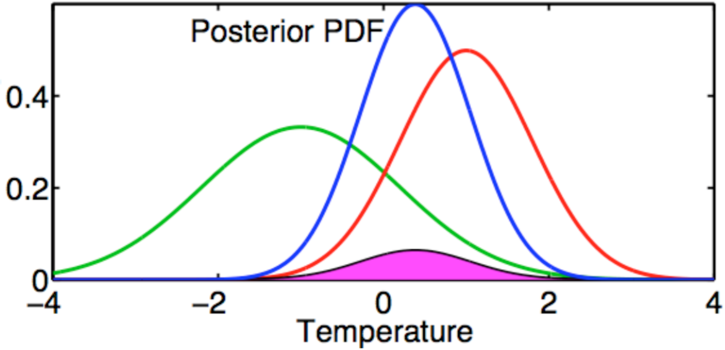
An example of improving prediction using data



An example of improving prediction using data



An example of improving prediction using data



Kalman filter

For linear systems, Kalman filter provides the linear quadratic estimation of unknown signal.
Applications : navigation and control, time series analysis, signal processing, robotic motions and etc.

- Prediction from forecast x^f with variance r^f
- observation y with variance r^o
- Estimation x^a with variance r^a

$$
\begin{aligned}
x^a &= x^f + K(y - x^f) & (1) \\
r^a &= (1 - K)r^f & (2)
\end{aligned}
$$

where K is the Kalman gain

$$K = \frac{r^f}{r^o + r^f}$$

Nonlinear systems

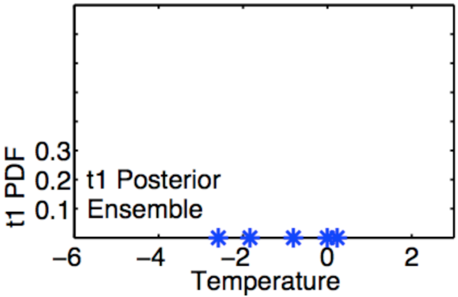
How we obtain the current forecast variance from the previous posterior variance? That is, how we propagate the variance in time?

- For linear systems, there is a simple method to calculate the change in the variance.
- For nonlinear systems, no simple way.

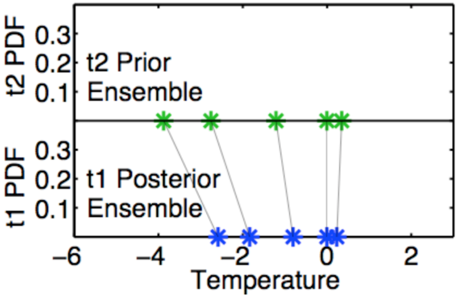
There are several approaches for nonlinear systems. The most obvious approach is the Monte Carlo approach which uses several samples to estimate the uncertainty.

The more samples we have, the more accurate estimation we obtain.

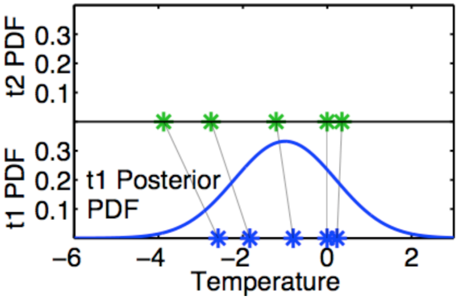
From posterior ensemble at t_1



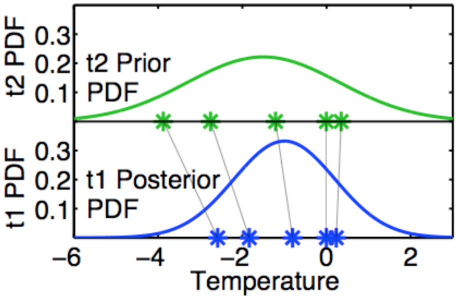
Advance each member to time t_2

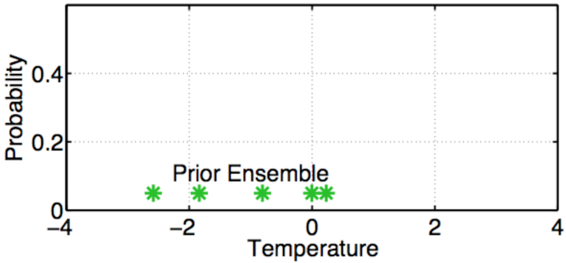


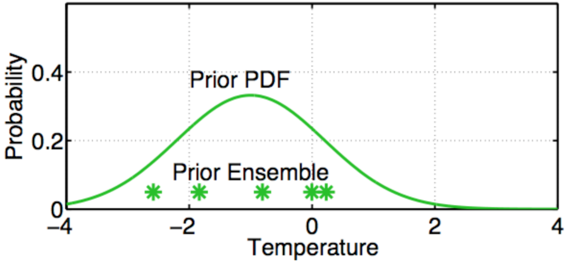
Same as advancing continuous PDF at t_1



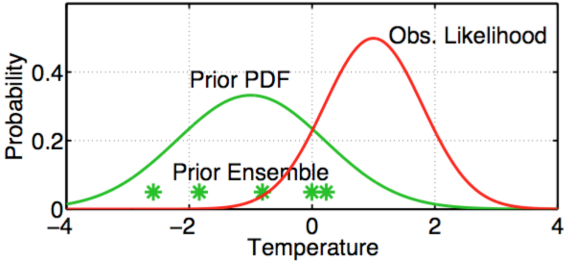
Same as advancing continuous PDF at t_1 to t_2



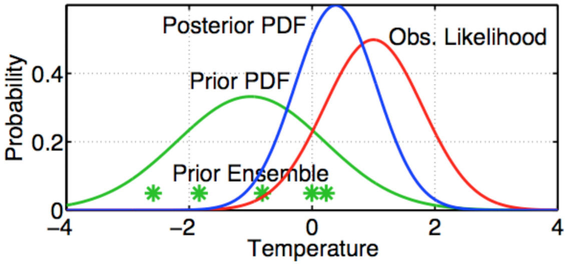

Probability
0.4
0.2
0
Prior Ensemble
–4
–2
0
2
4
Temperature



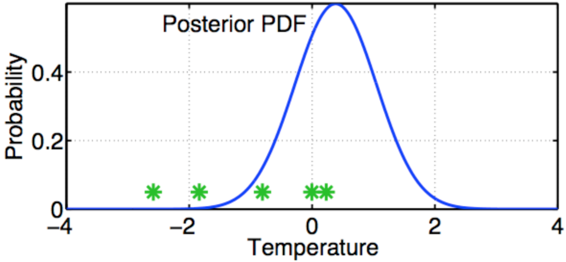
Fit a Gaussian to the sample.



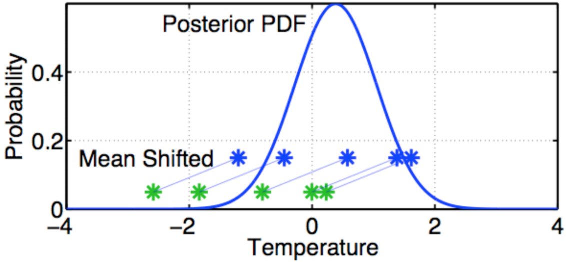
Get the observation likelihood.



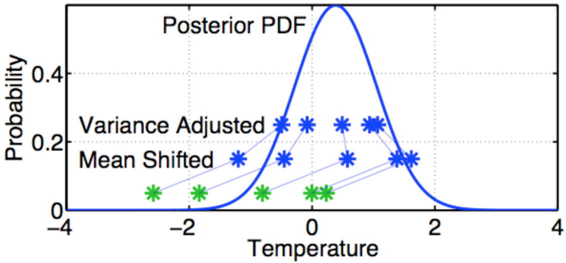
Compute the continuous posterior PDF.



Use a deterministic algorithm to adjust the ensemble.



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

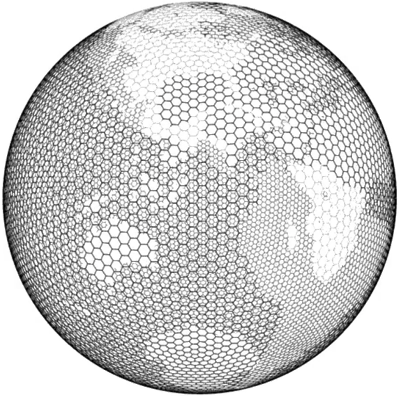


First, shift the ensemble to have the exact mean of the posterior. Second, linearly contract to have the exact variance of the posterior. Sample statistics are identical to Kalman filter.

High-dimensional state variable $\mathbf{x}$

Typical problems in science and engineering contains state variables in high-dimensional spaces.

For example, in a simulation of the earth system with a space of 10 km, there are more than 5,000,000 state variables (considering only surface)



Problems in the high-dimensional state estimation and prediction

- Huge computational resources to run forecast
  - ☐ parallel computations using High Performance Computing
  - ☐ reduced-order forecast models

Problems in the high-dimensional state estimation and prediction

- Huge computational resources to run forecast
  - ☐ parallel computations using High Performance Computing
  - ☐ reduced-order forecast models
- Inversion of matrices
- Spase observations

Inversion of a matrix

Bayesian update for high-dimensional spaces

- Prediction from forecast $\mathbf{x^f}$ with variance $\mathbf{R^f}$
- observation $\mathbf{y}$ with variance $\mathbf{R^o}$
- Estimation $\mathbf{x^a}$ with variance $\mathbf{R^a}$

$$\mathbf{x^a} = \mathbf{x^f} + \mathbf{K}(\mathbf{y} - \mathbf{x^f}) \tag{3}$$

$$\mathbf{r^a} = (\mathbf{I} - \mathbf{K})\mathbf{r^f} \tag{4}$$

where $\mathbf{K}$ is the Kalman gain

$$\mathbf{K} = \mathbf{R^f}(\mathbf{R^o} + \mathbf{R^f})^{-1}, \quad \text{matrix inversion!}$$

Sparsity of observations

It is very expensive to have many observations. Thus, in general, only very sparse observations are available compared to the full state variables.
In numerical weather prediction, for example, observations are dense over land but very sparse over ocean.

Sparsity of observations

It is very expensive to have many observations. Thus, in general, only very sparse observations are available compared to the full state variables.
In numerical weather prediction, for example, observations are dense over land but very sparse over ocean.

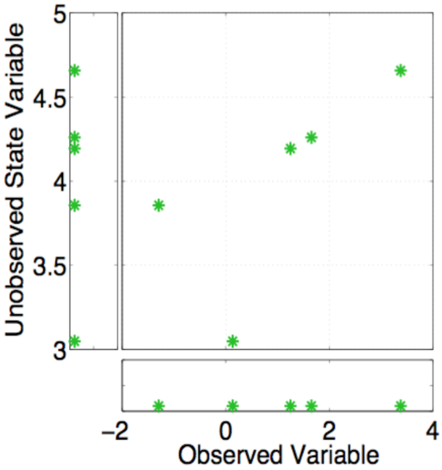
Question : How we handle unobserved variables?

Sparsity of observations

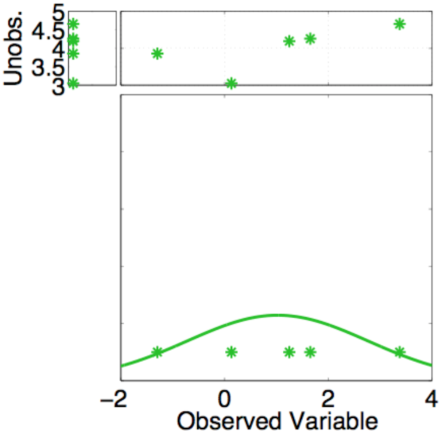
It is very expensive to have many observations. Thus, in general, only very sparse observations are available compared to the full state variables.
In numerical weather prediction, for example, observations are dense over land but very sparse over ocean.

Question : How we handle unobserved variables?

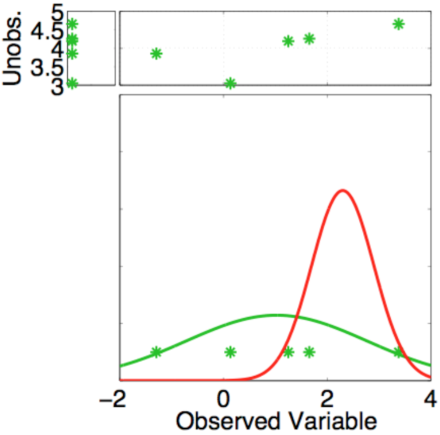
Answer : Use statistical information between the observed and unobserved variables, the correlation!



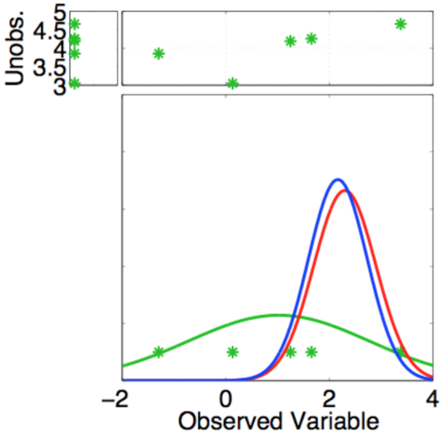
Assume that all we know is prior joint distribution. One variable is observed.



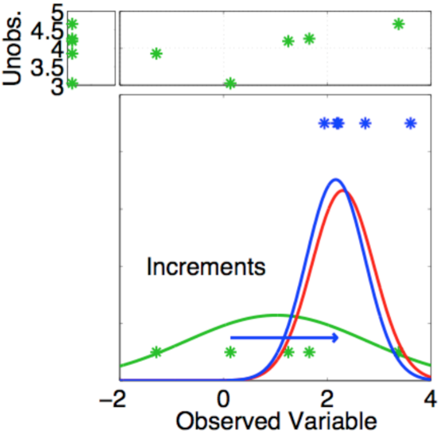
Update observed variable
with the previous method.



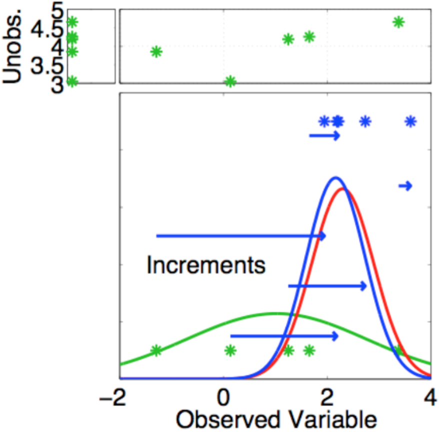
Update observed variable with the previous method.



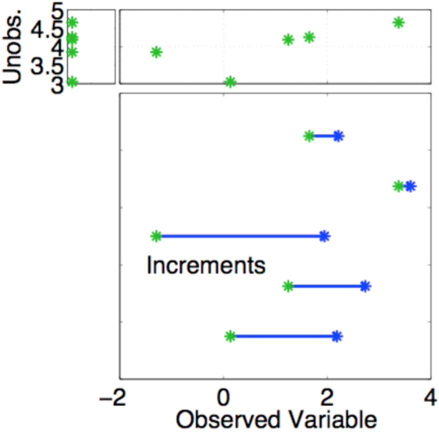
Update observed variable with the previous method.



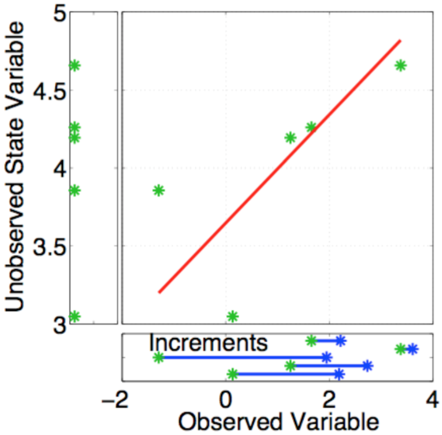
Compute increments for prior ensemble members of observed variable.



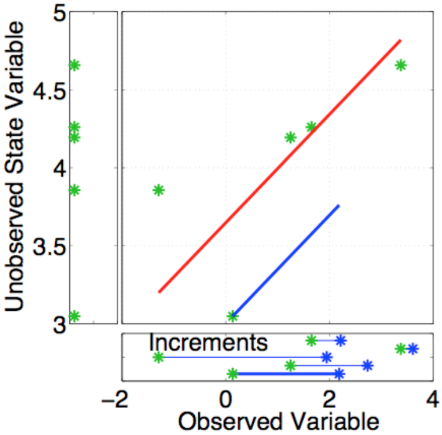
Compute increments for prior ensemble members of observed variable.



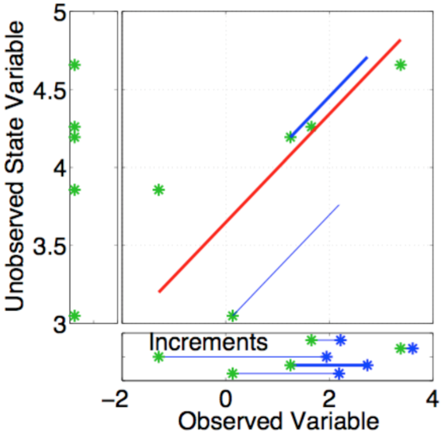
Using only increments guarantees that if observation had no impact on observed variable, unobserved variable is unchanged (high desirable).



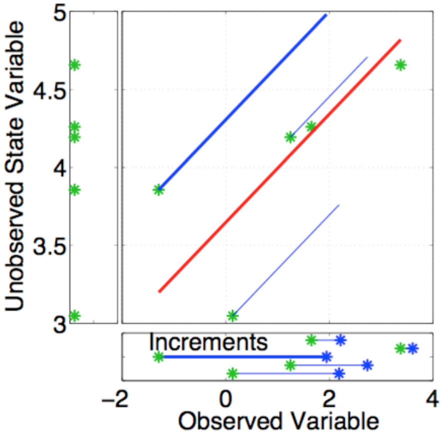
Use least squares for the unobserved variable. Equivalent to linear regression.



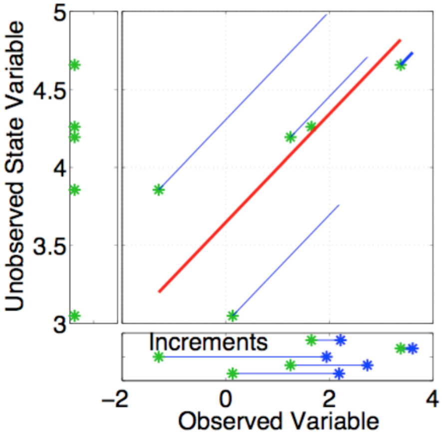
Next regress the observed variable increments onto increments for the unobserved variable. Equivalent to first finding image of increments in joint space.



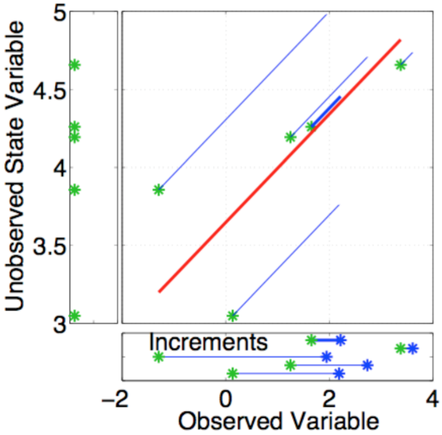
Next regress the observed variable increments onto increments for the unobserved variable. Equivalent to first finding image of increments in joint space.



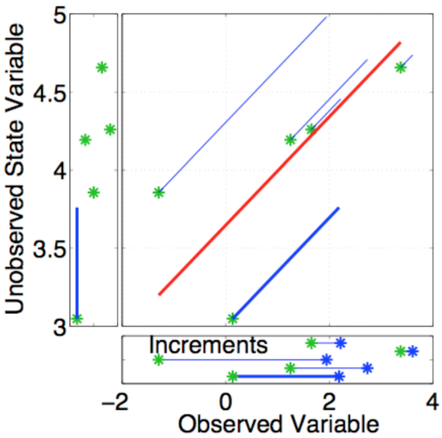
Next regress the observed variable increments onto increments for the unobserved variable. Equivalent to first finding image of increments in joint space.



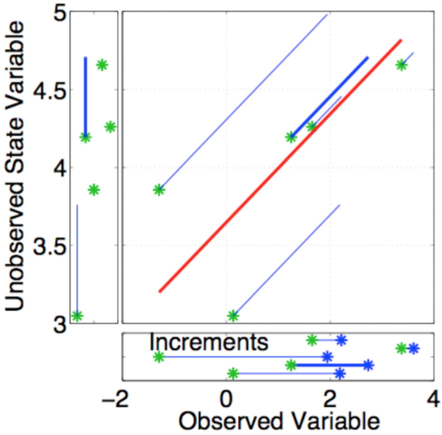
Next regress the observed variable increments onto increments for the unobserved variable. Equivalent to first finding image of increments in joint space.



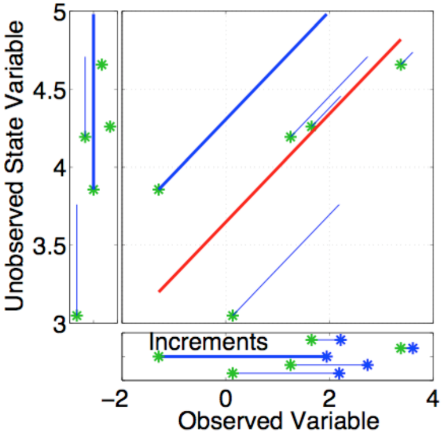
Next regress the observed variable increments onto increments for the unobserved variable. Equivalent to first finding image of increments in joint space.



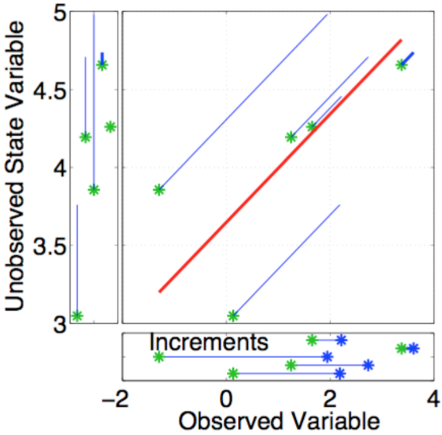
Then projecting from joint space onto unobserved priors.



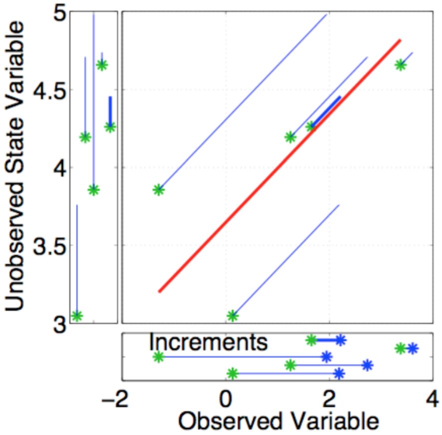
Then projecting from joint space onto unobserved priors.



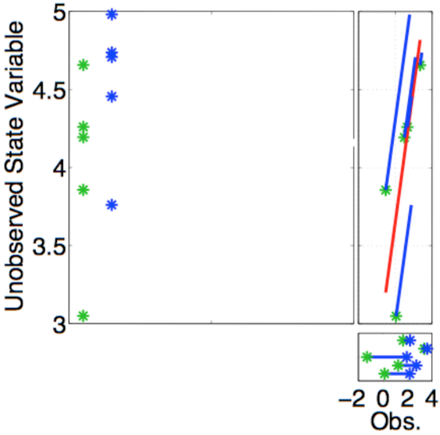
Then projecting from joint space onto unobserved priors.



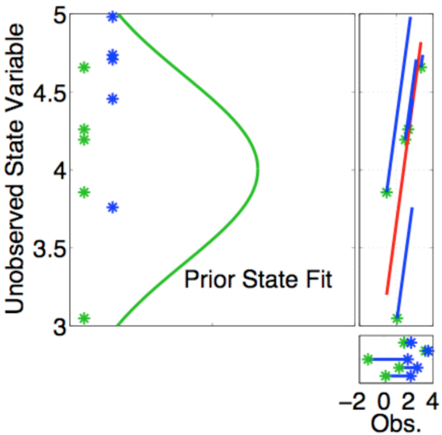
Then projecting from joint space onto unobserved priors.



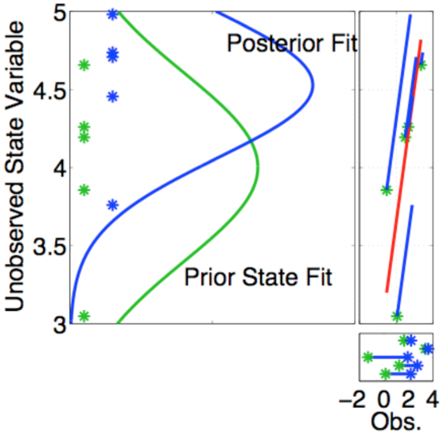
Then projecting from joint space onto unobserved priors.



Now have an updated ensemble for the unobserved variable.



Fitting Gaussians shows that mean and variance have changed.

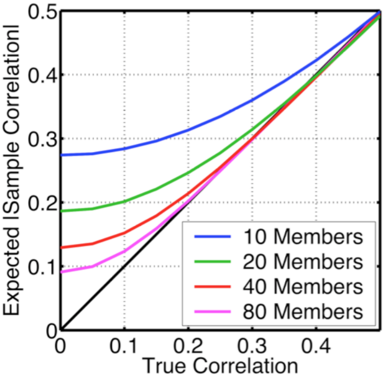


Other features of the prior distribution may also have changed.

Sampling error - spurious correlations

A small size of samples makes spurious correlations.

a Matlab test

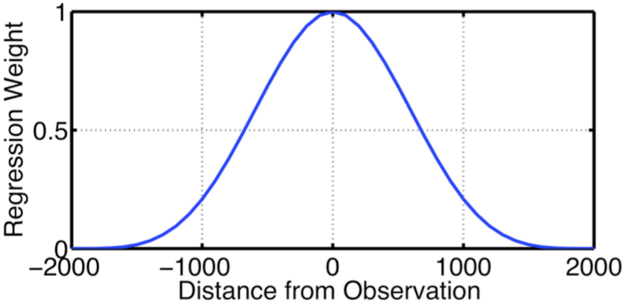


Question : Is it possible to handle the spurious correlation without increasing the number of samples?

Question : Is it possible to handle the spurious correlation without increasing the number of samples?
Hint : Adjacent variables have high correlations while distant variables have small or zero correlations.

Question : Is it possible to handle the spurious correlation without increasing the number of samples?
Hint : Adjacent variables have high correlations while distant variables have small or zero correlations.
Answer : Localization! Rescale the sampled correlation based on the distance from observations.

Question : Is it possible to handle the spurious correlation without increasing the number of samples?
Hint : Adjacent variables have high correlations while distant variables have small or zero correlations.
Answer : Localization! Rescale the sampled correlation based on the distance from observations.



Advantages of localization

- Handle the spurious correlation of the sample covariance

Advantages of localization

- Handle the spurious correlation of the sample covariance
- Stabilize the filter method

Advantages of localization

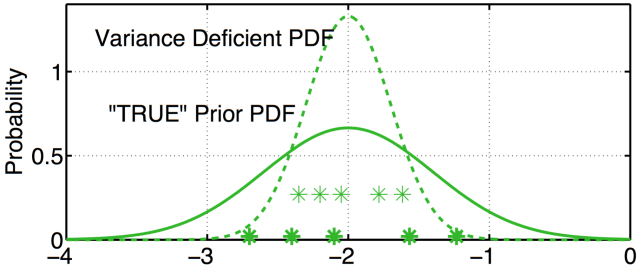
- Handle the spurious correlation of the sample covariance
- Stabilize the filter method
- Parallelization of the matrix inversion in the Kalman gain matrix

A problem which cannot be addressed by increasing samples

Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.

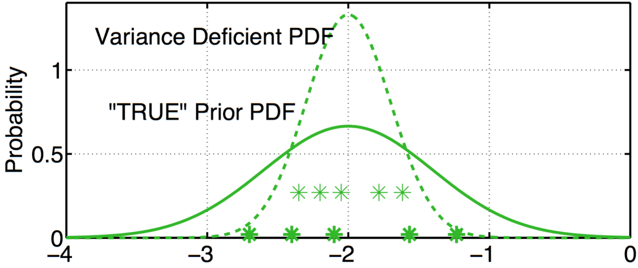
A problem which cannot be addressed by increasing samples

Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.



A problem which cannot be addressed by increasing samples

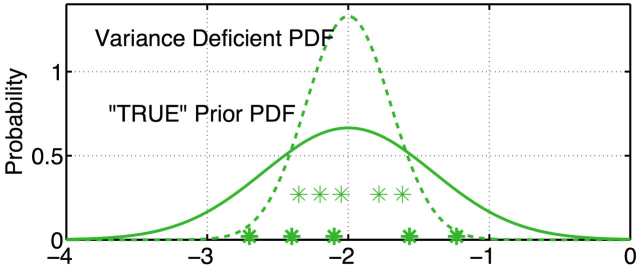
Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.



This can lead to 'filter divergence' :

A problem which cannot be addressed by increasing samples

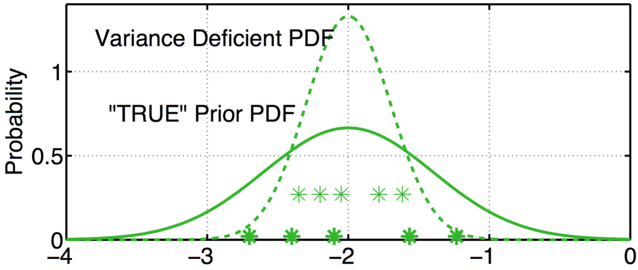
Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.



This can lead to 'filter divergence' : too small prior covariance

A problem which cannot be addressed by increasing samples

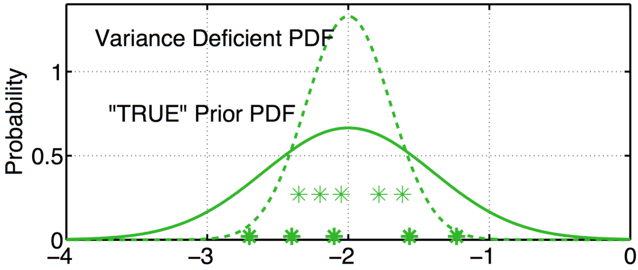
Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.



This can lead to 'filter divergence' : too small prior covariance $\rightarrow$ prior is too confident

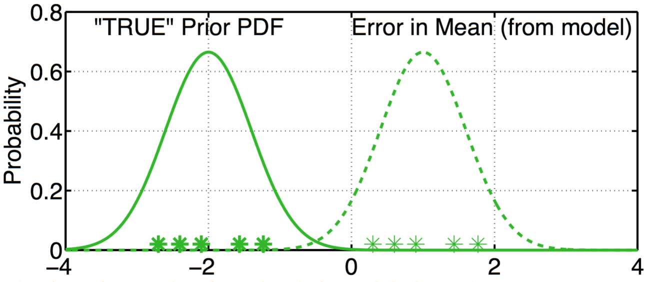
A problem which cannot be addressed by increasing samples

Prediction models have model errors (approximation errors) which can lead to biased mean or insufficient prior variance.



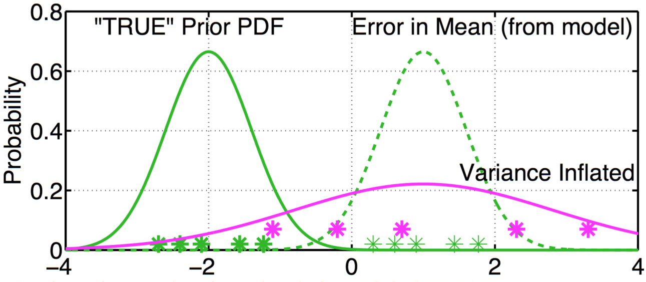
This can lead to 'filter divergence' : too small prior covariance $\rightarrow$ prior is too confident $\rightarrow$ observation is ignored

Biased mean and confident forecast



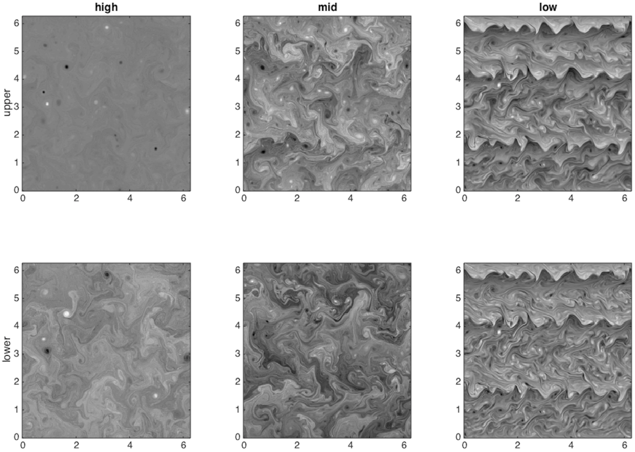
Covariance inflation

Increase the uncertainty of the prediction!



Application to two-layer quasigeostrophic turbulent flows

Describes the ocean flows at high, mid and low latitudes with strong coherent structures and zonal jets.



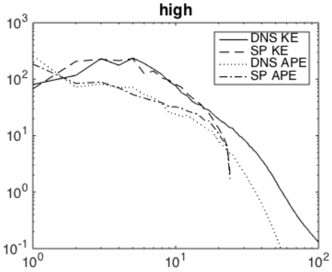
True signal has 256×256 state variables for both layers.

We use

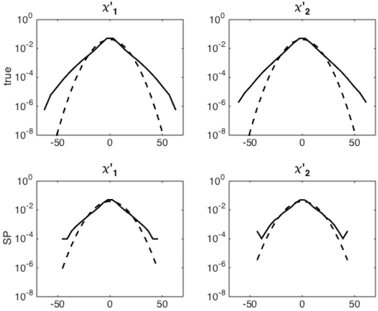
- A reduced-order forecast model with only 48×48 state variables for both layers; 250 times cheaper than the full resolution model.
- 48×48 observations of the upper layer variable, which is very sparse compared to the full resolution 256×256.

We use

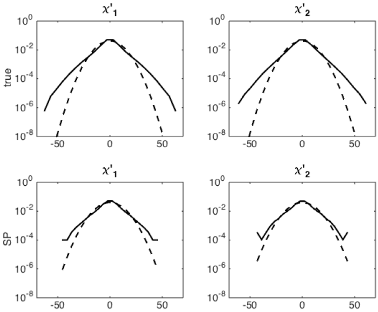
- A reduced-order forecast model with only 48×48 state variables for both layers; 250 times cheaper than the full resolution model.
- 48×48 observations of the upper layer variable, which is very sparse compared to the full resolution 256×256.



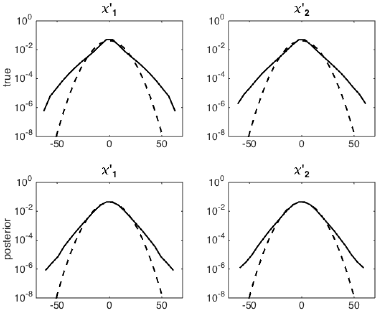
without data assimilation



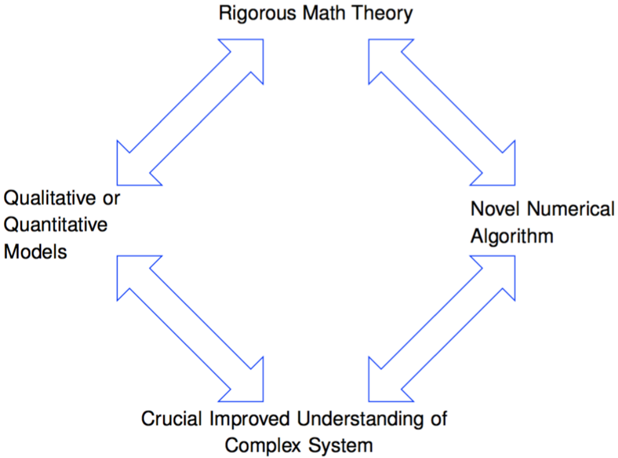
without data assimilation



with data assimilation



Modern Applied Science Paradigm



Research questions

- What happens if the state variables are not directly observed? That is, the observation can mix state variables at different spatial locations.

Research questions

- What happens if the state variables are not directly observed? That is, the observation can mix state variables at different spatial locations.
- Is accurate observation better than noisy observation?

Research questions

- What happens if the state variables are not directly observed? That is, the observation can mix state variables at different spatial locations.
- Is accurate observation better than noisy observation?
- What is the minimum size of observations for stable filter performance. Is plentiful observation better than sparse observation?

Research questions

- What happens if the state variables are not directly observed? That is, the observation can mix state variables at different spatial locations.
- Is accurate observation better than noisy observation?
- What is the minimum size of observations for stable filter performance. Is plentiful observation better than sparse observation?
- Are many samples always good?

Thank you!