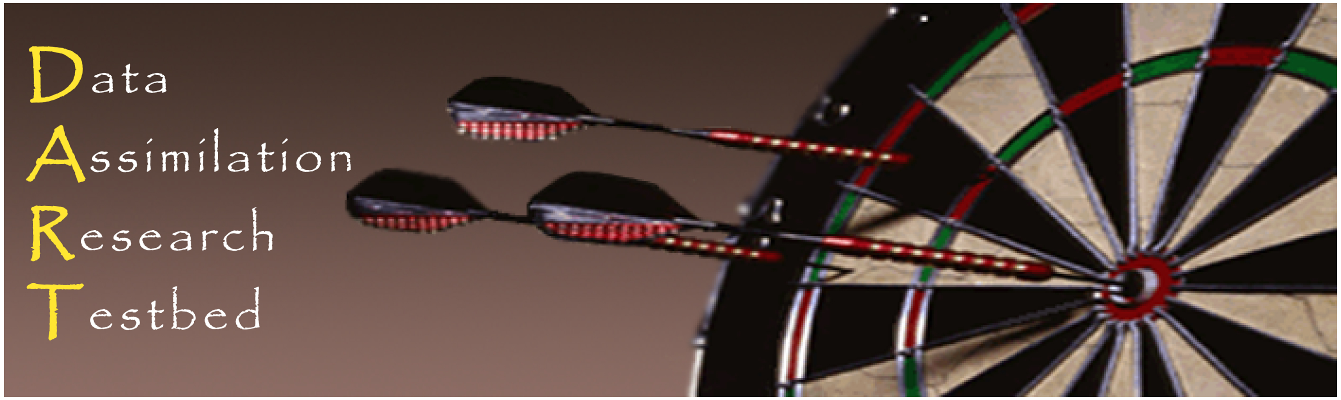
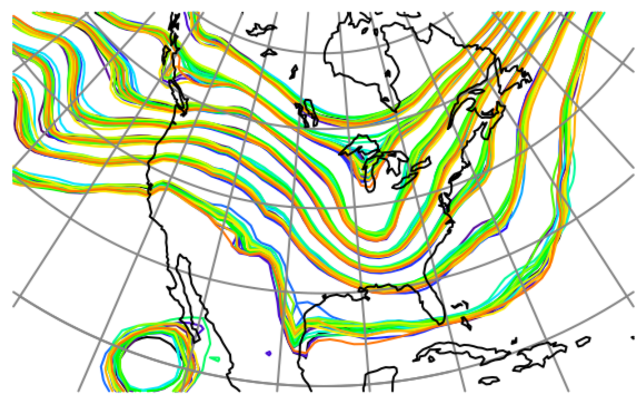


DART Tutorial Section 1:
Filtering For a One Variable System



The National Center for Atmospheric Research is sponsored by the National Science Foundation.
Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR UCAR | National Center for Atmospheric Research

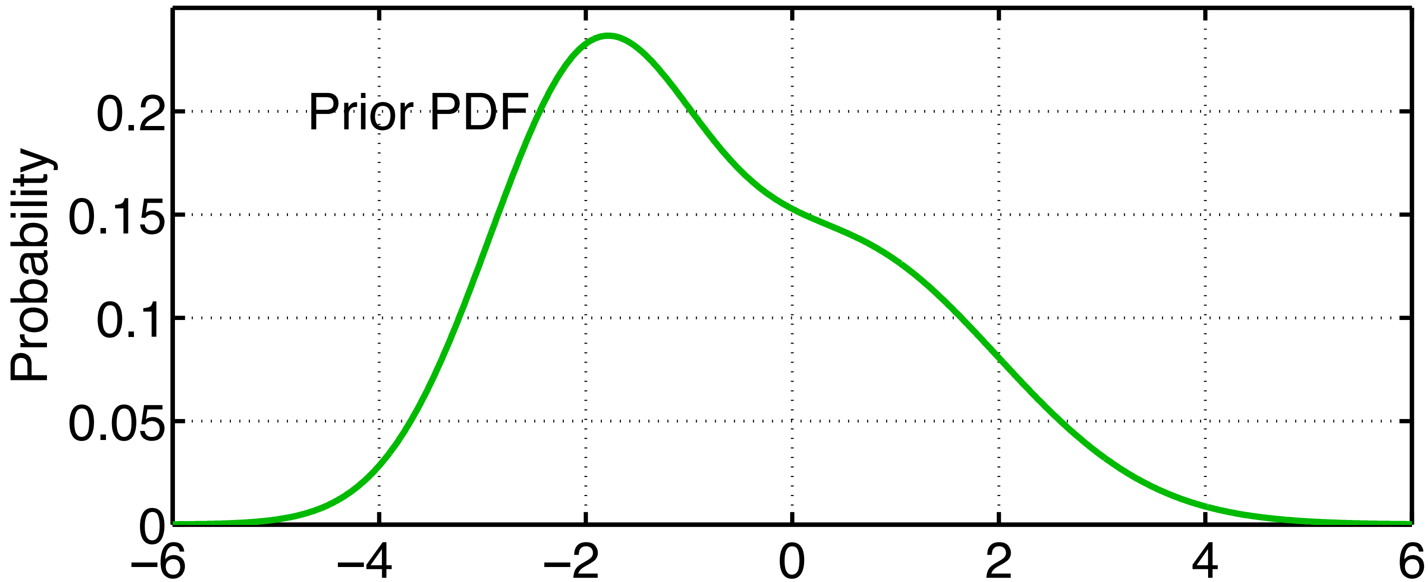
Introduction

This series of tutorial presentations is designed to introduce both basic Ensemble Kalman filter theory and the Data Assimilation Research Testbed Community Facility for Ensemble Data Assimilation.

There is significant overlap with the DART_LAB tutorial that is also part of the DART subversion checkout. If you have already studied DART_LAB, feel free to skip through the redundant theory slides. However, doing the exercises in all sections of this tutorial is recommended in order to learn the best ways to use the DART system.

Bayes' Rule

$$p\left(A \mid BC\right) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



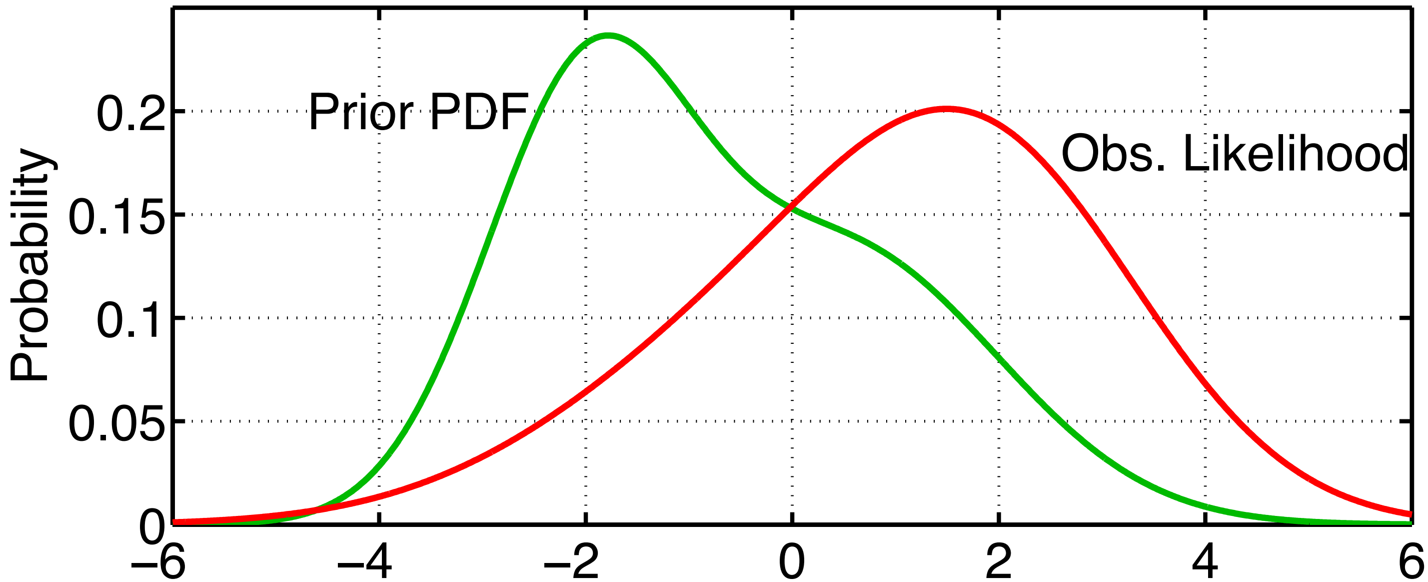
A : Prior Estimate based on all previous information, *C*.

B : An additional observation.

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

Bayes' Rule

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



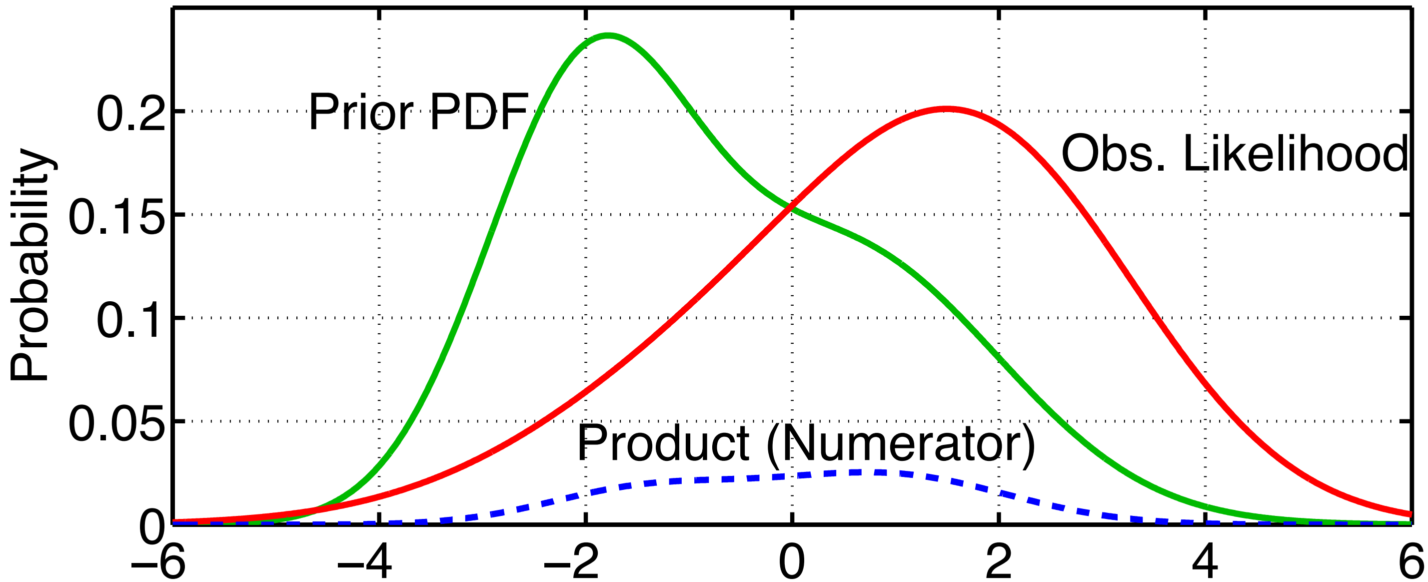
A : Prior Estimate based on all previous information, C.

B : An additional observation.

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

Bayes' Rule

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



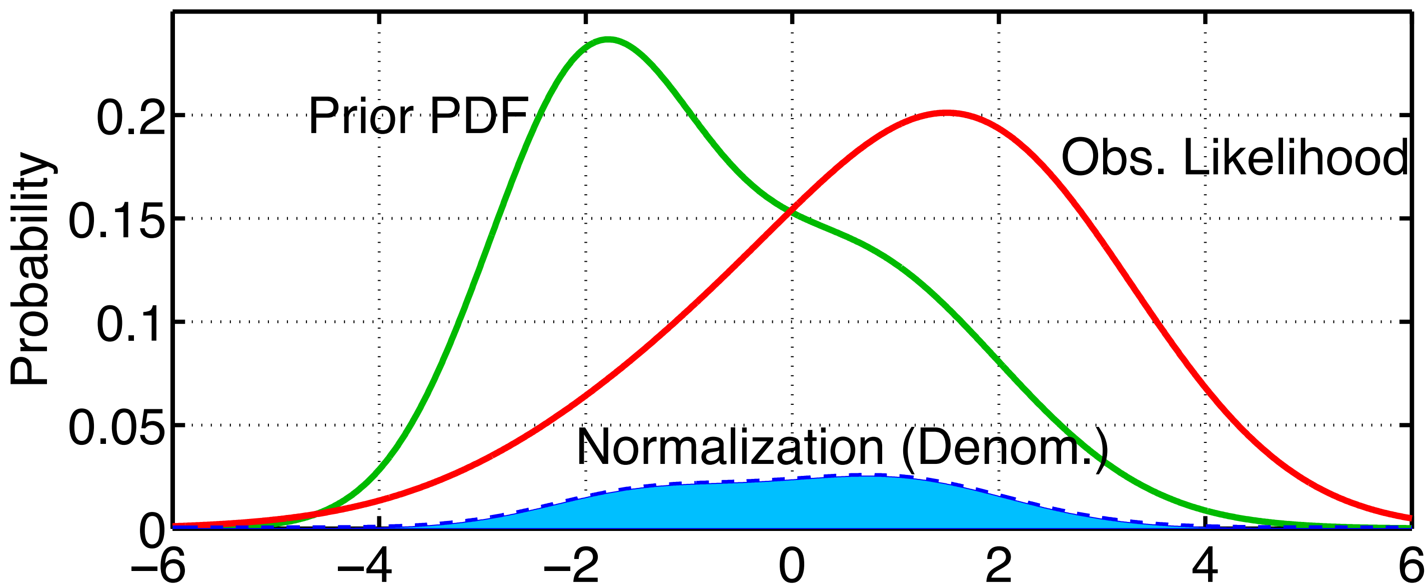
A : Prior Estimate based on all previous information, C.

B : An additional observation.

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

Bayes' Rule

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



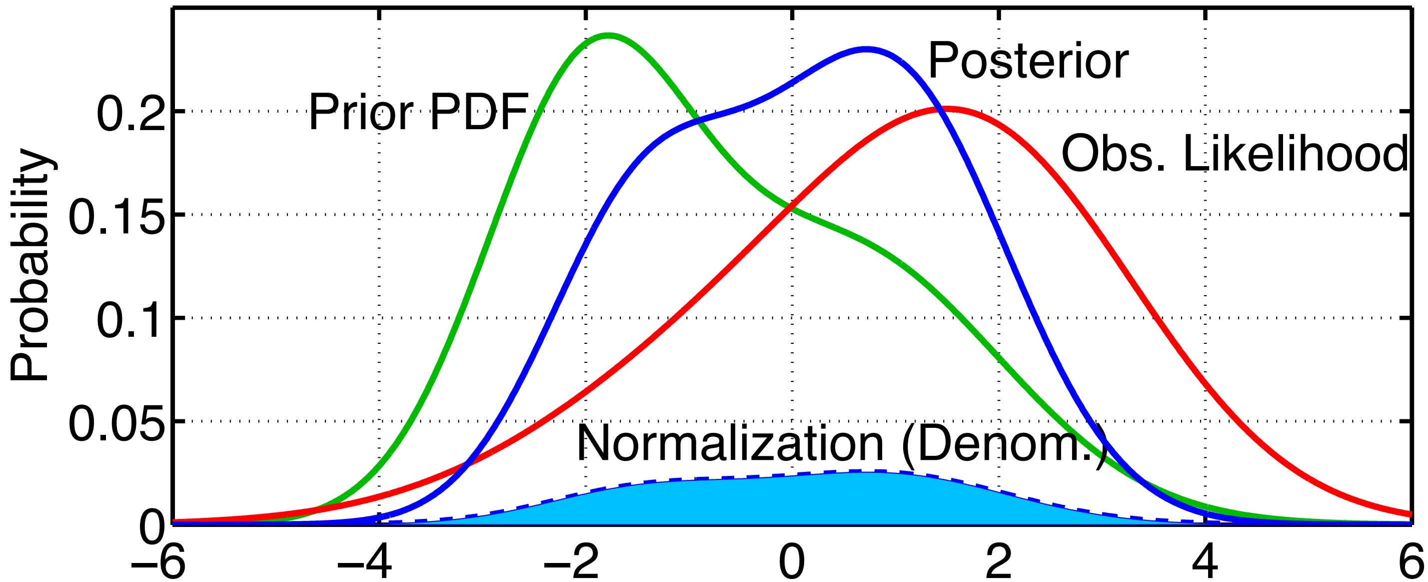
A : Prior Estimate based on all previous information, C.

B : An additional observation.

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

Bayes' Rule

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



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

B : An additional observation.

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

Color Scheme

Green == Prior

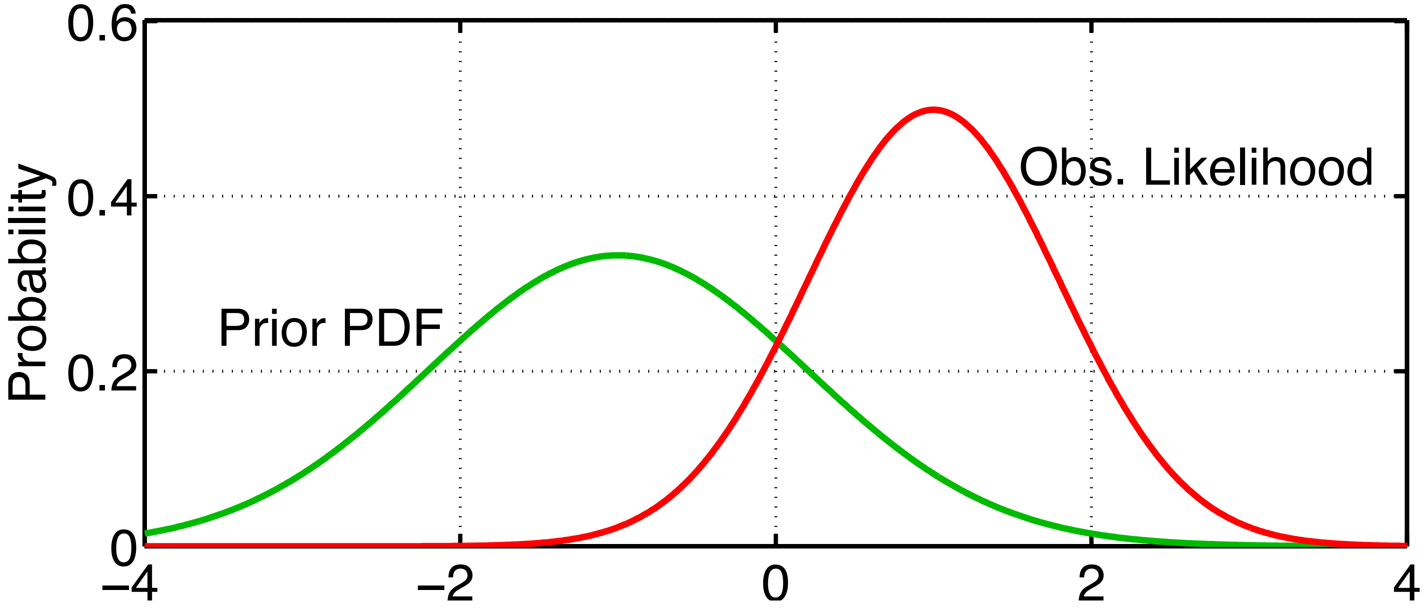
Red == Observation

Blue == Posterior

The same color scheme is used throughout ALL Tutorial materials.

Product of Two Gaussians

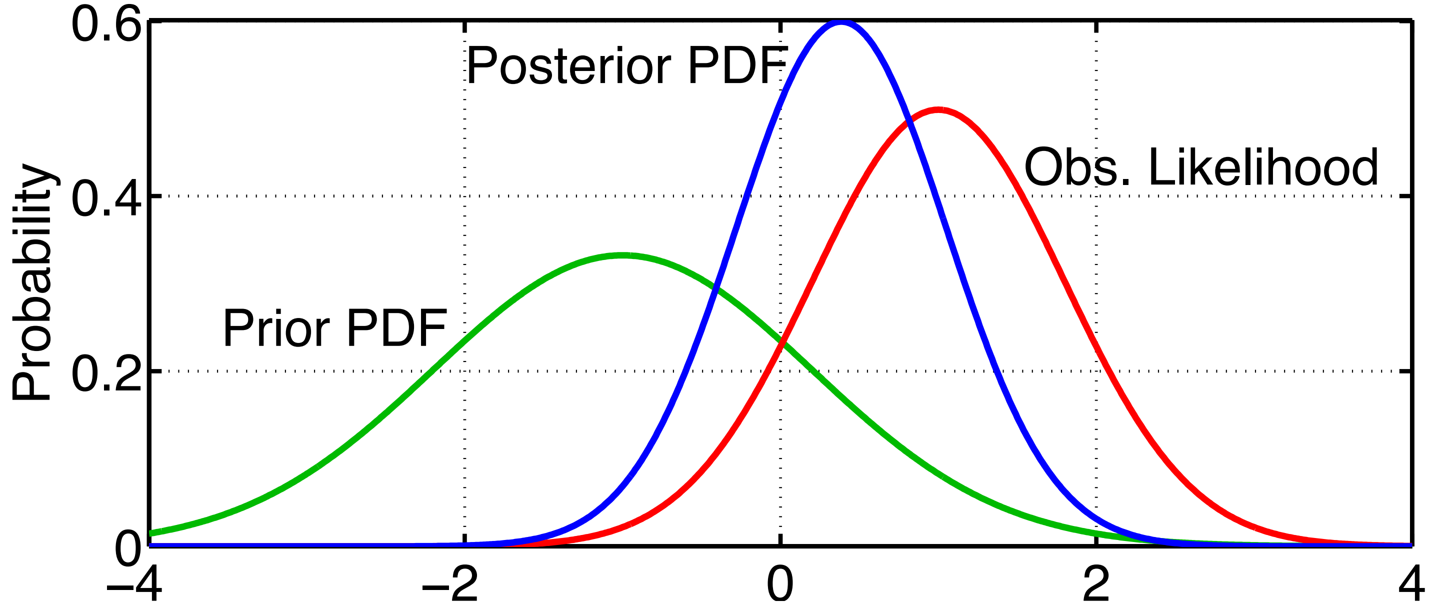
$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$



Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

This product is closed for Gaussian distributions.



Product of Two Gaussians

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

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

Product of Two Gaussians

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

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

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

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

Product of Two Gaussians

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

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

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

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

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

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

Product of Two Gaussians

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

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

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

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

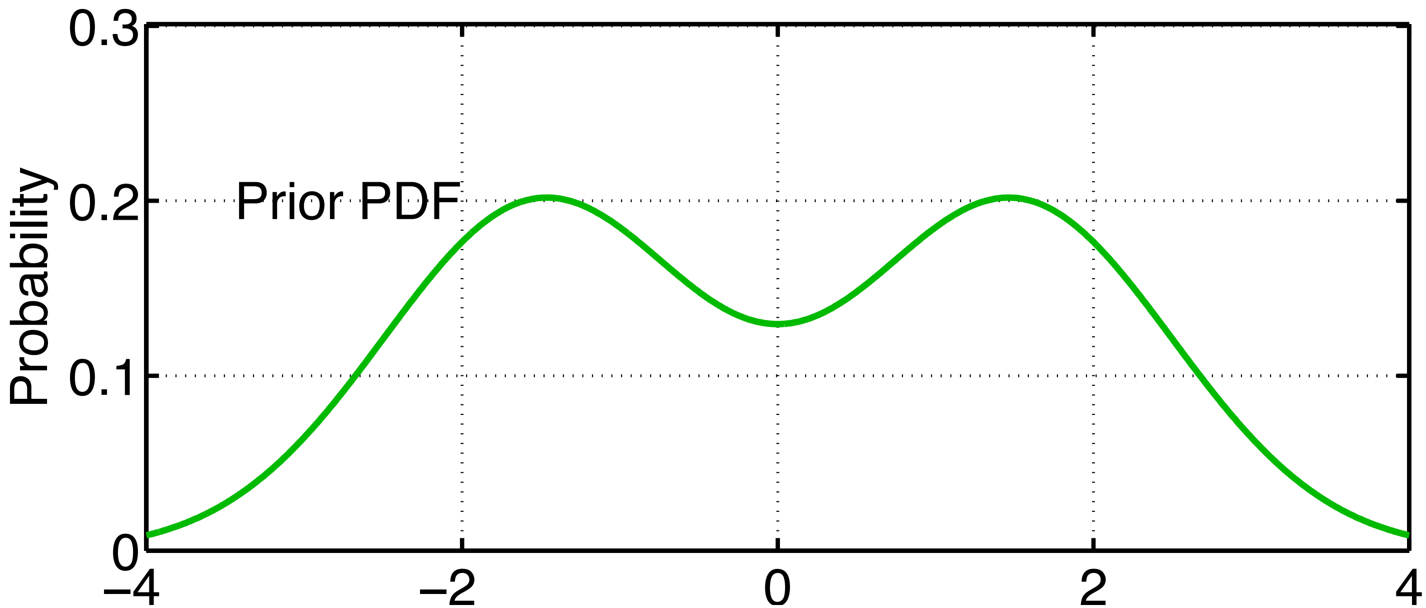
Weight: $$c = \frac{1}{(2\Pi)^{d/2}\left|\Sigma_1 + \Sigma_2\right|^{1/2}} \exp\left\{-\frac{1}{2}\left[(\mu_2 - \mu_1)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_2 - \mu_1)\right]\right\}$$

Easy to derive for 1-D Gaussians; just do products of exponentials.

Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

This product is closed for Gaussian distributions.

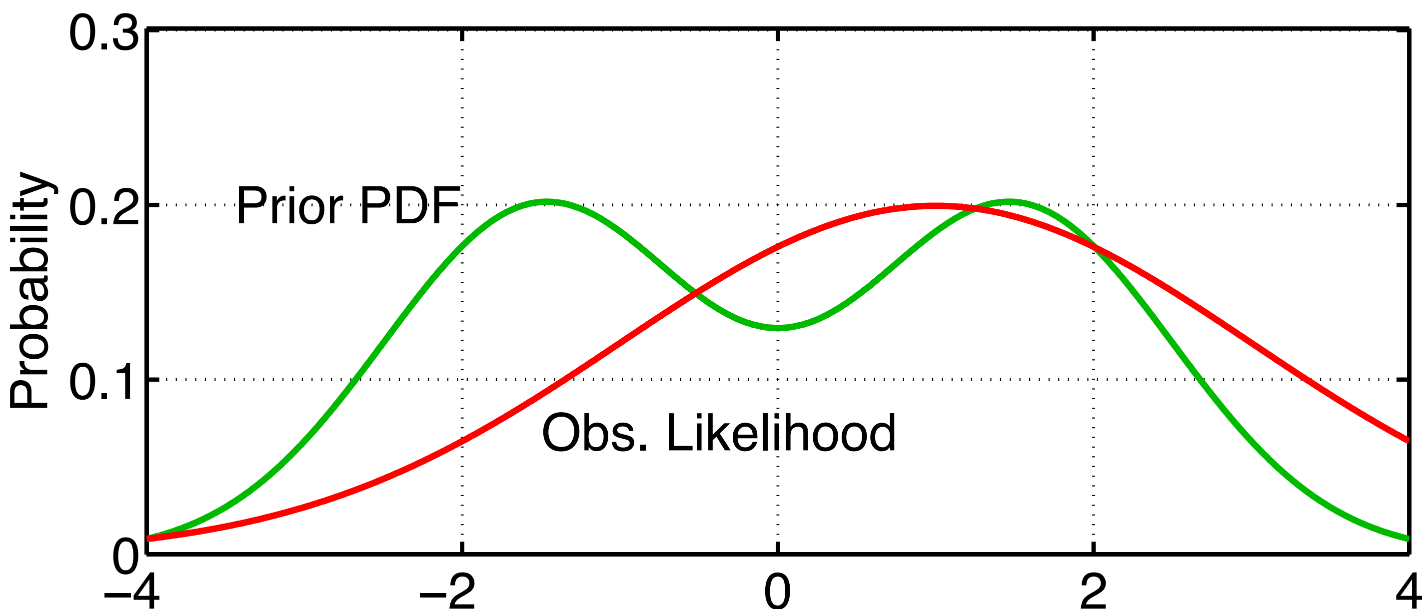


There are other families of functions for which it is closed …
But, for general distributions, there's no analytical product.

Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

This product is closed for Gaussian distributions.

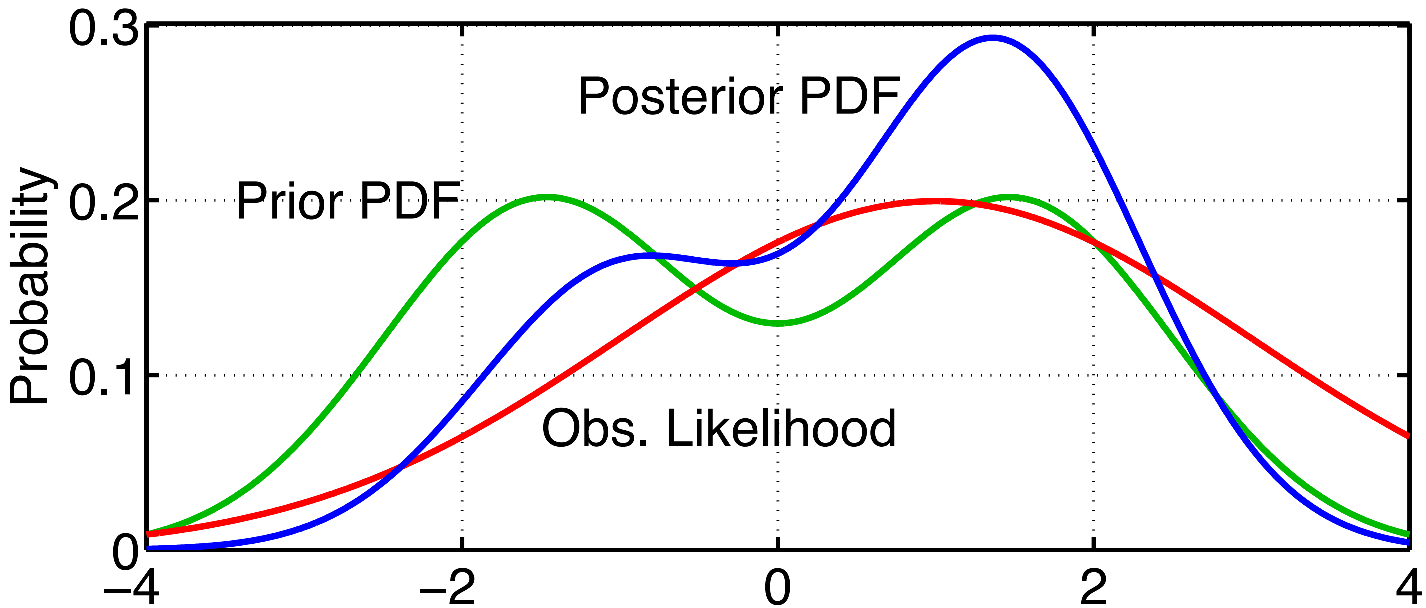


There are other families of functions for which it is closed …
But, for general distributions, there's no analytical product.

Product of Two Gaussians

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

This product is closed for Gaussian distributions.

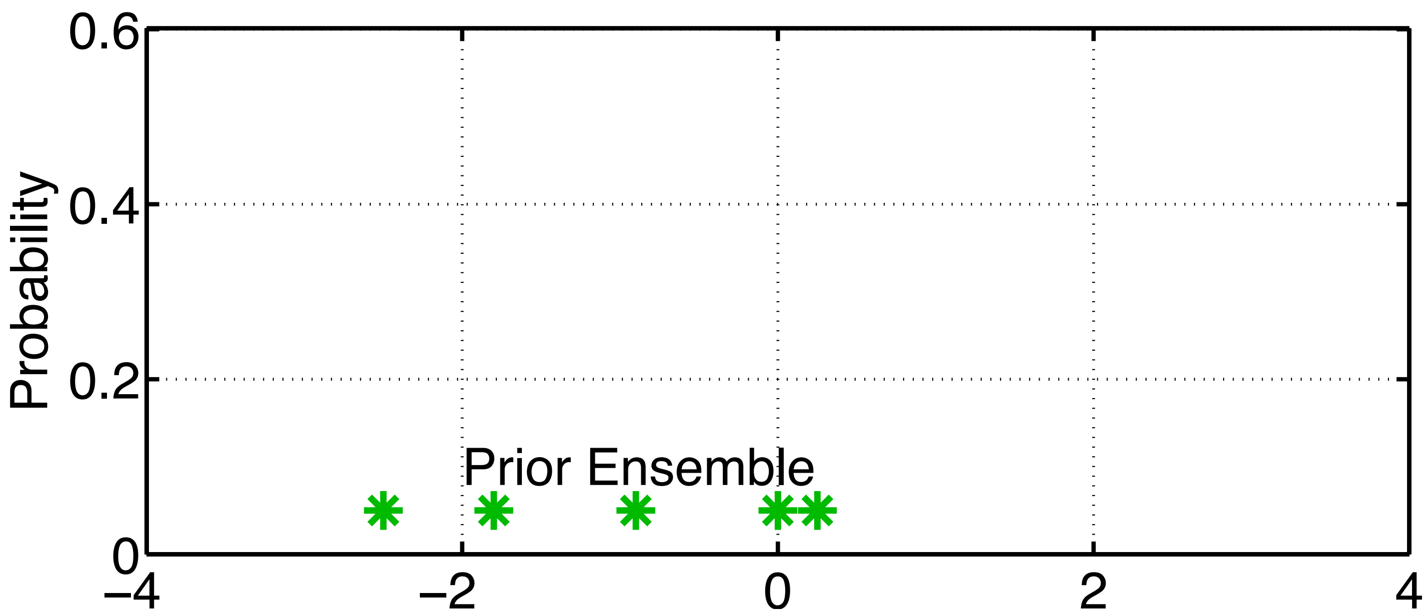


There are other families of functions for which it is closed …
But, for general distributions, there's no analytical product.

Product of Two Gaussians

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

Ensemble filters: Prior is available as finite sample.

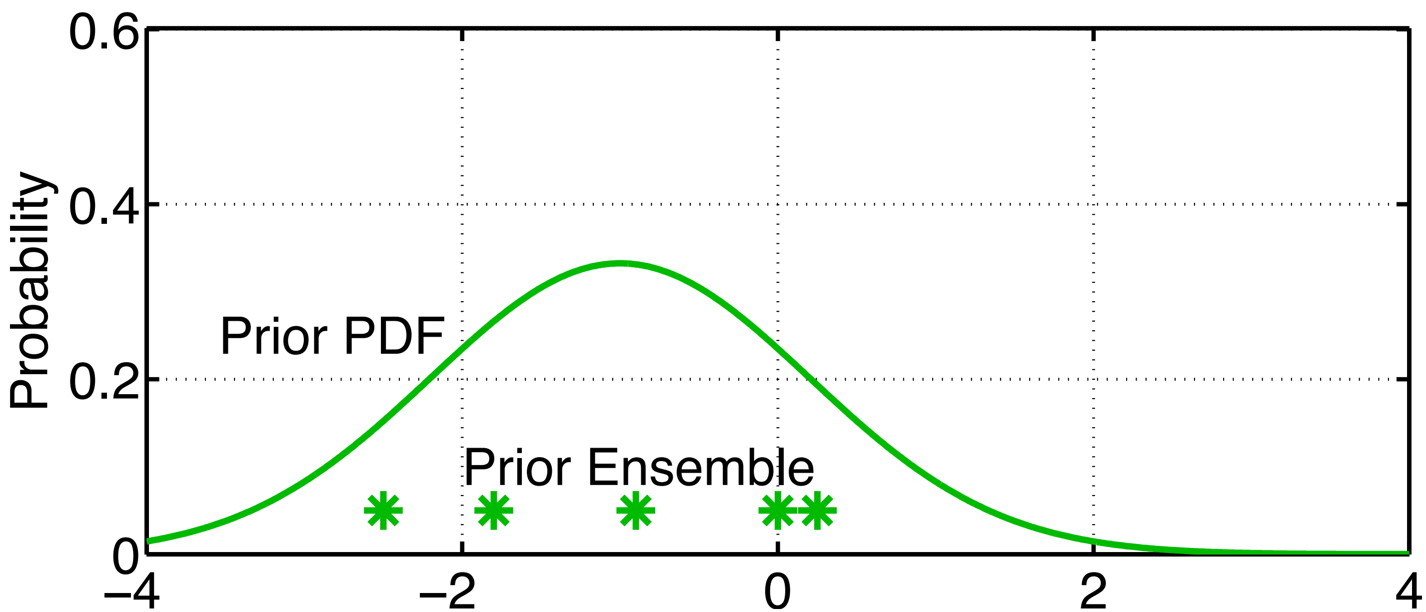


Don't know much about properties of this sample.
May naively assume it is random draw from 'truth'.

Product of Two Gaussians

$$\color{blue}{p(A|BC)} = \frac{\color{red}{p(B|AC)}\color{green}{p(A|C)}}{p(B|C)} = \frac{\color{red}{p(B|AC)}\color{green}{p(A|C)}}{\int p(B|x)p(x|C)dx}$$

How can we take product of sample with continuous likelihood?

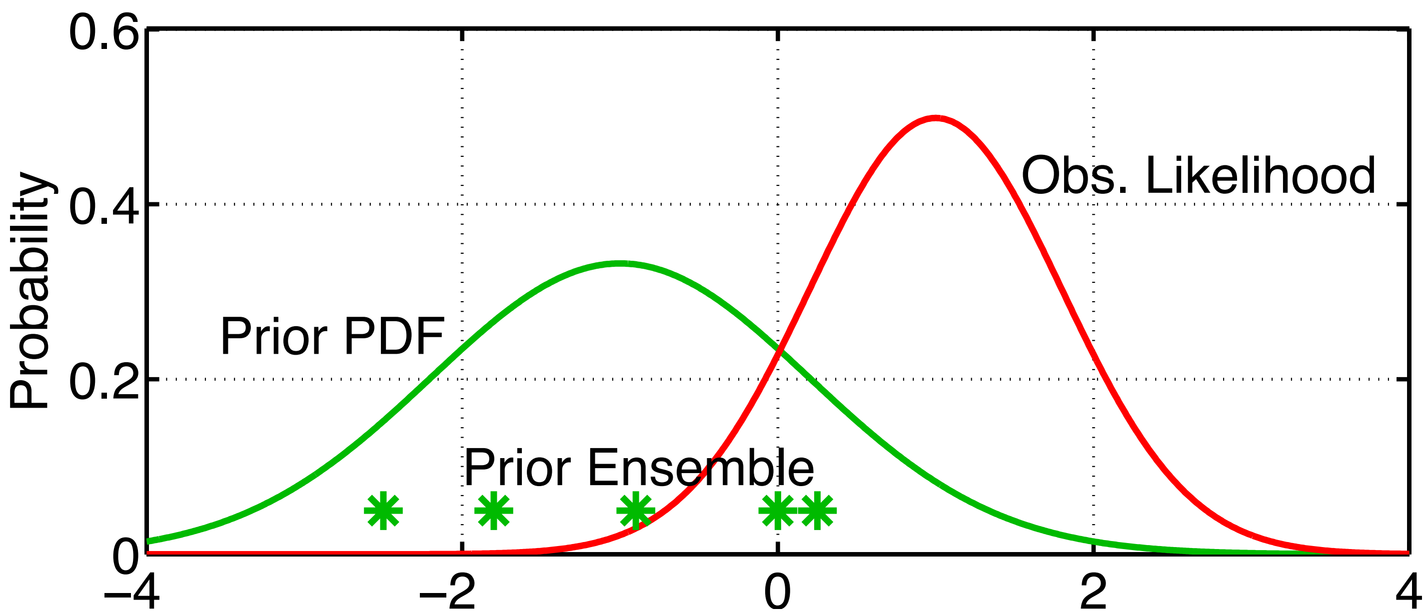


Fit a continuous (Gaussian for now) distribution to sample.

Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

Observation likelihood usually continuous (nearly always Gaussian).

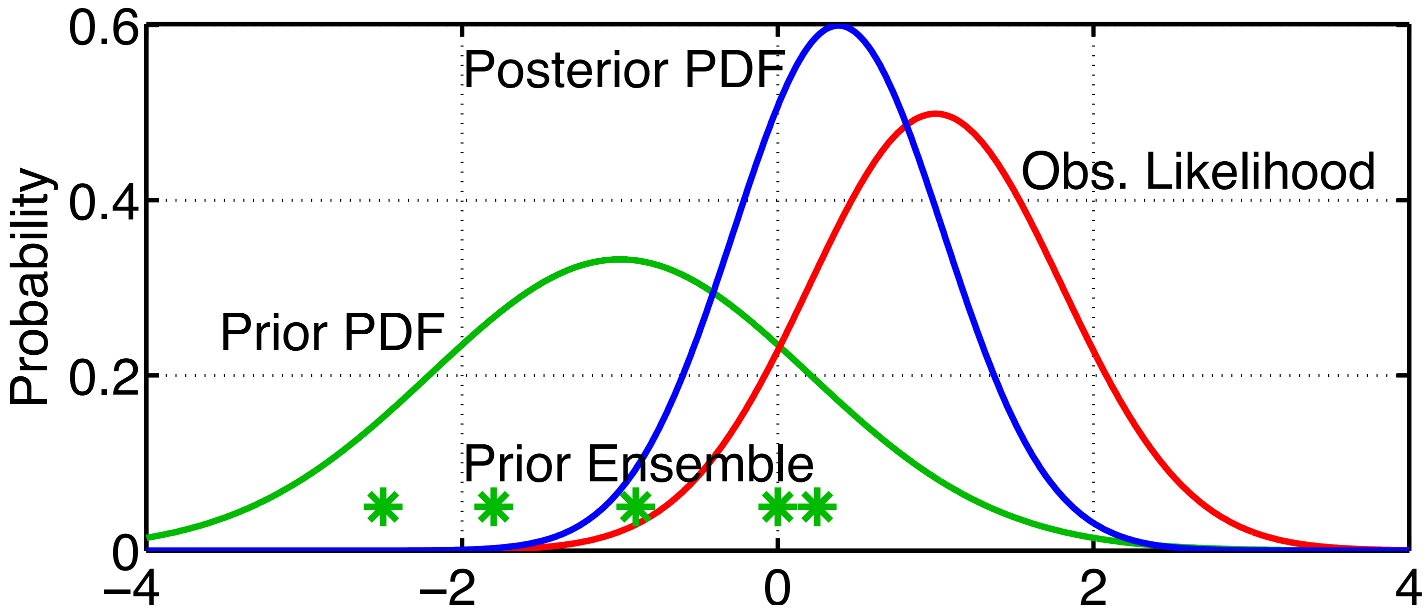


If Obs. Likelihood isn't Gaussian, can generalize methods below.
For instance, can fit set of Gaussian kernels to obs. likelihood.

Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

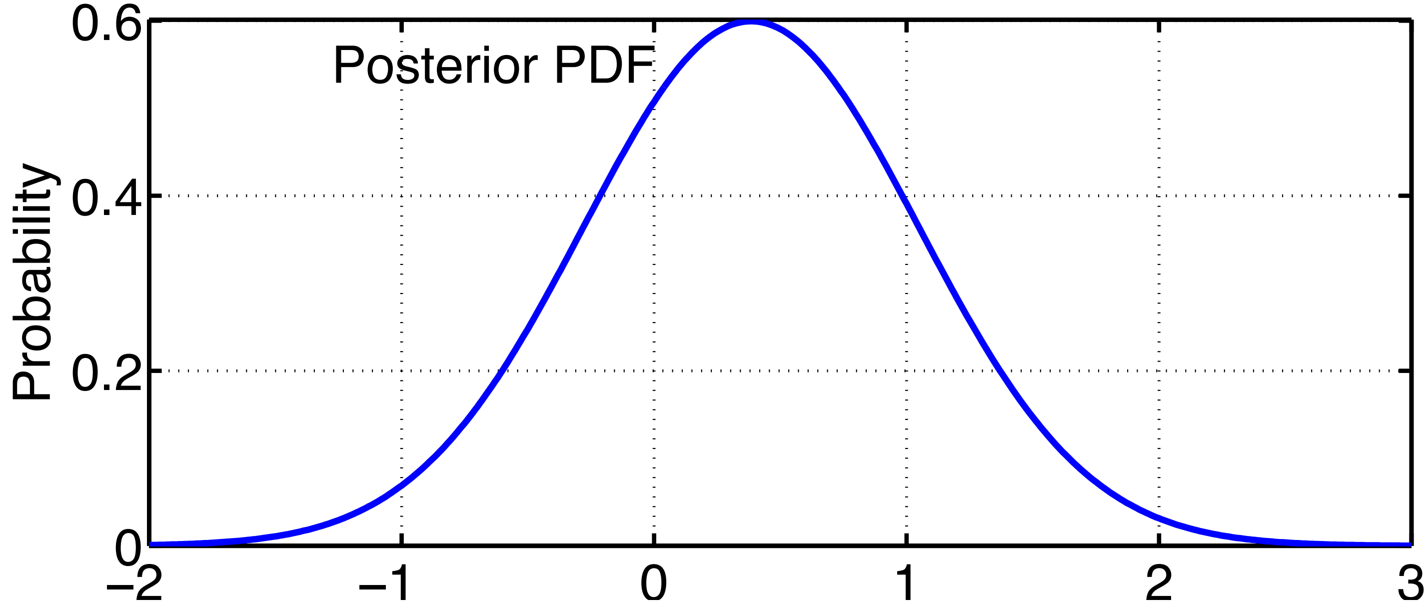
Product of prior Gaussian fit and Obs. likelihood is Gaussian.



Computing continuous posterior is simple.
BUT, need to have a SAMPLE of this PDF.

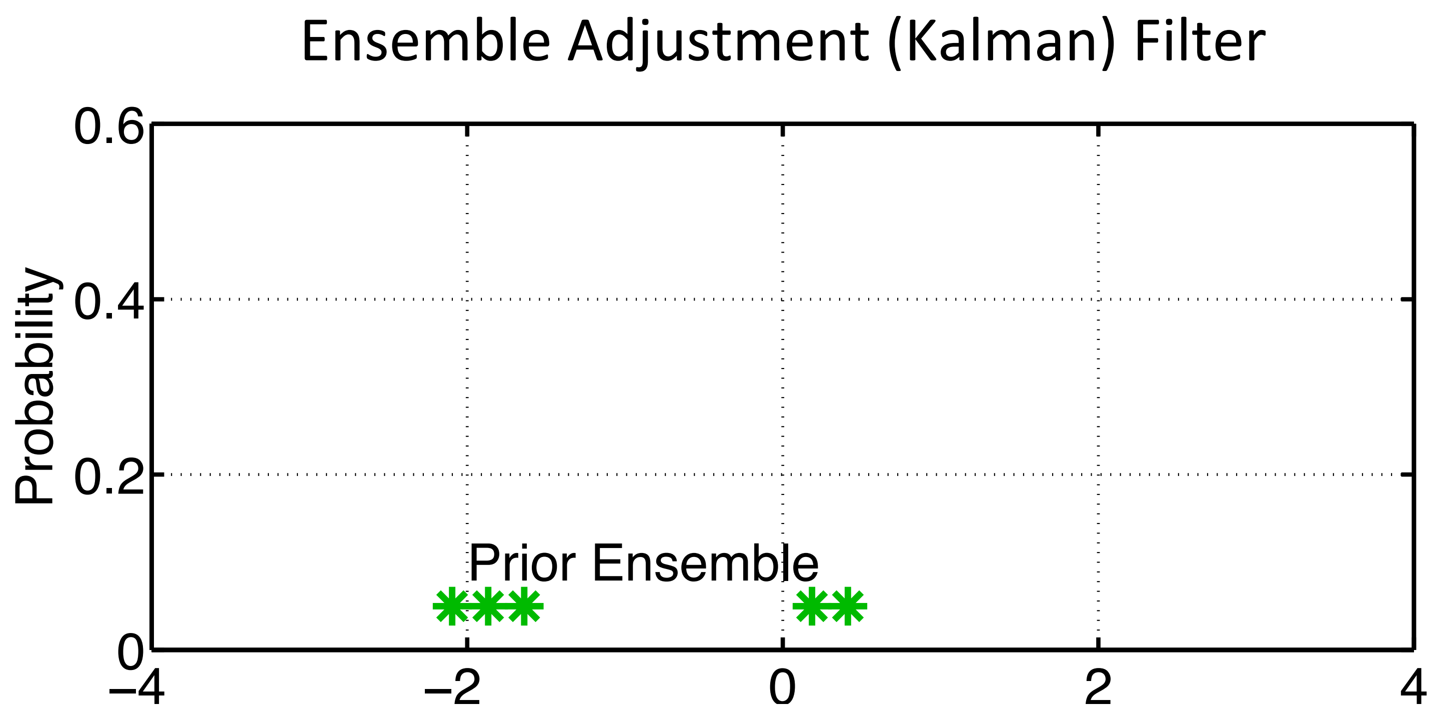
Sampling Posterior PDF

There are many ways to do this.

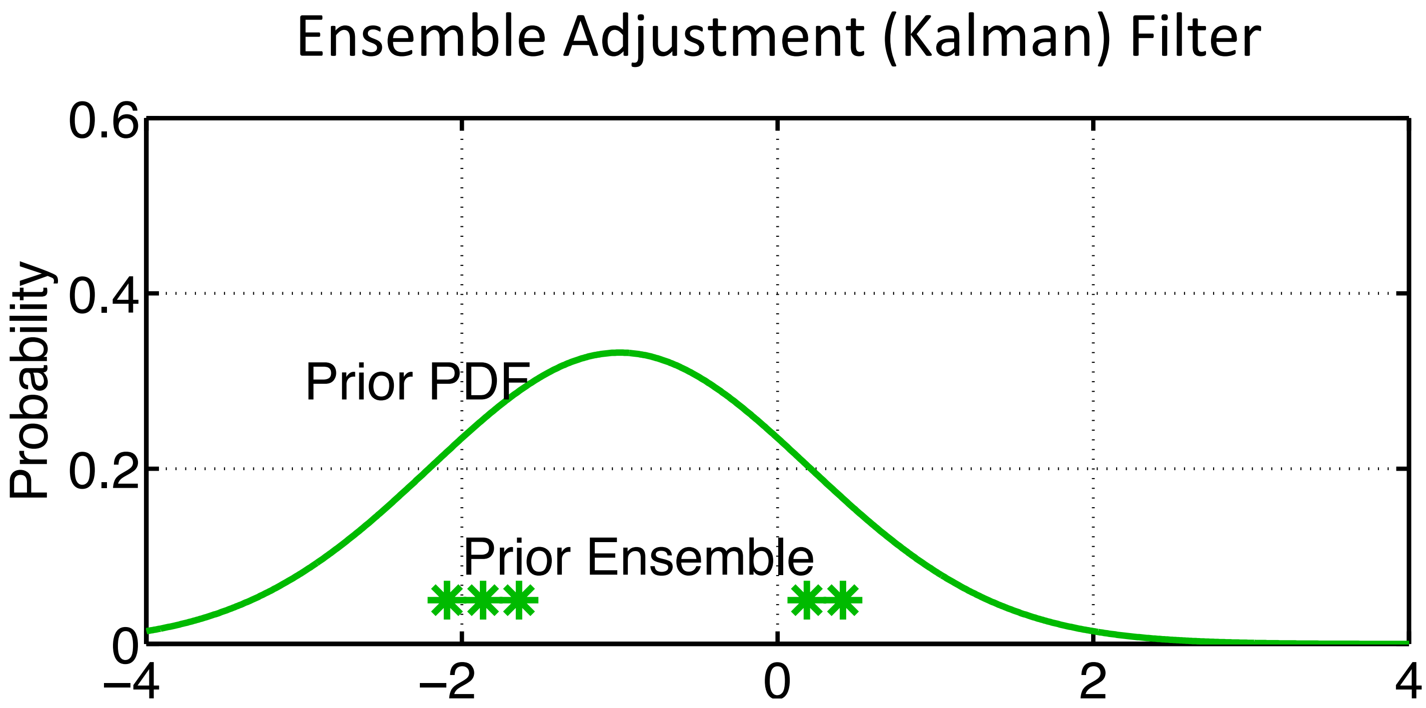


Exact properties of different methods may be unclear.
Trial and error still best way to see how they perform.
Will interact with properties of prediction models, etc.

Sampling Posterior PDF

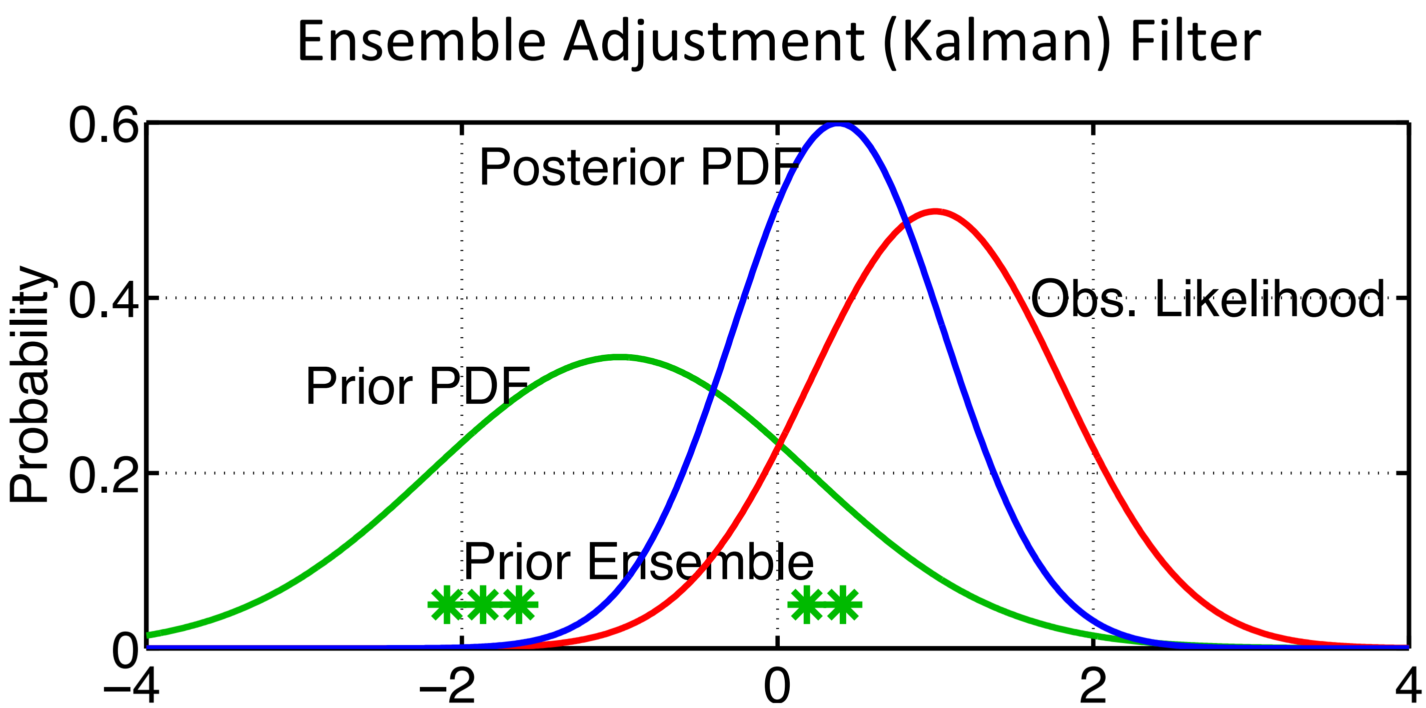


Sampling Posterior PDF



Again, fit a Gaussian to sample.

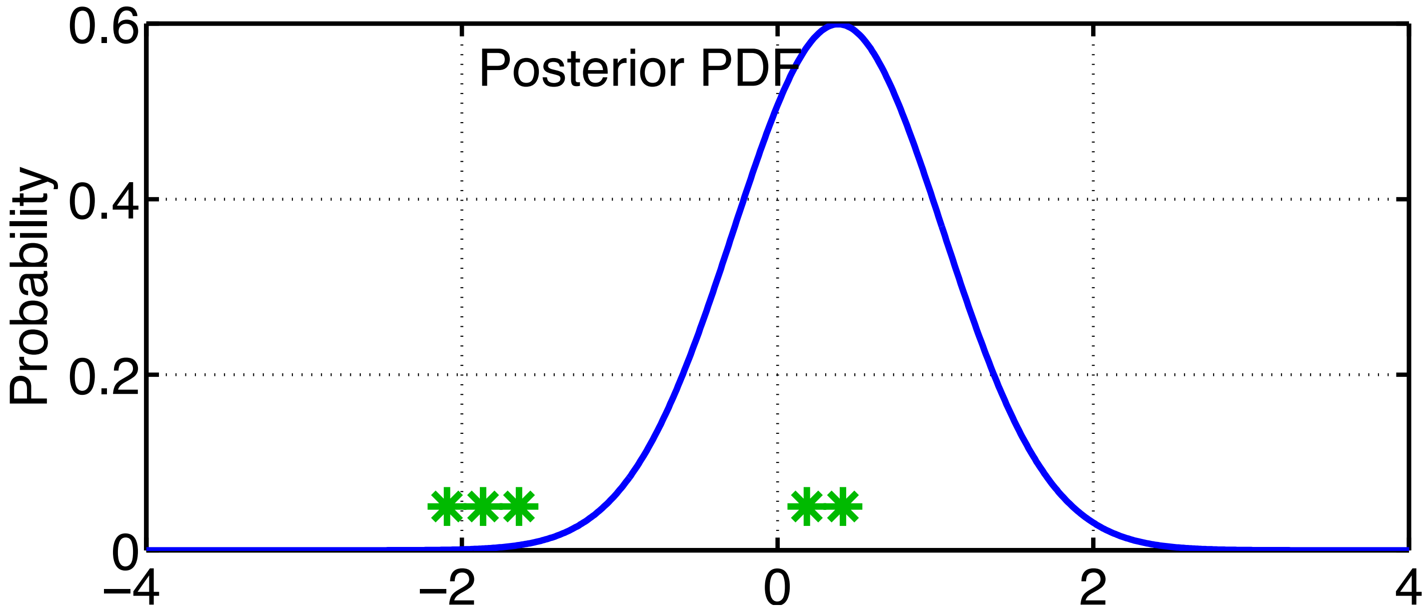
Sampling Posterior PDF



Compute posterior PDF (same as previous algorithms).

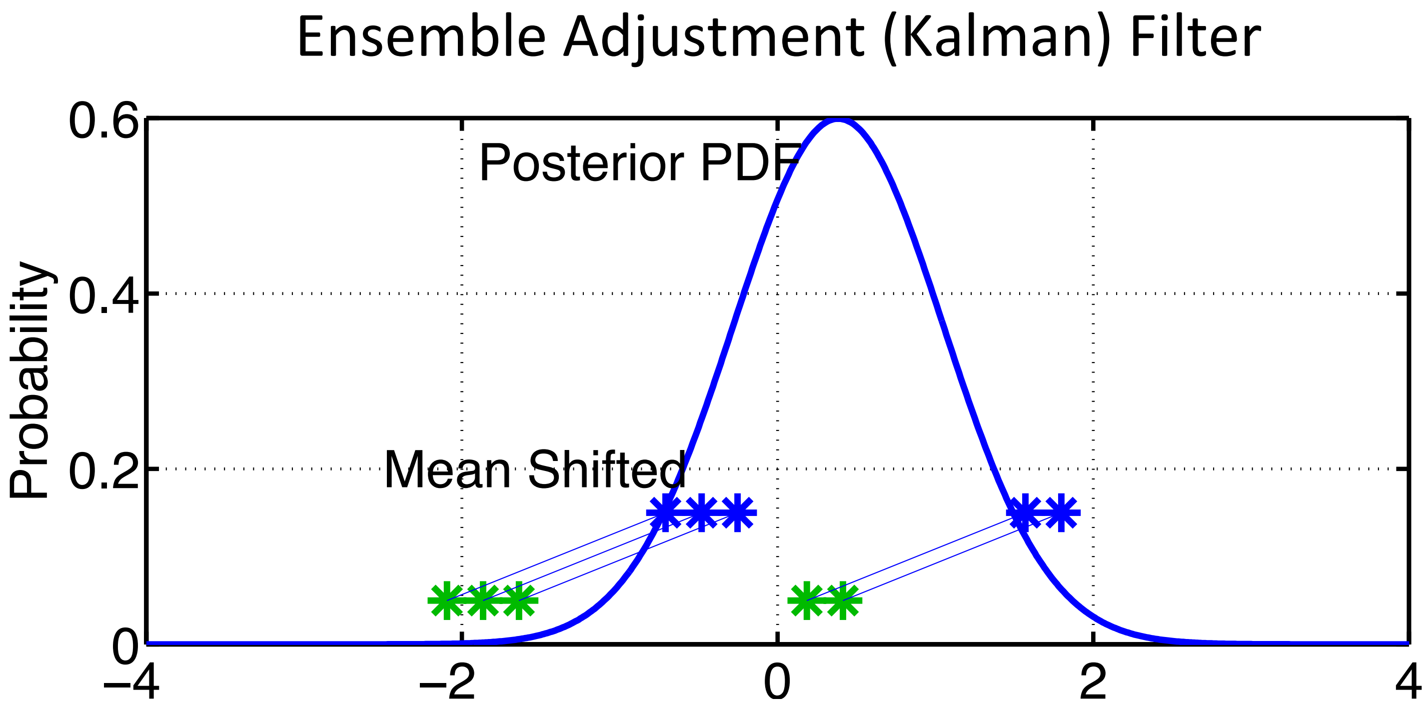
Sampling Posterior PDF

Ensemble Adjustment (Kalman) Filter



Use deterministic algorithm to 'adjust' ensemble.

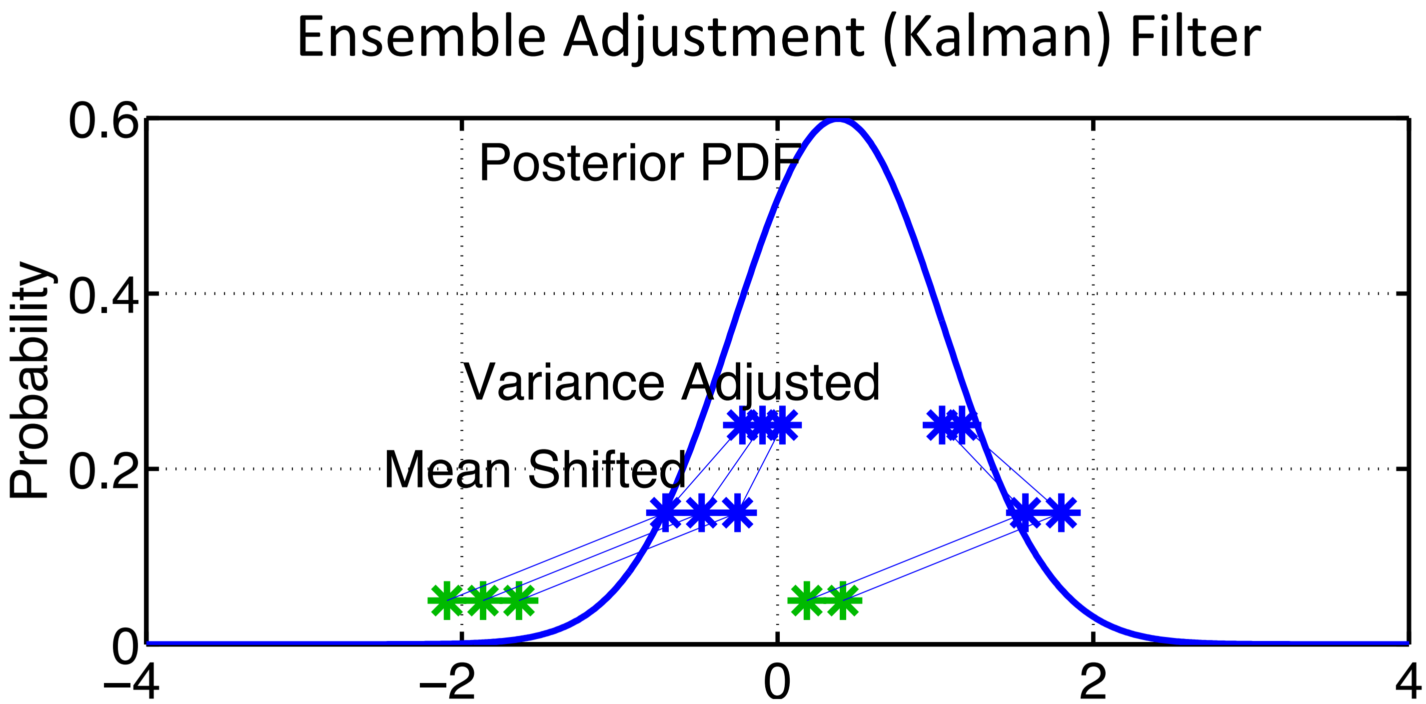
Sampling Posterior PDF



Use deterministic algorithm to 'adjust' ensemble.

1. 'Shift' ensemble to have exact mean of posterior.

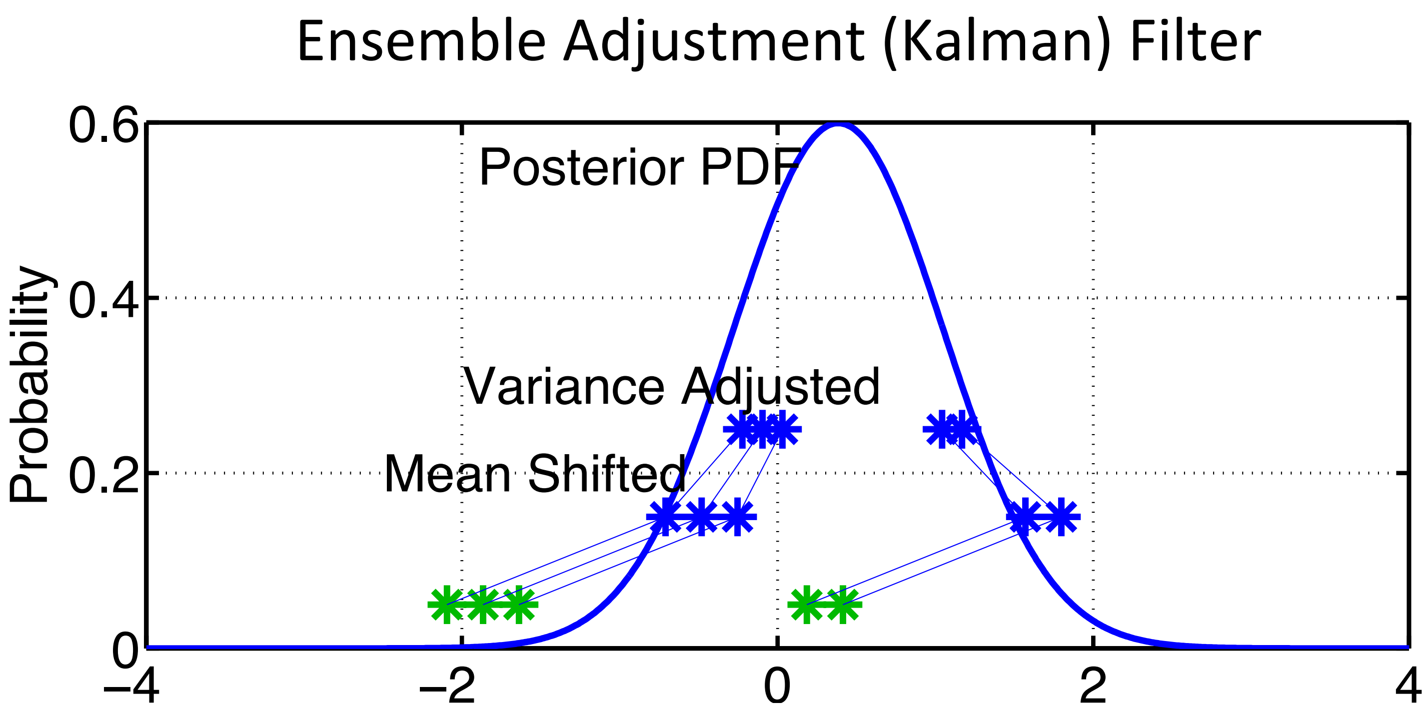
Sampling Posterior PDF



Use deterministic algorithm to ‘adjust’ ensemble.

1. ‘Shift’ ensemble to have exact mean of posterior.
2. Use linear contraction to have exact variance of posterior.

Sampling Posterior PDF



$$x_i^u = \left(x_i^p - \overline{x}^p\right) \bullet \left(\sigma^u / \sigma^p\right) + \overline{x}^u$$

i = 1,..., ensemble size.

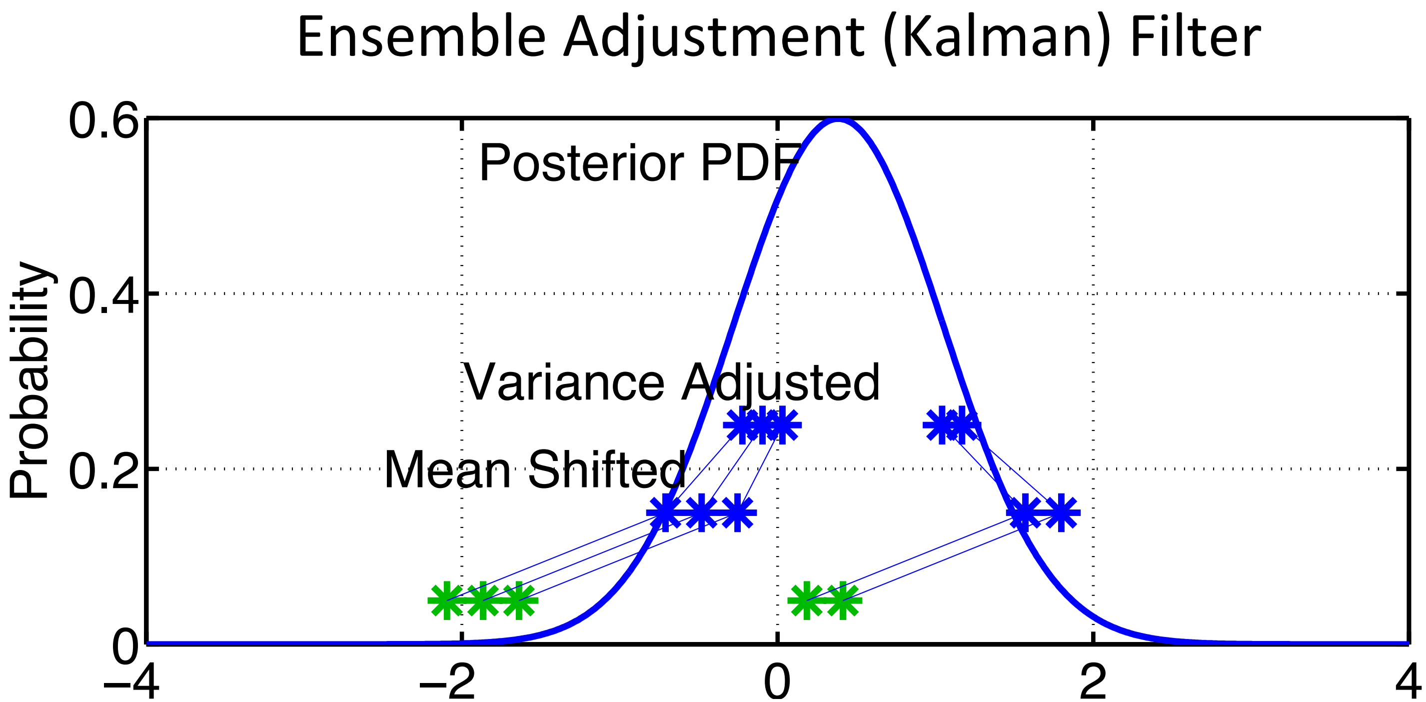
p is prior,

u is update (posterior),

σ is standard deviation,

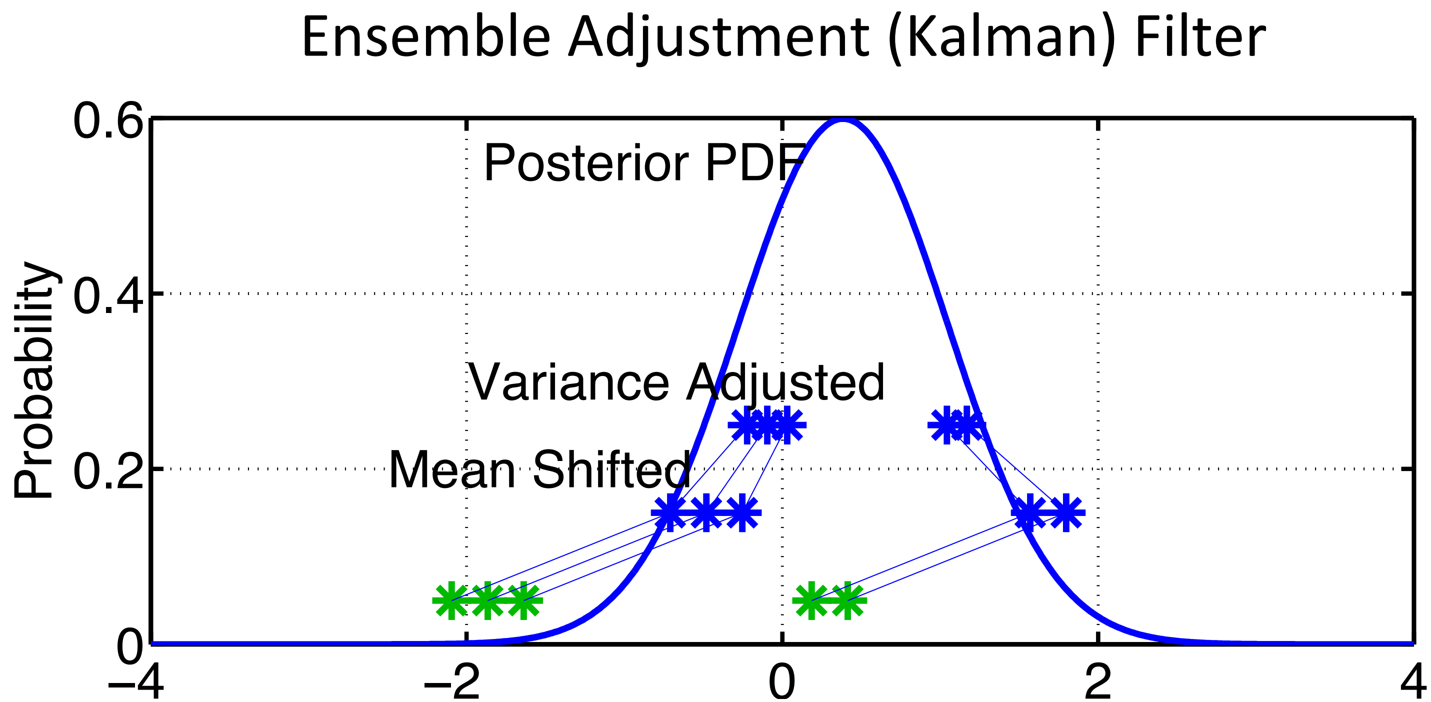
overbar is ensemble mean.

Sampling Posterior PDF



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

Sampling Posterior PDF



There are a variety of other ways to deterministically adjust ensemble.
Class of algorithms sometimes called deterministic square root filters.

1st look at DART Diagnostics

cd models/lorenz_63/work	in your DART sandbox.
csh workshop_setup.csh	Does stuff you'll learn to do later.
matlab -nodesktop	

Output from a DART assimilation in 3-variable model.
20 member ensemble.
Observations of each variable once every '6 hours'; error variance 8.
Observation ONLY impacts its own state variable.
For assimilation, looks like 3 independent single variable problems.
Model advance between assimilations isn't independent.

Initial ensemble members are random selection from long model run.
Initial error should be an upper bound (random guess).

1st look at DART Diagnostics

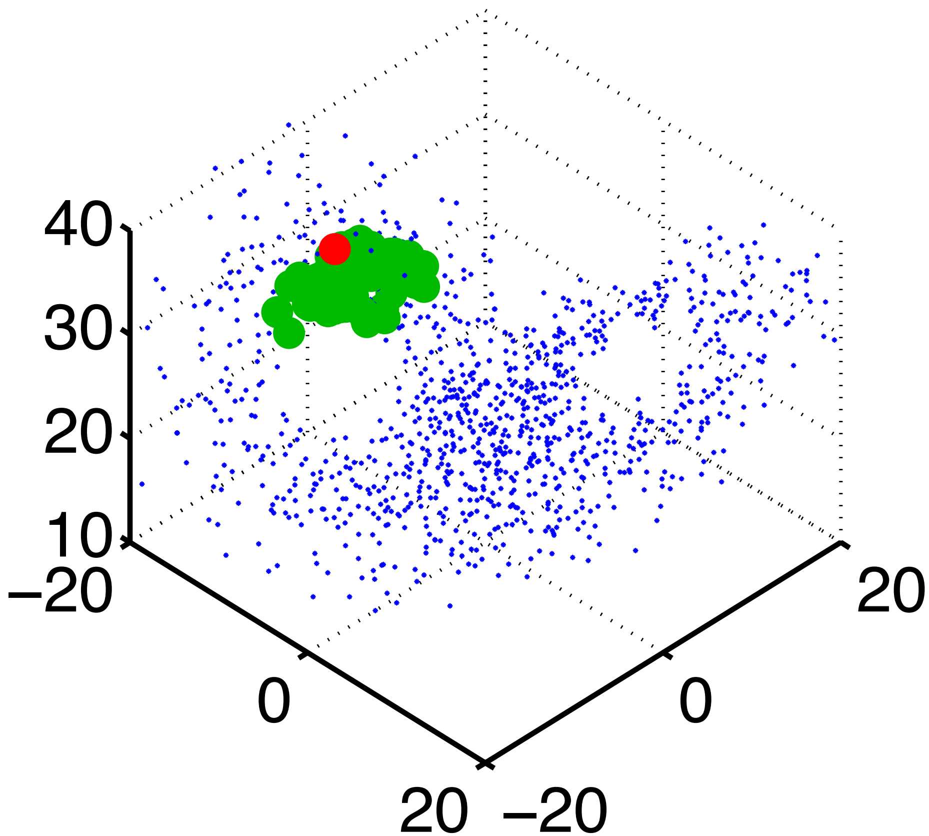
Try the following matlab commands. Each will ask you to:

Input name of true model trajectory file:
<cr> for perfect_output.nc
Input name of ensemble trajectory file:
<cr> for preassim.nc

Just select the default files by hitting carriage return for all matlab exercises.

plot_total_err	time series of distance between prior ensemble mean and truth in blue; spread: average prior distance between ensemble members and mean in red. (Record total values of total error and spread for later.)
plot_ens_mean_time_series	time series of truth in blue; ensemble mean prior.
plot_ens_time_series	also includes prior ensemble members.

Simple Example: Lorenz-63
3-variable chaotic model



Observation in red.

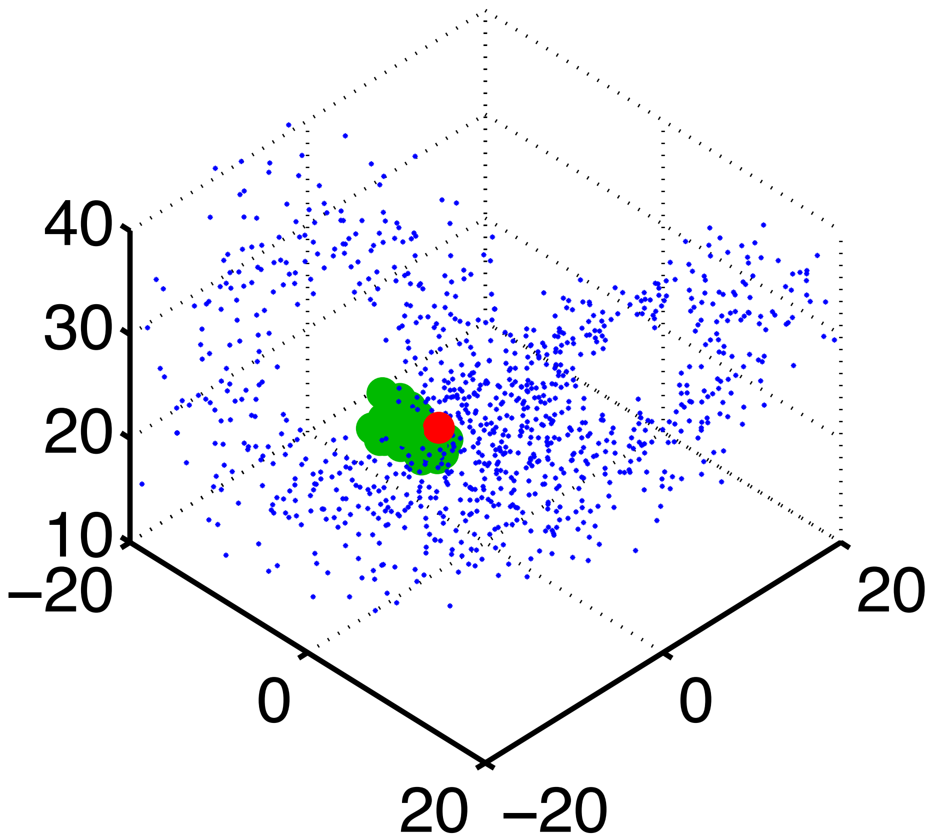
Prior ensemble in green.

Observing all three state variables.

Obs. Error variance = 4.0.

Four 20-member ensembles.

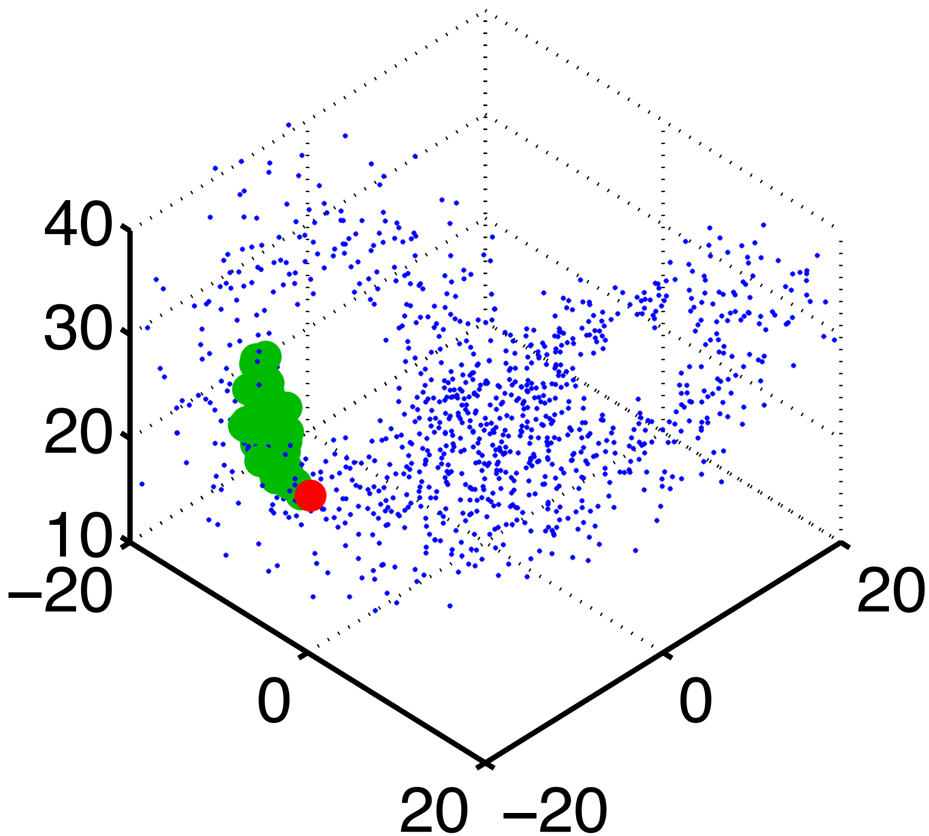
Simple Example: Lorenz-63 3-variable chaotic model



Observation in red.

Prior ensemble in green.

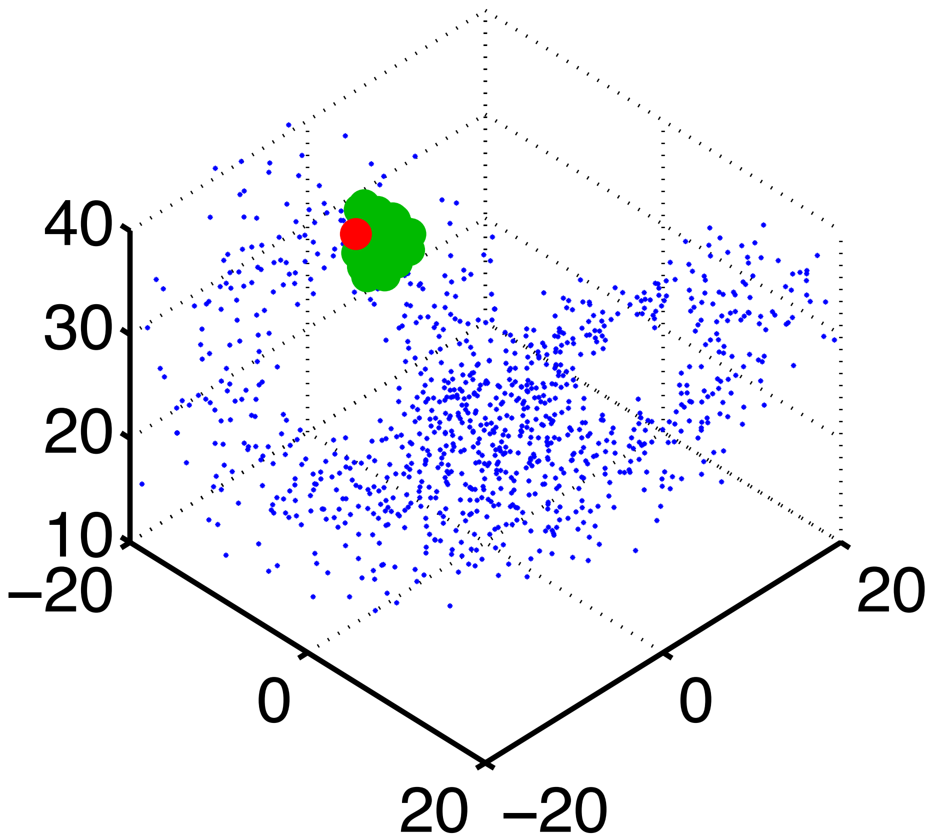
Simple Example: Lorenz-63
3-variable chaotic model



Observation in red.

Prior ensemble in green.

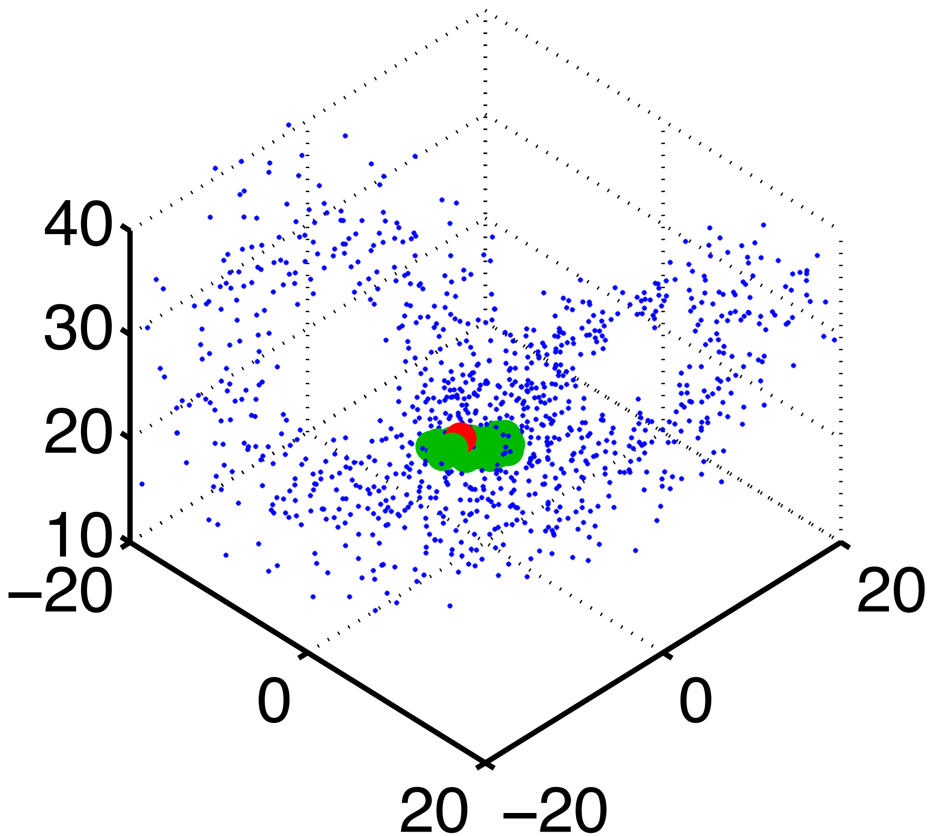
Simple Example: Lorenz-63
3-variable chaotic model



Observation in red.

Prior ensemble in green.

Simple Example: Lorenz-63
3-variable chaotic model

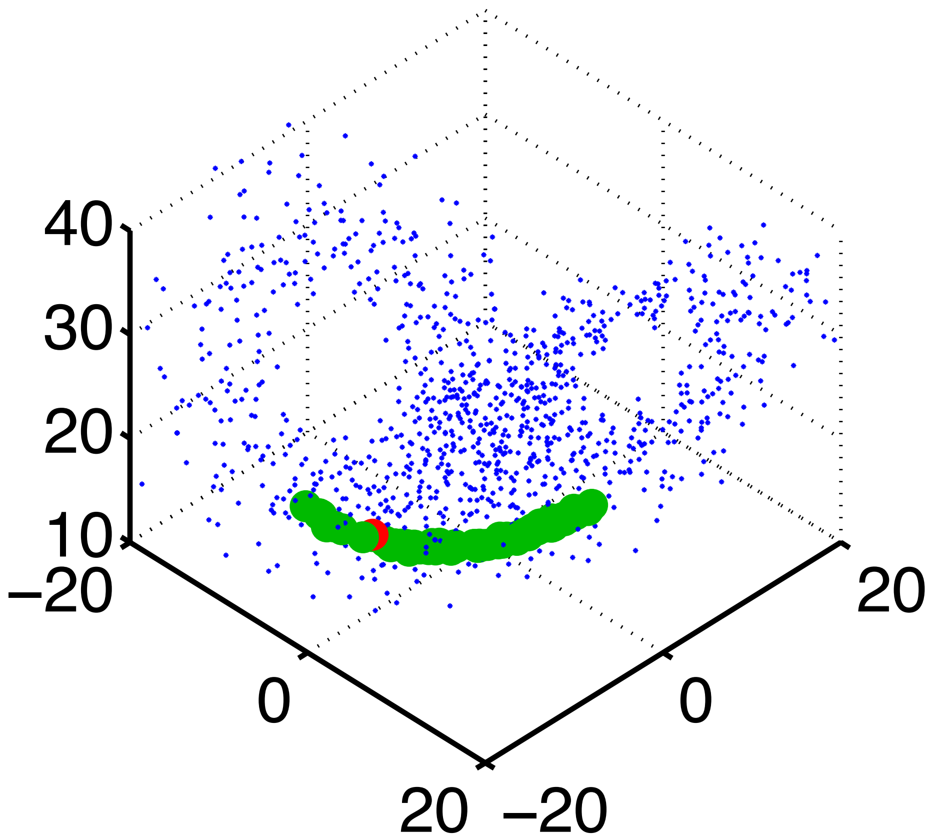


Observation in red.

Prior ensemble in green.

Ensemble is passing through an unpredictable region.

Simple Example: Lorenz-63
3-variable chaotic model

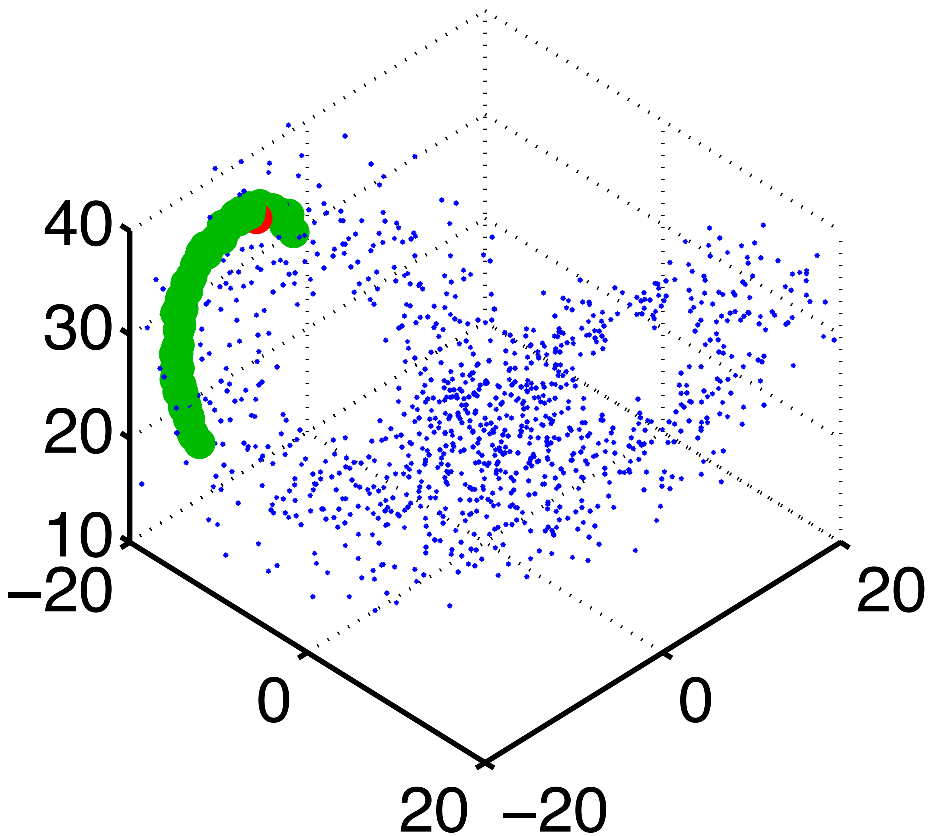


Observation in red.

Prior ensemble in green.

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

Simple Example: Lorenz-63 3-variable model



Observation in red.

Prior ensemble in green.

Using DART Diagnostics

Using DART diagnostics from the simple Lorenz-63 assimilation:

Can you see evidence of enhanced uncertainty?

Where does this occur?

Does the ensemble appear to be consistent with the truth?
(Is the truth normally inside the ensemble range?)

DART Tutorial Index to Sections