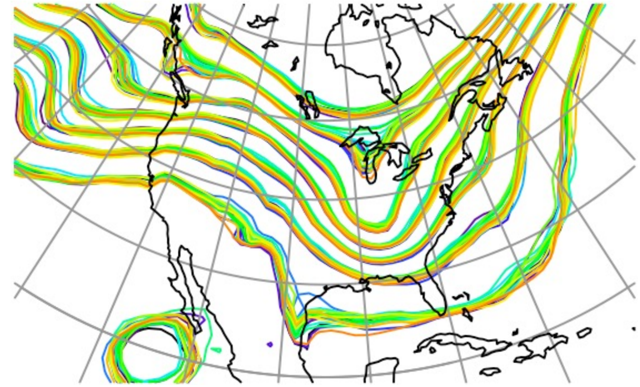
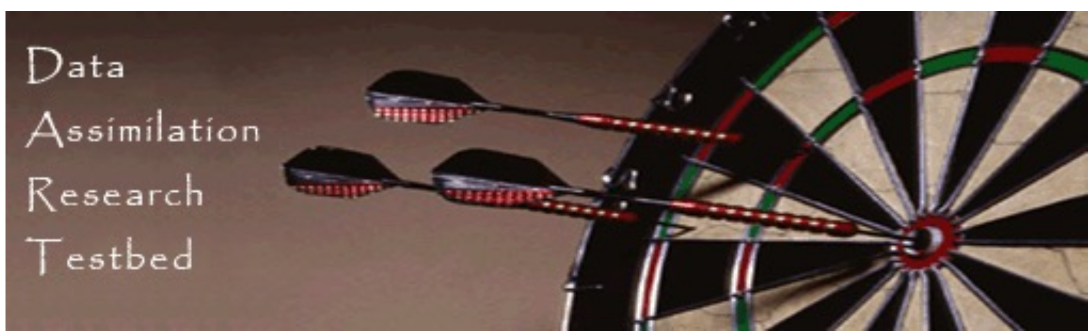
 

A General Ensemble Filtering Framework Using Quantiles

Jeff Anderson, NCAR Data Assimilation Research Section
PRE-RELEASE, Please do NOT Share without permission



Outline

- Review of Ensemble Kalman Filters

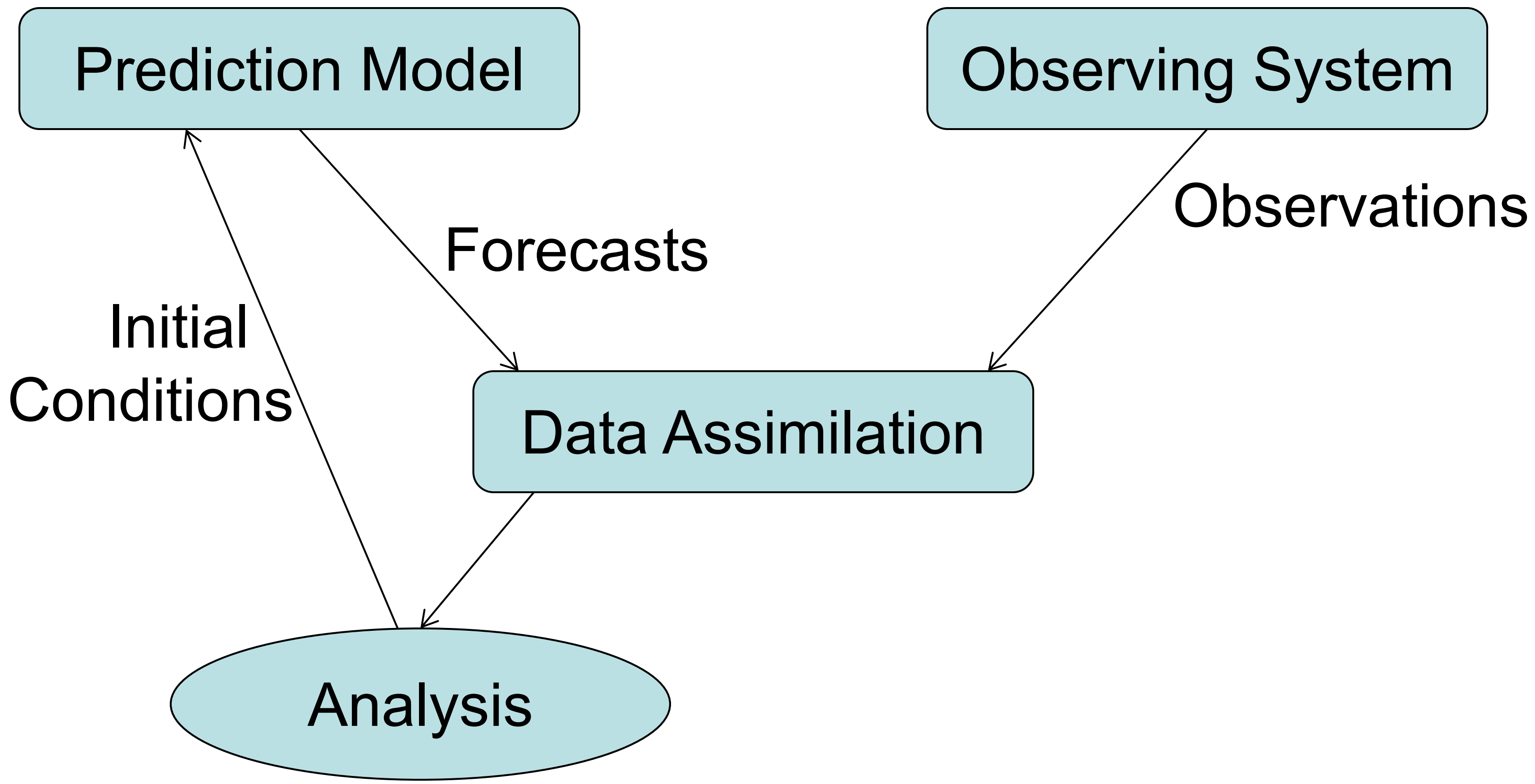
- General Ensemble Filtering Framework using Quantiles
  - Can fit arbitrary continuous prior to ensemble
  - Product with a likelihood gives continuous posterior
  - Posterior ensemble with same quantiles as prior

- Useful families for continuous priors and likelihoods

- Idealized examples for priors useful for earth science
  - Normal: identical to existing ensemble Kalman filters
  - Rank histogram
  - Sum of normal kernels
  - Gamma / inverse gamma
  - Beta

- Extension to multivariate application

Building a Forecast System with Data Assimilation



Schematic of a Sequential Ensemble Filter

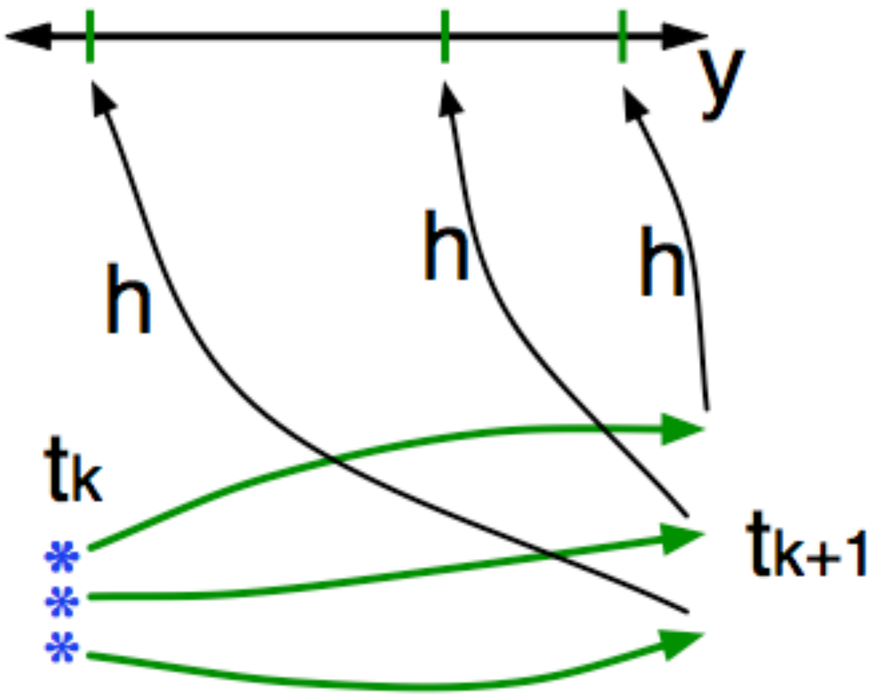
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)

Schematic of a Sequential Ensemble Filter

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

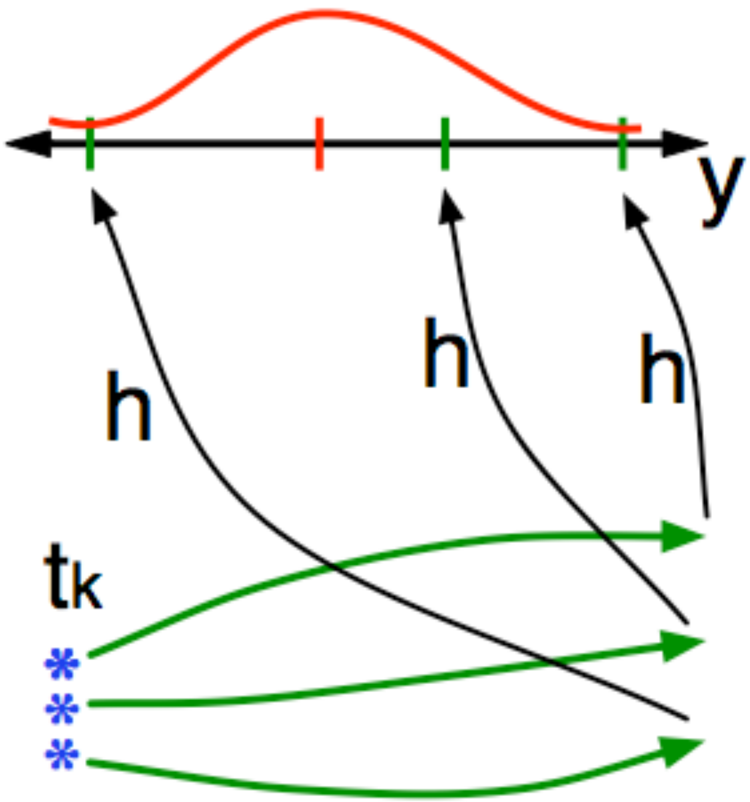


Theory: observations from instruments with uncorrelated errors can be done sequentially.

Can think about single observation without (too much) loss of generality.

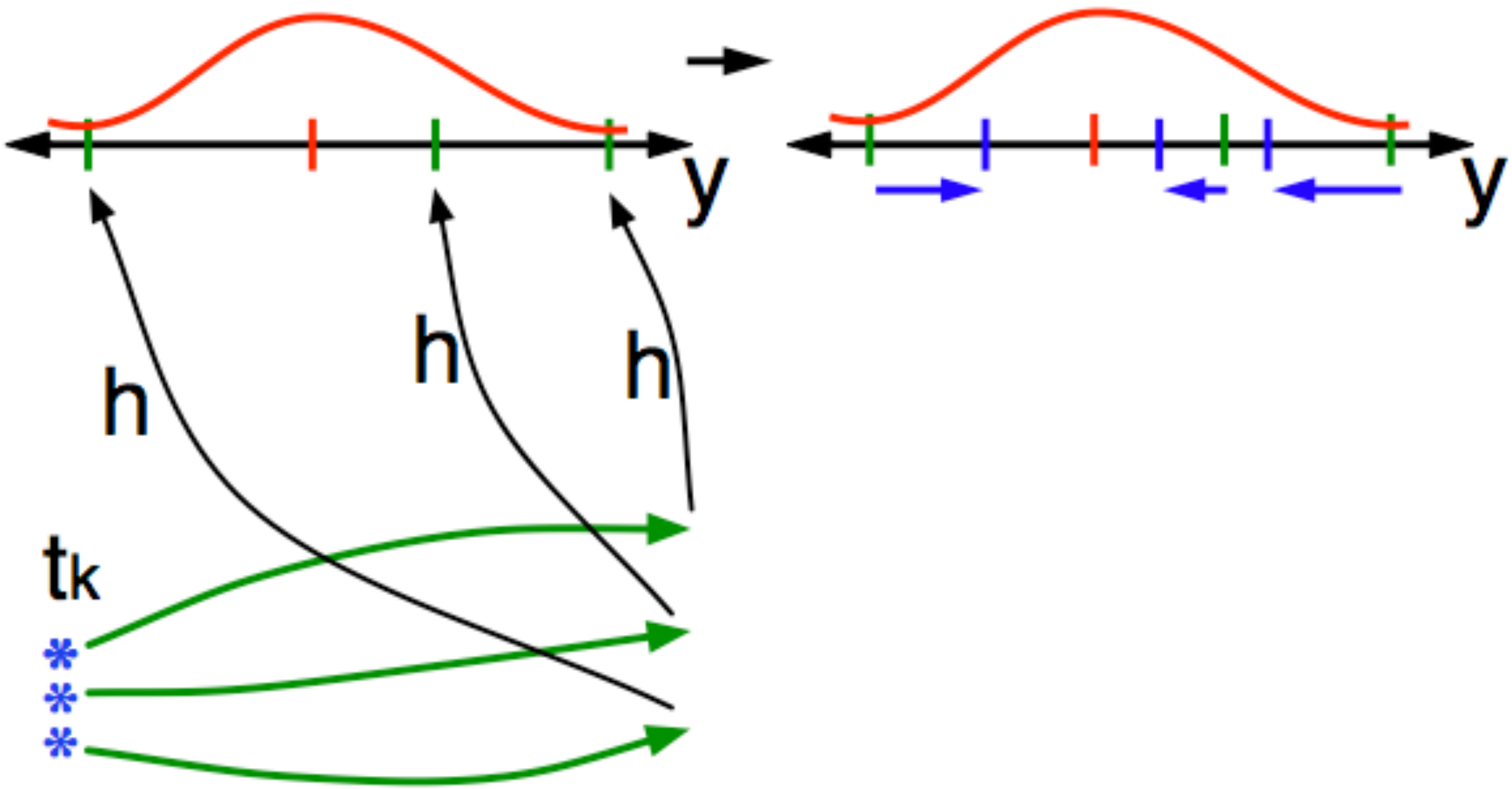
Schematic of a Sequential Ensemble Filter

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



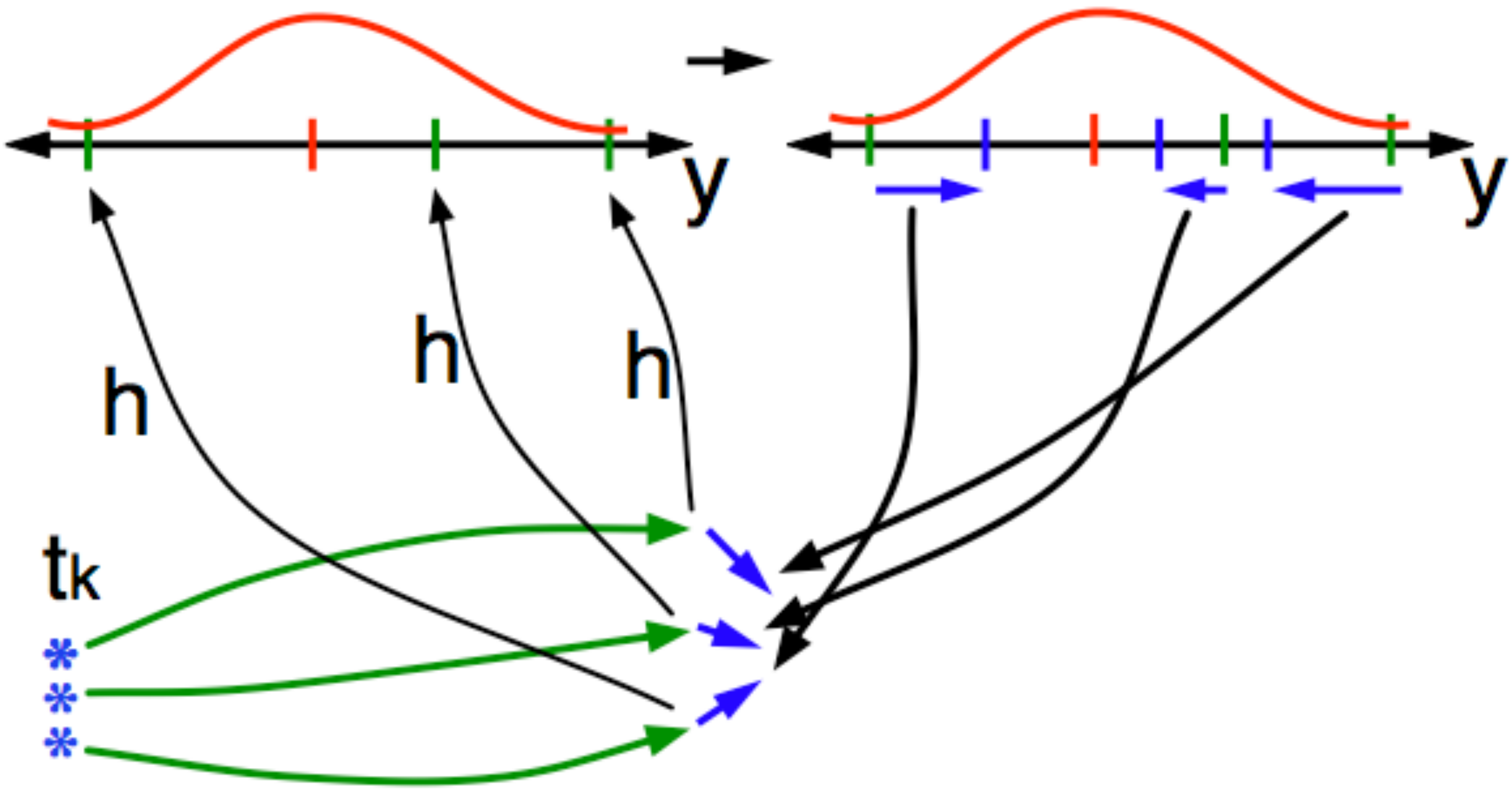
Schematic of a Sequential Ensemble Filter

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



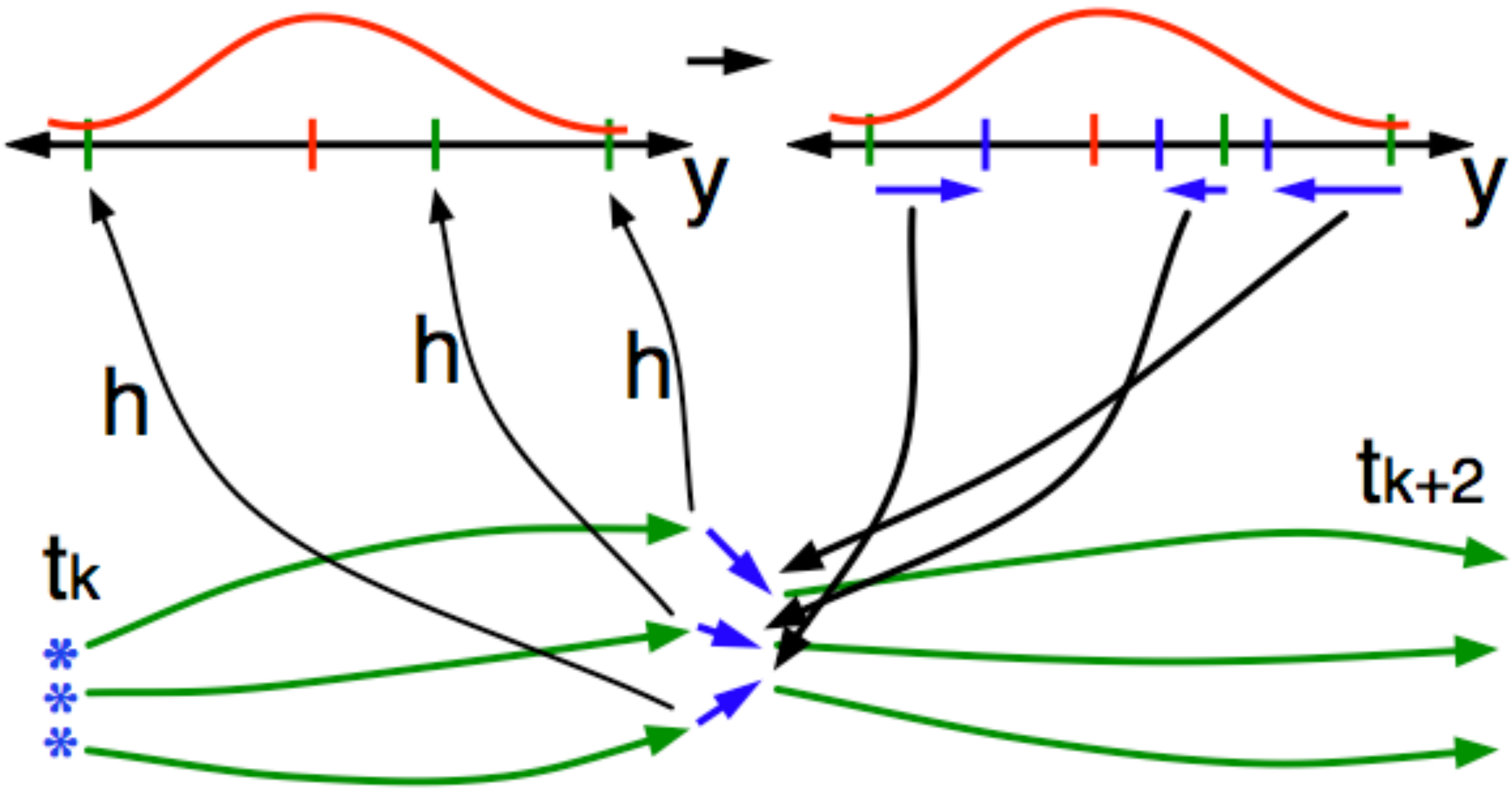
Schematic of a Sequential Ensemble Filter

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



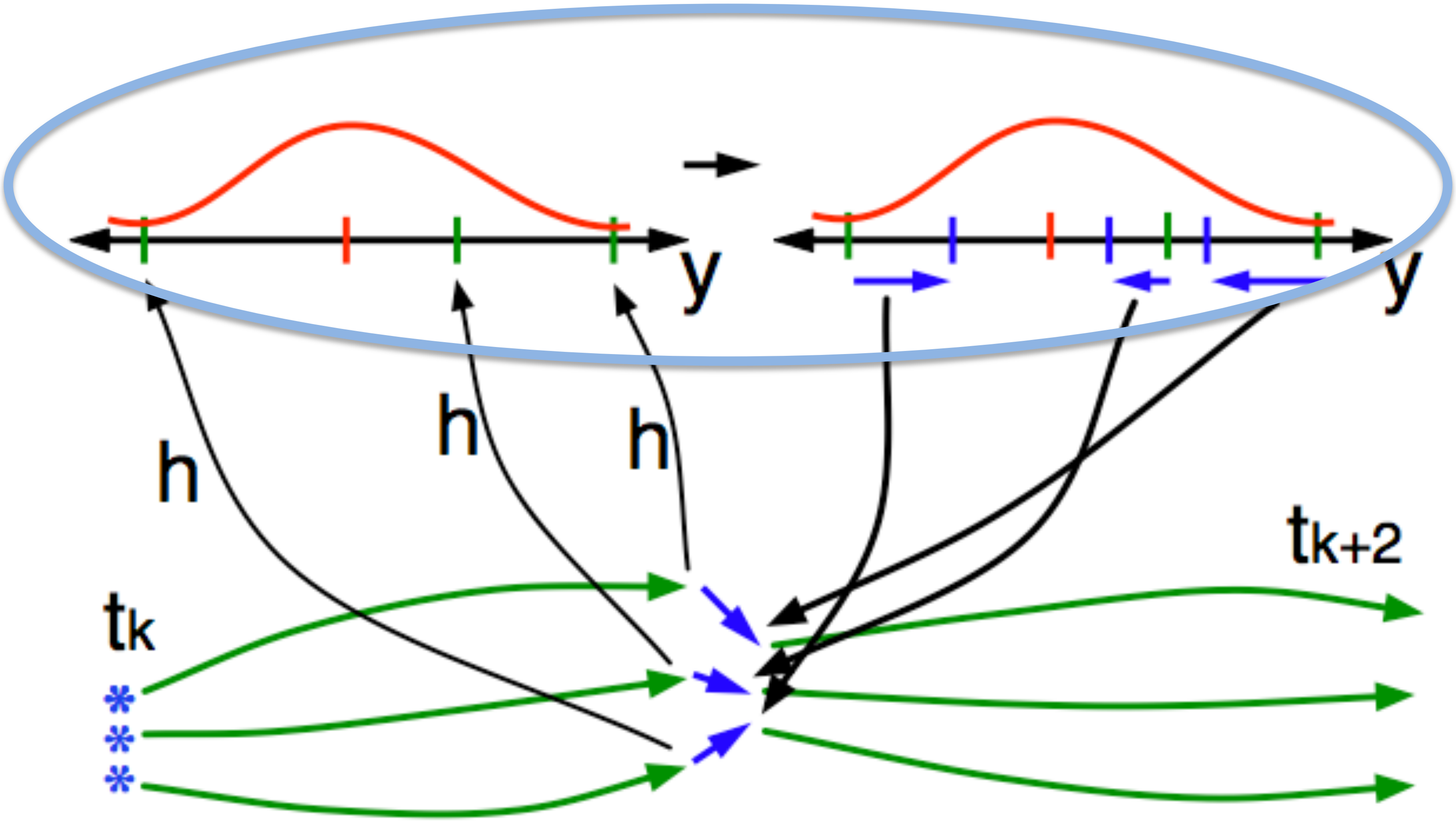
Schematic of a Sequential Ensemble Filter

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

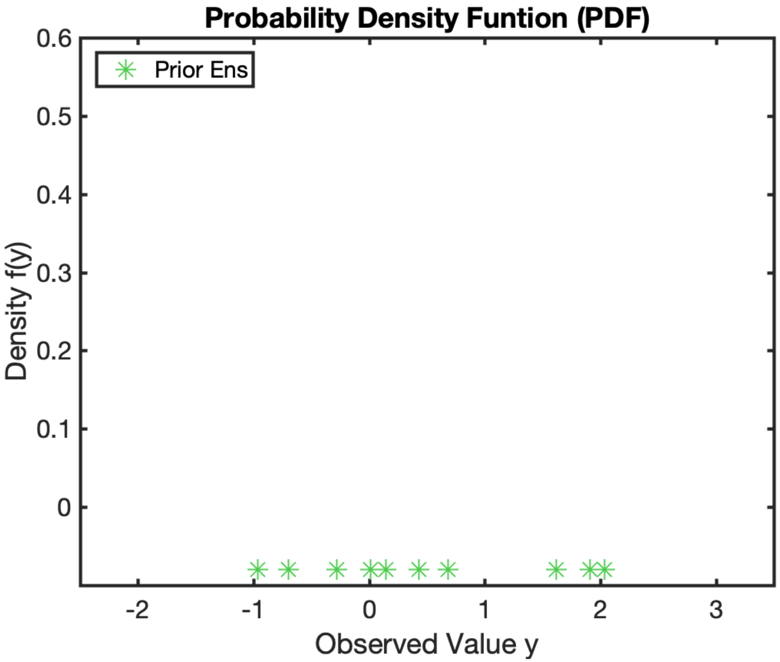


Schematic of a Sequential Ensemble Filter

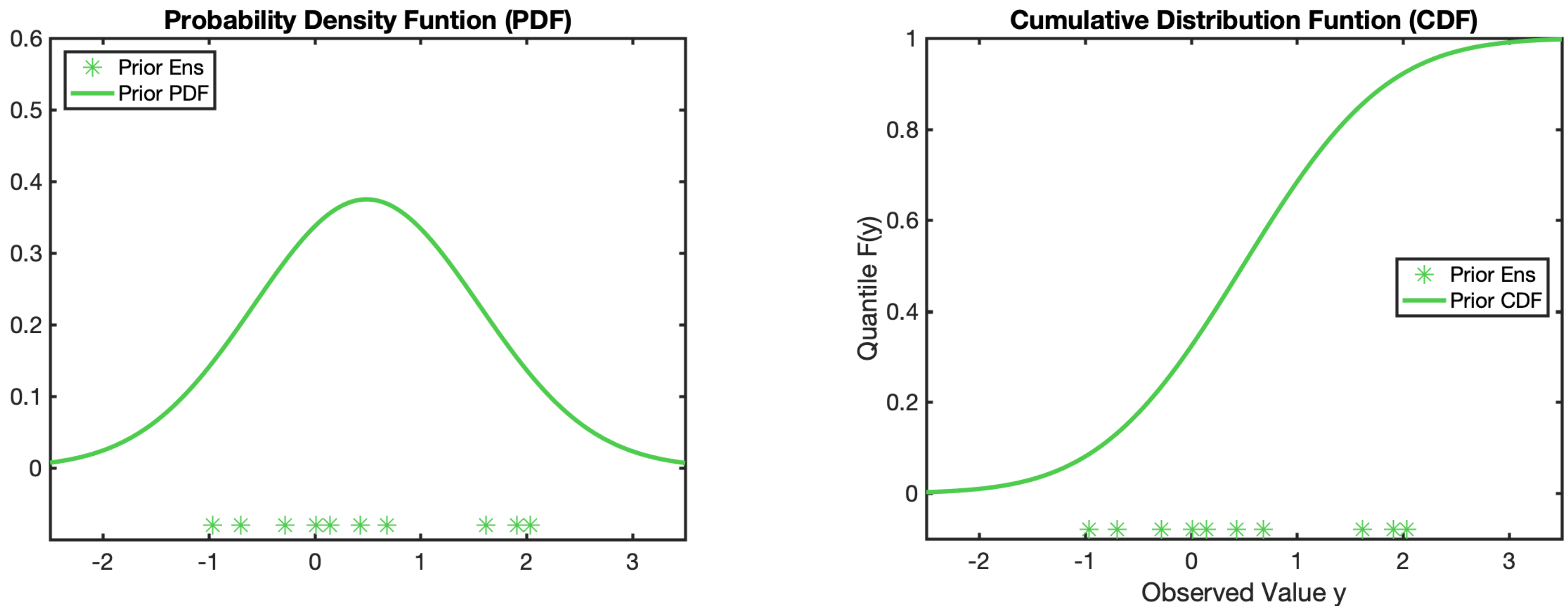
First part focuses on the scalar problem for an observed variable, y.
All other model state variables updated by (linear or rank) regression.



Given a prior ensemble estimate of an observed quantity, y



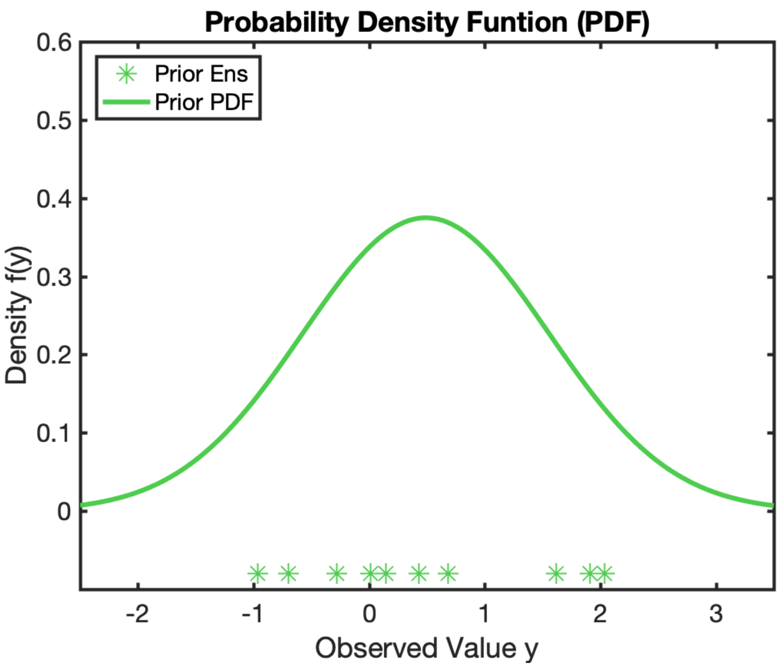
Fit a continuous PDF from an appropriate distribution family and find the corresponding CDF

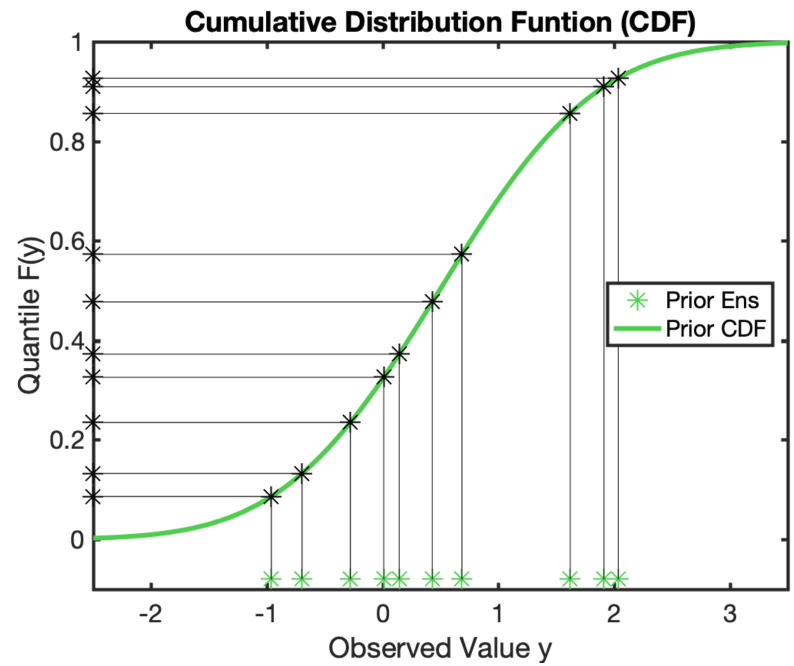


This example uses a normal PDF

Generalized Ensemble Filter Framework using Quantiles

Compute the quantile of ensemble members;
just the value of CDF evaluated for each member.

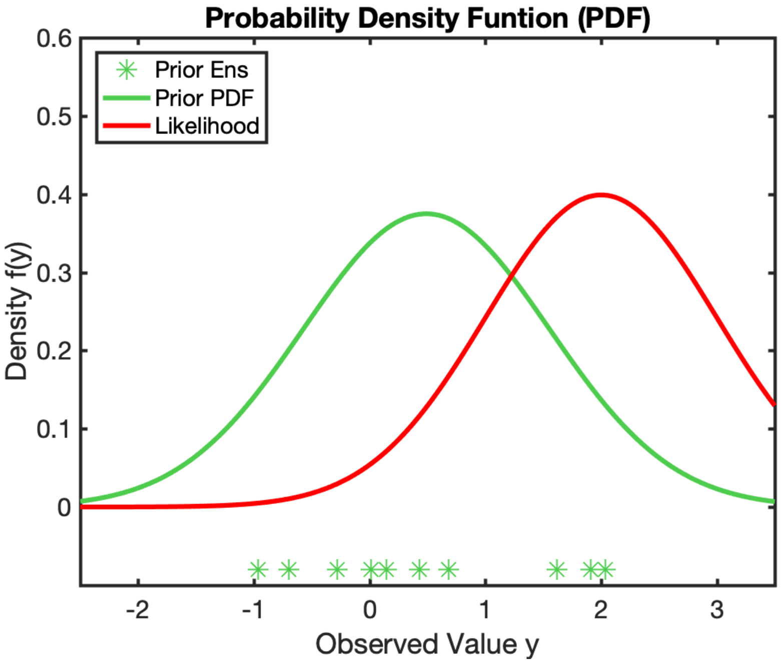


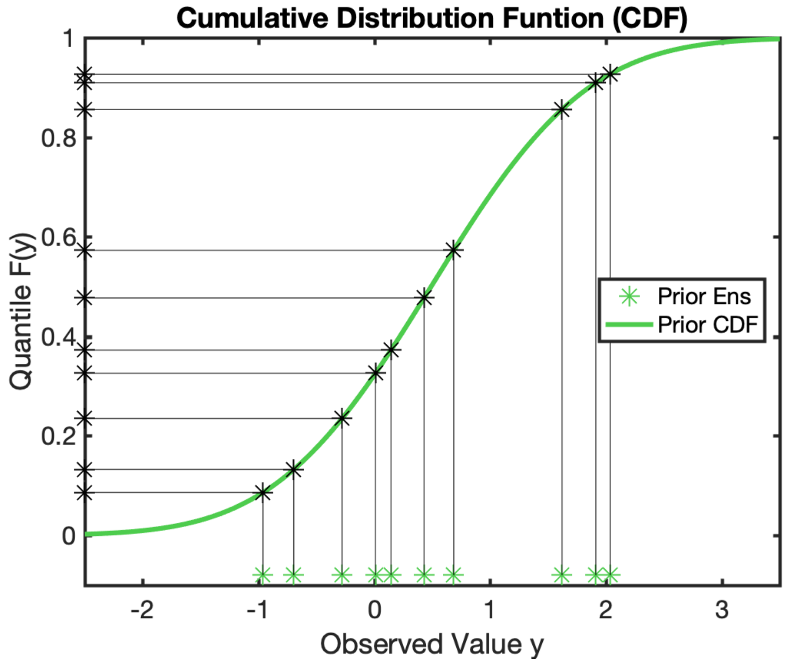


This example uses a normal PDF

Generalized Ensemble Filter Framework using Quantiles

Continuous likelihood for this observation.

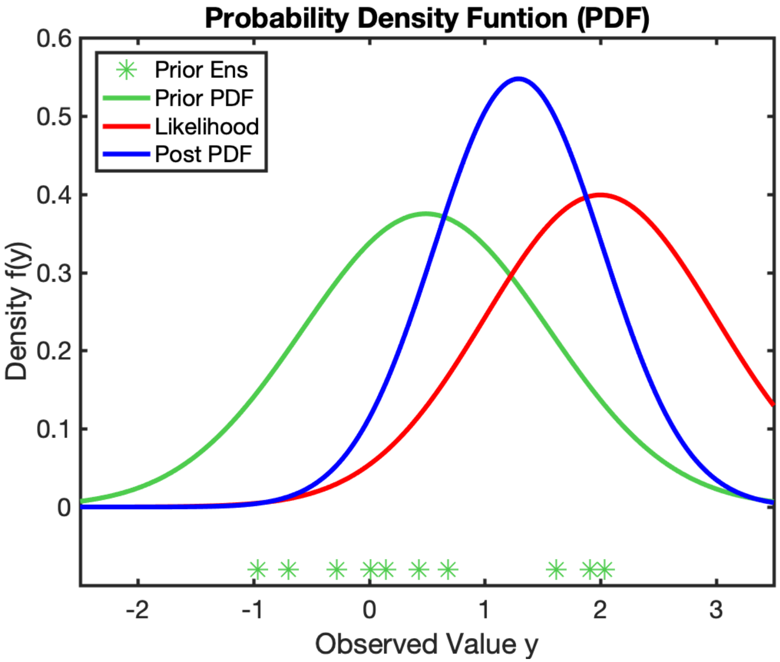


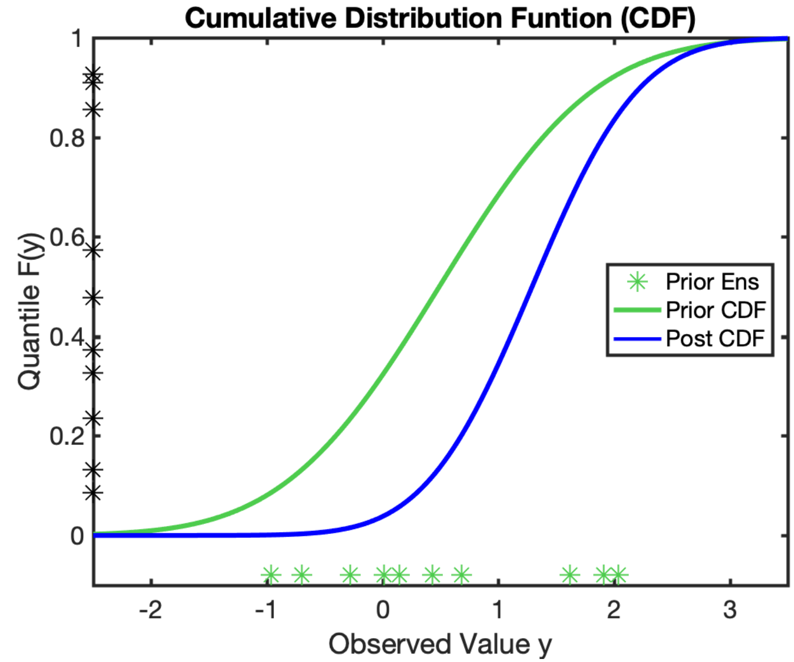


This example uses a normal PDF

Data Assimilation Research Testbed

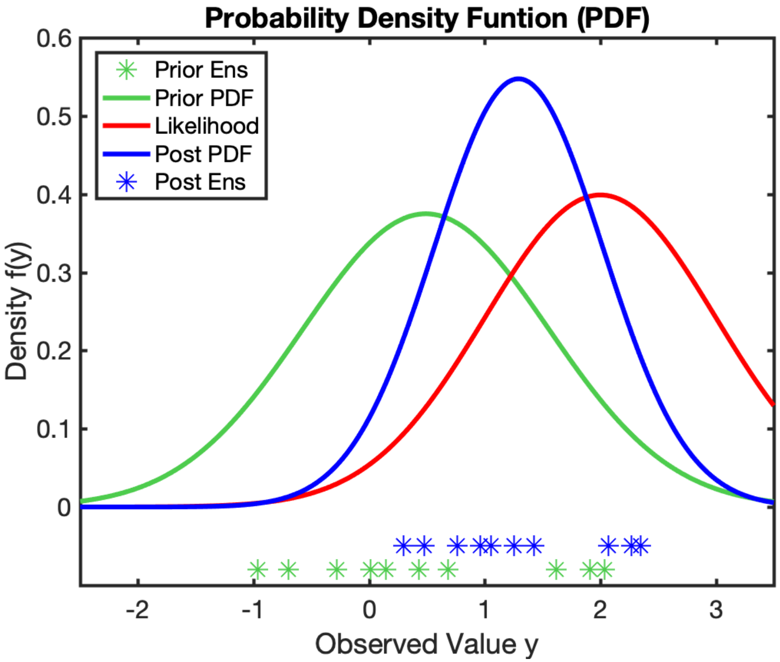
Bayes tells us that the continuous posterior PDF is the product of the continuous likelihood and prior.

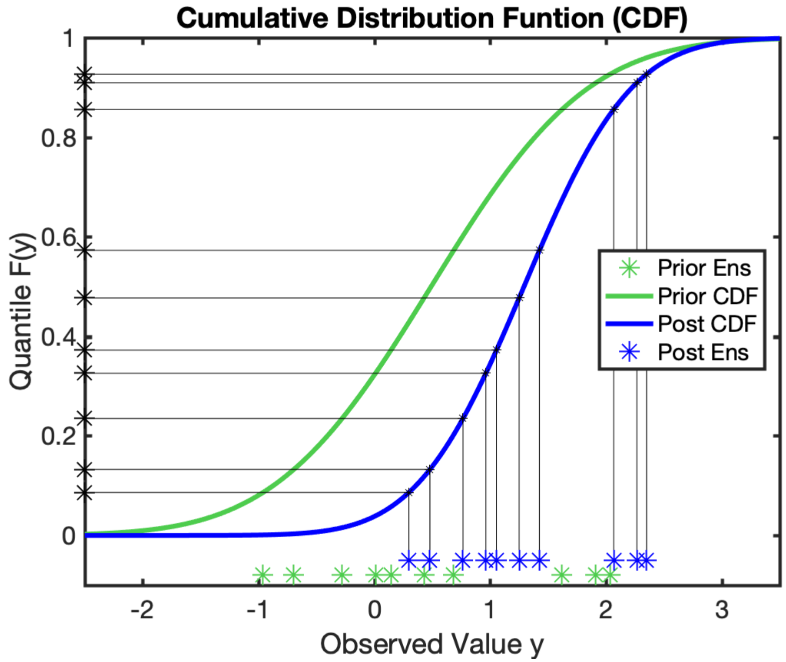




Normal times normal is normal.

Posterior ensemble members have same quantiles as prior.
This is quantile function, inverse of posterior CDF.

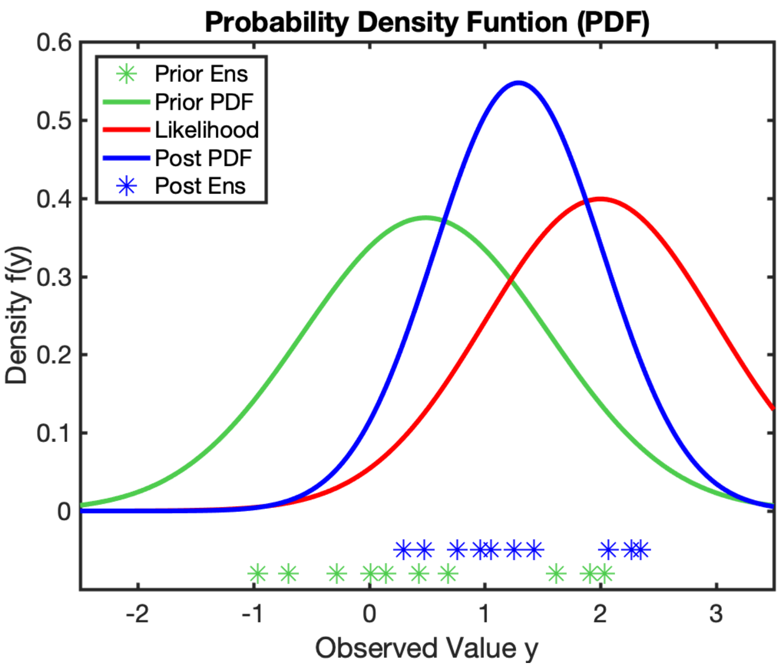


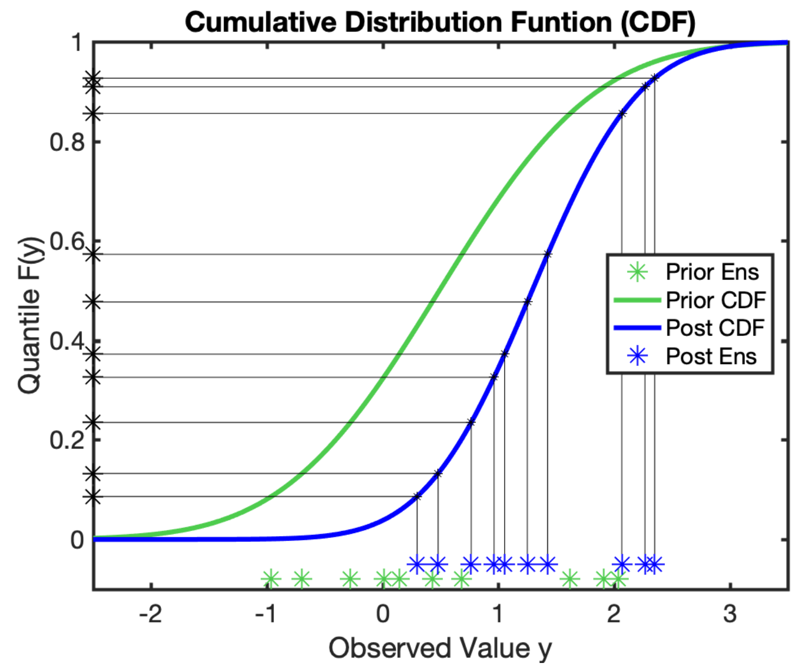


This example uses a normal PDF

Generalized Ensemble Filter Framework using Quantiles

For normal prior and likelihood, this is identical to existing deterministic Ensemble Adjustment Kalman Filter (EAKF)





Useful families for continuous priors and likelihoods

Different families of distributions for continuous priors and likelihoods can lead to analytic continuous posterior.

This is similar to the notion of conjugate priors for estimating parameters of distributions.

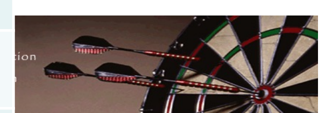
A list of prior / likelihood pairs that may be useful for scientific application follows.

Useful families for continuous priors and likelihoods

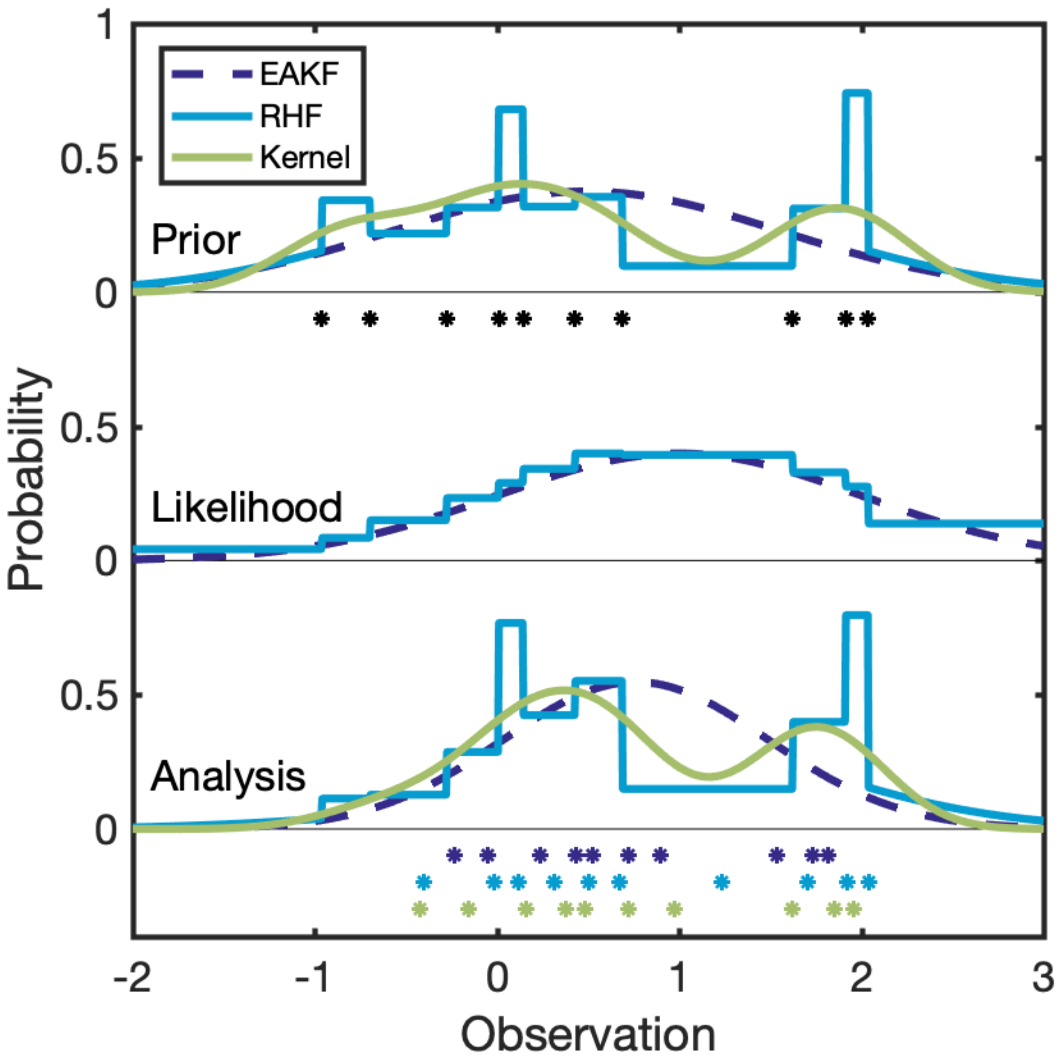
Prior	Likelihood	Posterior
Normal	Normal	Normal
Lognormal	Lognormal	Lognormal
Gamma	Gamma	Gamma
Inverse Gamma	Inverse Gamma	Inverse Gamma
Beta	Beta	Beta
Beta prime	Beta prime	Beta prime
Exponential	Exponential	Exponential
Pareto	Pareto	Pareto
Genl. Gamma given p	Genl. Gamma given p	Genl. Gamma given p
Any	Uniform	Any
Gamma	Poisson	Gamma

Useful families for continuous priors and likelihoods (2)

Prior	**Likelihood**	**Posterior**
Delta function	Any	Delta function
Skew normal	Normal	Skew normal
Truncated normal	Normal	Truncated normal
Any	Piecewise constant	Piecewise weighted
Rank histogram	Any	Rank histogram (except tails)
Huber	Huber	Piecewise normal and exponential
Weighted sum of two normals	Normal	Weighted sum of two normals
Sum of N normals same variance	Normal	Weighted sum of N normals same variance
Jeffreys	Various	Various



Examples: EAKF, Rank Histogram, Sum of Normal Kernels



Three different continuous priors are shown for the same prior ensemble.

Continuous likelihood is normal. Piecewise constant approximation is used for RHF.

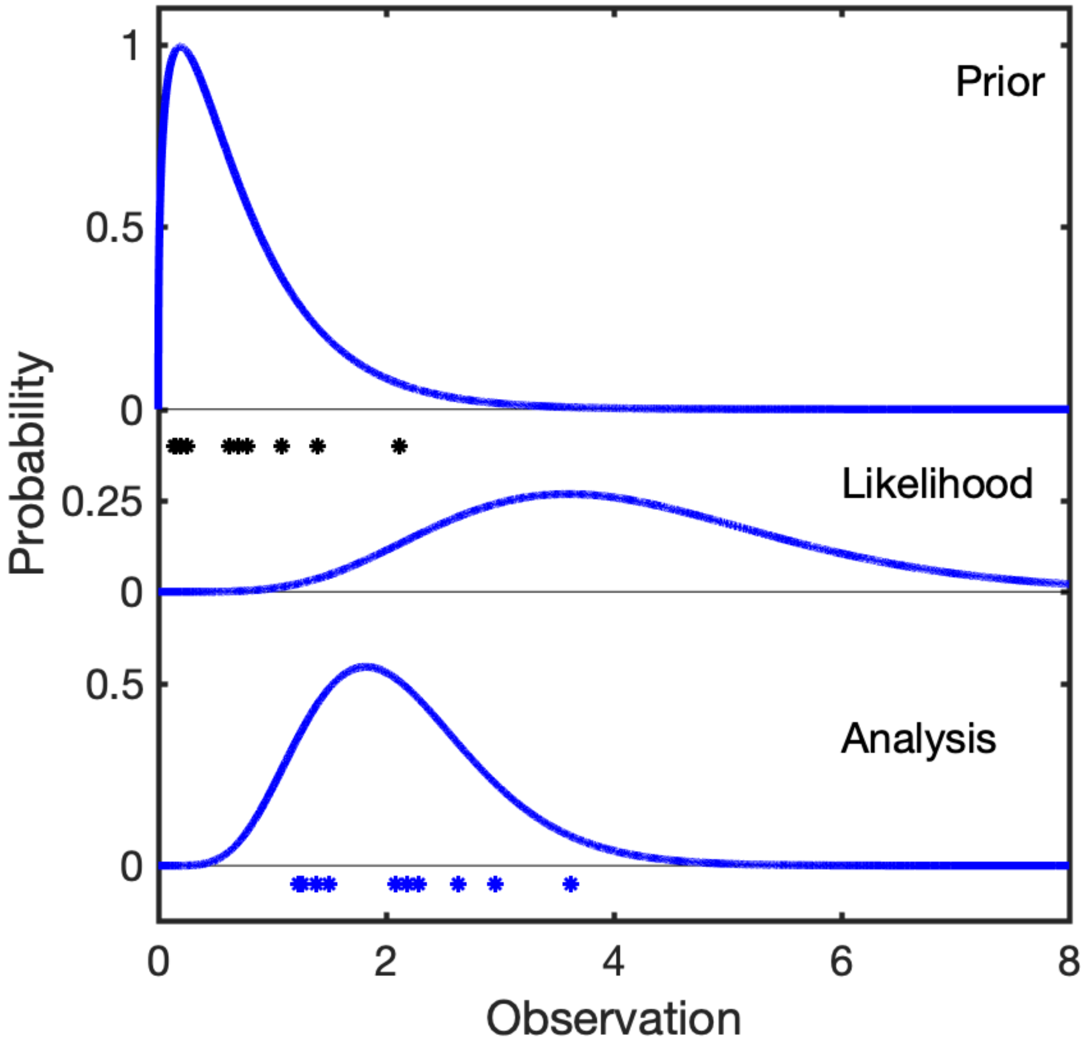
Posterior ensembles differ qualitatively.

NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

NSF

Data Assimilation Research Testbed

Example: Gamma prior, Gamma Likelihood

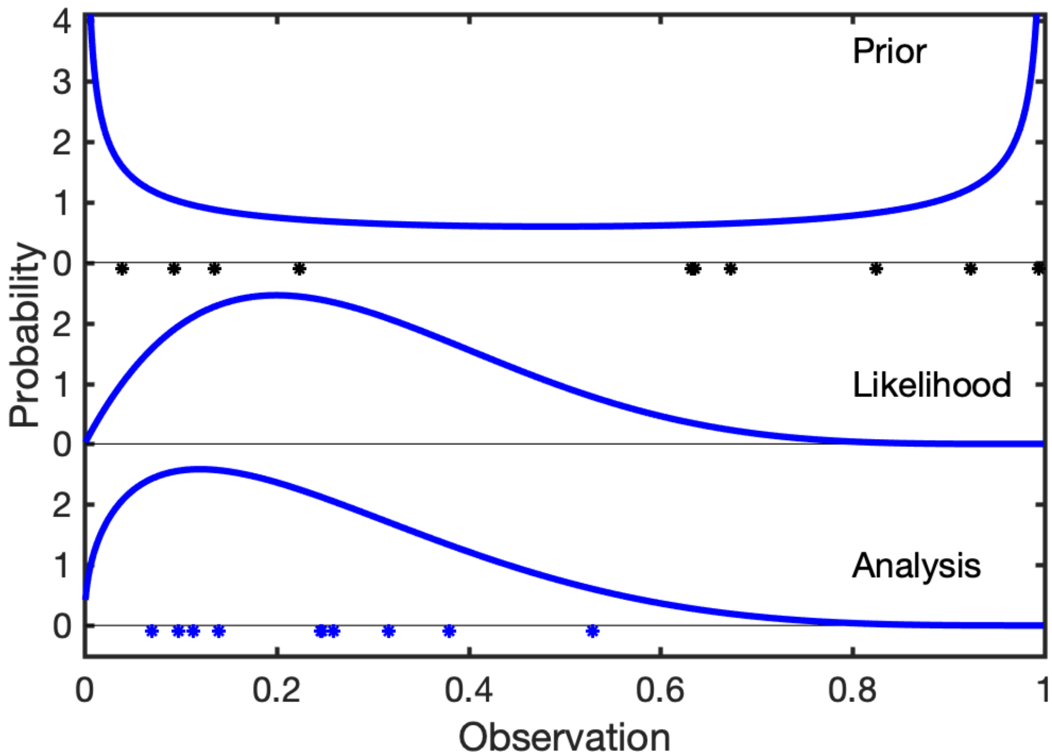


Physical quantities may be bounded. For instance, amount of water vapor is non-negative.

Gamma prior enforces non-negativity.

Gamma likelihood leads to gamma posterior.

Example: Beta Prior and Likelihood



Sea ice concentration is bounded between 0 and 1.

A beta distribution can enforce these bounds.

A beta likelihood leads to a beta posterior.

Statistical Foundation of GEFFQ

There is statistical support for this method.

The Kolmogorov-Smirnov statistic is the same for the prior ensemble/continuous PDF as for the posterior.

The method is also related to the Q-Q plot for comparing distributions.

Using regression to update state variables means we lose control over the analysis distribution.

Can't enforce boundedness for example.

Could apply GEFFQ directly to a state variable x,

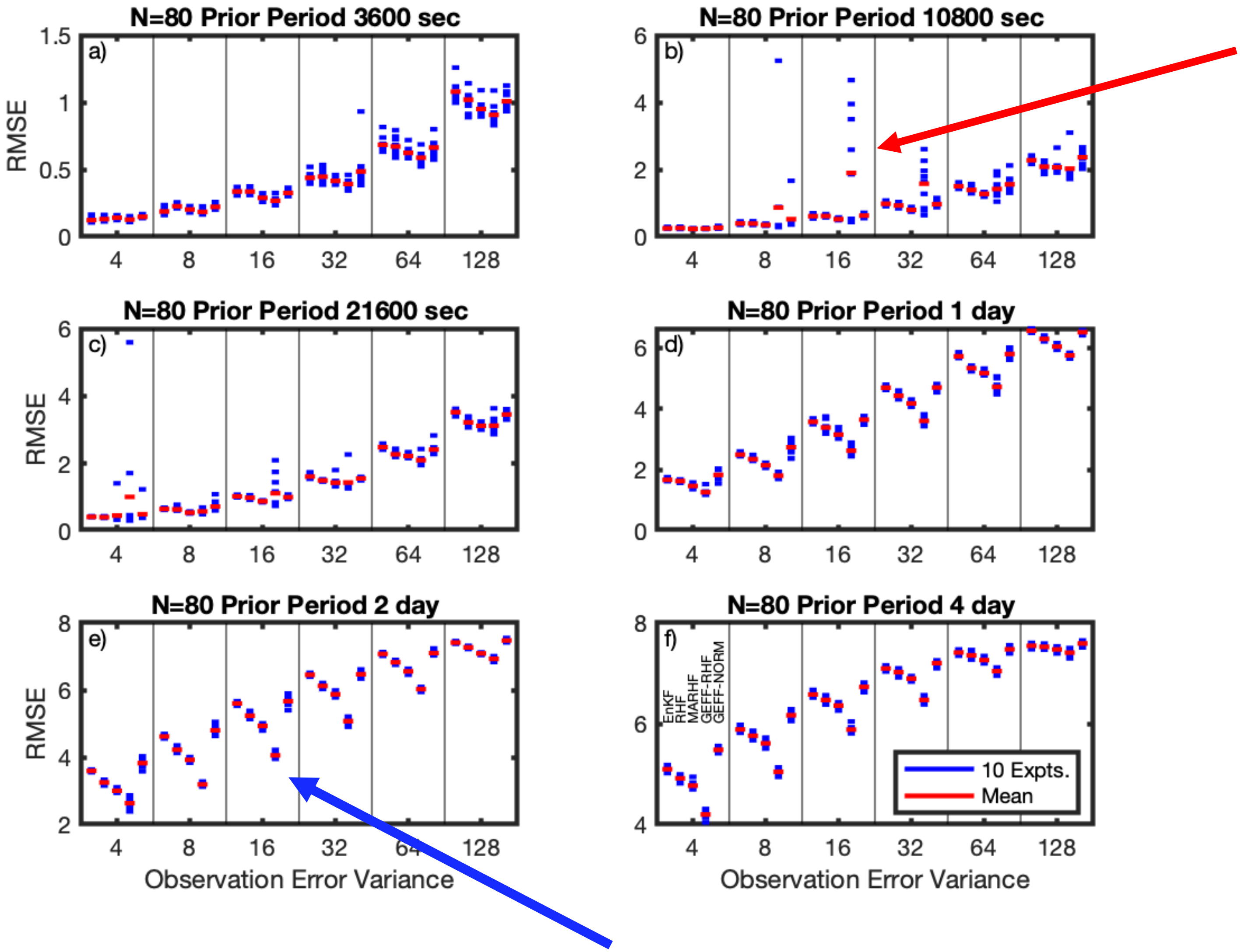
- Need a continuous likelihood as function of x,
- Only know likelihood for ensemble members,
- Piecewise constant approx. gives continuous likelihood.

This is same method used in the rank histogram filter (RHF).

Can directly update marginal ensemble for all state variables.

- Useful for reanalysis purposes, too.

Proof of Concept Results: L63



www.image.ucar.edu/DAReS/DART

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Arellano, A., 2009: *The Data Assimilation Research Testbed: A community facility.* BAMS, **90**, 1283—1296, doi: 10.1175/2009BAMS2618.1