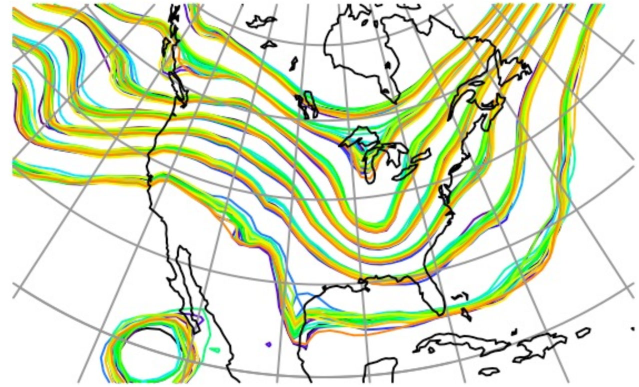
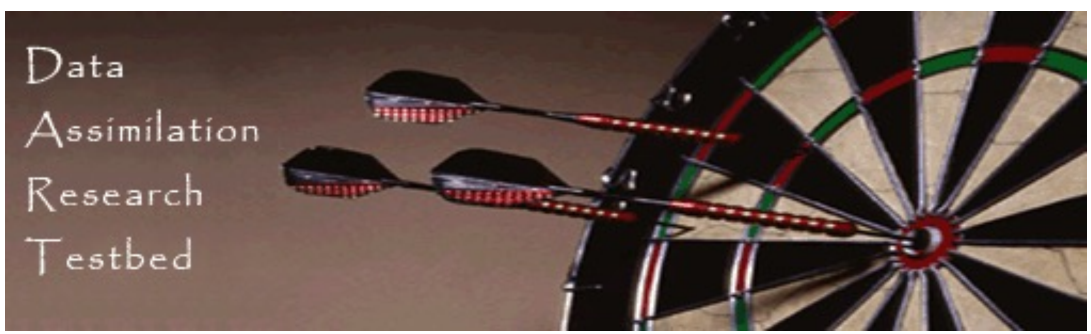


A Quantile Conserving Particle Filter with Likelihood Localization

Jeff Anderson, NCAR Data Assimilation Research Section



Outline

- Review of cycling ensemble Kalman filters
- Quantile Conserving Ensemble Filtering Framework
  - Can fit arbitrary continuous prior to ensemble
  - Product with a likelihood gives continuous posterior
  - Posterior ensemble with same quantiles as prior
- Quantile Conserving Particle Filter for scalar problem
- Extension to posterior marginals for multivariate problem
- Likelihood localization
- Cycling multivariate application using marginal adjustment

Schematic of a Sequential Ensemble Filter

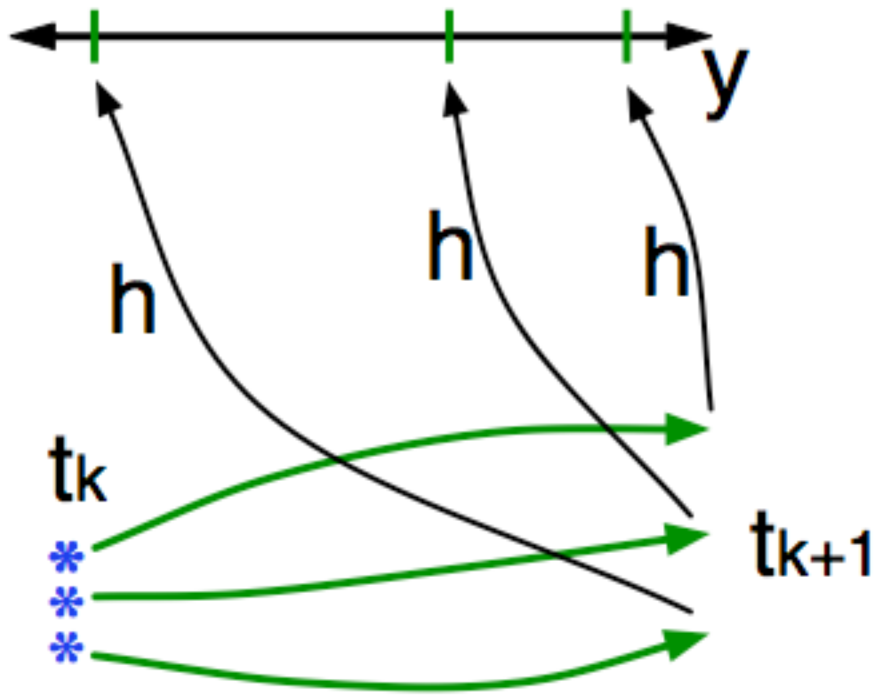
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

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.

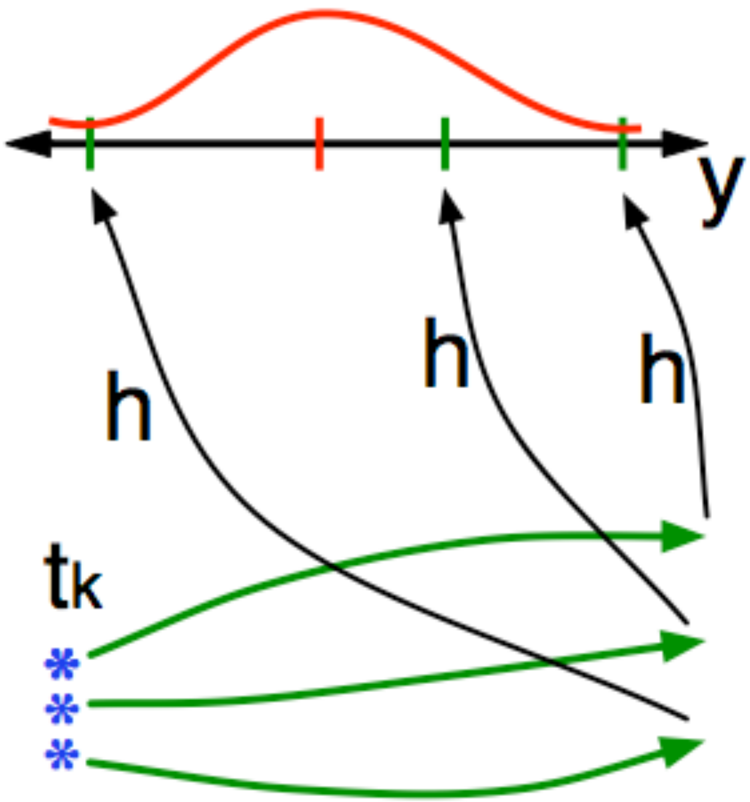


Theory: observations from instruments with uncorrelated errors can be done sequentially.

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

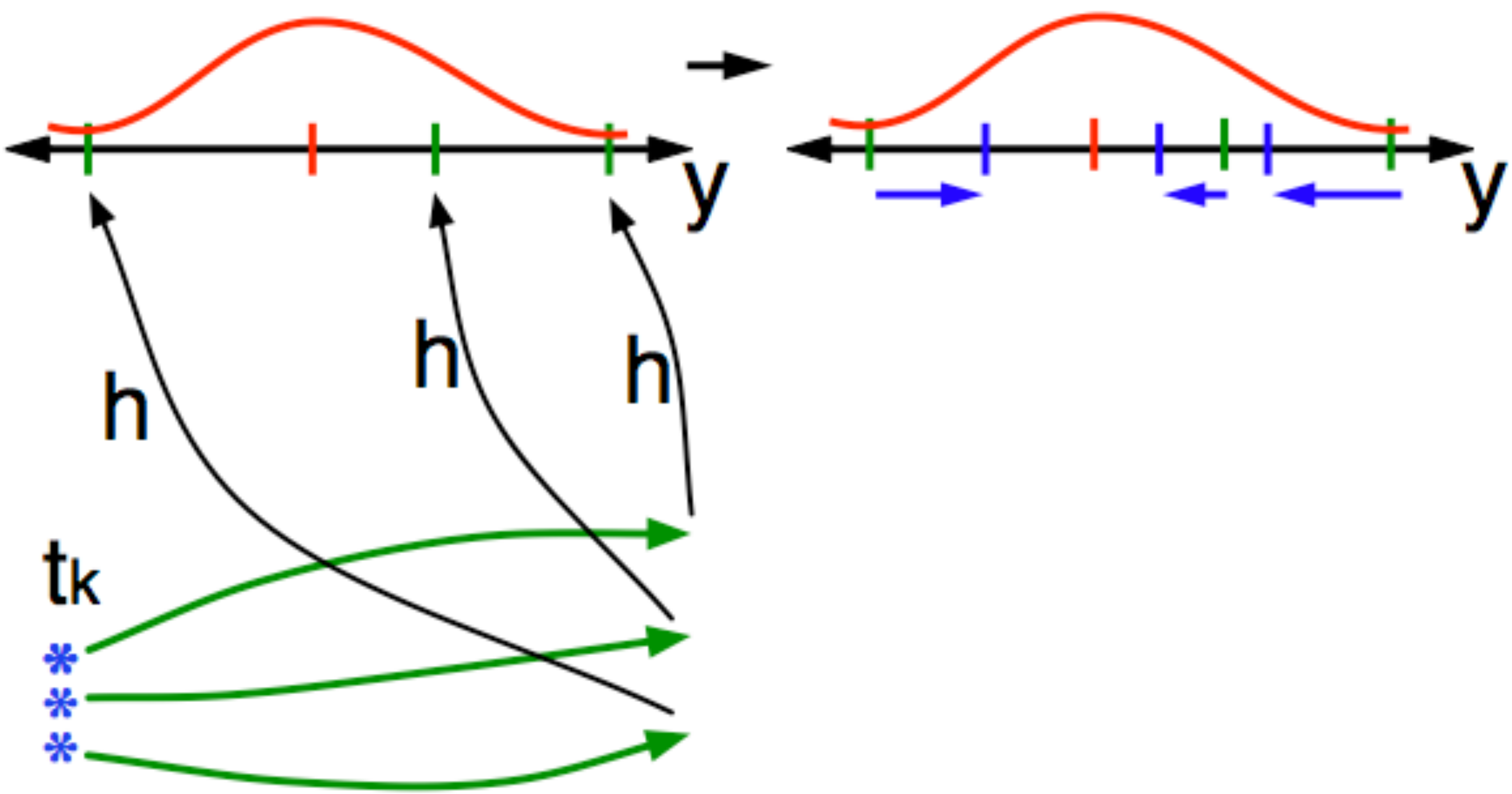
Schematic of a Sequential Ensemble Filter

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



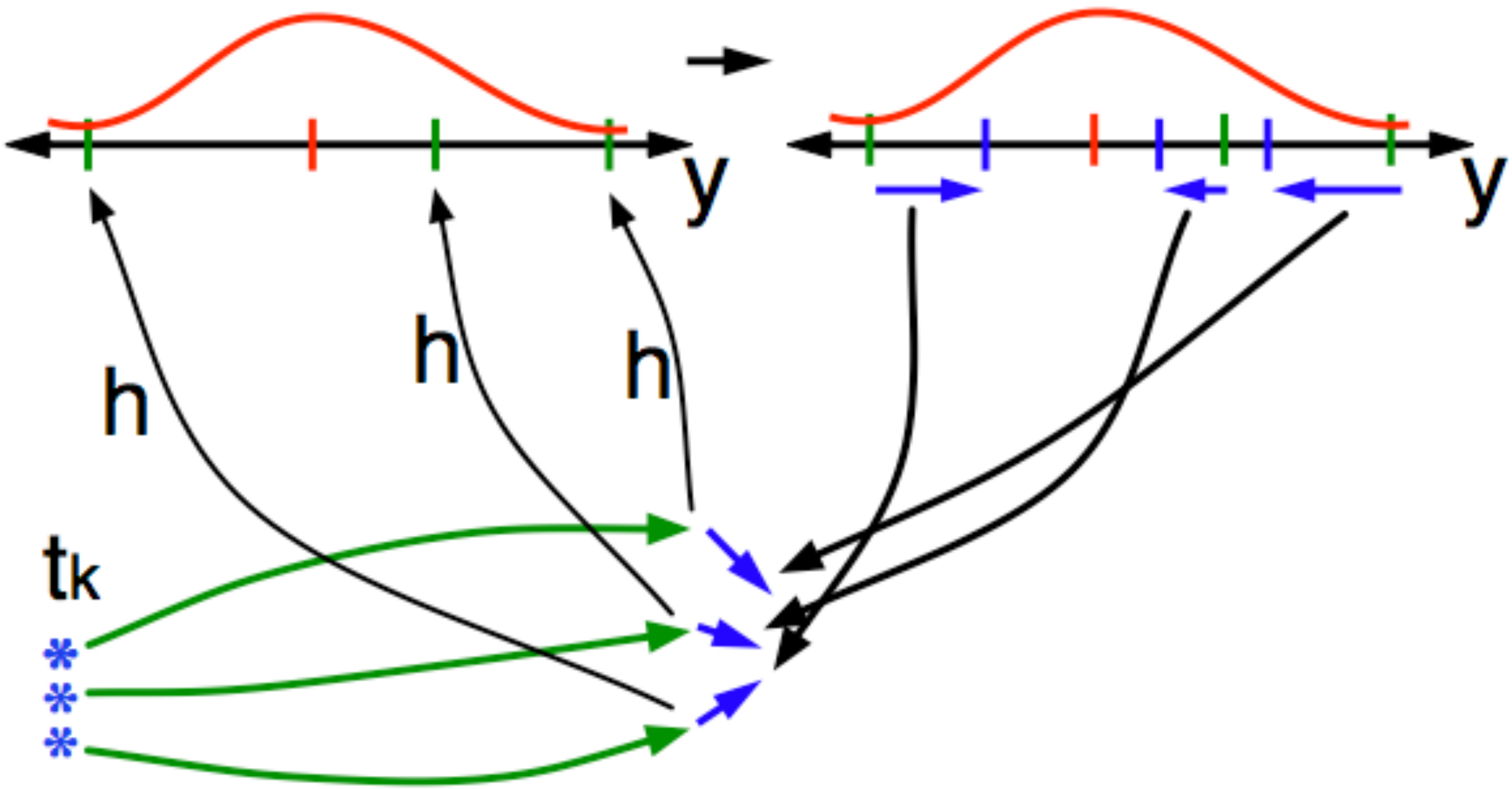
Schematic of a Sequential Ensemble Filter

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



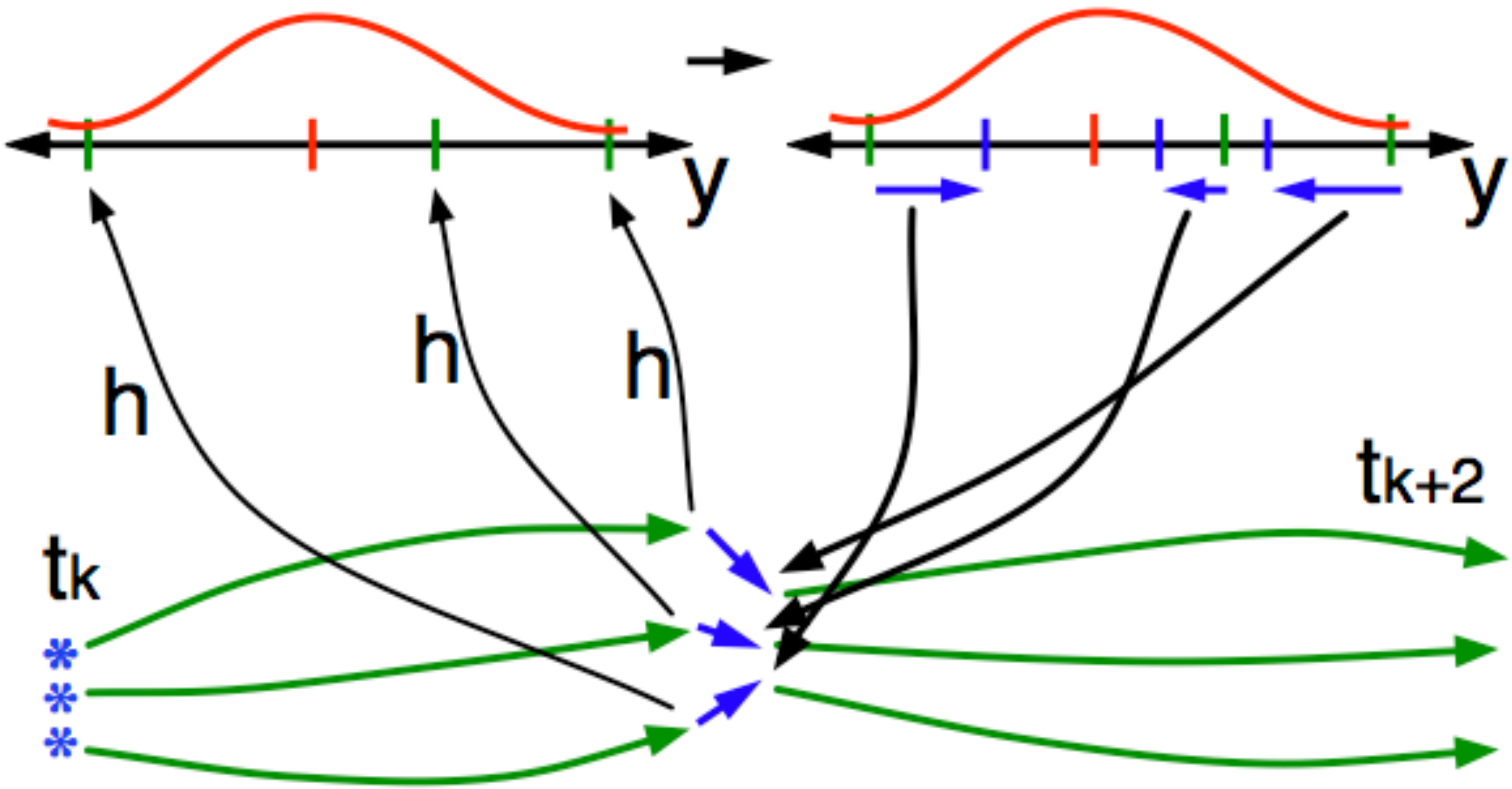
Schematic of a Sequential Ensemble Filter

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



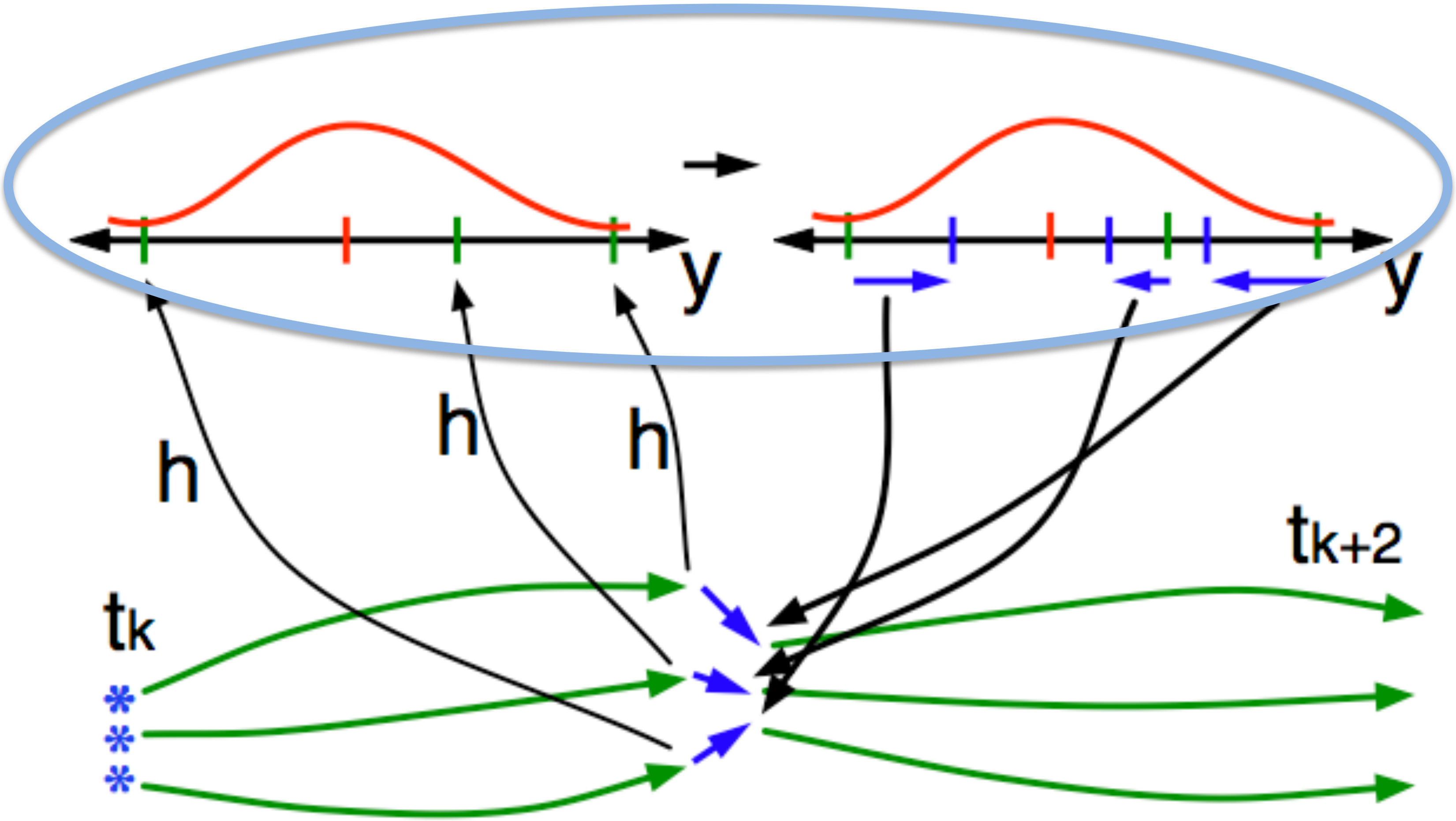
Schematic of a Sequential Ensemble Filter

6. When all ensemble members for each state variable are updated, integrate to time of next observation …

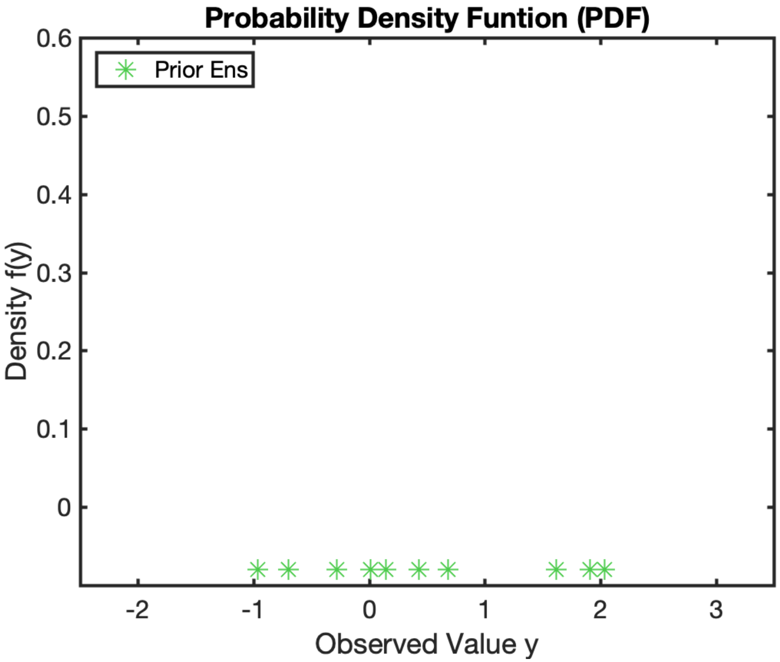


Schematic of a Sequential Ensemble Filter

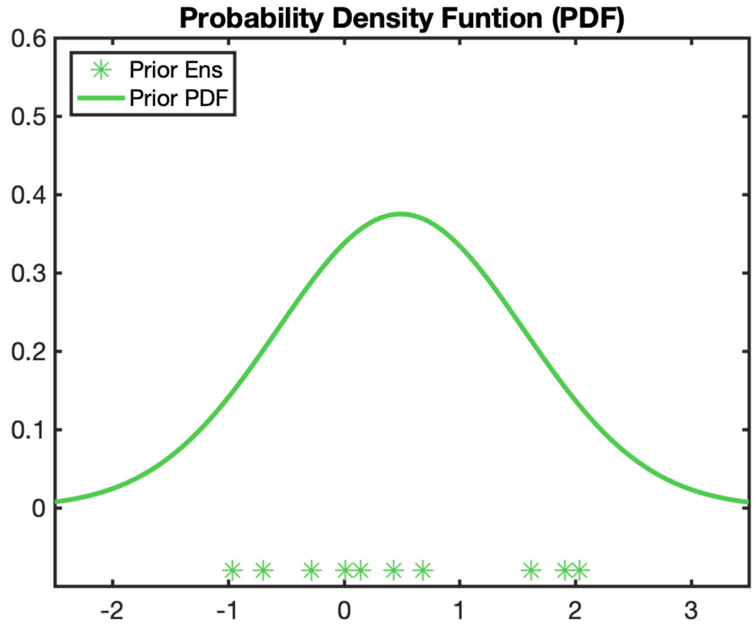
First part focuses on the scalar problem for an observed variable, y.
All other model state variables updated by (linear or rank) regression.

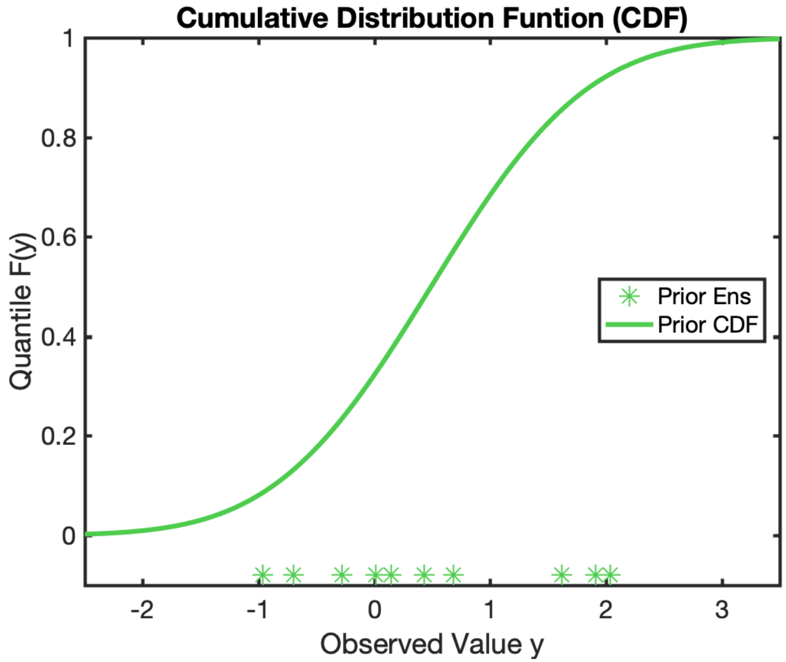


Given a prior ensemble estimate of an observed quantity, y



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

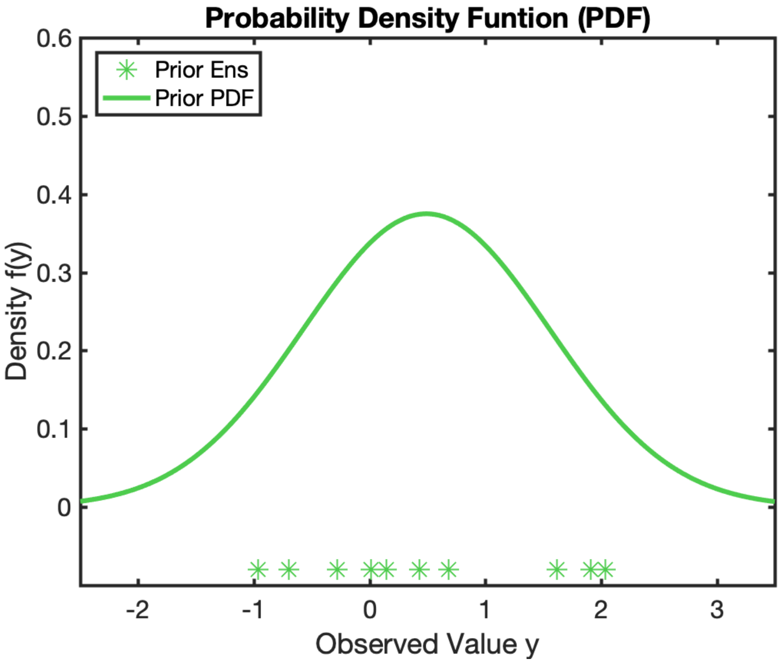


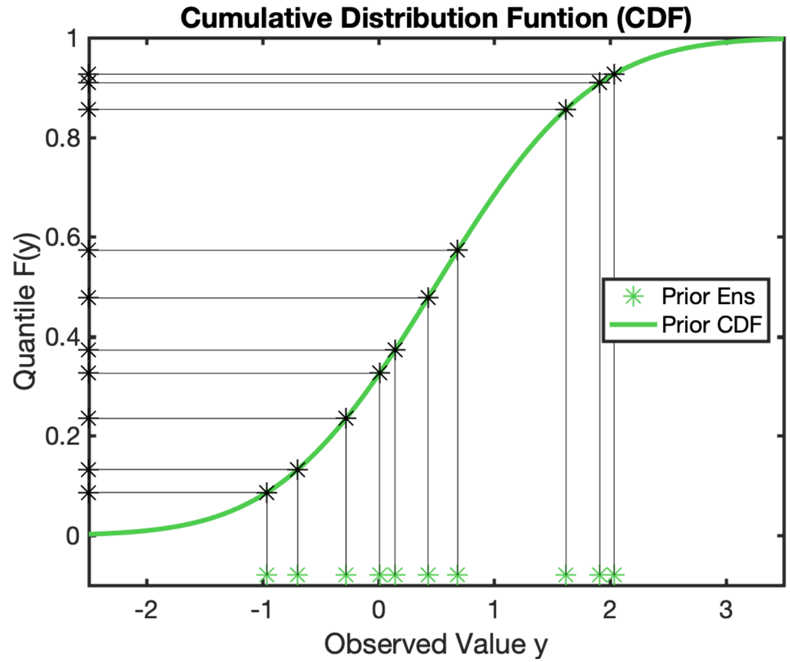


This example uses a normal PDF.

Quantile Conserving Ensemble Filter Framework

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

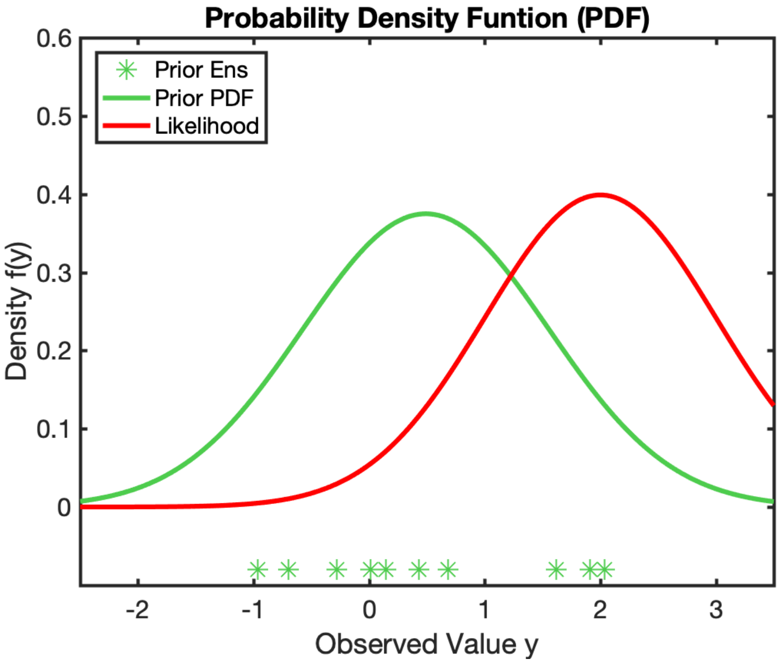


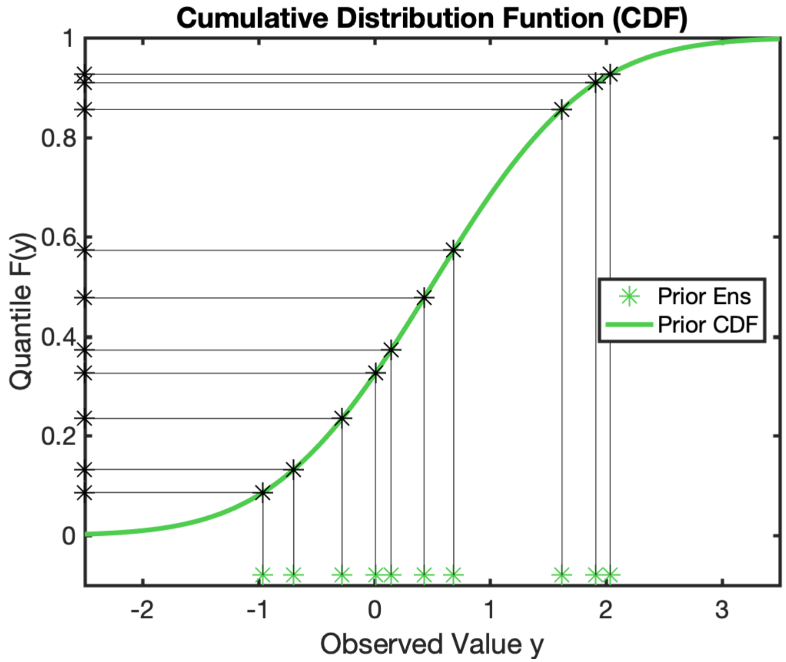


This example uses a normal PDF.

Quantile Conserving Ensemble Filter Framework

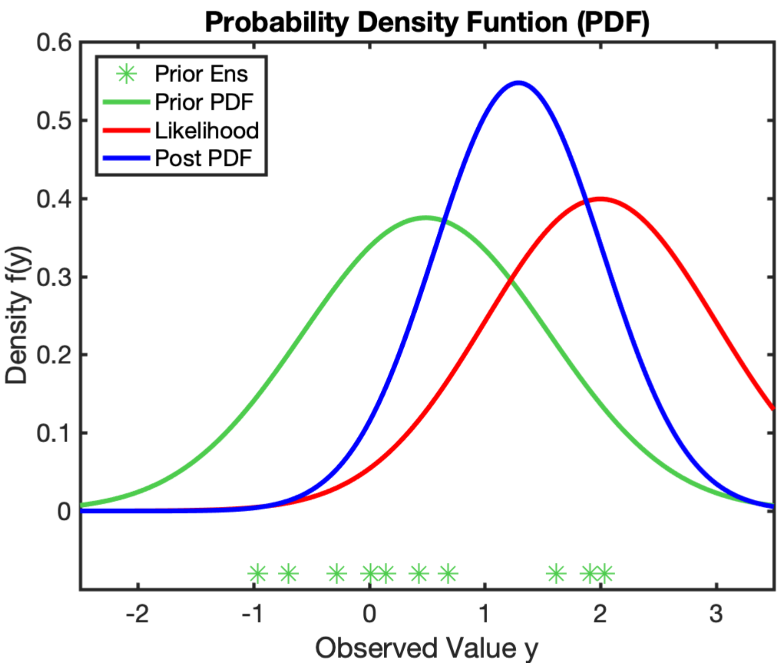
Continuous likelihood for this observation.

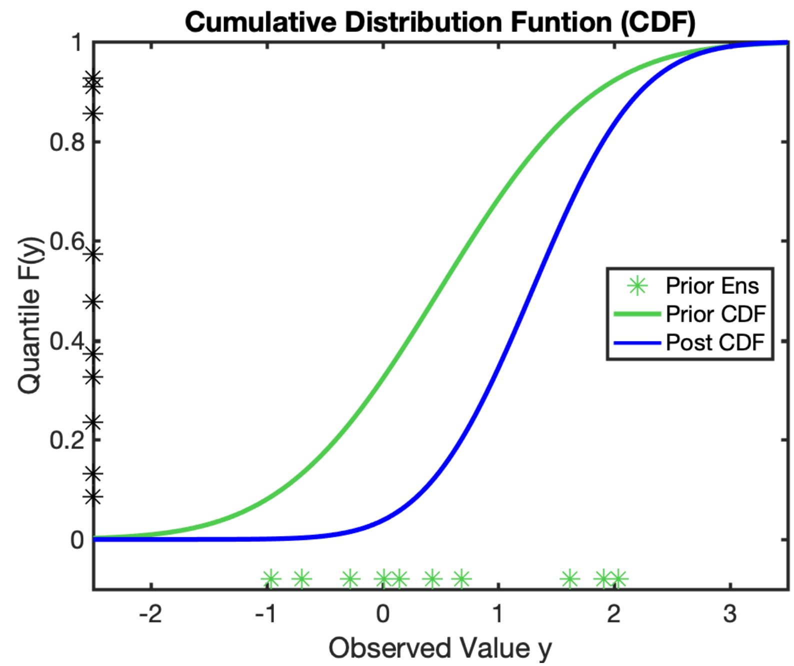




This example uses a normal PDF.

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

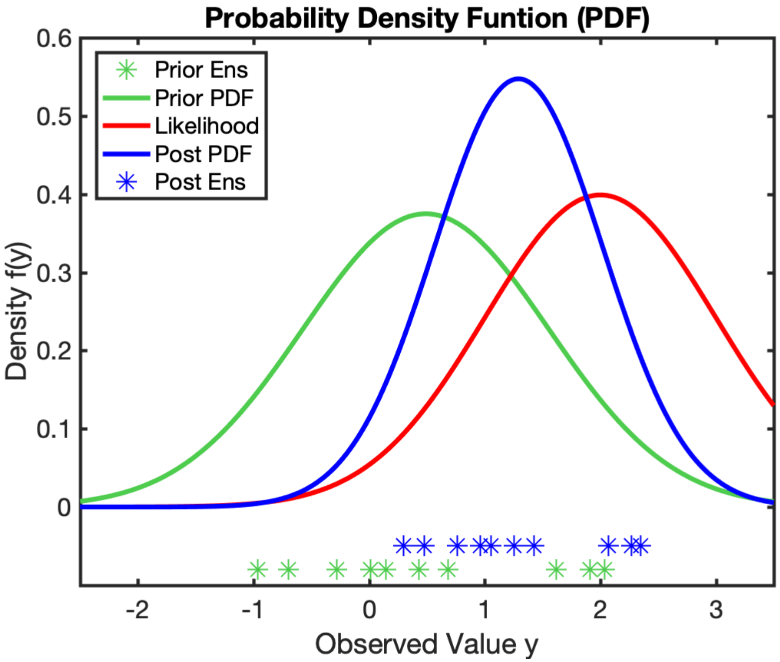


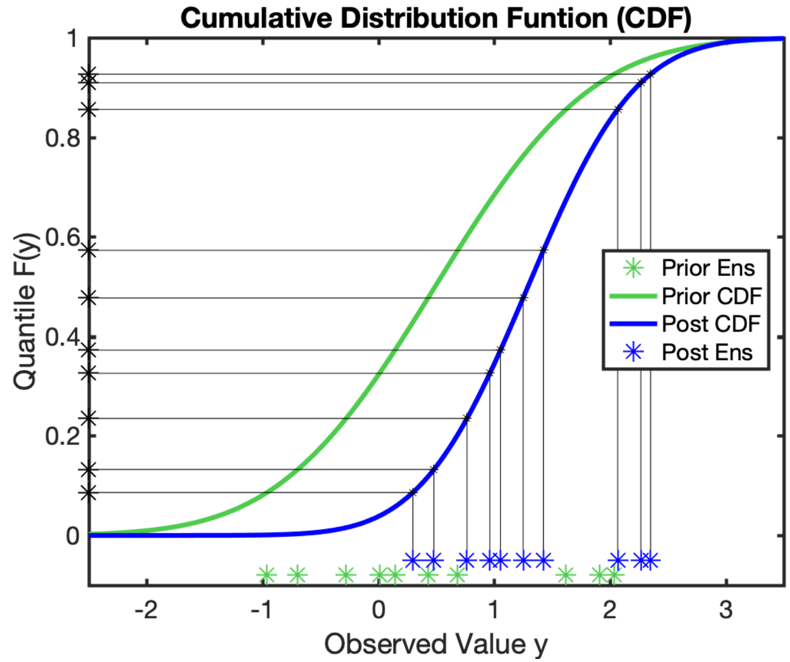


Normal times normal is normal.

Quantile Conserving Ensemble Filter Framework

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

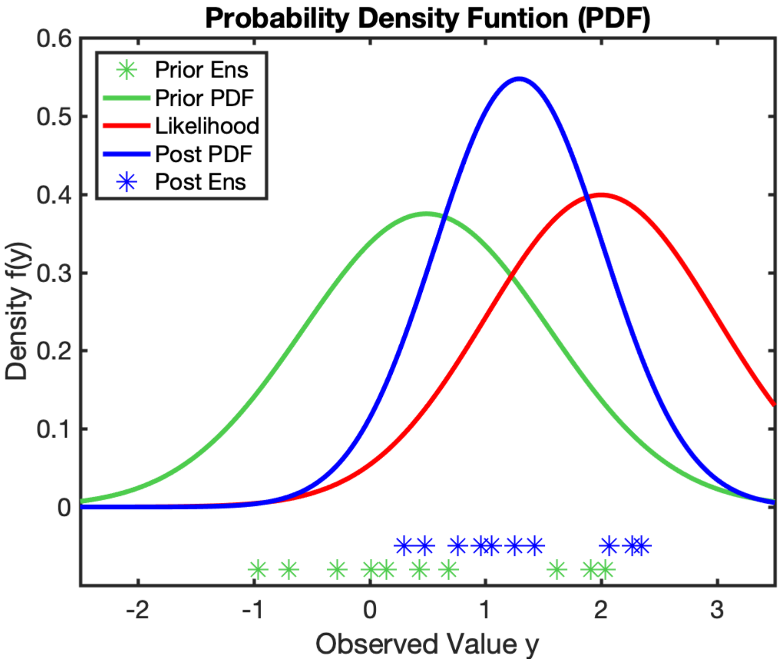


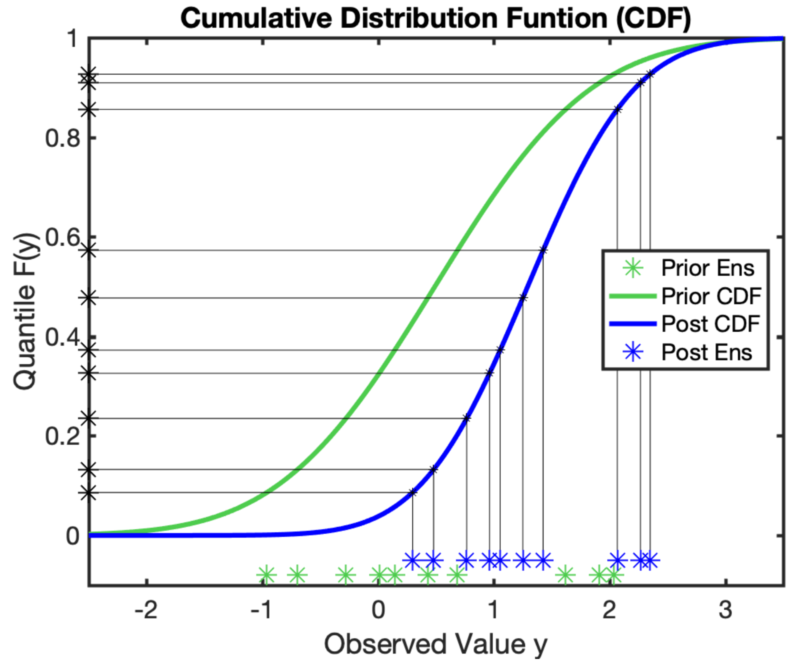


This example uses a normal PDF

Quantile Conserving Ensemble Filter Framework

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.

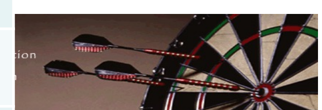
A list of prior / likelihood pairs that may be useful for scientific application follows.

Useful families for continuous priors and likelihoods

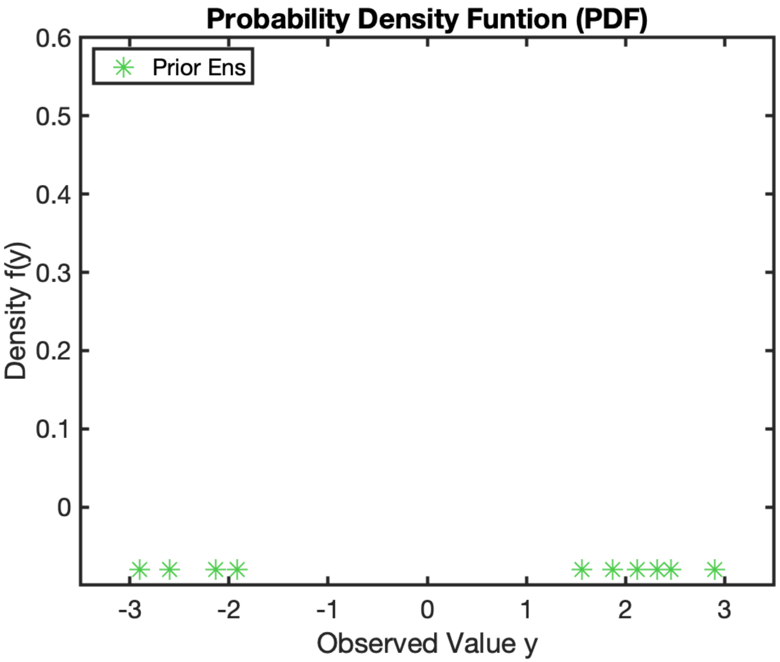
Prior	Likelihood	Posterior
Normal	Normal	Normal
Lognormal	Lognormal	Lognormal
Gamma	Gamma	Gamma
Inverse Gamma	Inverse Gamma	Inverse Gamma
Beta	Beta	Beta
Beta prime	Beta prime	Beta prime
Exponential	Exponential	Exponential
Pareto	Pareto	Pareto
Genl. Gamma given p	Genl. Gamma given p	Genl. Gamma given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

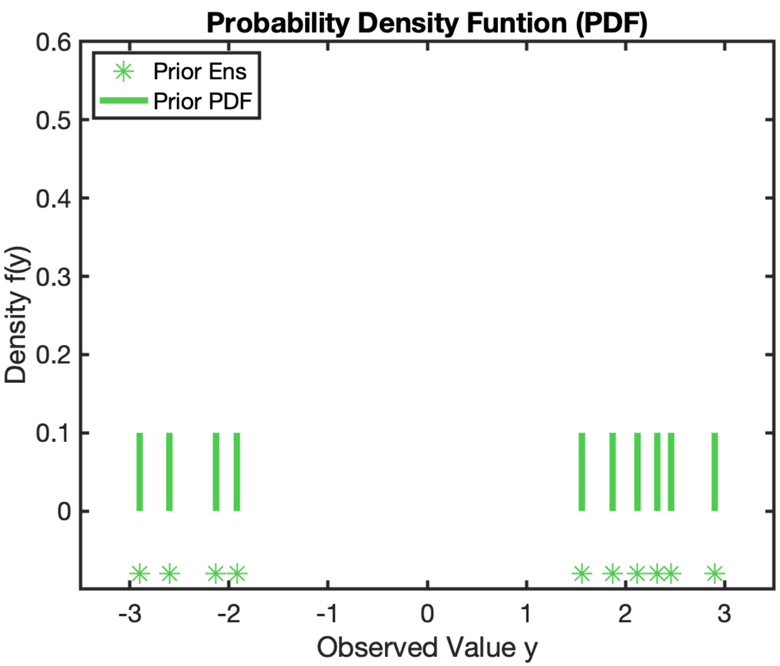


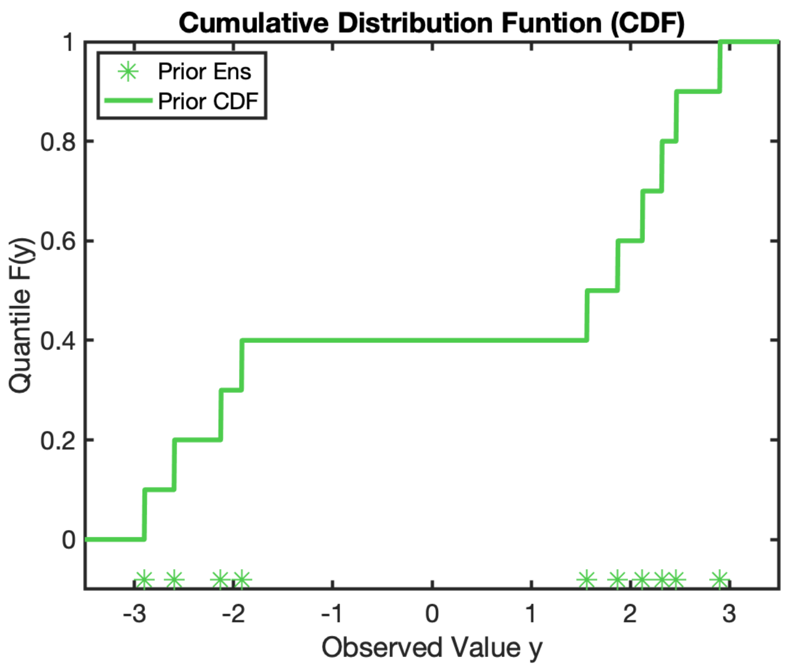
Given a prior ensemble estimate of an observed quantity, y



Quantile Conserving Particle Filter

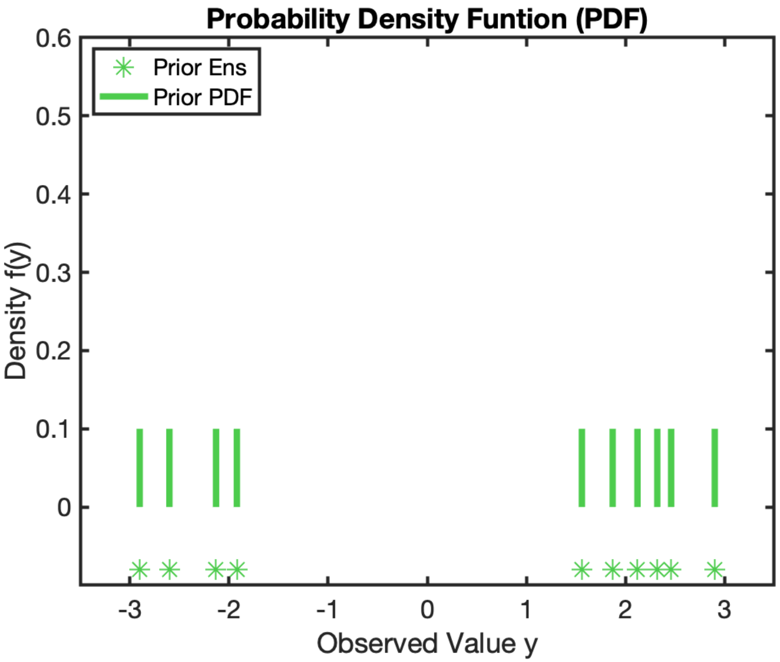
Fit a ~~continuous~~ particle filter PDF and find the corresponding CDF.

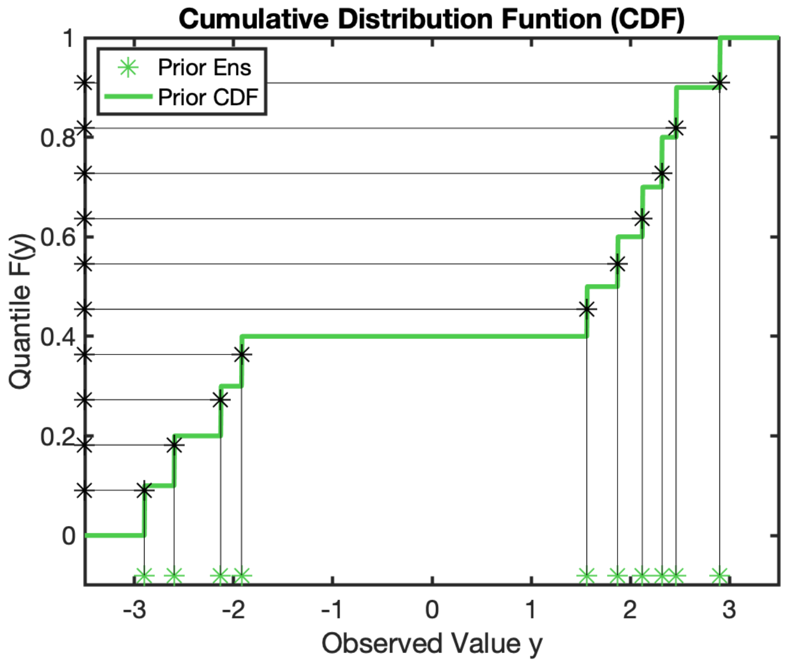




Quantile Conserving Particle Filter

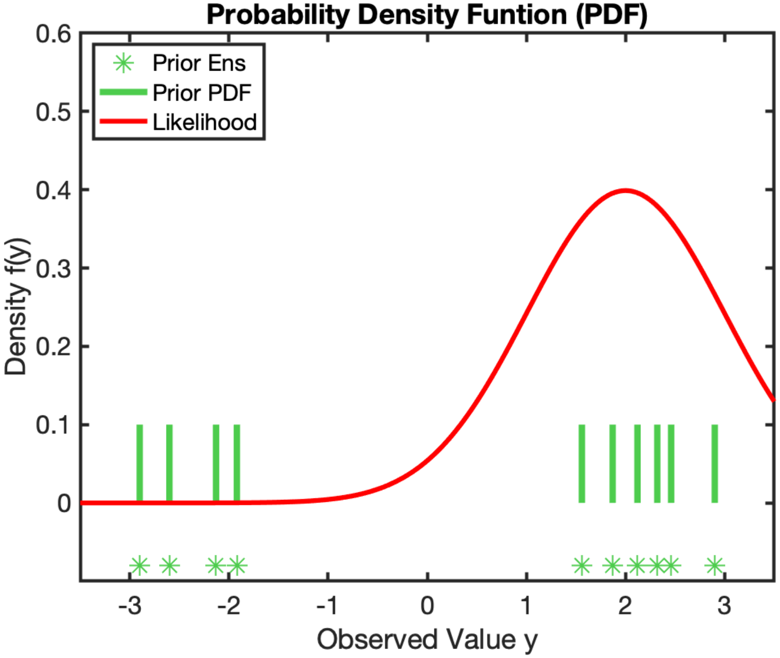
Quantile of ensemble members are defined to be uniformly distributed: $q_i = \frac{i}{ens\_size+1}$.

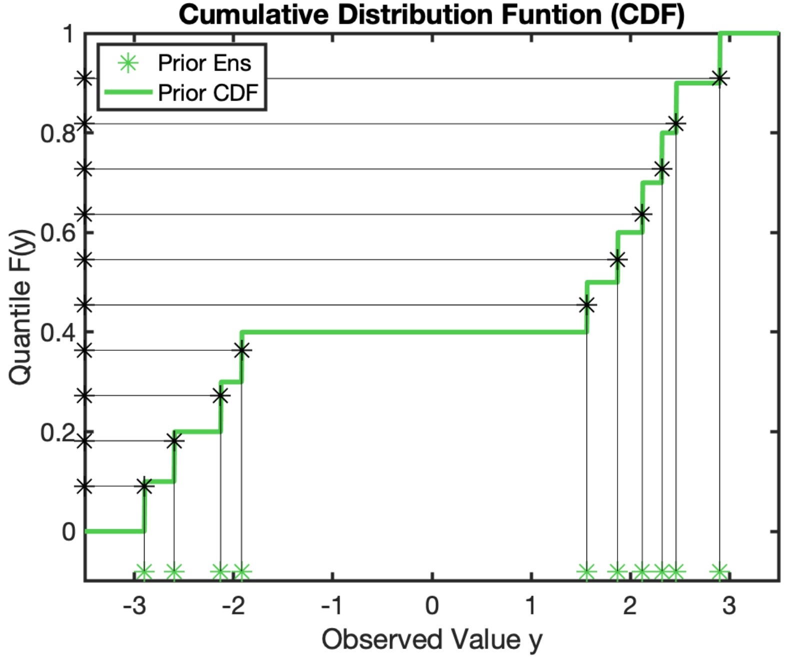




Quantile Conserving Particle Filter

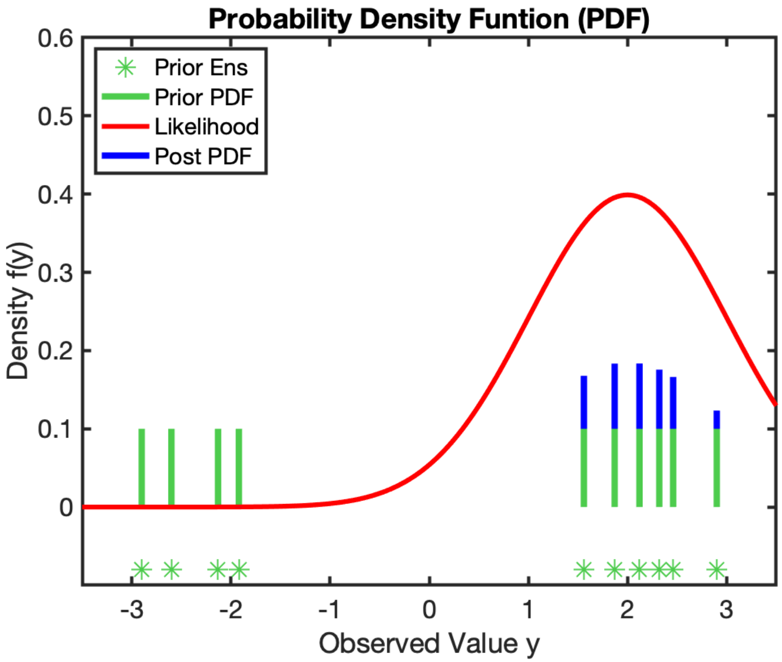
Continuous likelihood for this observation comes from observing system.

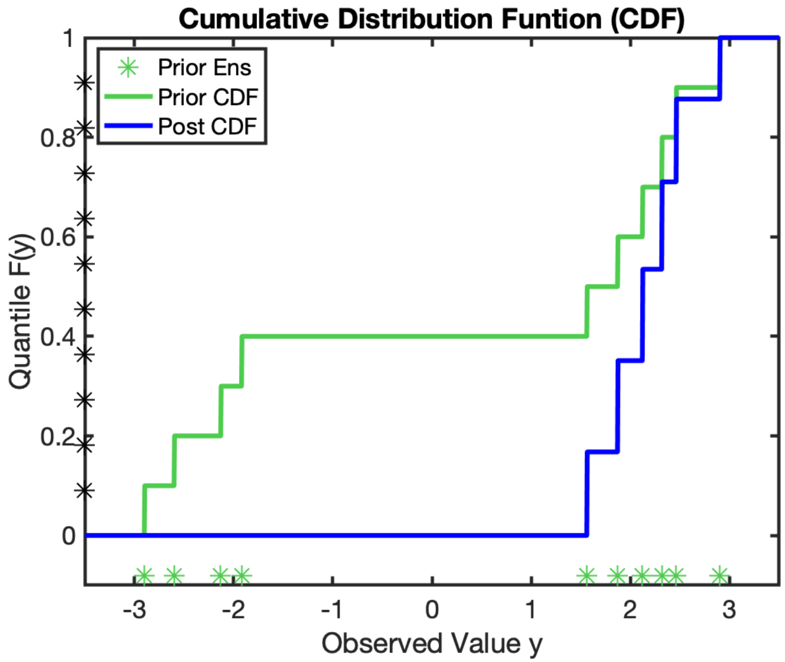




Quantile Conserving Particle Filter

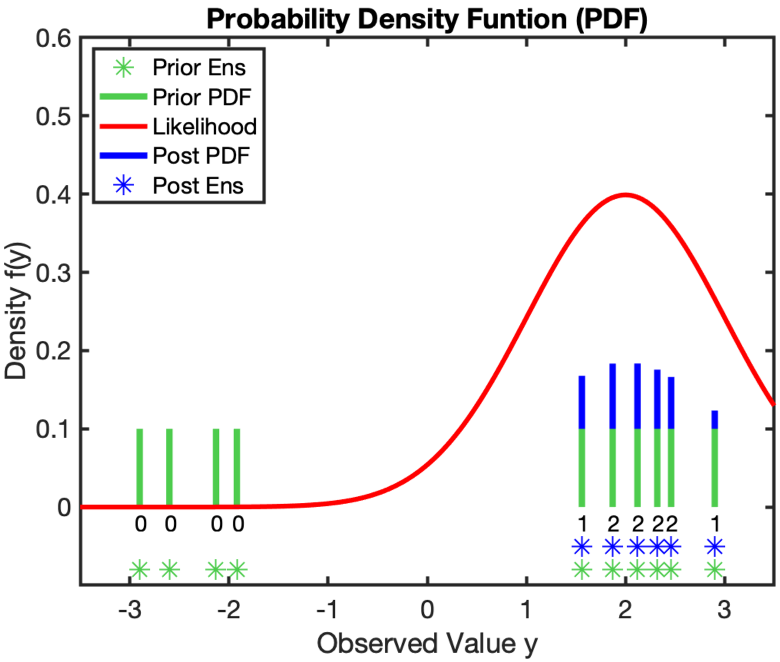
Posterior PDF is prior times normalized likelihood weights.
Get posterior CDF.

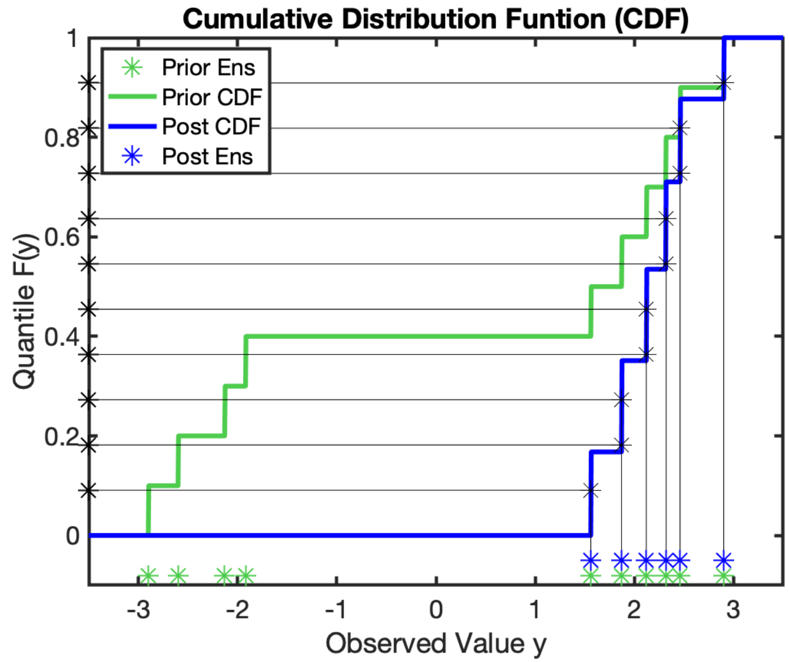




Quantile Conserving Particle Filter

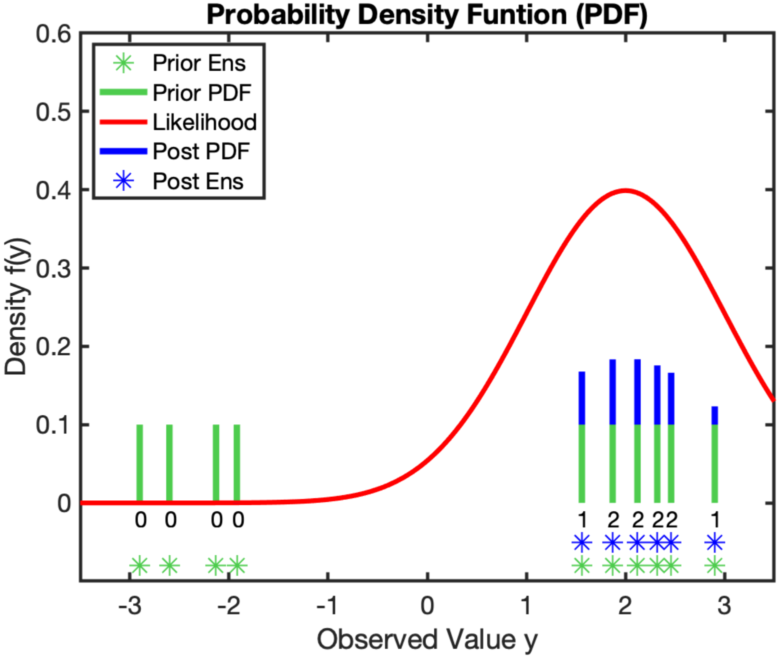
Posterior ensemble members have same quantiles as prior.
This is quantile function, inverse of posterior CDF.

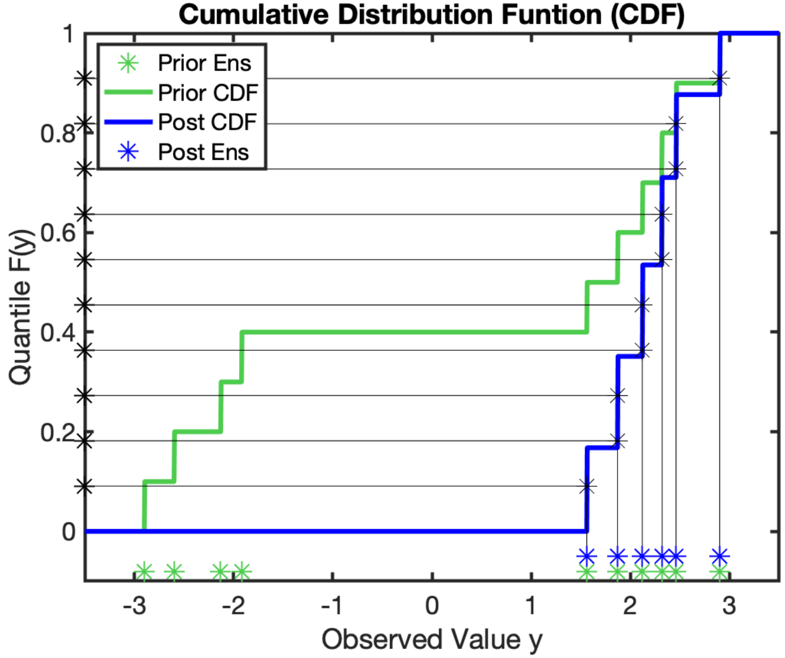




Quantile Conserving Particle Filter

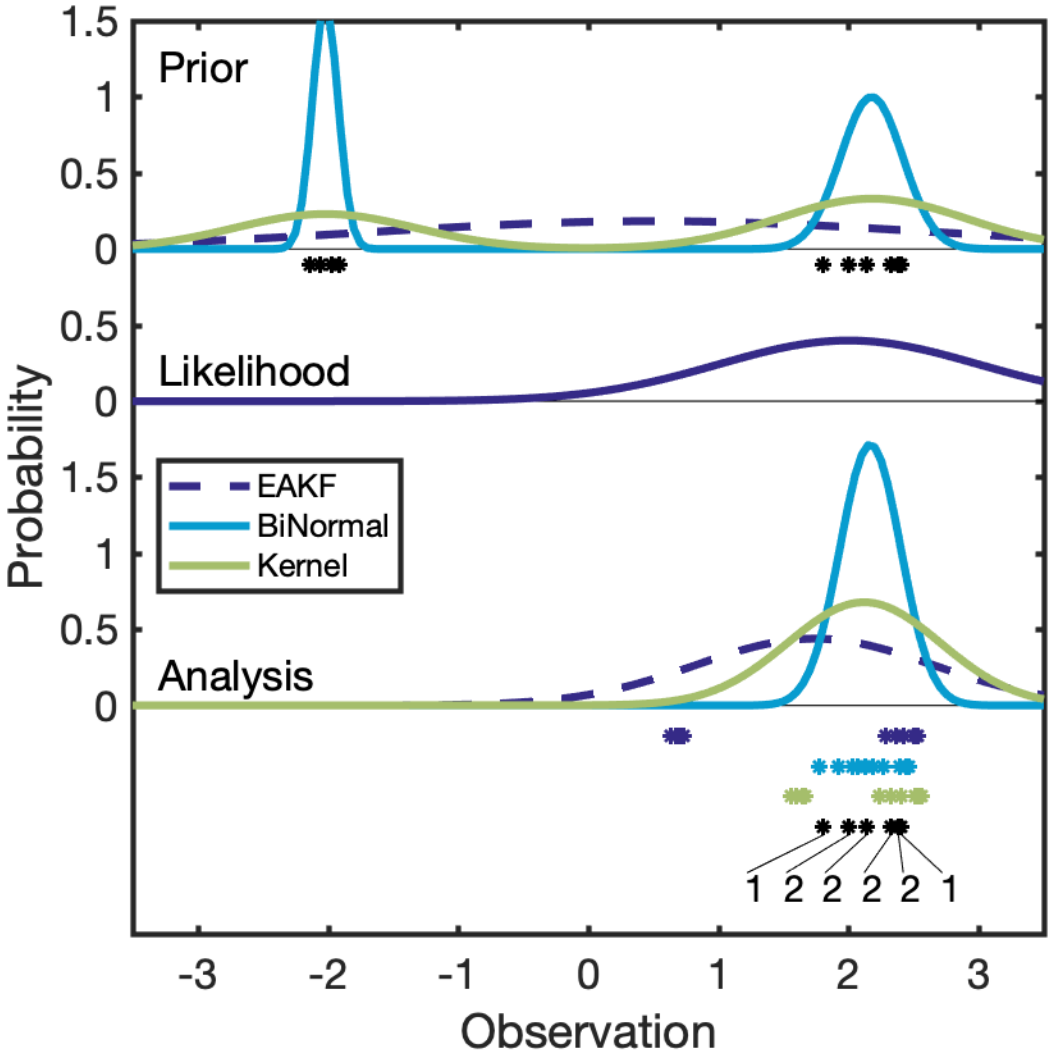
Posterior ensemble members have same quantiles as prior.
This is quantile function, inverse of posterior CDF.





This is a deterministic univariate 'particle filter'.
In a sense, it's an optimal transport 'particle filter'.

Comparison to Other QCEFF Prior Distribution Choices



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

1. Normal (EAKF),
2. Weighted binormal
3. Normal kernel centered on each ensemble member,
4. Particle.

Statistical Foundation of QCEFF

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.

Extension to multivariate application

Apply QCEFF directly to an unobserved state variable x,
Only need likelihood for ensemble members.

Can directly update marginal ensemble for any state variable,
Or any function of a state variable.

Need ‘localization’ for small ensemble success.

Note: without localization this converges to the same distribution as the stochastic particle filter in limit of large ensemble size.

Likelihood Localization for QCEFF

Standard localization of increments can lead to posterior members that are inconsistent with prior.

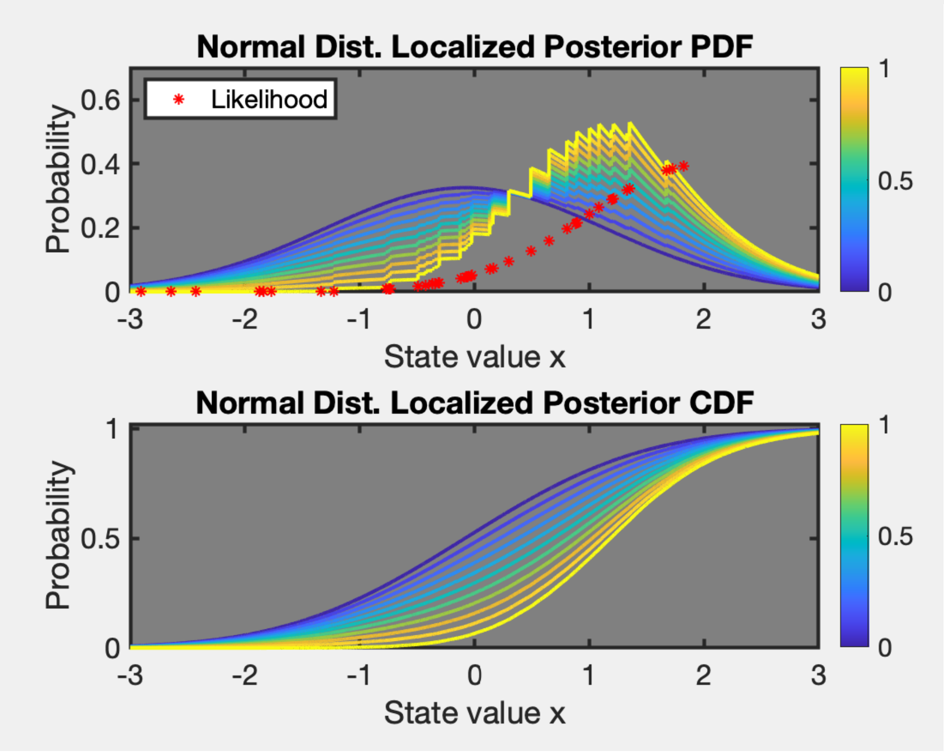
Likelihood localization adjusts the likelihood that gives the relationship between an observation and a state variable.

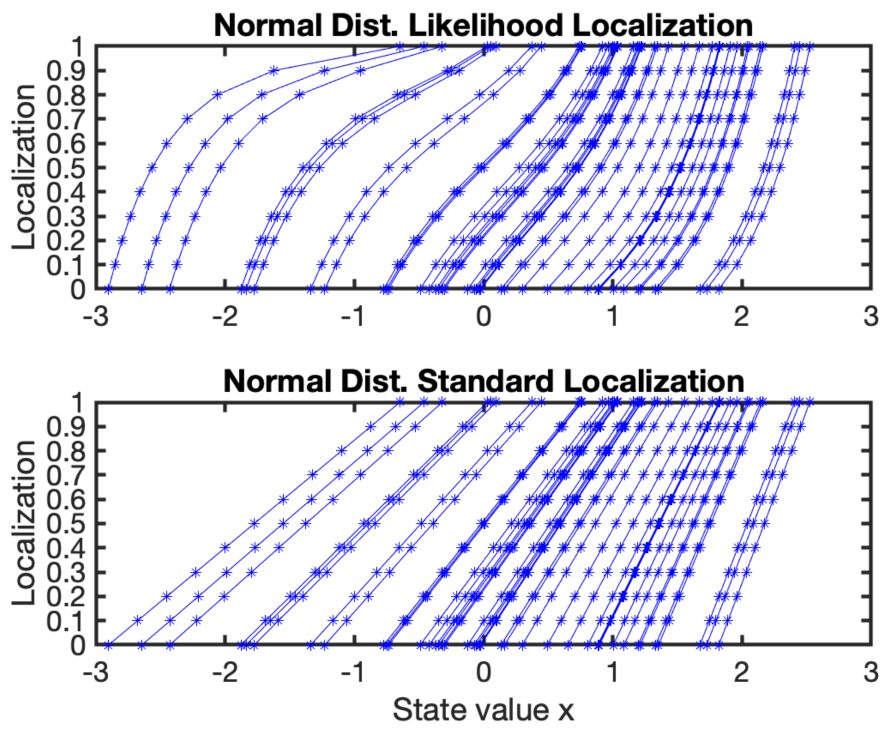
Localized likelihood gives analysis PDF that is weighted average of prior and unlocalized analysis using Euclidean distance metric.

Avoids analysis that is (obviously) inconsistent with prior or likelihood.

Likelihood Localization, Normal Prior

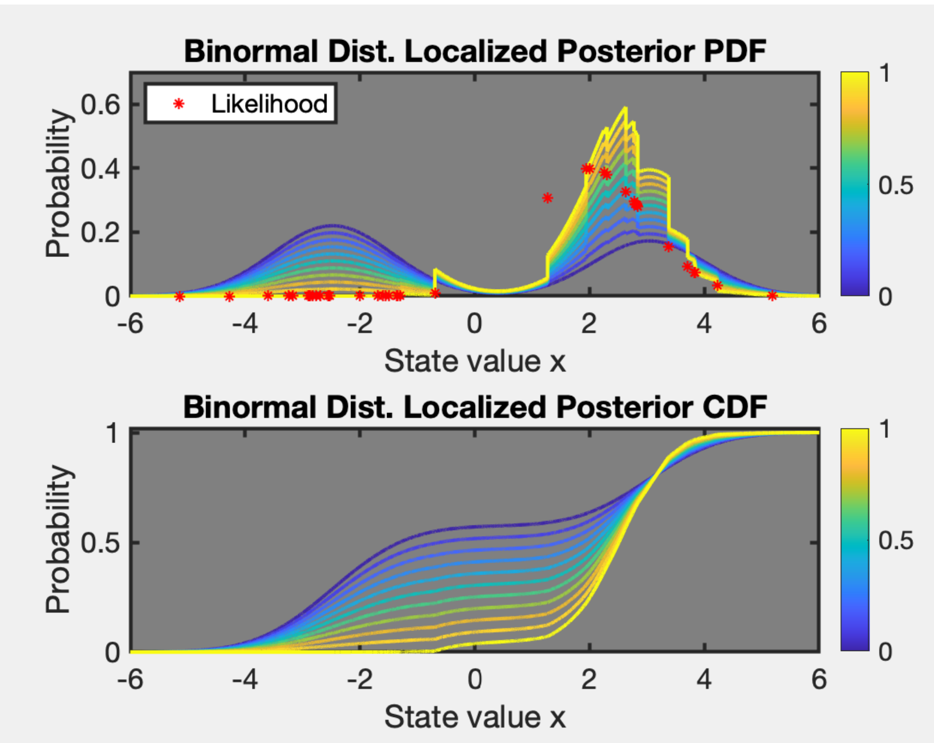
Even for normal prior, localization is quite different.

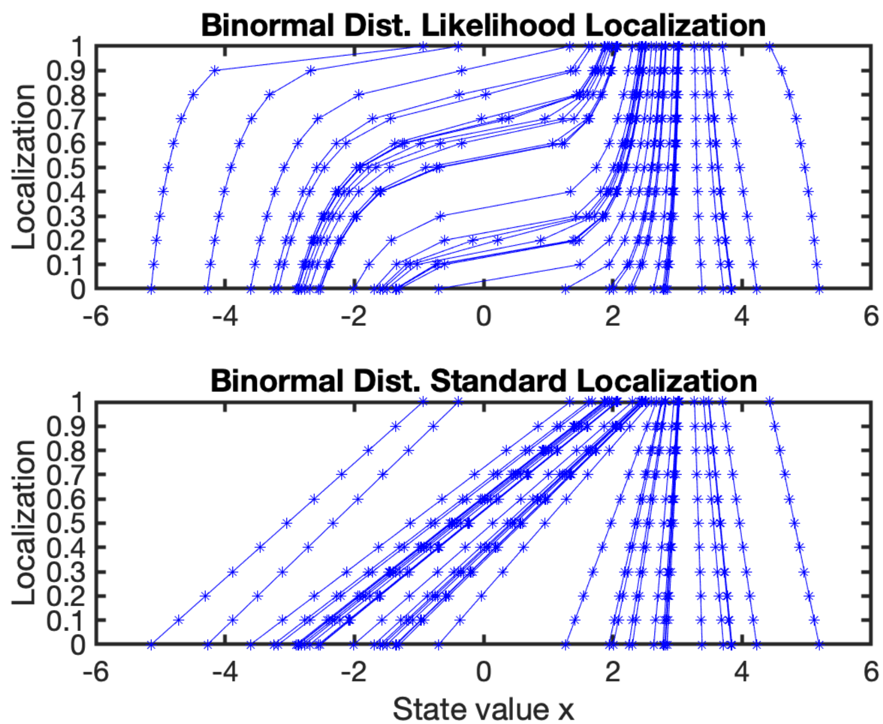




Likelihood Localization, Binormal Prior

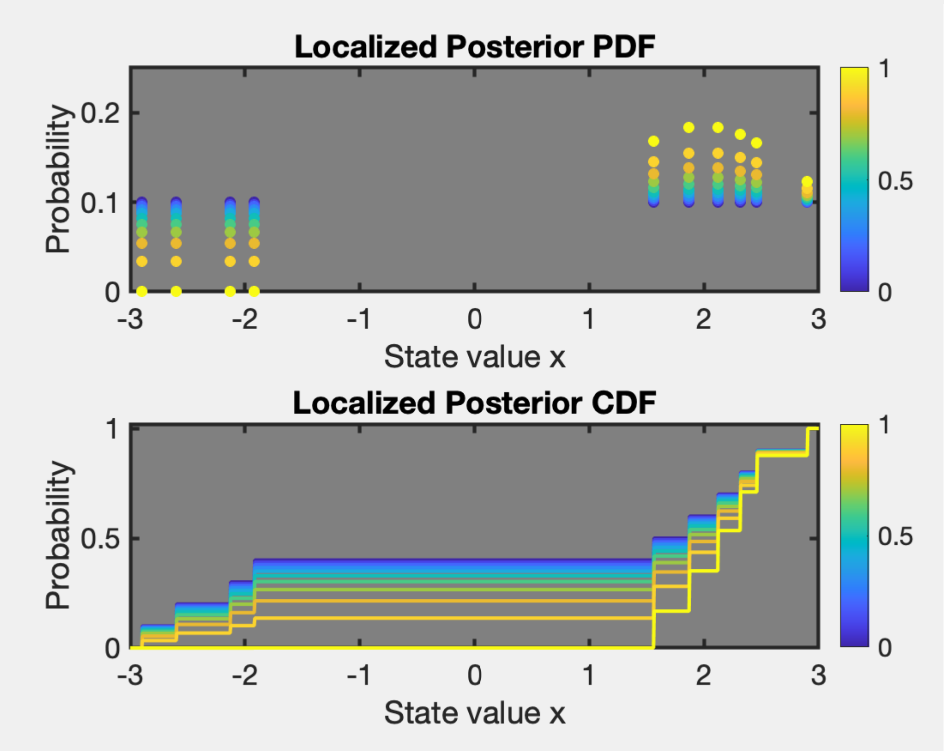
Binormal, bad behavior of standard localization apparent.
(Or could be good behavior, user must decide)

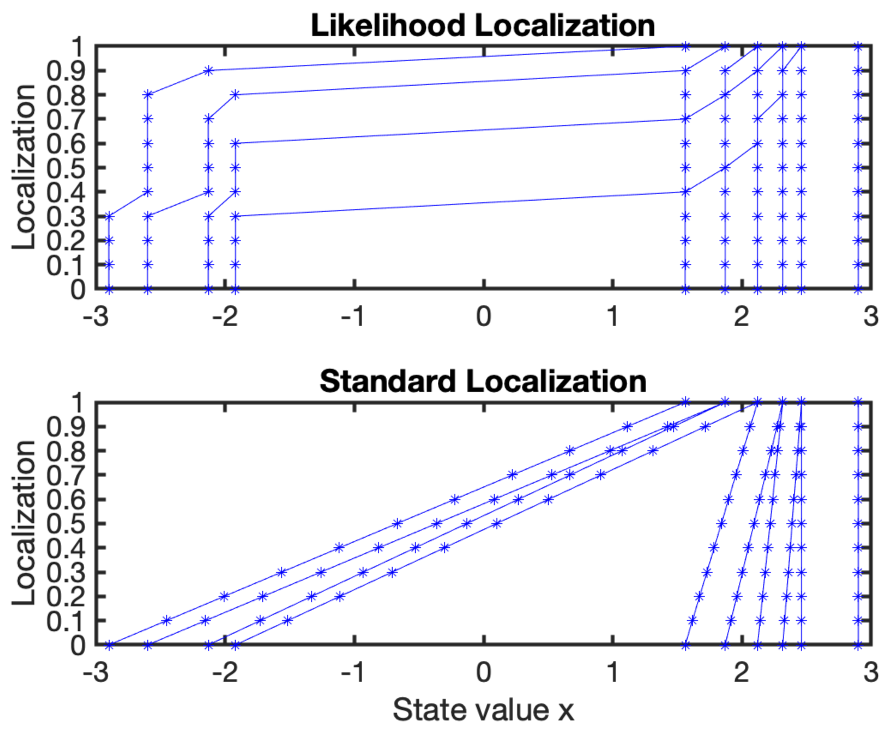




Likelihood Localization, Particle Filter

Likelihood localized posterior members are all selected from set of prior members.





Challenge: Multivariate Posterior Statistics

Marginal statistics with this method are good.

However, multivariate statistics like correlation can be poor with localization.

The rank order of ensemble for each state is unchanged.

Need some way to adjust the rank order to improve multivariate statistics.

Pairing Marginal Values to Ensemble Members

Marginal Adjustment method (MA-QCEFF):

Do standard state variable updates with linear regression to get preliminary posterior.

Get QCEFF particle filter posterior marginal for unobserved variable.

Modify rank statistics of posterior to be same as preliminary posterior.

    Ensemble member with smallest preliminary posterior value gets smallest posterior value…

Works well for many applications (but more expensive).

Finally, I may have a way to get around this ugliness…

Quantile Inflation

Need inflation to deal with a variety of error sources.

Want inflated ensemble to be consistent with distributions.

Example: Bounded quantities should stay bounded,
Other characteristics of distribution maintained.

Current inflation methods generally do not guarantee this.

Can implement a particle inflation scheme that gives inflated ensemble members that are all selected from a subset of the prior members.

Basically inflate the prior quantiles and invert.

Efficient Implementation of MA-QCEFF for State

QCEFF particle filter is more expensive:

Inverting CDF to get updated ensemble members,

As described, have to do this for every obs/state variable pair.

An efficient alternative:

Can get the combined likelihood from many observations,

Just multiply the continuous localized likelihoods together,

These are just for particles here, so multiplication is easy,

Only need one posterior computation for all obs,

If there are many obs per state, incremental cost is reduced.

Warning: There is still a cost increase!

Can mix and match.

Only do state space QCEFF for variables with non-normal priors.

Lorenz-96 Results

Preliminary qualitative results:

QCEFF Particle filter doesn't diverge with adaptive multiplicative inflation.

Particle filter is significantly worse than EAKF, EnKF, RHF for approximately normal cases.

Competitive with RHF for highly non-normal cases with bounded observations (better than EAKF, EnKF).

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