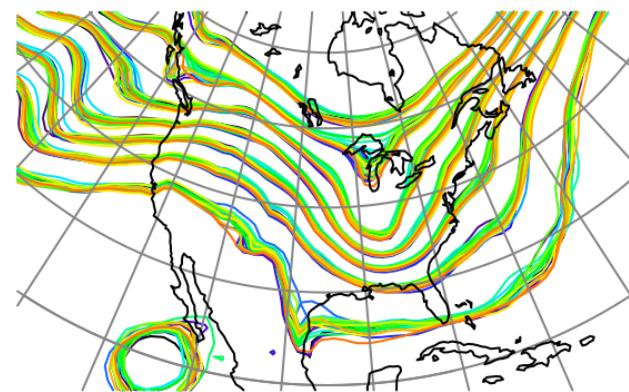


Data  
Assimilation  
Research  
Testbed



## DART Tutorial Section 1: Filtering For a One Variable System



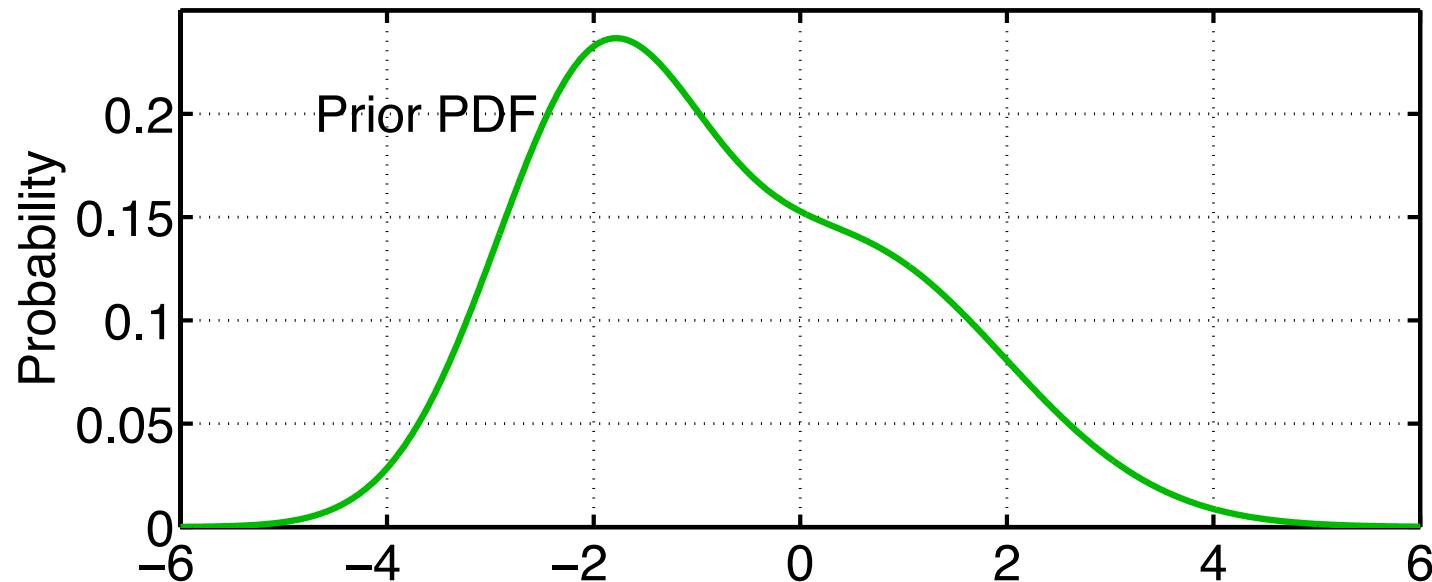
The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

©UCAR 2014

NCAR | National Center for  
UCAR | Atmospheric Research

# Bayes' Rule

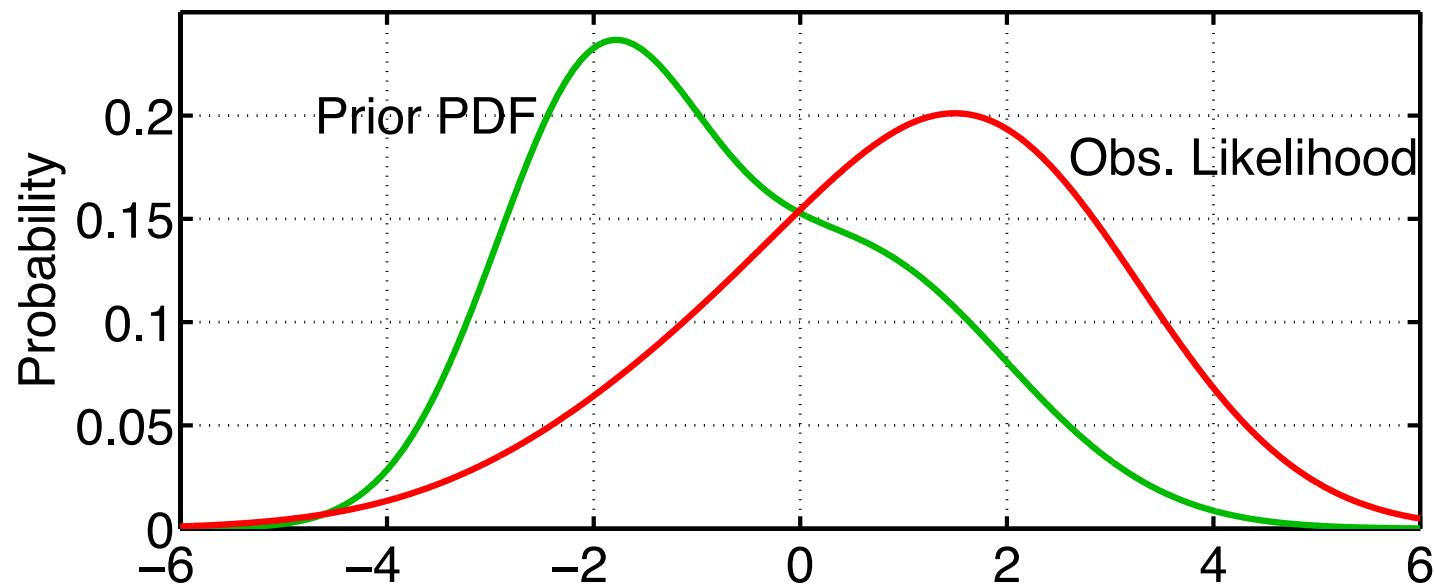
$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



- 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

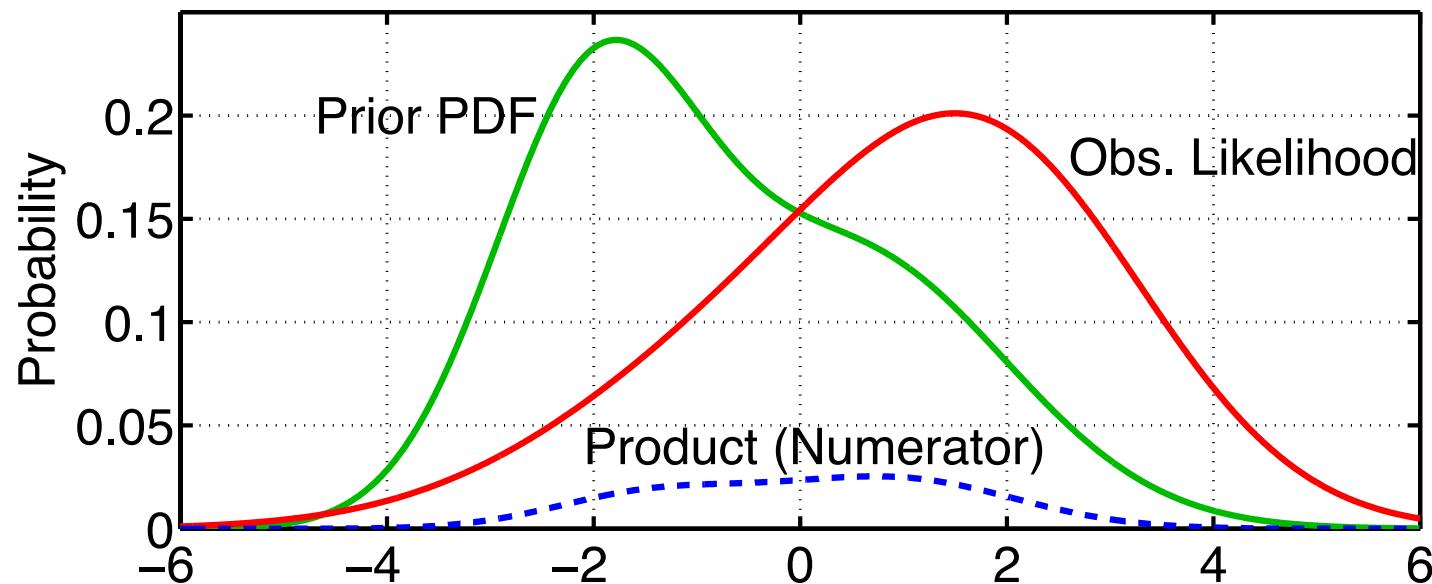
$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



- $A$  : Prior Estimate based on all previous information,  $C$ .  
 $\underline{B}$  : An additional observation.  
 $p(A|BC)$  : Posterior (updated estimate) based on  $C$  and  $B$ .

# Bayes' Rule

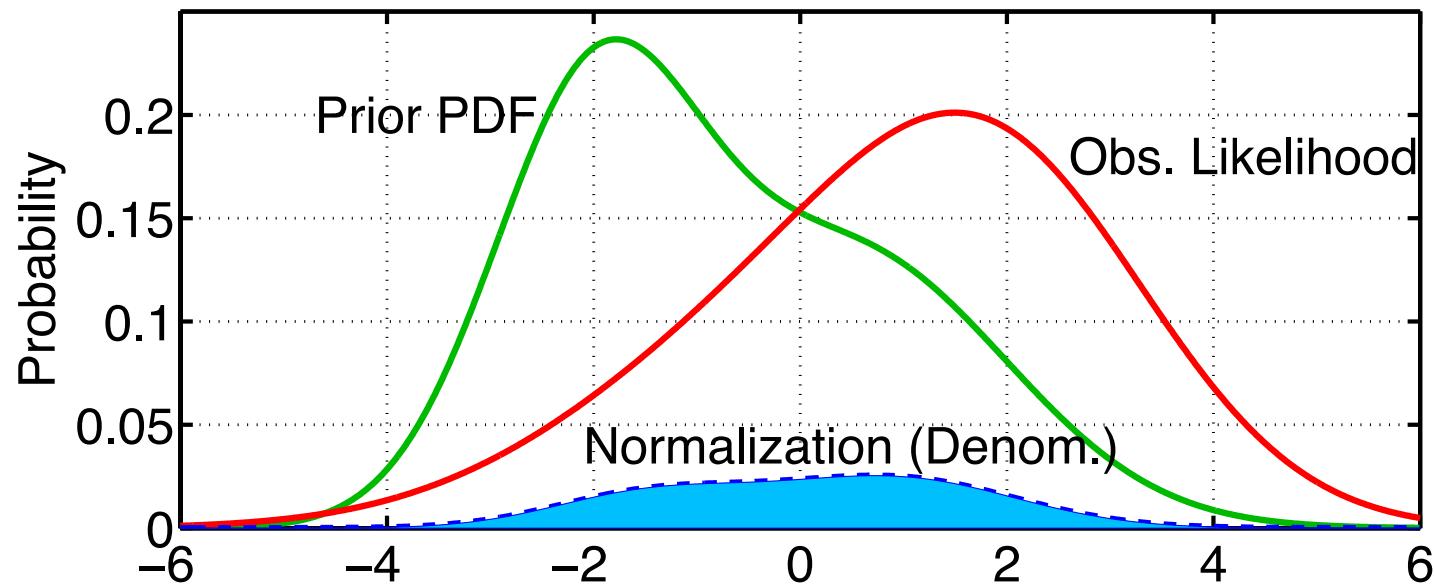
$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



- $A$  : 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

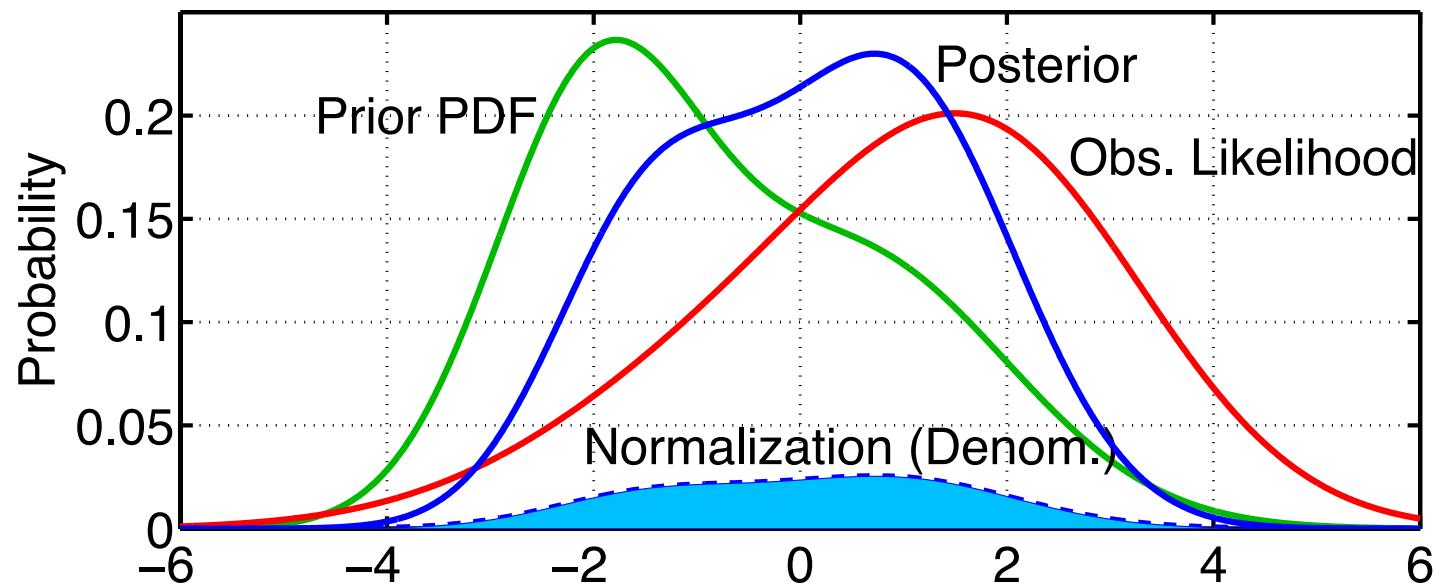
$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



- $A$  : 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

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



- $A$  : Prior Estimate based on all previous information,  $C$ .  
 $B$  : An additional observation.  
 $p(A|BC)$  : Posterior (updated estimate) based on  $C$  and  $B$ .

# Color Scheme

Green == Prior

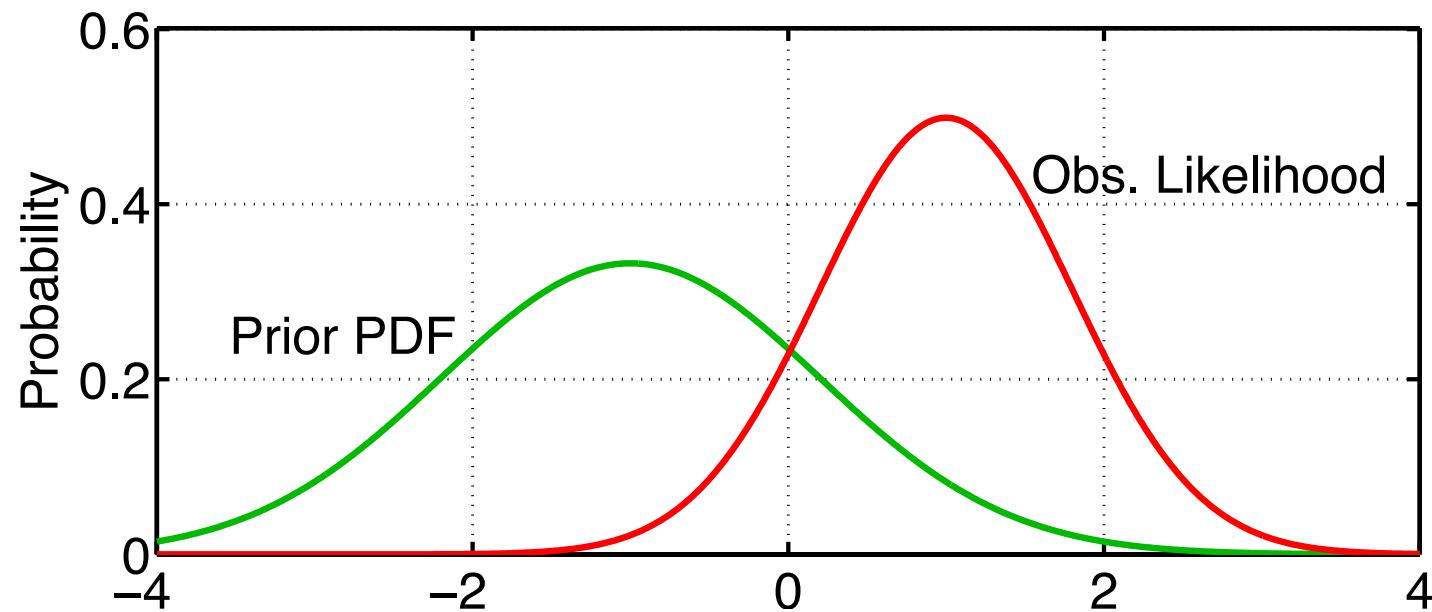
Red == Observation

Blue == Posterior

The same color scheme is used throughout ALL Tutorial materials.

# Product of Two Gaussians

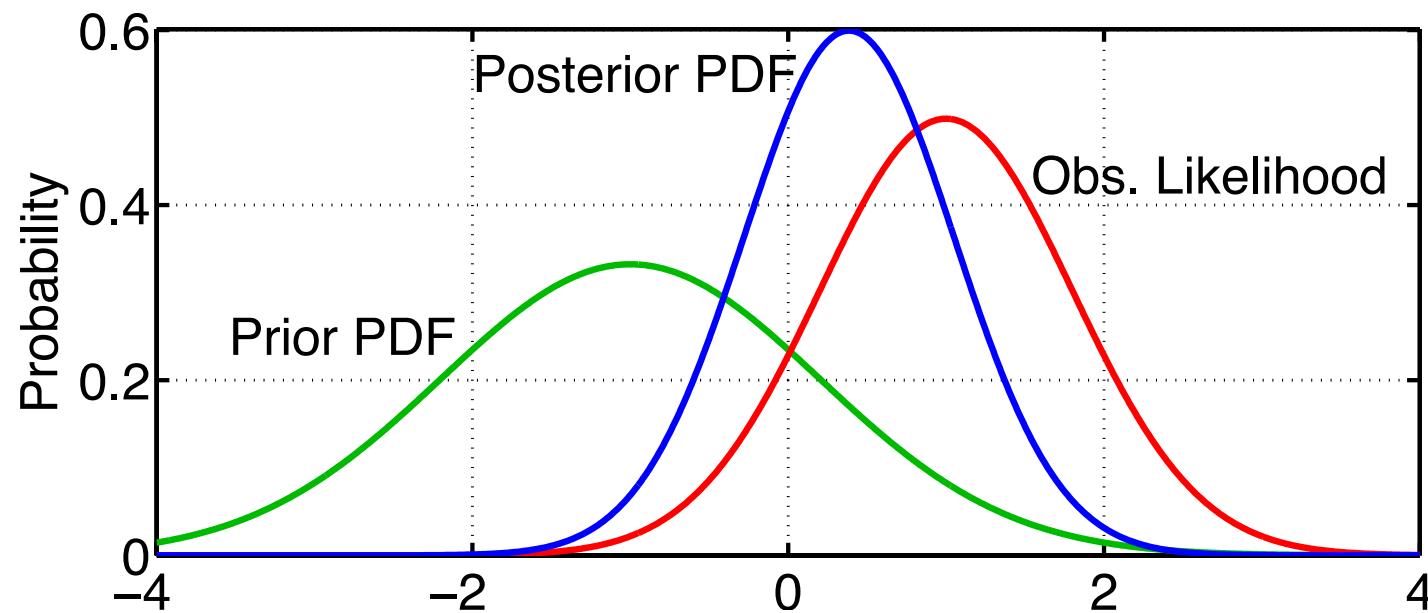
$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$



# 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.



# Product of Two Gaussians

Product of d-dimensional normals with means  $\mu_1$  and  $\mu_2$  and covariance matrices  $\Sigma_1$  and  $\Sigma_2$  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  $\mu_1$  and  $\mu_2$  and covariance matrices  $\Sigma_1$  and  $\Sigma_2$  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  $\mu_1$  and  $\mu_2$  and covariance matrices  $\Sigma_1$  and  $\Sigma_2$  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:  $c = \frac{1}{(2\pi)^{d/2} |\Sigma_1 + \Sigma_2|^{1/2}} \exp \left\{ -\frac{1}{2} \left[ (\mu_2 - \mu_1)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_2 - \mu_1) \right] \right\}$

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

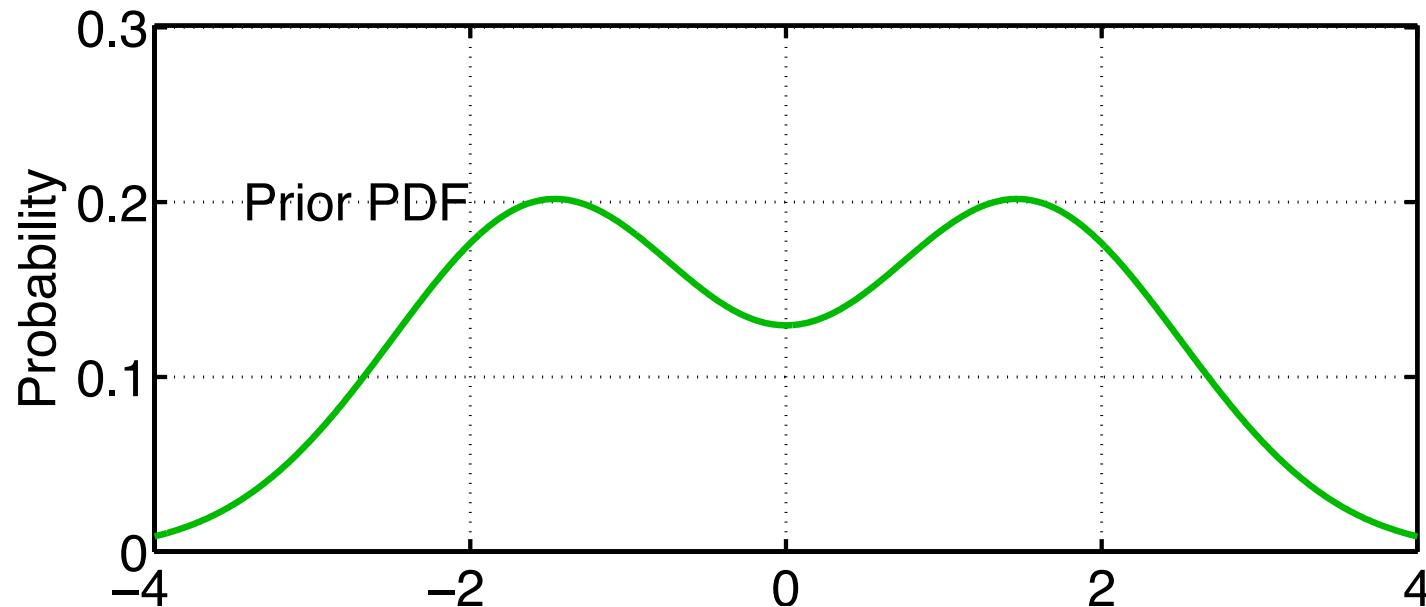
Weight:  $c = \frac{1}{(2\pi)^{d/2} |\Sigma_1 + \Sigma_2|^{1/2}} \exp \left\{ -\frac{1}{2} \left[ (\mu_2 - \mu_1)^T (\Sigma_1 + \Sigma_2)^{-1} (\mu_2 - \mu_1) \right] \right\}$

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

# Product of Two Gaussians

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

This product is closed for Gaussian distributions.

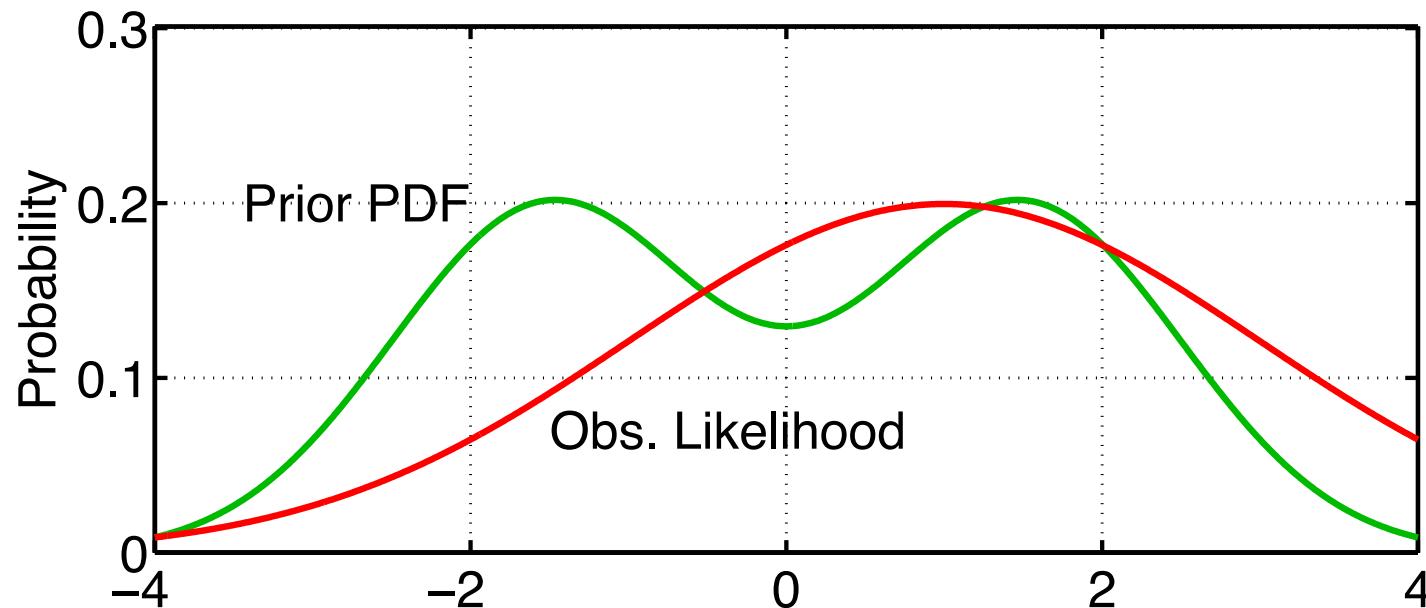


There are other families of functions for which it is closed ...  
But, for general distributions, there's no analytical product.

# Product of Two Gaussians

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

This product is closed for Gaussian distributions.

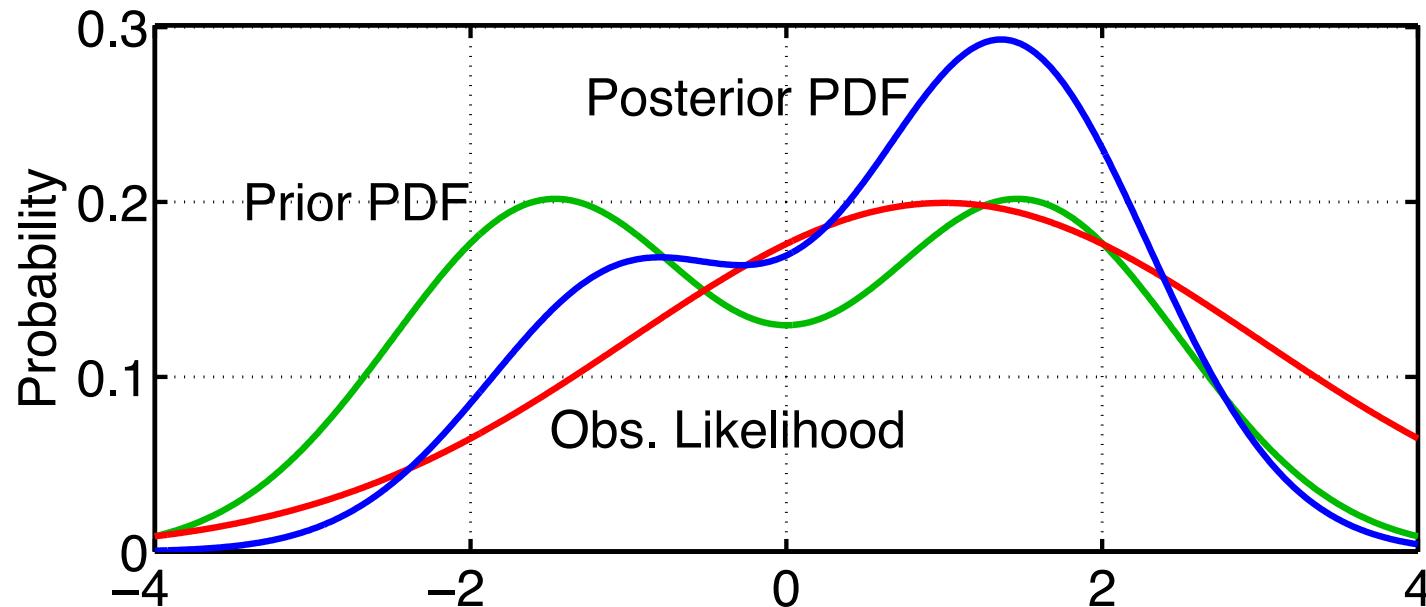


There are other families of functions for which it is closed ...  
But, for general distributions, there's no analytical product.

# 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.

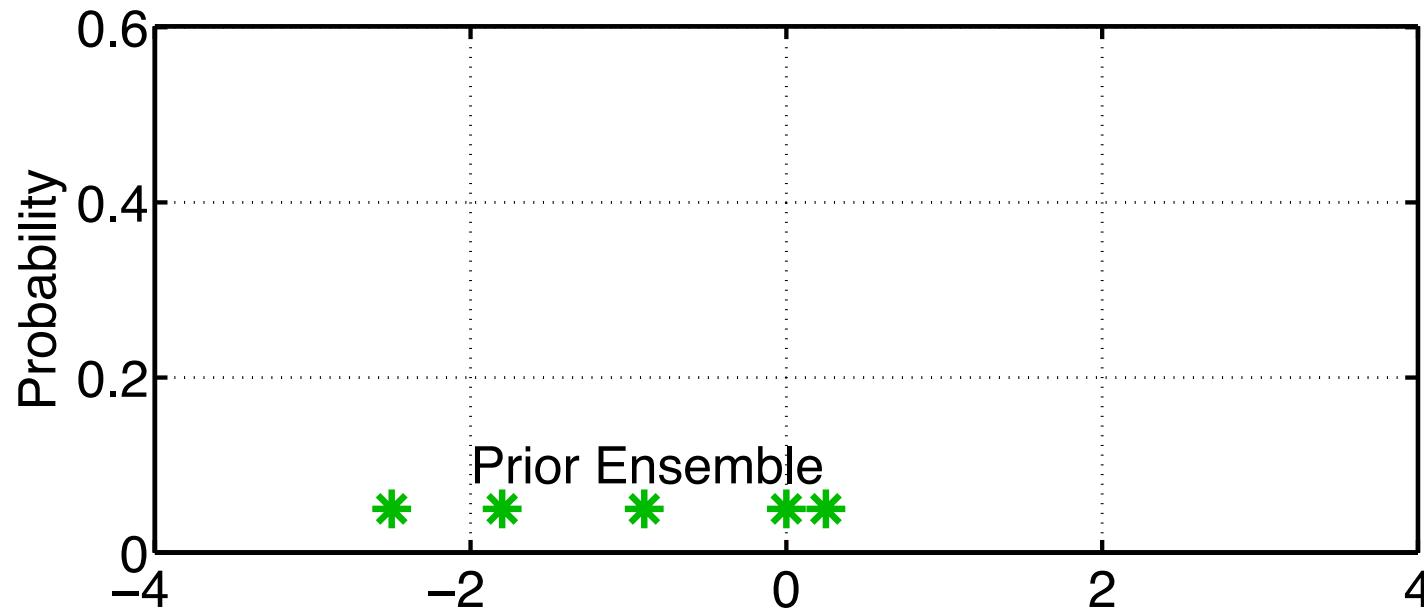


There are other families of functions for which it is closed ...  
But, for general distributions, there's no analytical product.

# 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.

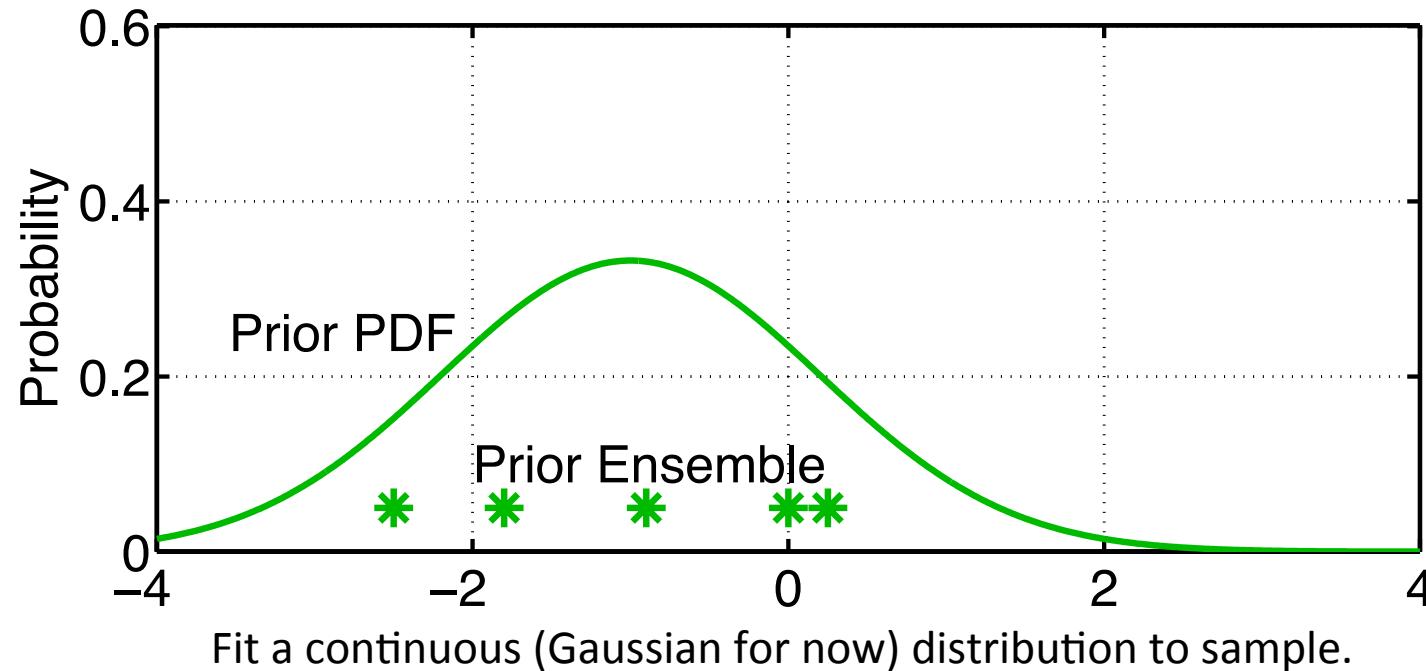


Don't know much about properties of this sample.  
May naively assume it is random draw from 'truth'.

# Product of Two Gaussians

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

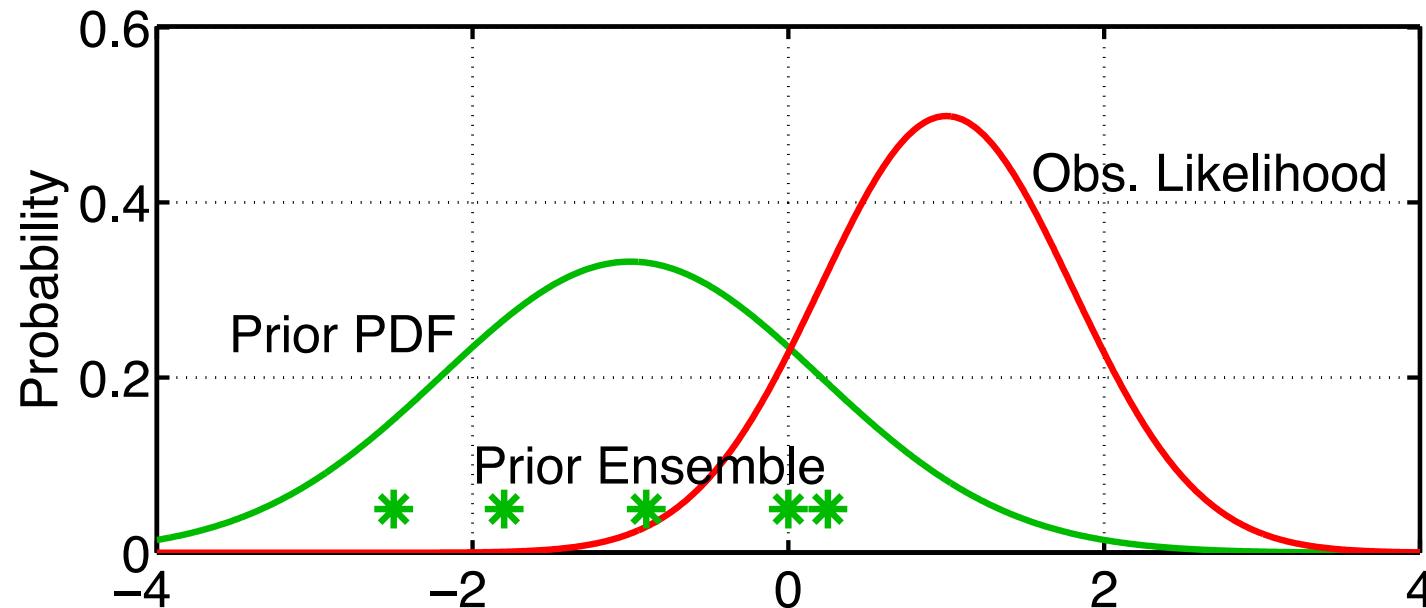
How can we take product of sample with continuous likelihood?



# Product of Two Gaussians

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

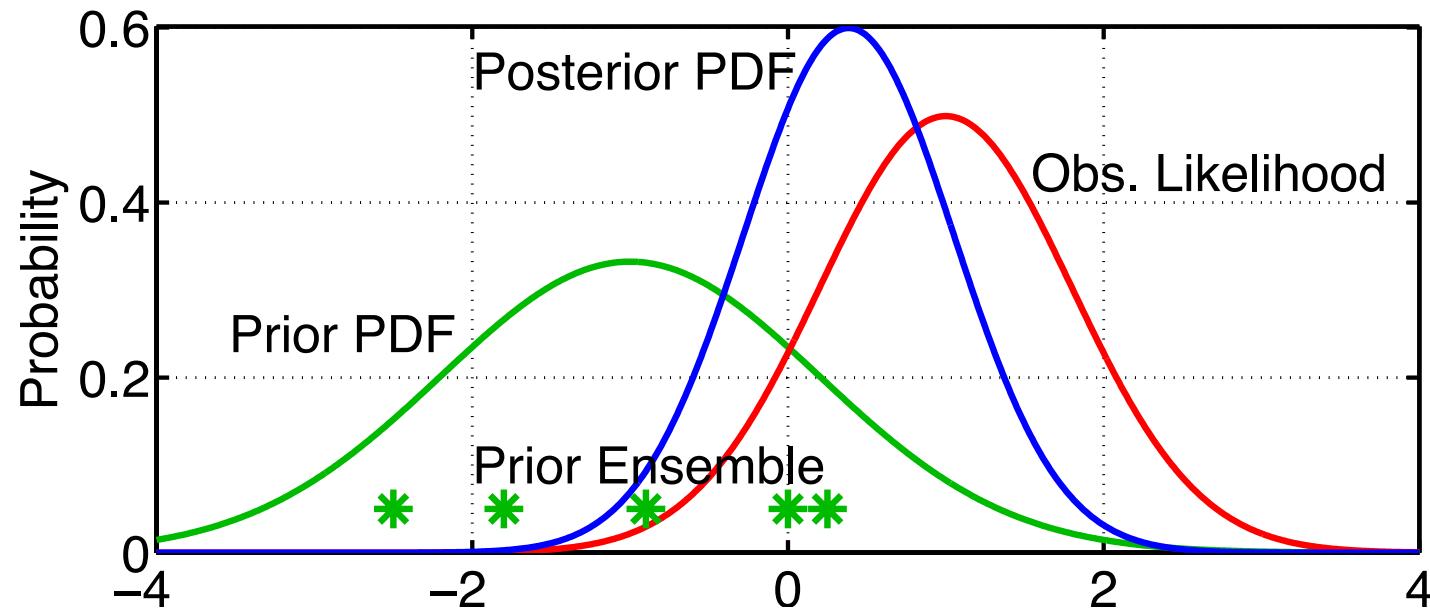
Observation likelihood usually continuous (nearly always Gaussian).



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

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

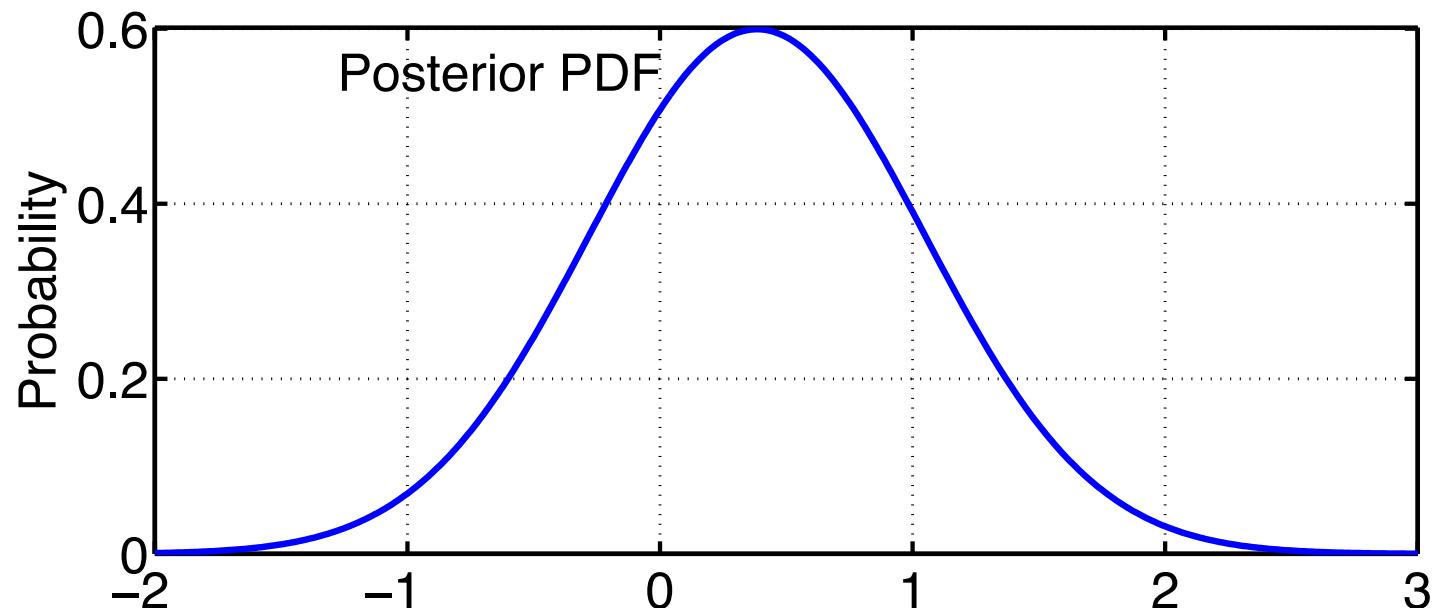
Product of prior Gaussian fit and Obs. likelihood is Gaussian.



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

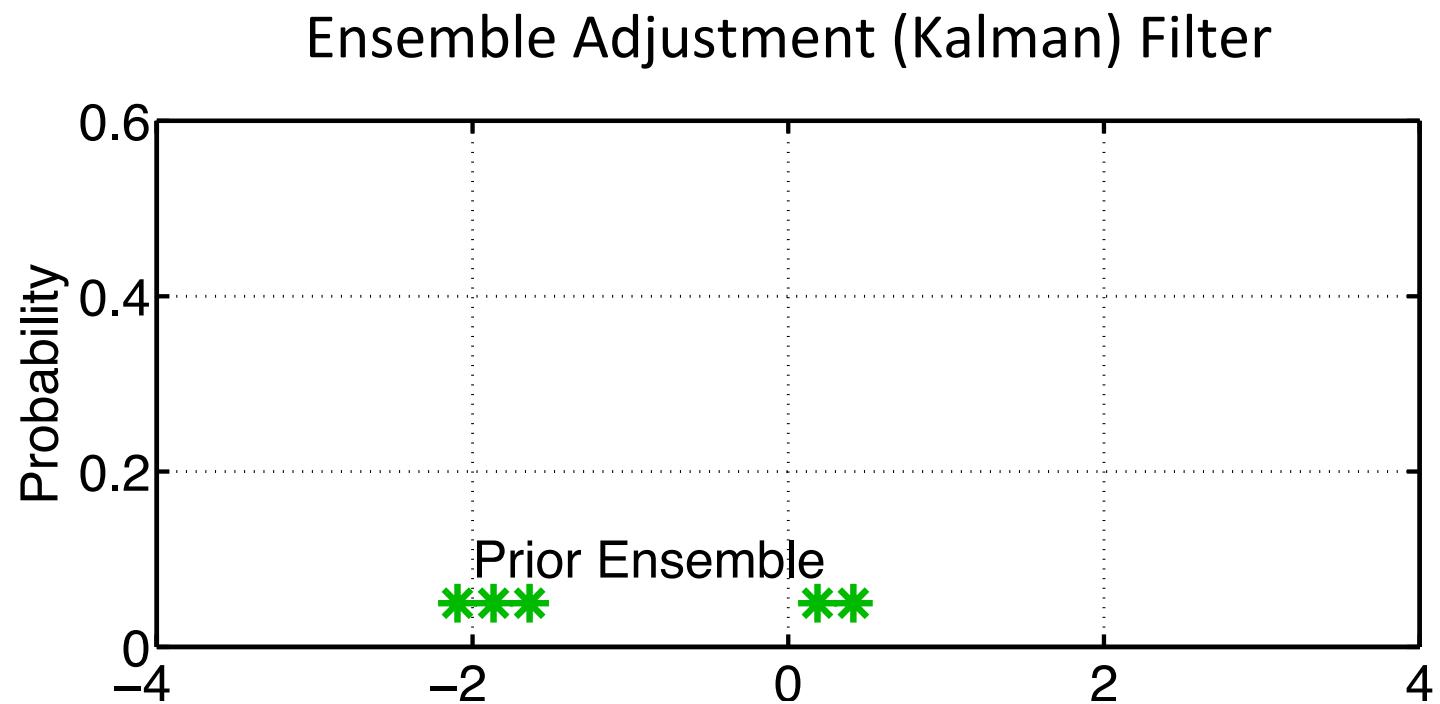
# Sampling Posterior PDF

There are many ways to do this.

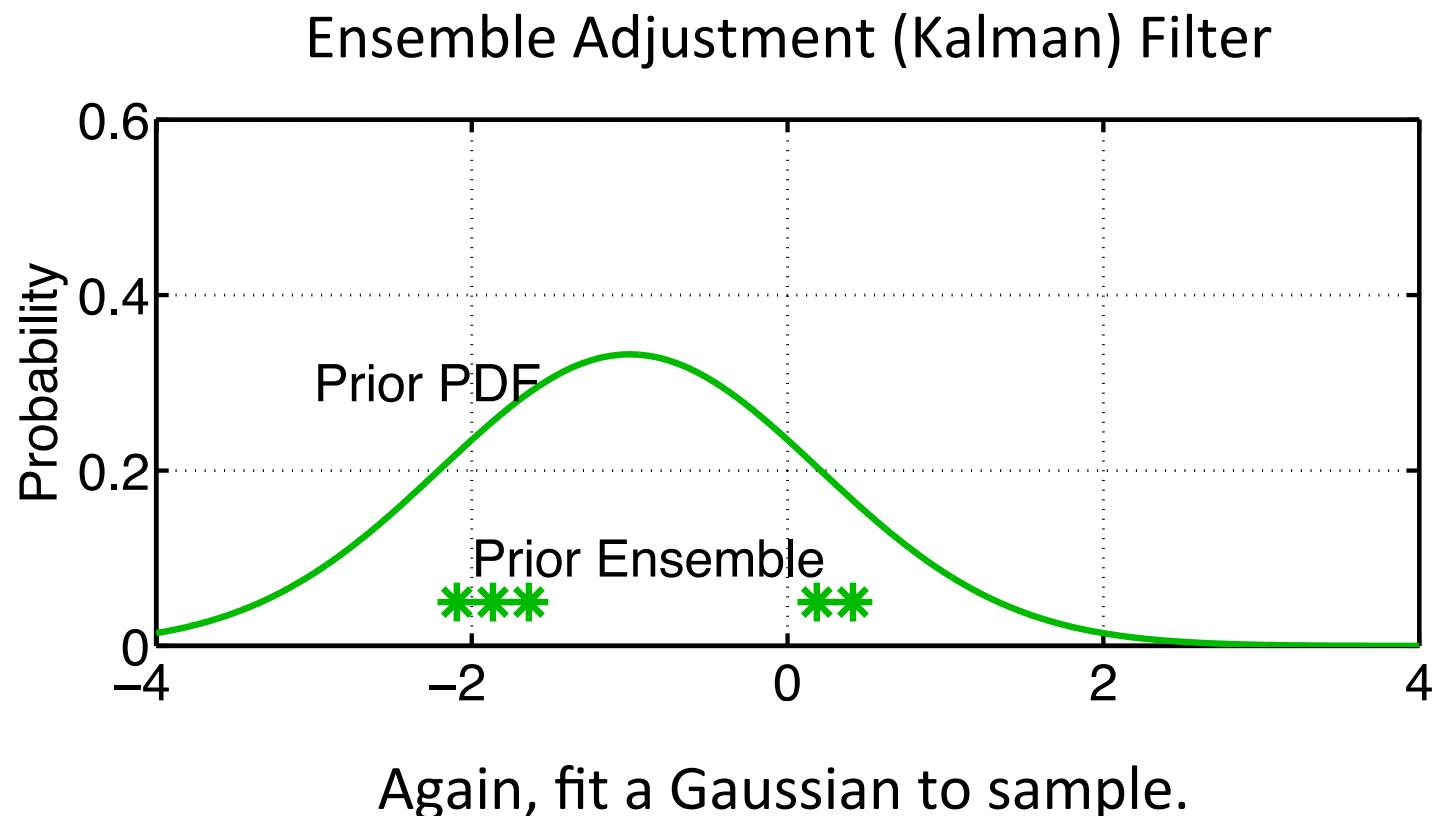


Exact properties of different methods may be unclear.  
Trial and error still best way to see how they perform.  
Will interact with properties of prediction models, etc.

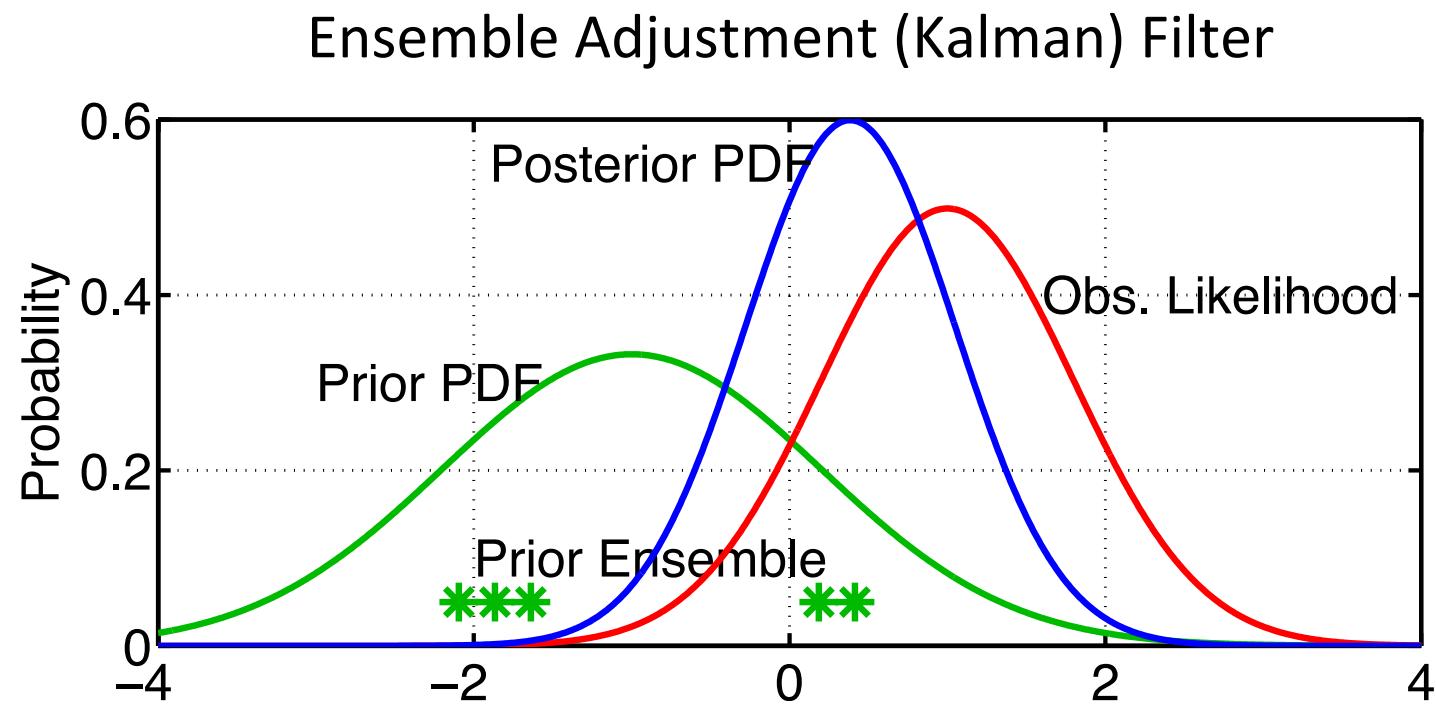
# Sampling Posterior PDF



# Sampling Posterior PDF

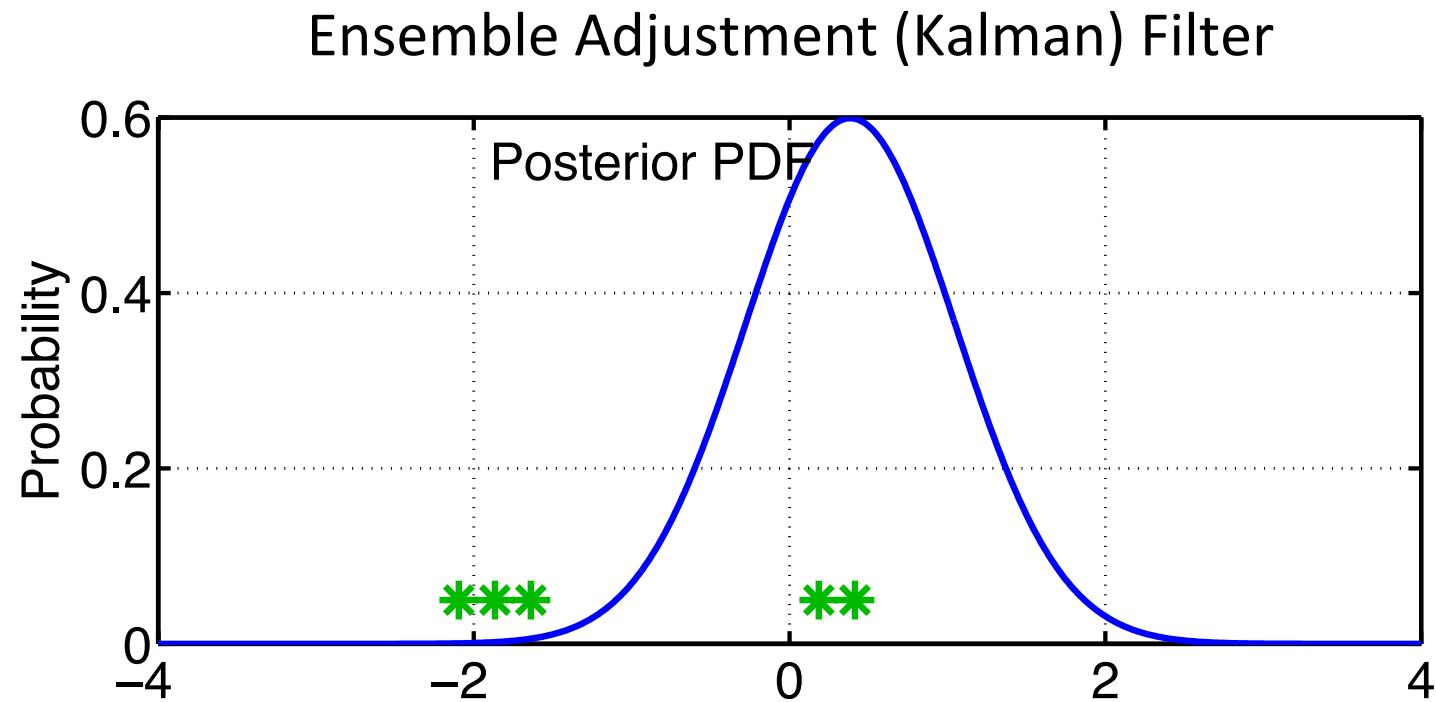


# Sampling Posterior PDF



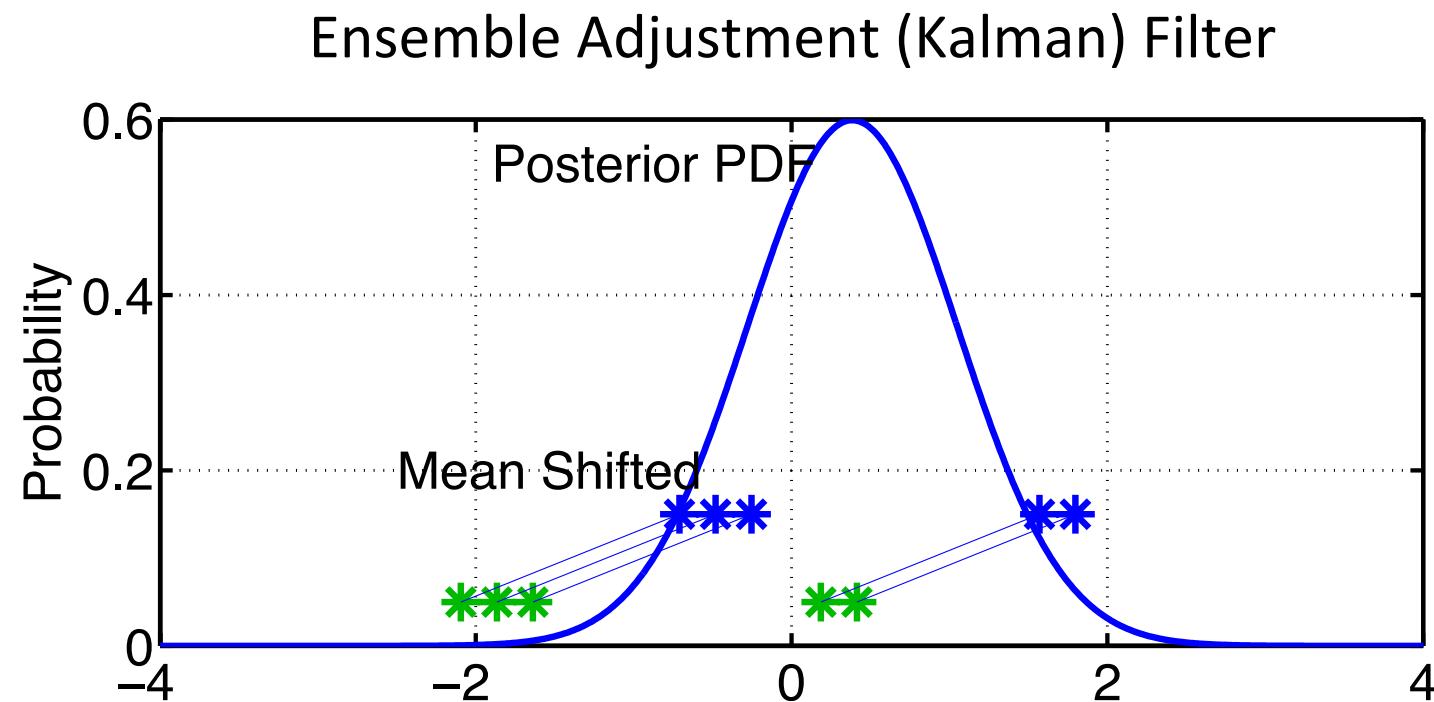
Compute posterior PDF (same as previous algorithms).

# Sampling Posterior PDF



Use deterministic algorithm to 'adjust' ensemble.

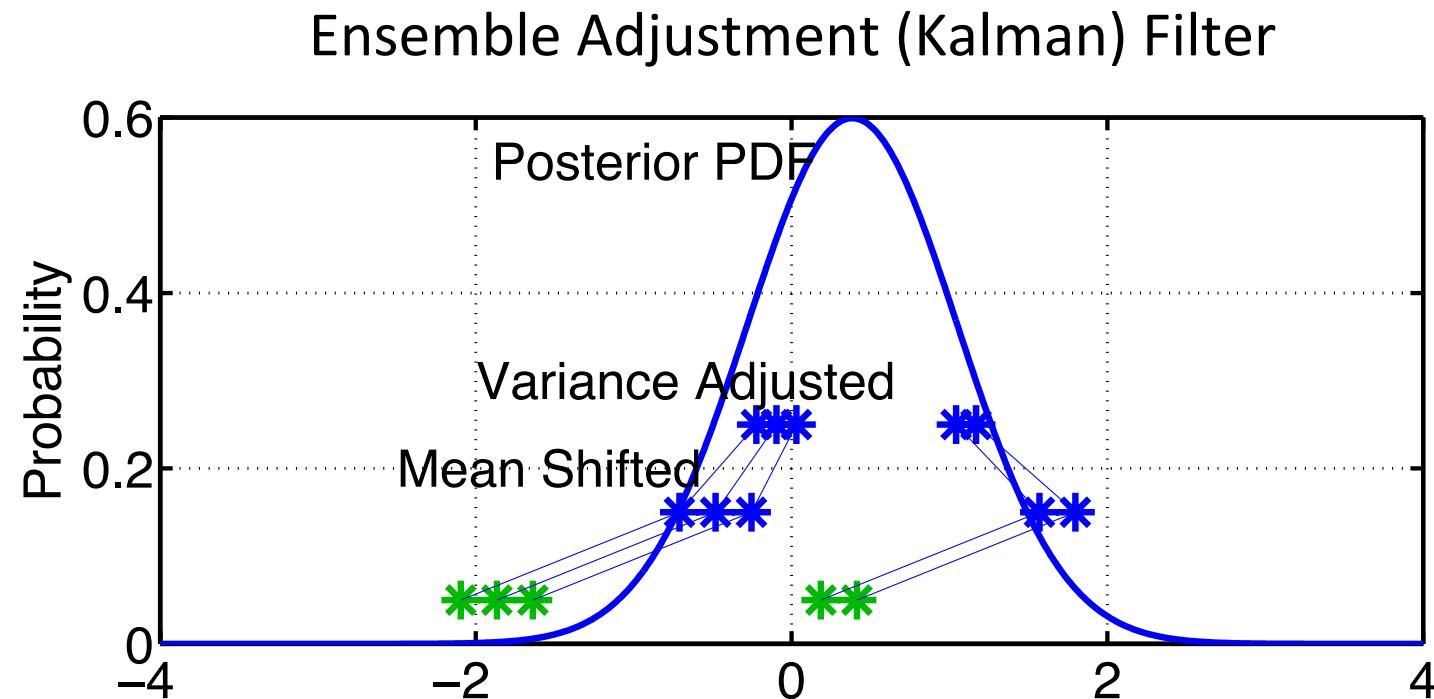
# Sampling Posterior PDF



Use deterministic algorithm to ‘adjust’ ensemble.

1. ‘Shift’ ensemble to have exact mean of posterior.

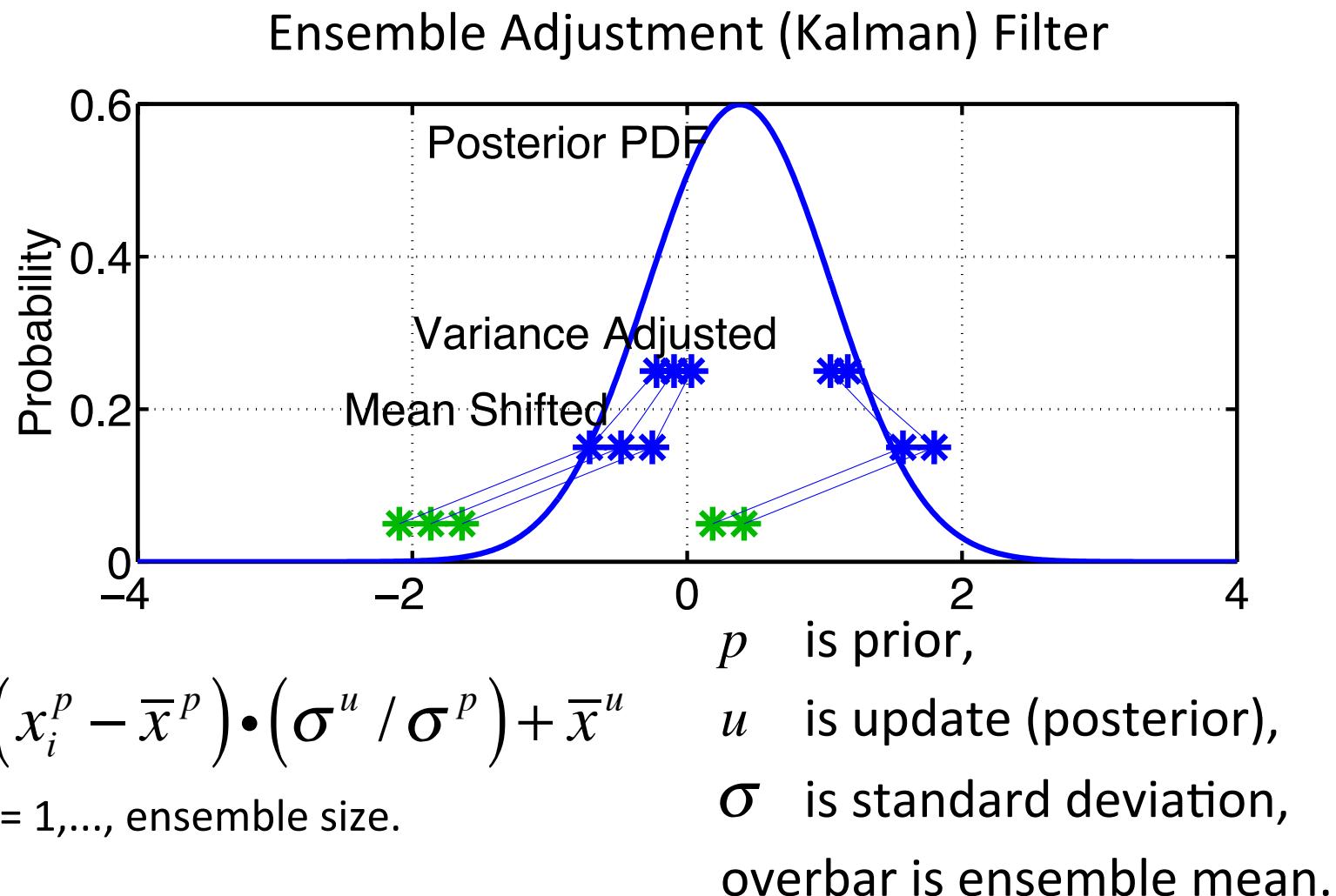
# Sampling Posterior PDF



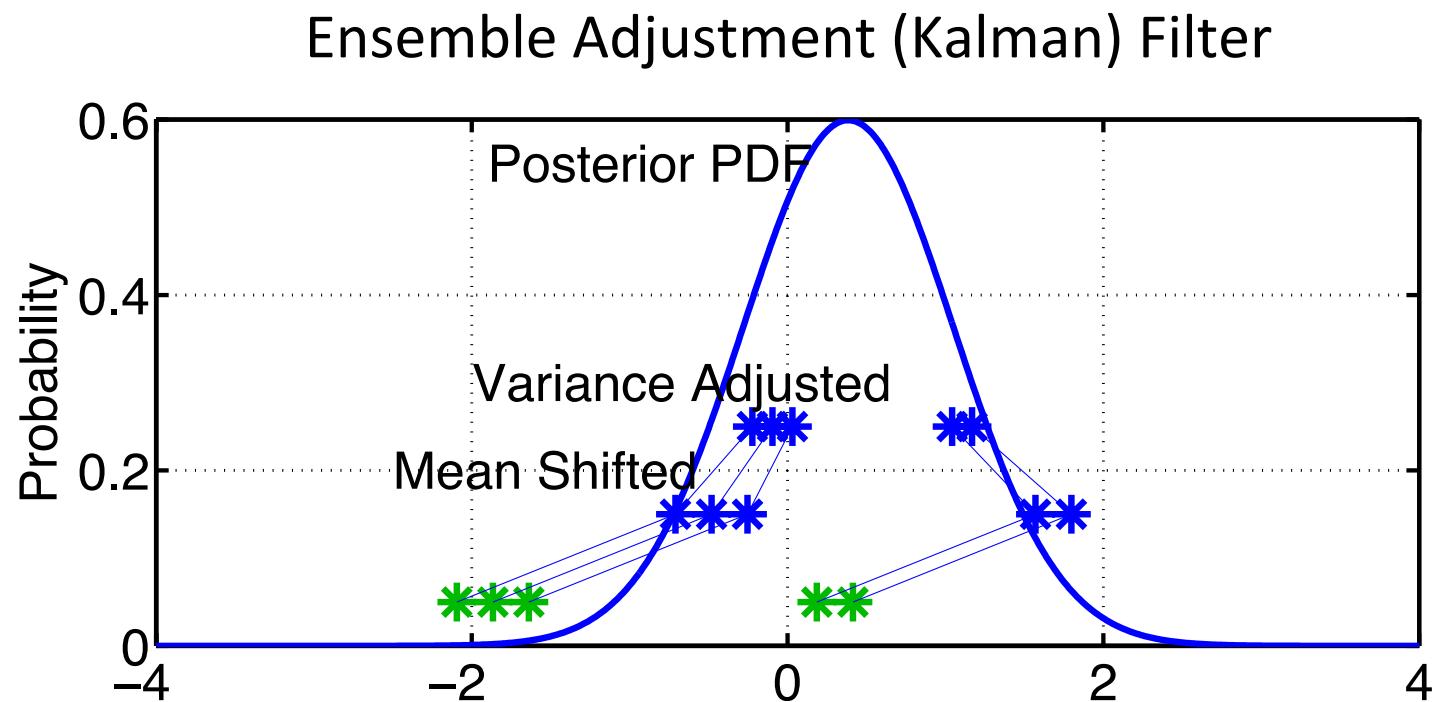
Use deterministic algorithm to ‘adjust’ ensemble.

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

# Sampling Posterior PDF

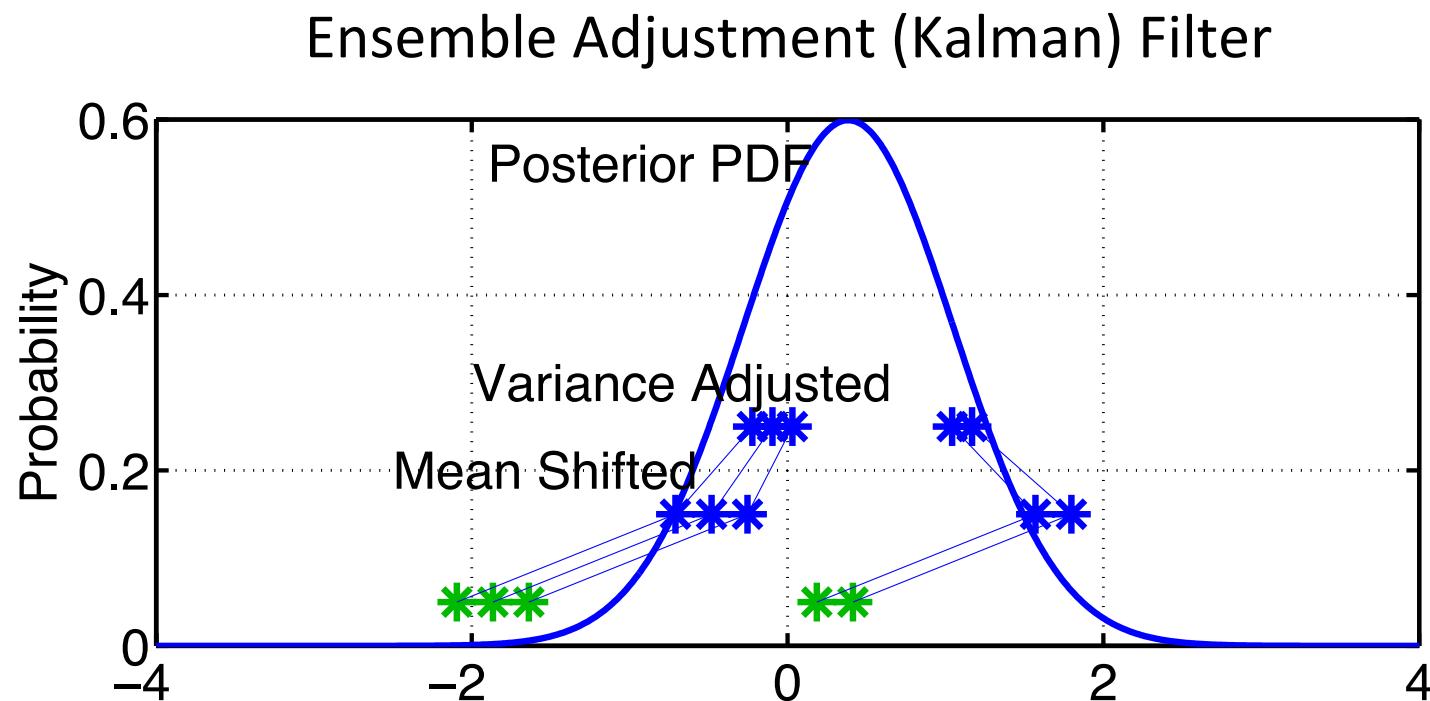


# Sampling Posterior PDF



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.

# 1<sup>st</sup> 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).

# 1<sup>st</sup> look at DART Diagnostics

Try the following matlab commands:

***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.)**

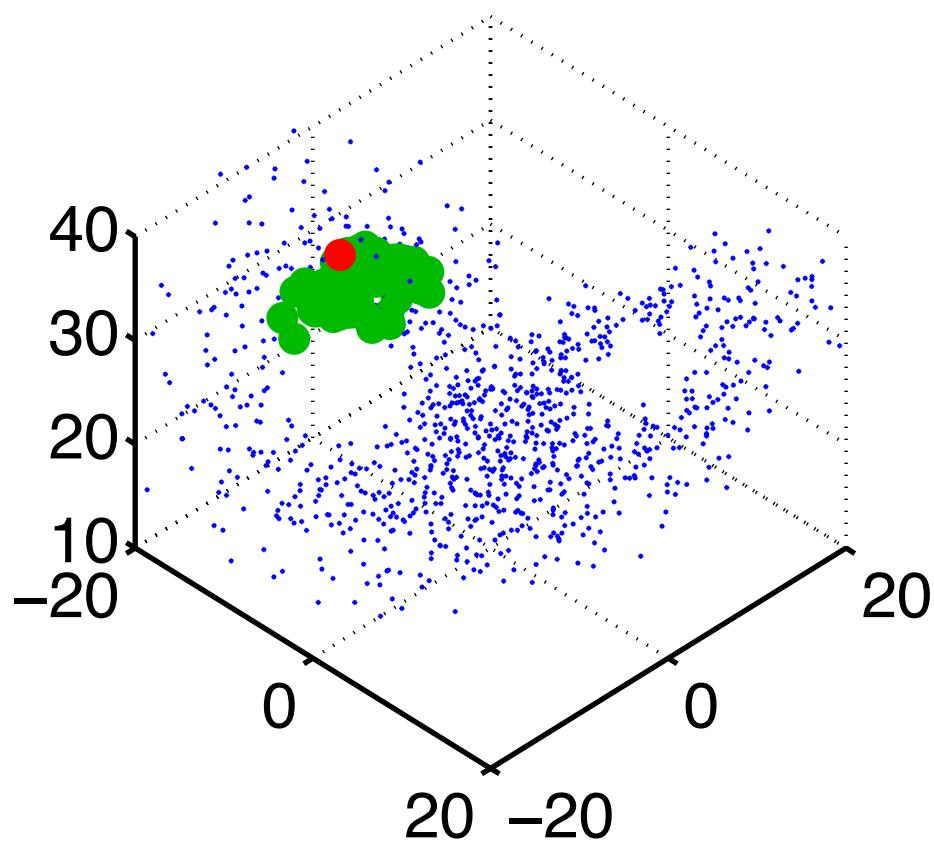
***plot\_ens\_mean\_time\_series***

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.

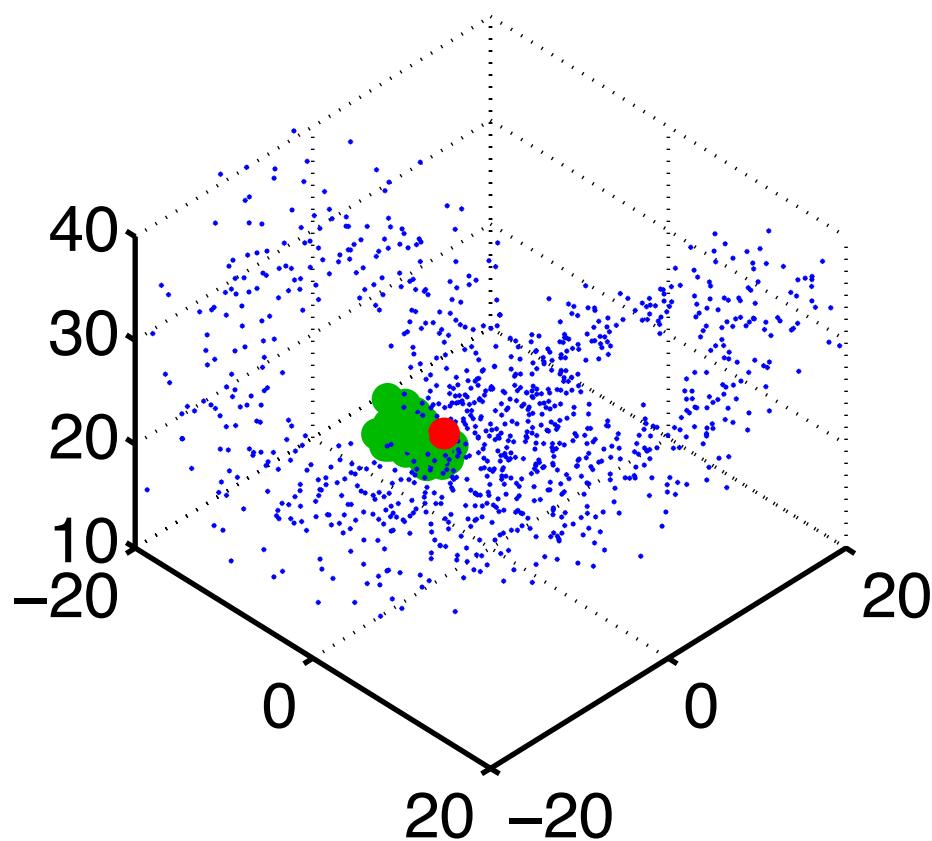
Prior ensemble in green.

Observing all three state variables.

Obs. Error variance = 4.0.

Four 20-member ensembles.

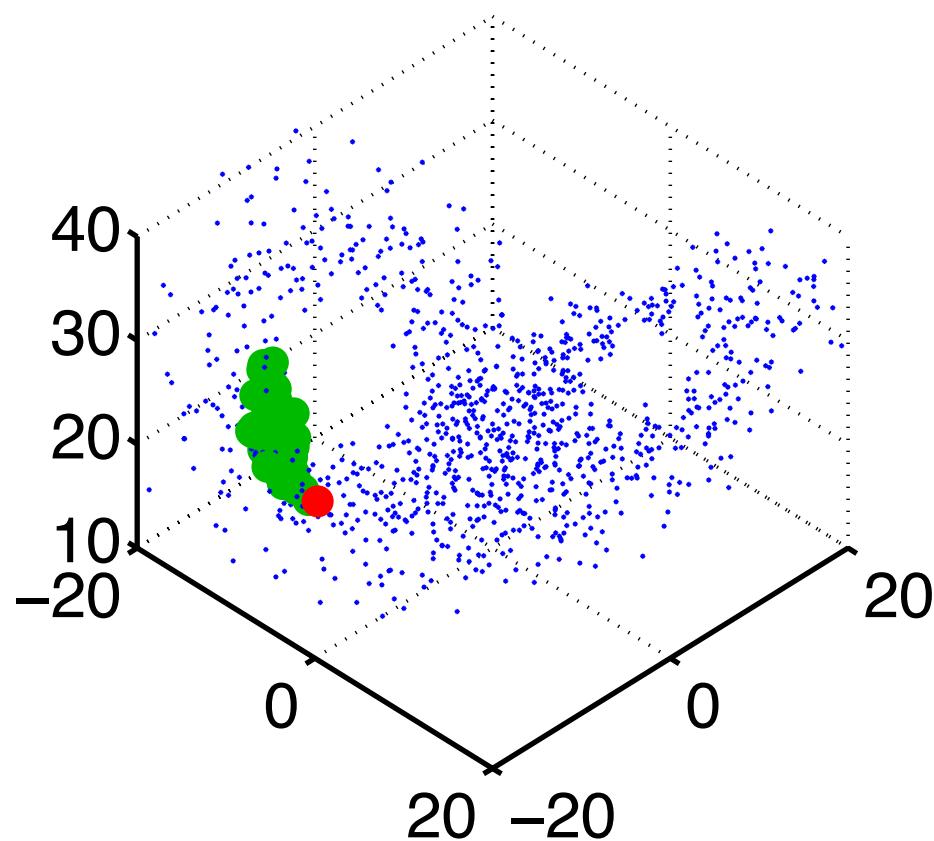
# Simple Example: Lorenz-63 3-variable chaotic model



Observation in red.

Prior ensemble in green.

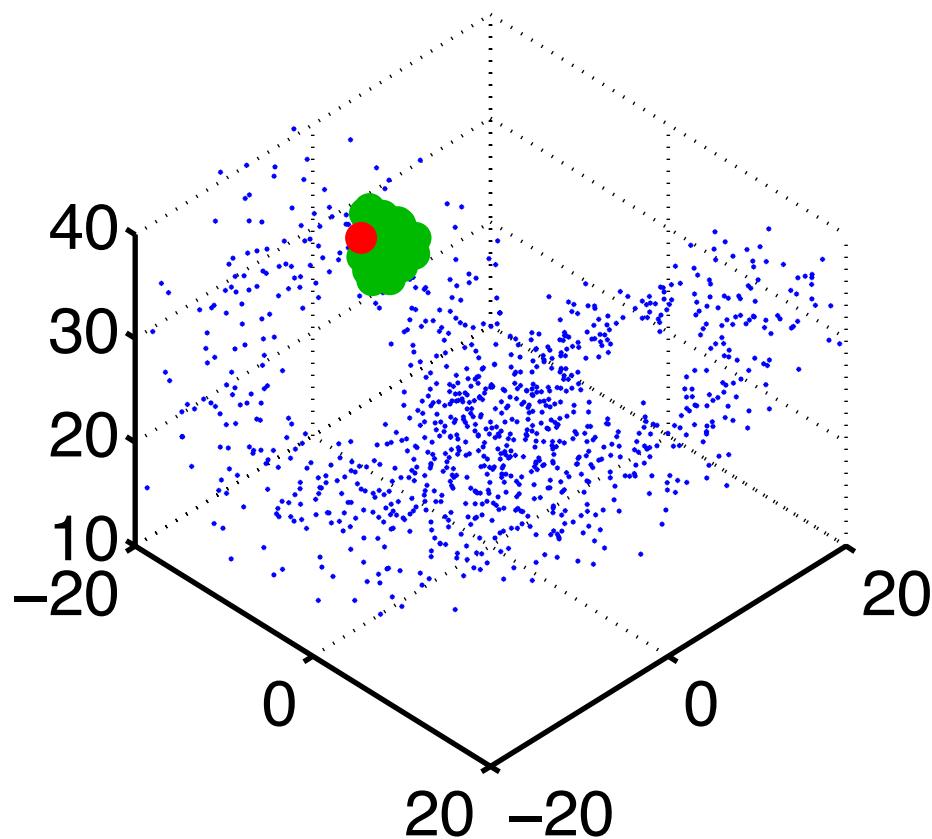
# Simple Example: Lorenz-63 3-variable chaotic model



Observation in red.

Prior ensemble in green.

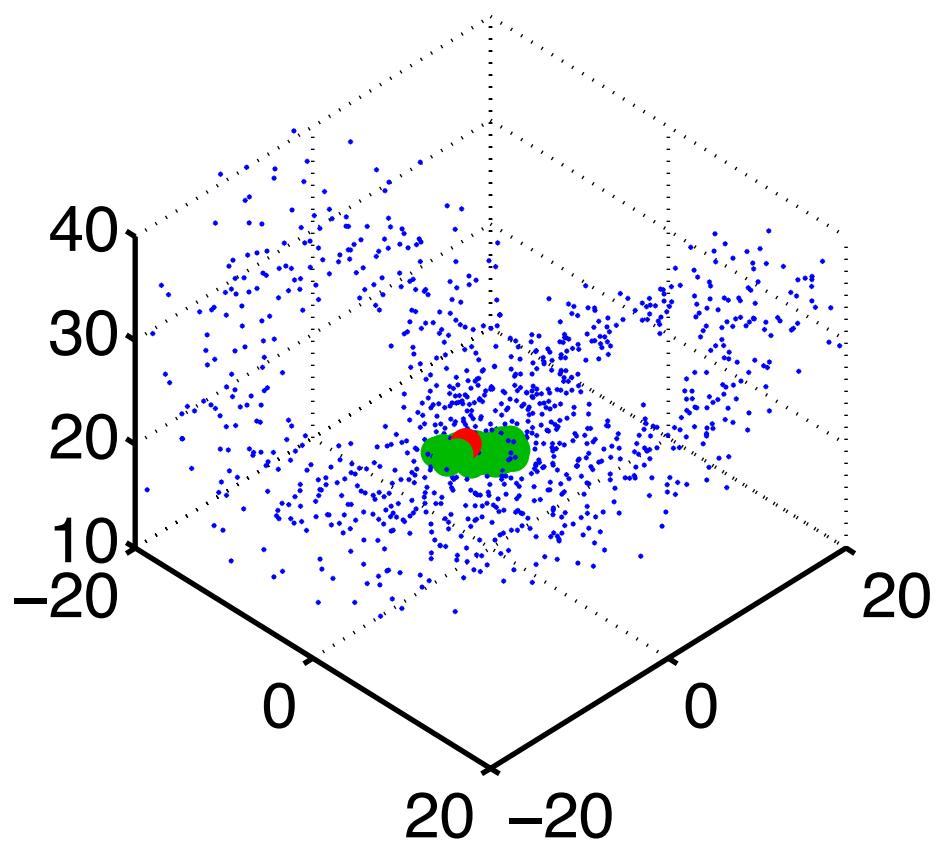
# Simple Example: Lorenz-63 3-variable chaotic model



Observation in red.

Prior ensemble in green.

# Simple Example: Lorenz-63 3-variable chaotic model

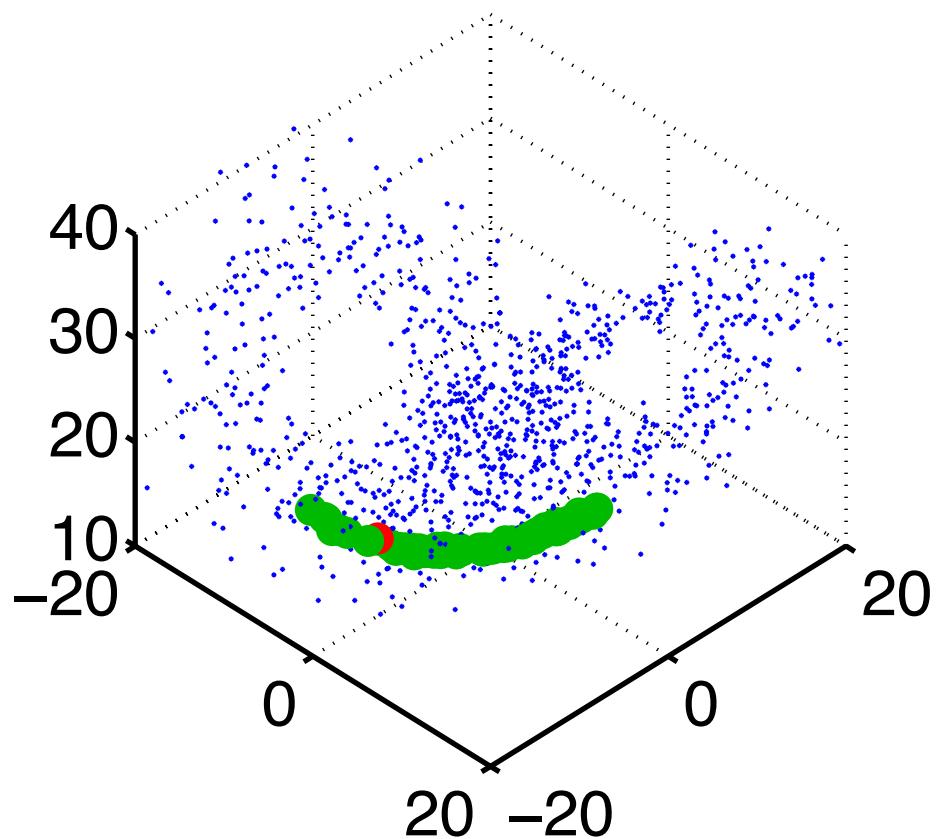


Observation in red.

Prior ensemble in green.

Ensemble is passing through an unpredictable region.

# Simple Example: Lorenz-63 3-variable chaotic model

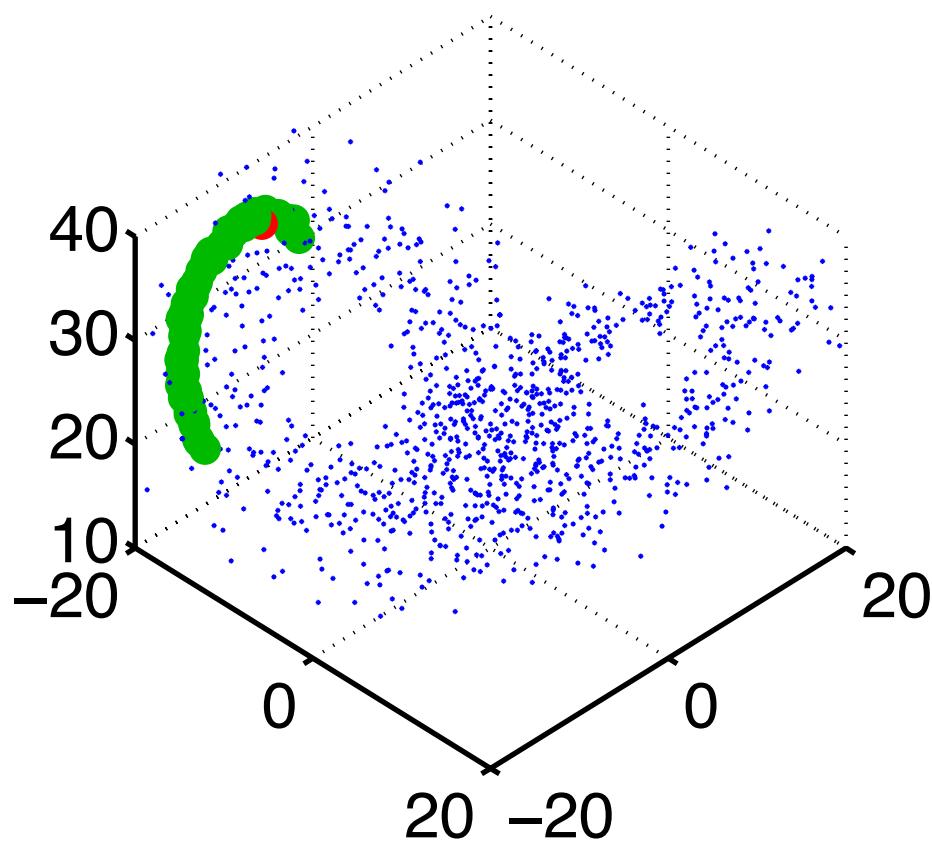


Observation in red.

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

1. Filtering For a One Variable System
2. The DART Directory Tree
3. DART Runtime Control and Documentation
4. How should observations of a state variable impact an unobserved state variable?  
Multivariate assimilation.
5. Comprehensive Filtering Theory: Non-Identity Observations and the Joint Phase Space
6. Other Updates for An Observed Variable
7. Some Additional Low-Order Models
8. Dealing with Sampling Error
9. More on Dealing with Error; Inflation
10. Regression and Nonlinear Effects
11. Creating DART Executables
12. Adaptive Inflation
13. Hierarchical Group Filters and Localization
14. Quality control
15. DART Experiments: Control and Design
16. Diagnostic Output
17. Creating Observation Sequences
18. Lost in Phase Space: The Challenge of Not Knowing the Truth
19. DART-Compliant Models and Making Models Compliant
20. Model Parameter Estimation
21. Observation Types and Observing System Design
22. Parallel Algorithm Implementation
23. Location module design (not available)
24. Fixed lag smoother (not available)