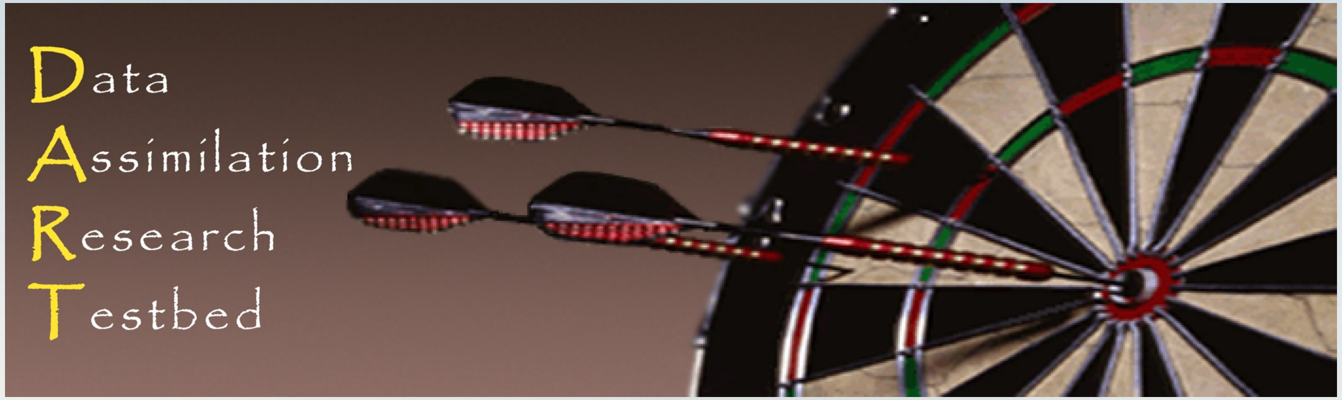
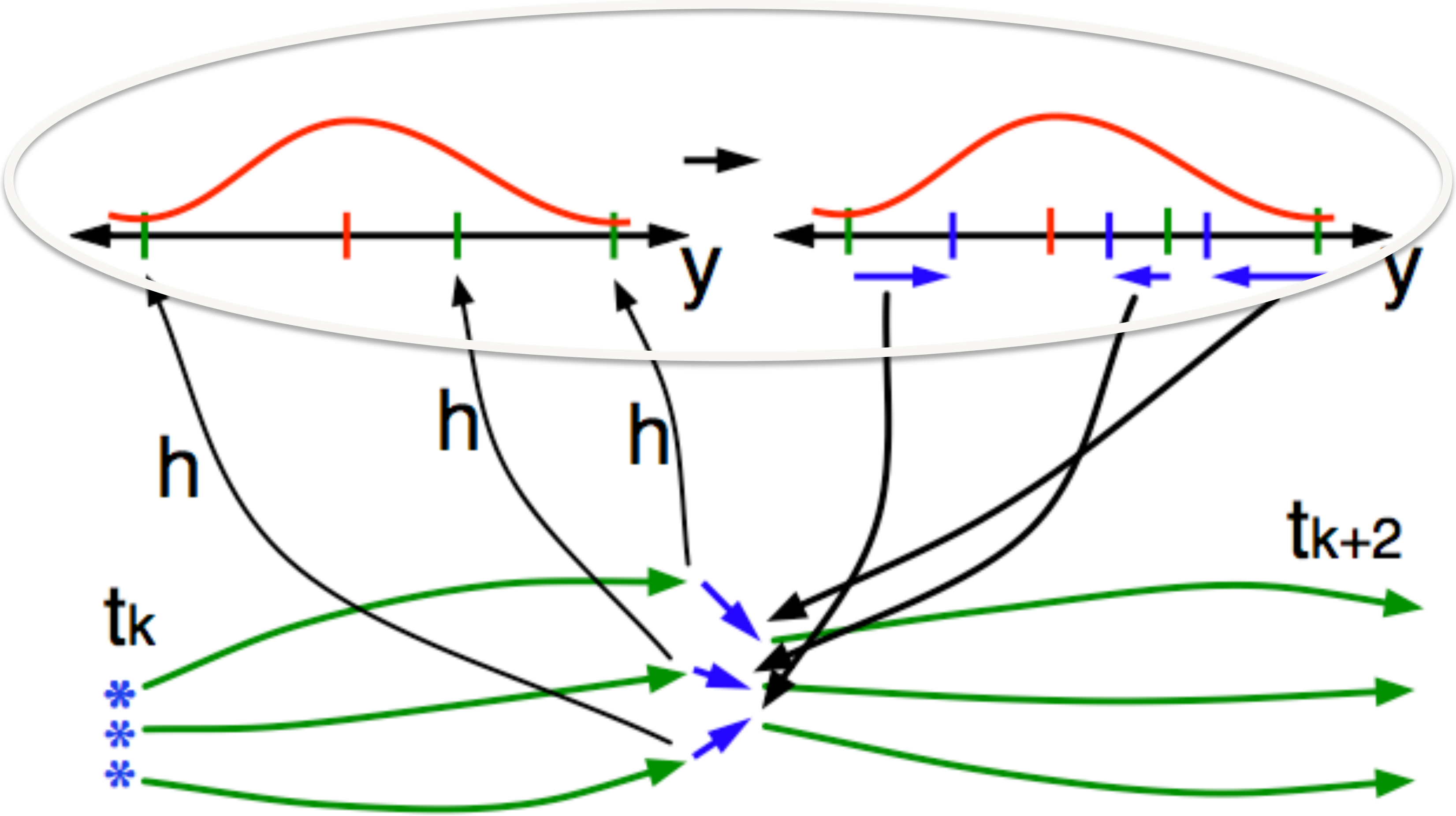


DART_LAB Tutorial Section 4: Nonlinear and Non-Gaussian Extensions

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

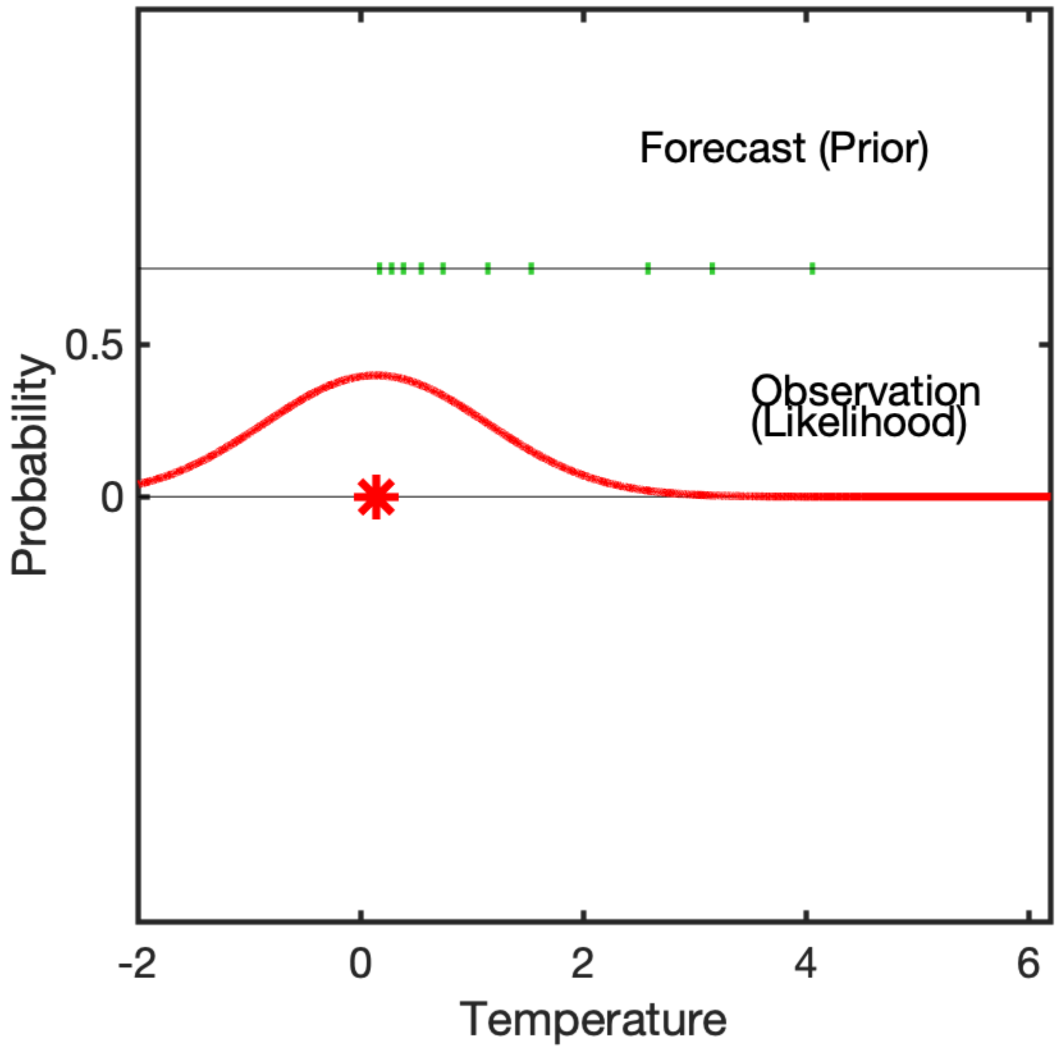
Quantile Conserving Ensemble Filters in Observation Space

DART now provides nearly general solutions for this step:
(Anderson, 2022, MWR*150*, 1061-1074).



The Mesa Lab: Weather can Impact the Commute

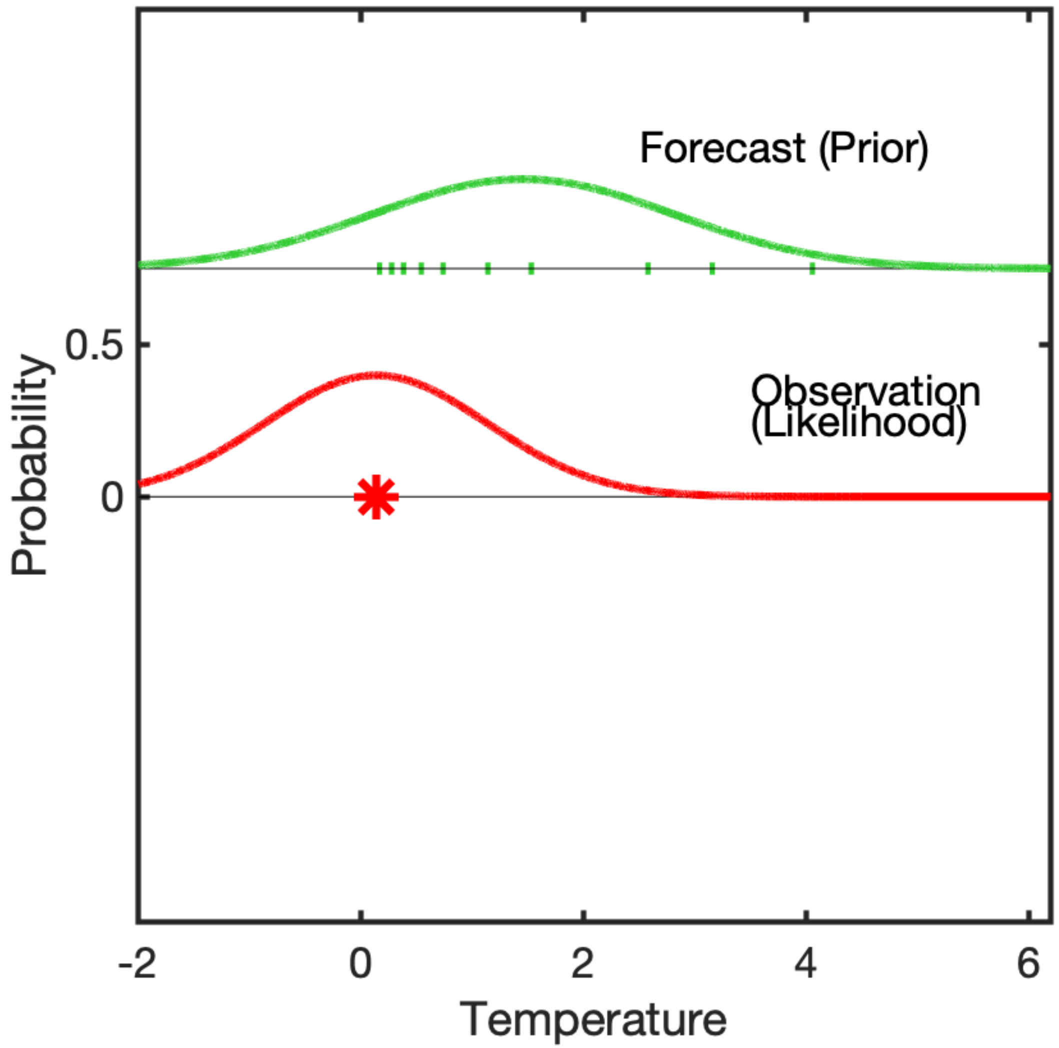
Should I Worry About Ice Going Down the Hill?



Have 10 forecasts of NCAR temperature.

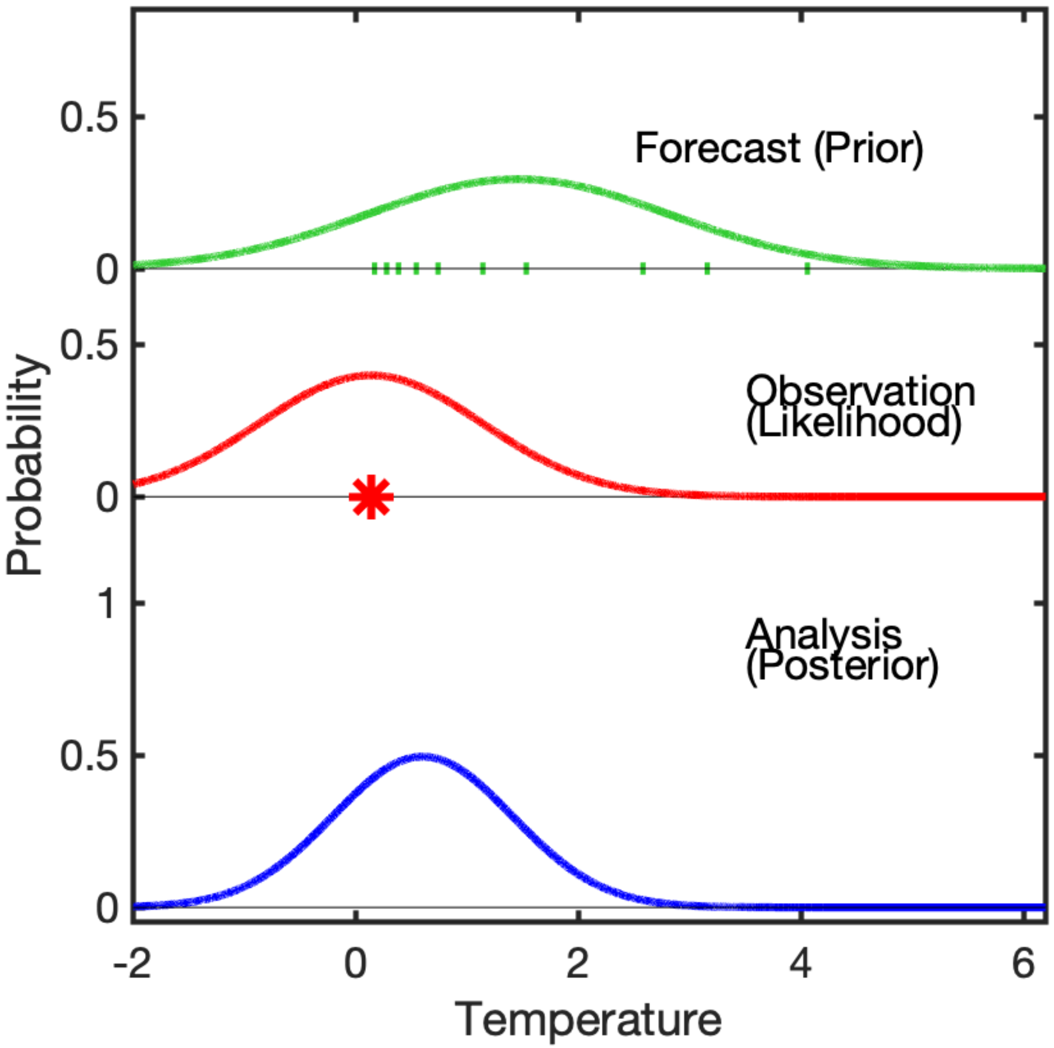
Use Bayes to combine with uncertain NCAR temperature observation.

Should I Worry About Ice Going Down the Hill?



Standard Ensemble Filter:
Fit a normal to the forecast ensemble.

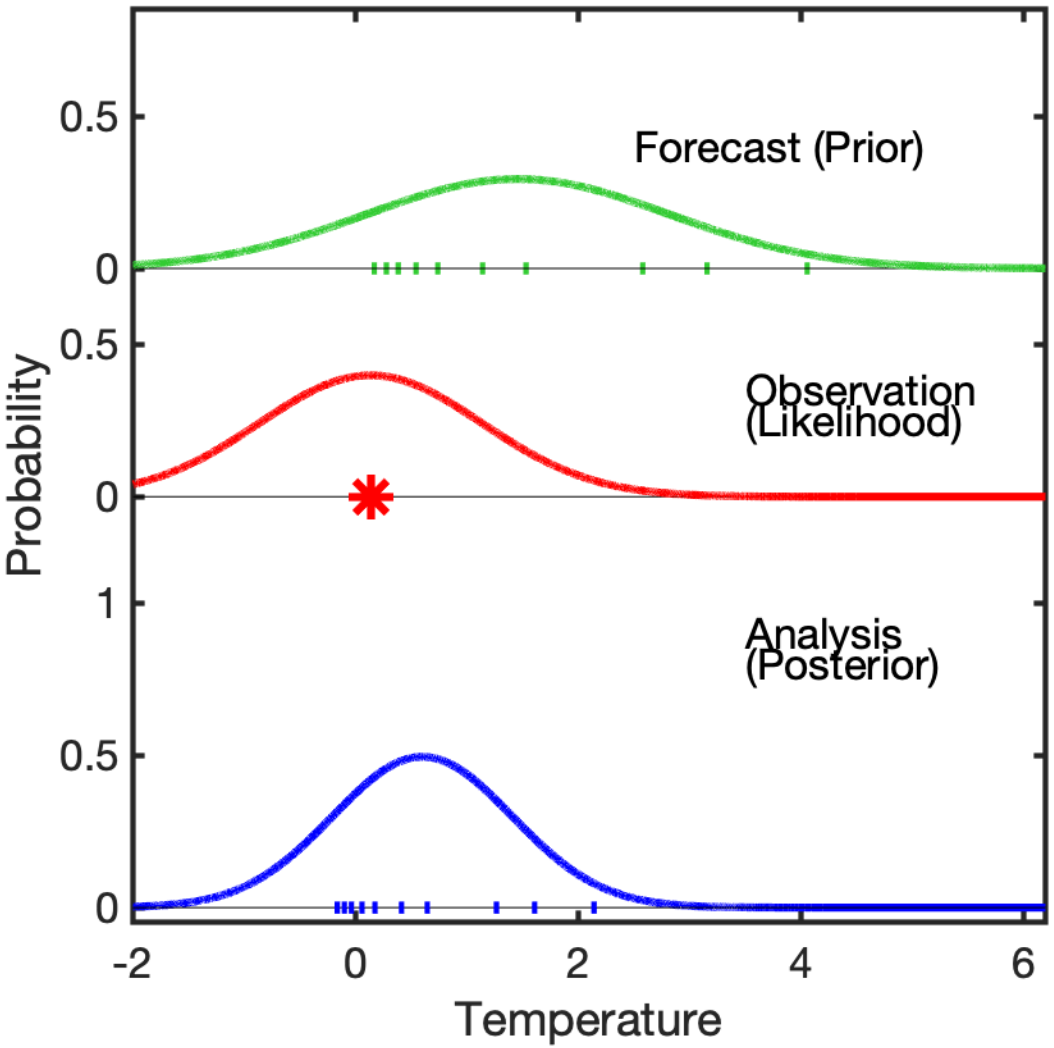
Should I Worry About Ice Going Down the Hill?



Bayes product gives continuous normal posterior.

$$P\left(\mathbf{x}_{t_k}|\mathbf{Y}_k\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{k-1}\right)}{Normalization}$$

Should I Worry About Ice Going Down the Hill?

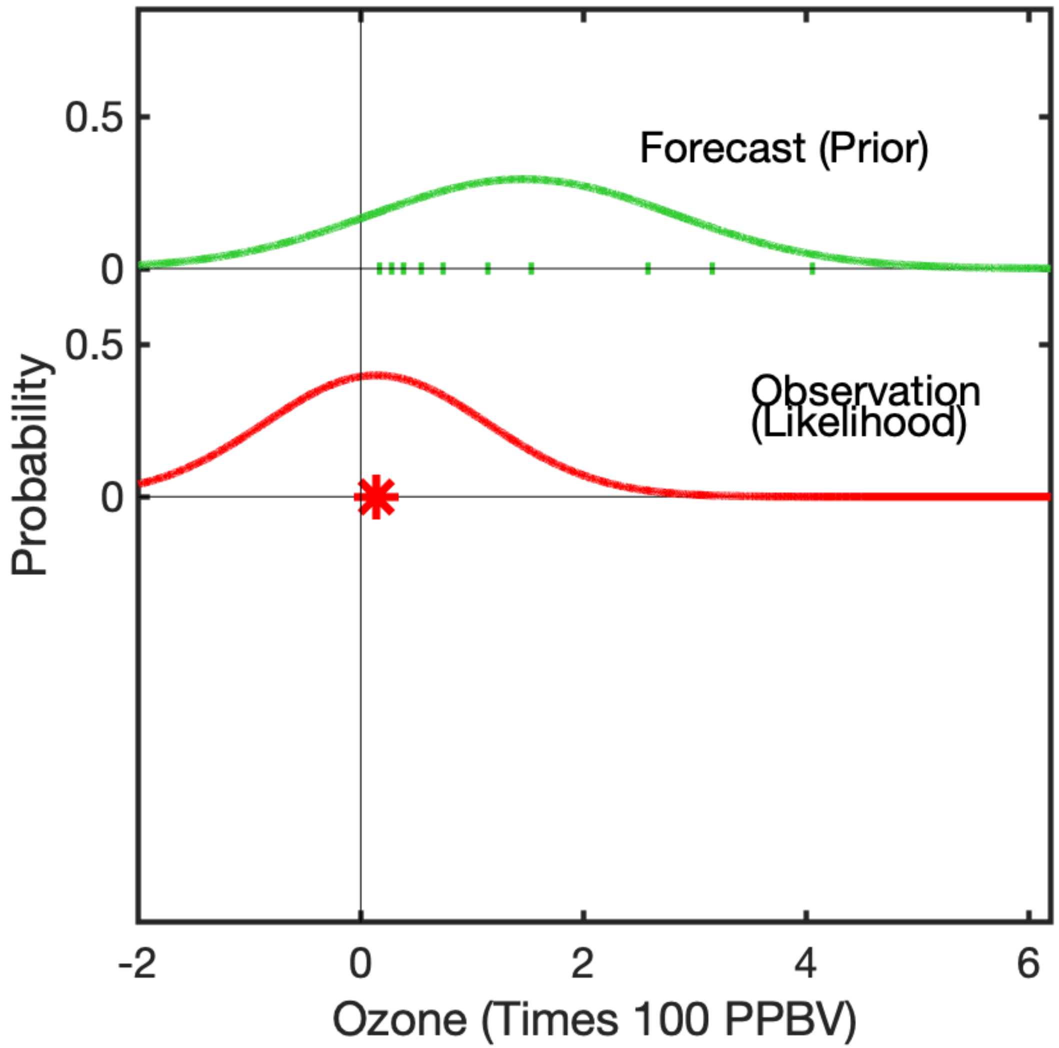


Get a posterior ensemble.

At one time, we only knew how to do this for normal distributions.

Normal may work okay for applications like NWP.

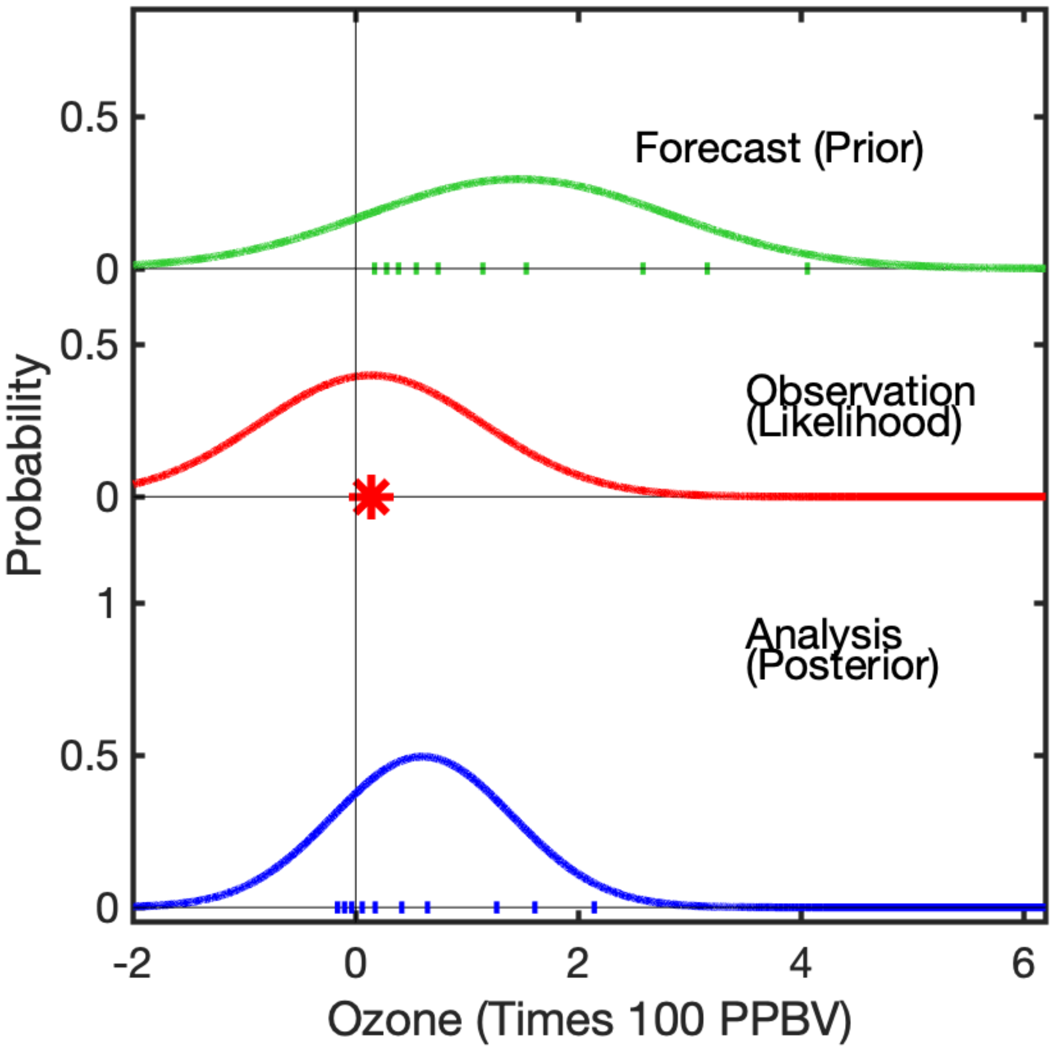
Should I Worry About Air Quality Going Down the Hill?



Forecast model knows ozone must be positive.

Fitting a normal leads to probability of negative.

Should I Worry About Air Quality Going Down the Hill?

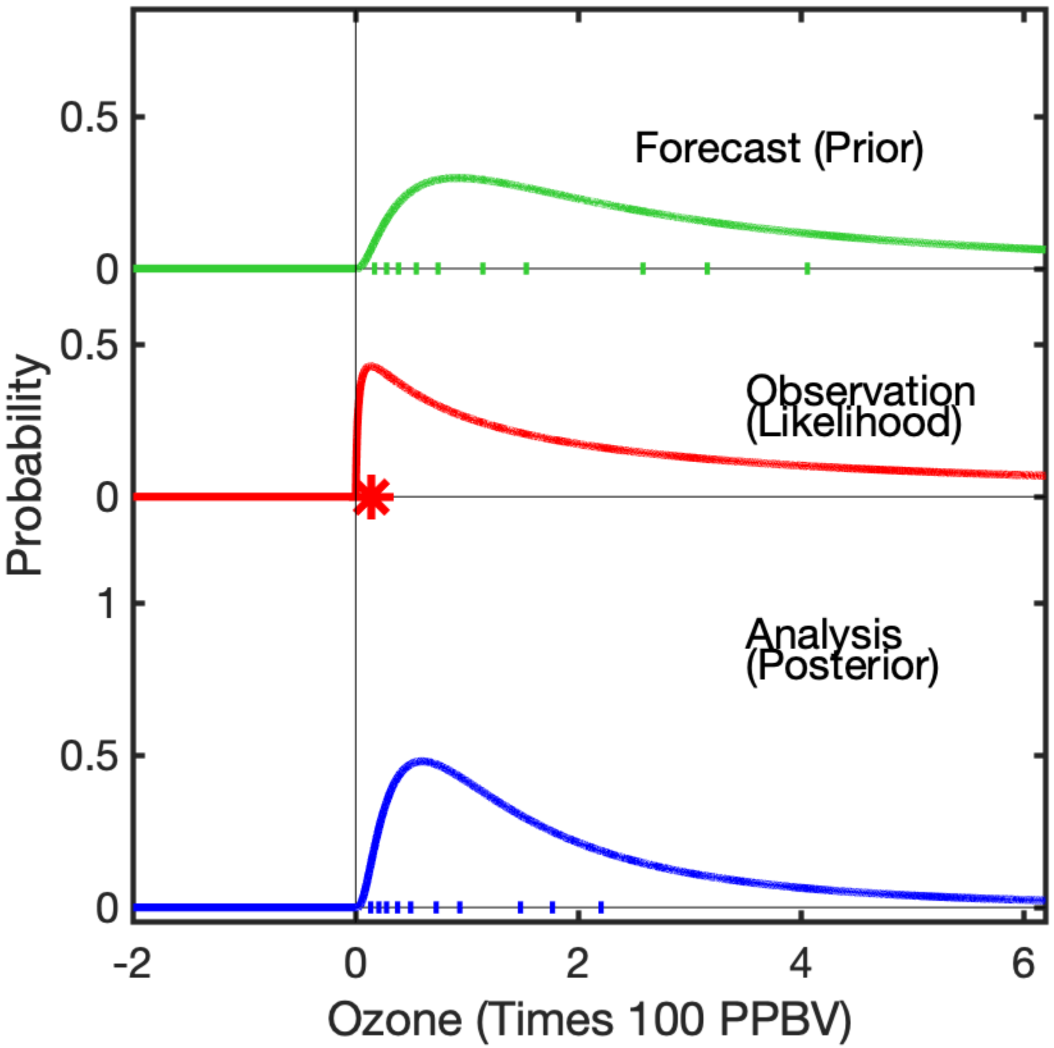


Doing the DA can lead to negative ensemble members.

What does that mean? Not sure, but nothing good.

Putting these back into model to make new forecasts is a problem, too.

Should I Worry About Air Quality Going Down the Hill?

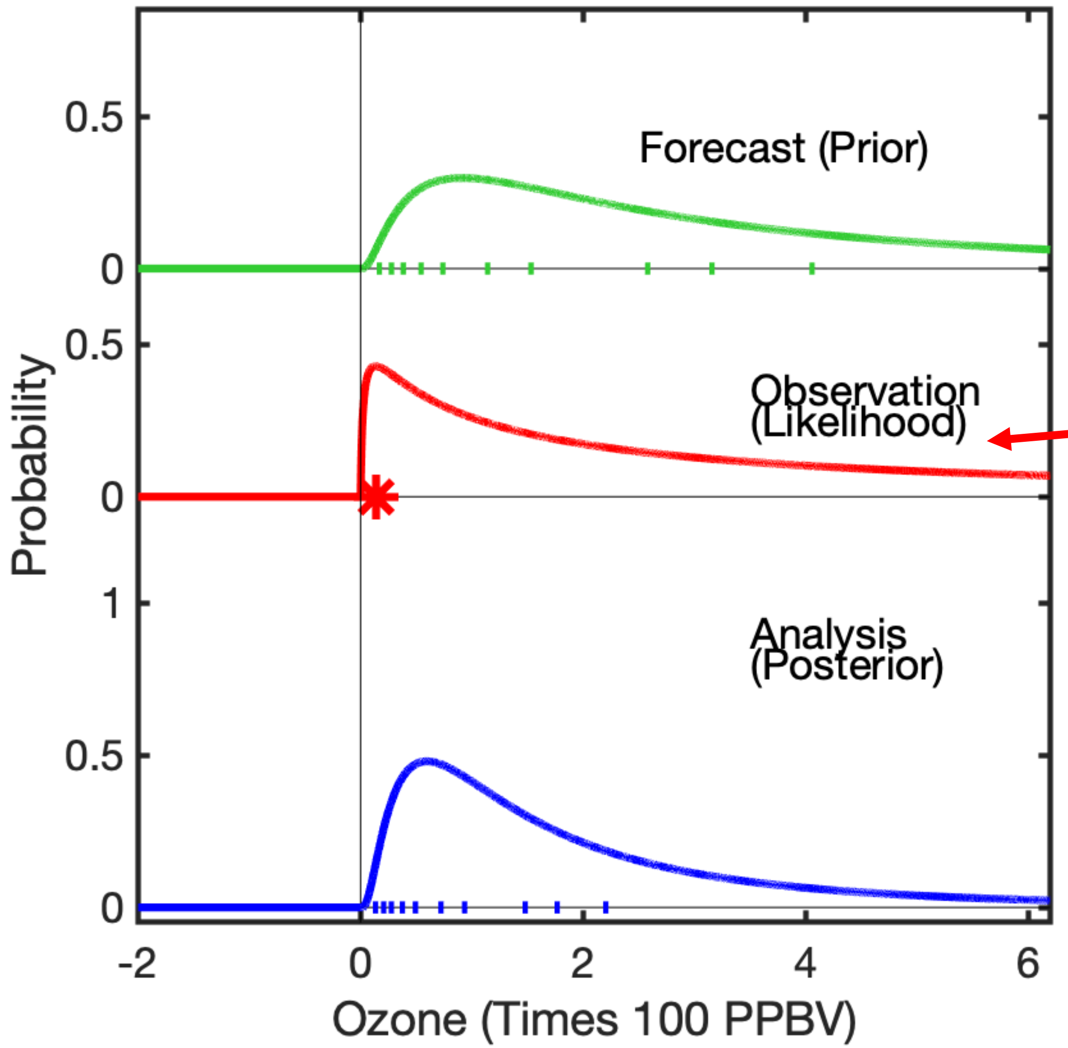


Now can do any distribution using quantile conserving ensemble algorithms.

Example: Gamma for bounded quantity like ozone.

Posterior ensemble no longer crazy.

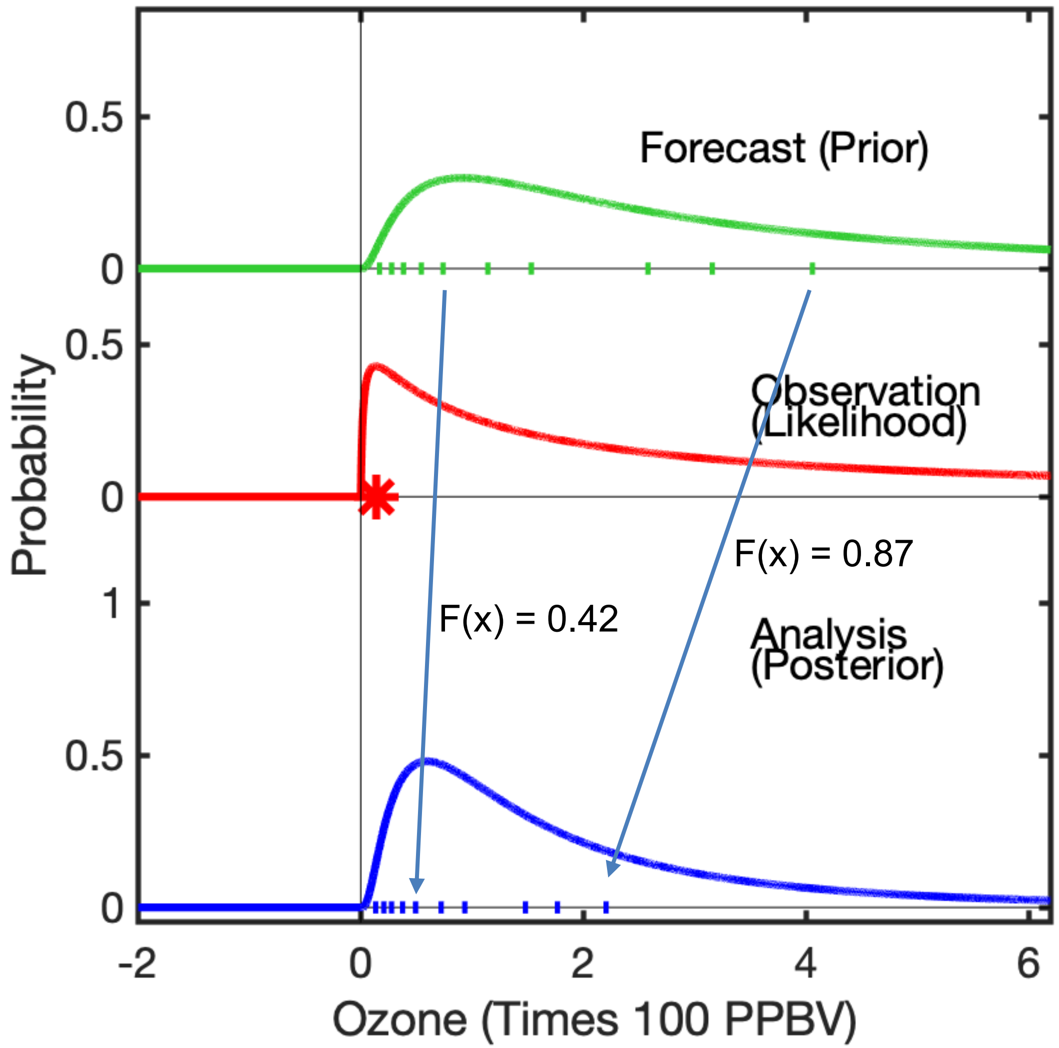
Should I Worry About Air Quality Going Down the Hill?



Now can do any distribution using quantile conserving ensemble algorithms.

Can now use much more general information about observation error from instrument experts. Nice collaborations are possible.

Quantile Conserving Ensemble Filter Framework



How to select ensembles for the analysis distribution?

Conserve quantiles from the prior ensemble.

F is the cumulative distribution function (CDF) for the prior or analysis continuous distribution. It gives the quantile.

Introduction: Quantile Conserving Ensemble Filter Framework

- Ensemble Kalman filters are effective but make implicit assumptions about normal distributions

- Present a generalization for ensemble ~~Kalman~~ filters that can use arbitrary univariate distributions

Key (very simple) Innovation

- Getting continuous PDF from ensemble is often simple.
  - Sample mean, variance for many distributions.

- Getting ‘nice’ ensemble from continuous PDF has been harder.

A Solution: Conserve quantiles of ensemble to sample a modified PDF.

Posterior ensemble quantiles are the same as prior ensemble quantiles.

Quantile Conserving Ensemble Filter Framework

1. Pick any appropriate continuous PDF given a prior ensemble.

2. Get the corresponding CDF, F^p .

3. Compute quantiles of ensemble members, $F^p(x_n^p), \quad n = 1, \cdots, N.$

4. Modify the PDF (filter, inflate, localize, whatever).

5. Get the modified analysis CDF, F^a .

6. Updated ensemble conserves quantiles,

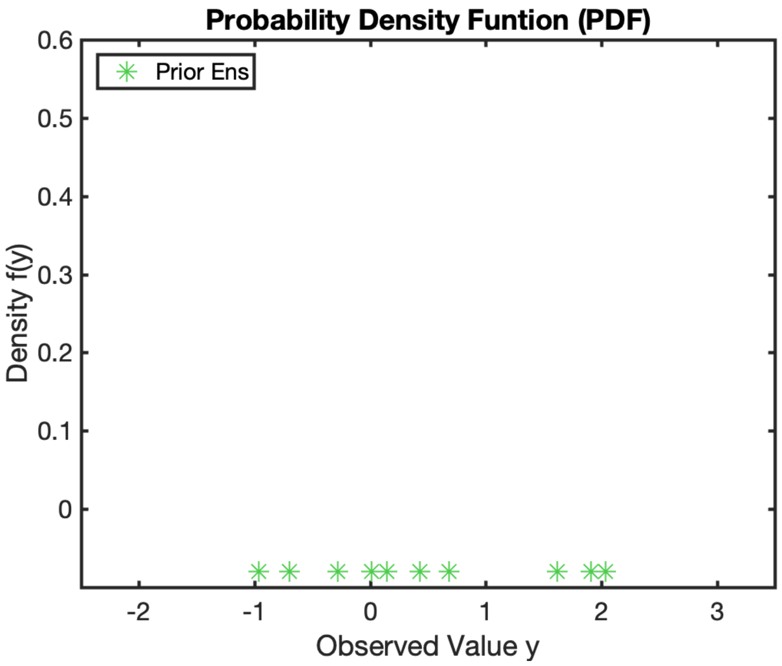
$$x_n^a = (F^a)^{-1}\left[F^p\left(x_n^p\right)\right], \; n = 1, \cdots, N \; .$$

Generalized inverse if F^a is not invertible

$$(F^a)^{-1}(y) = min\{x: y \leq F^a(x)\}.$$

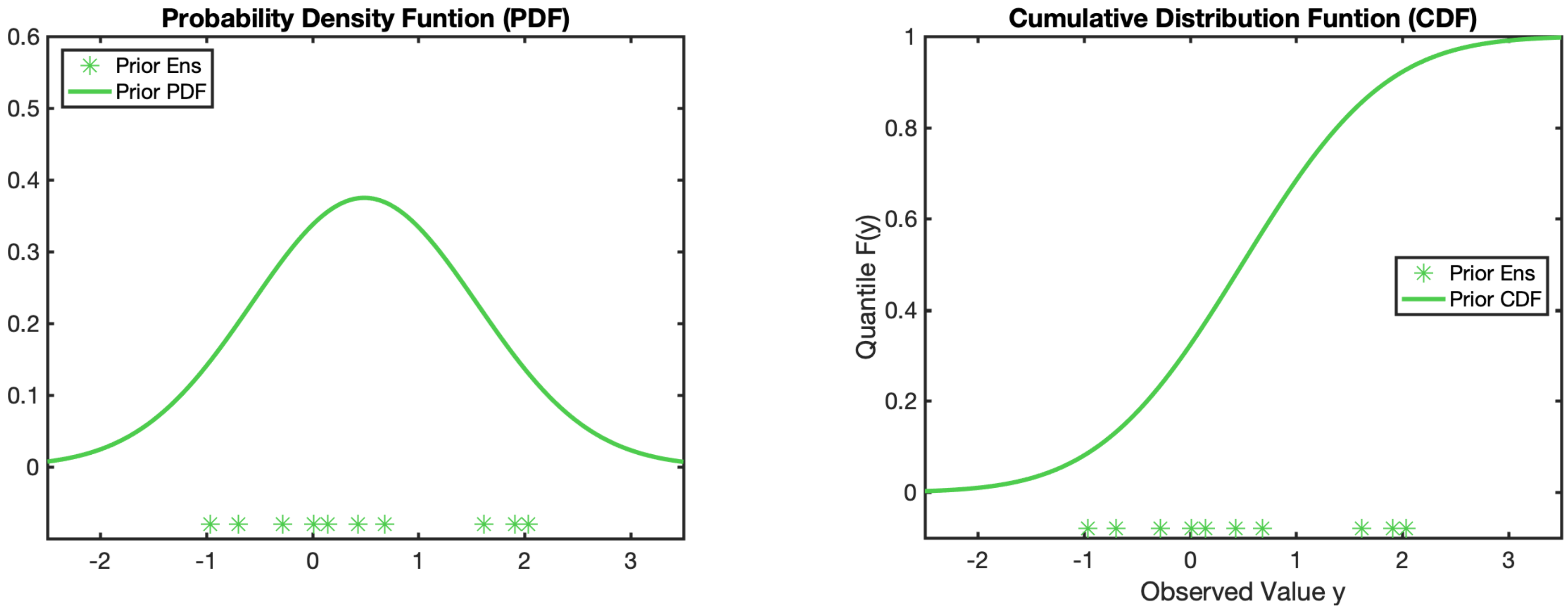
Application 1: Bayesian filtering for an observed variable

Given a prior ensemble estimate of an observed quantity, y,



Application 1: Bayesian filtering for an observed variable

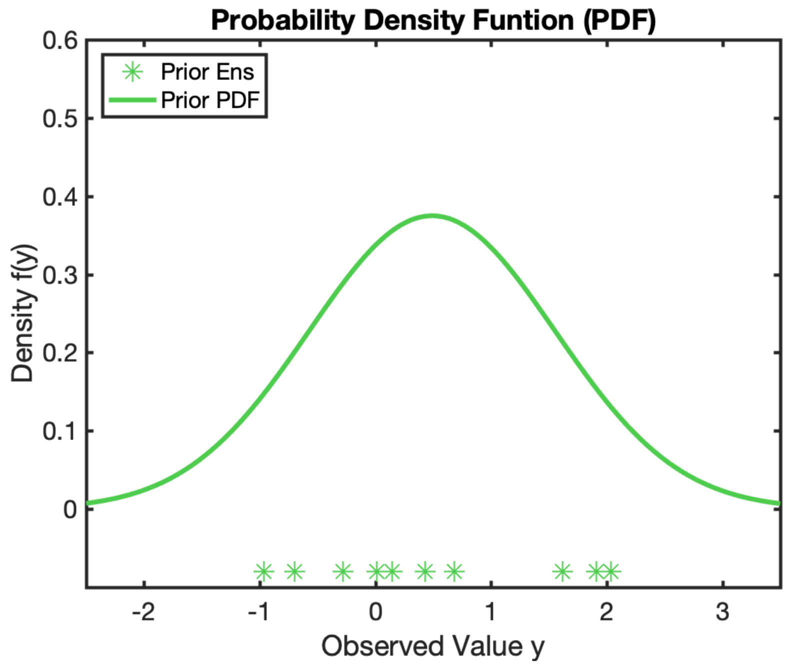
Fit a continuous PDF from an appropriate distribution family and find the corresponding CDF.

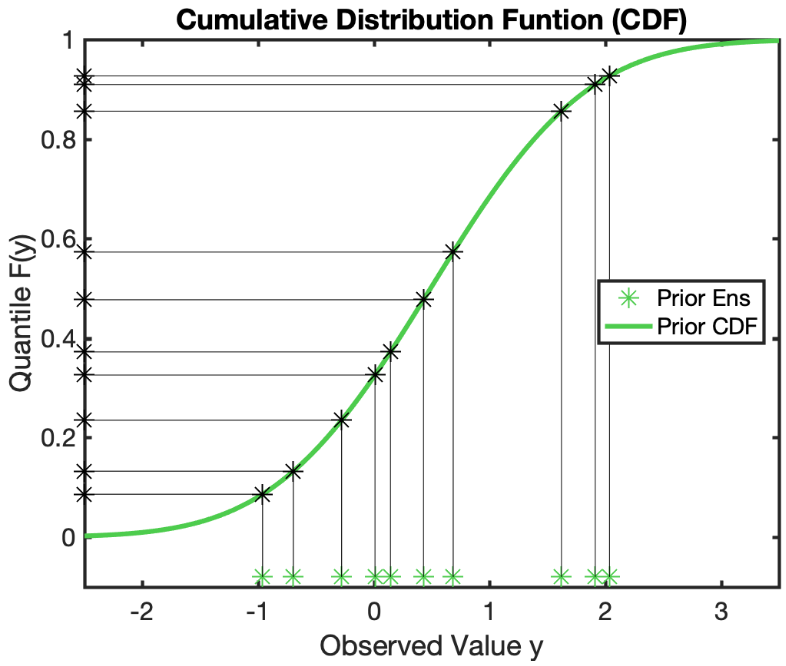


This example uses a normal PDF.

Application 1: Bayesian filtering for an observed variable

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

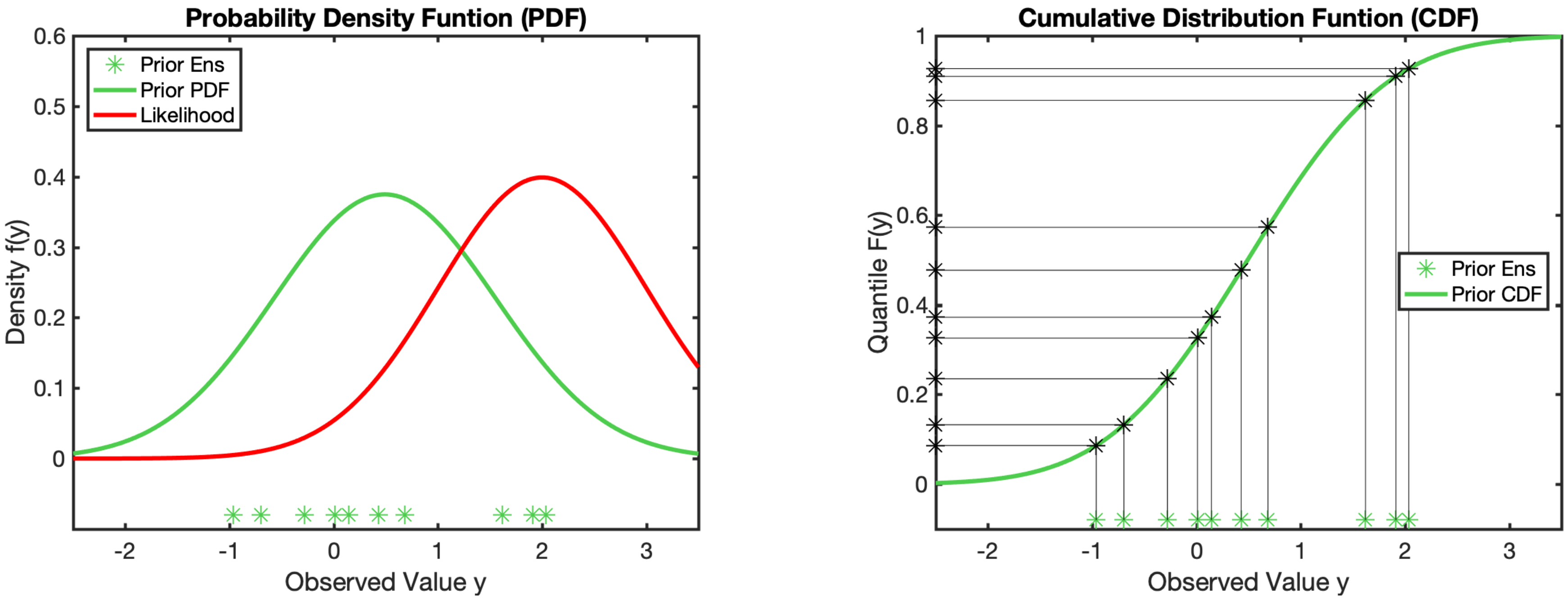




This example uses a normal PDF.

Application 1: Bayesian filtering for an observed variable

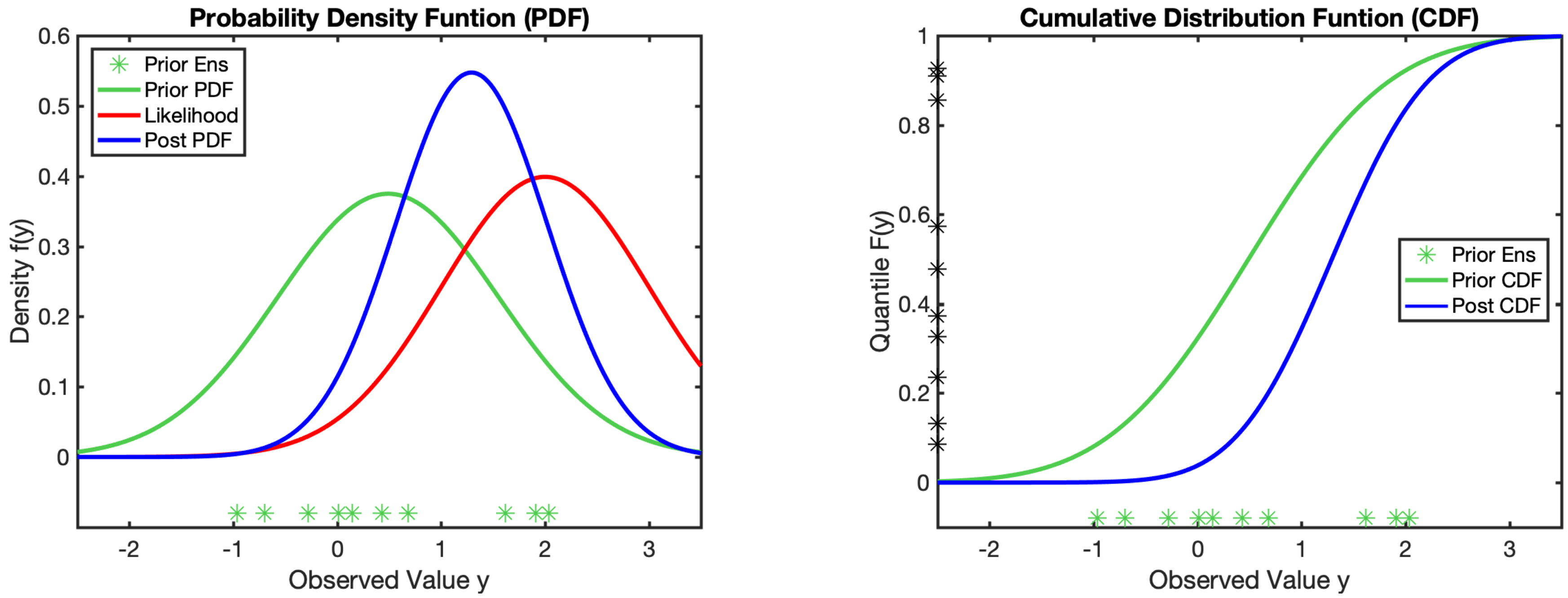
Continuous likelihood for this observation.



This example uses a normal PDF.

Application 1: Bayesian filtering for an observed variable

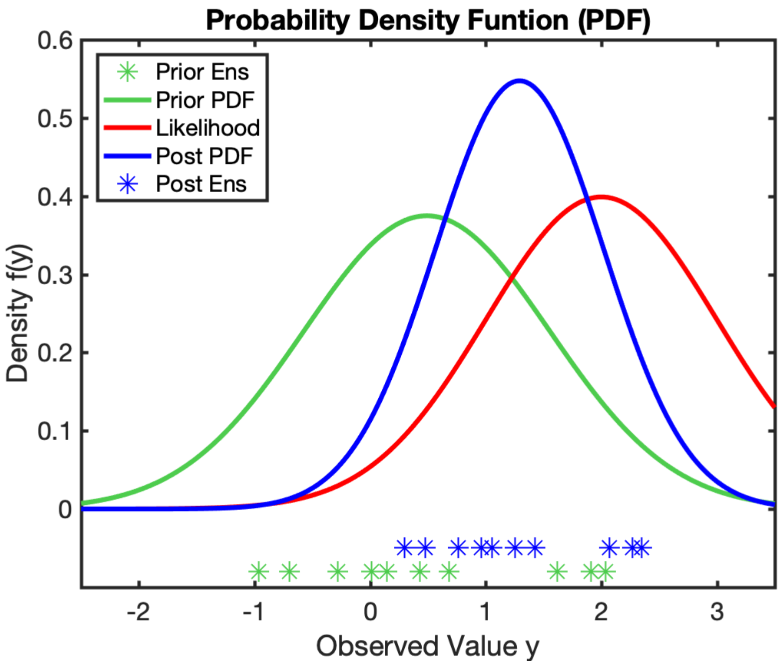
Bayes tells us that the continuous posterior PDF is the product of the continuous likelihood and prior.

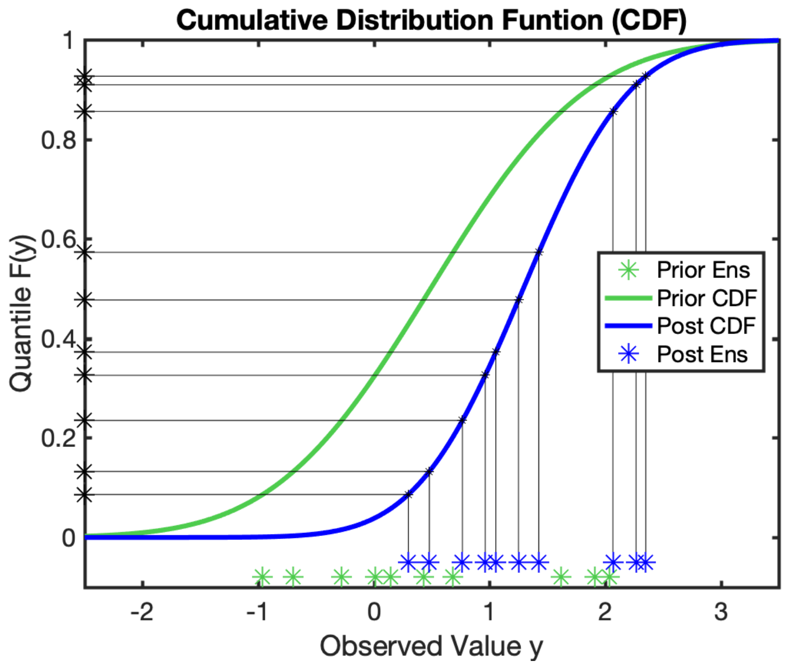


Normal times normal is normal.

Application 1: Bayesian filtering for an observed variable

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

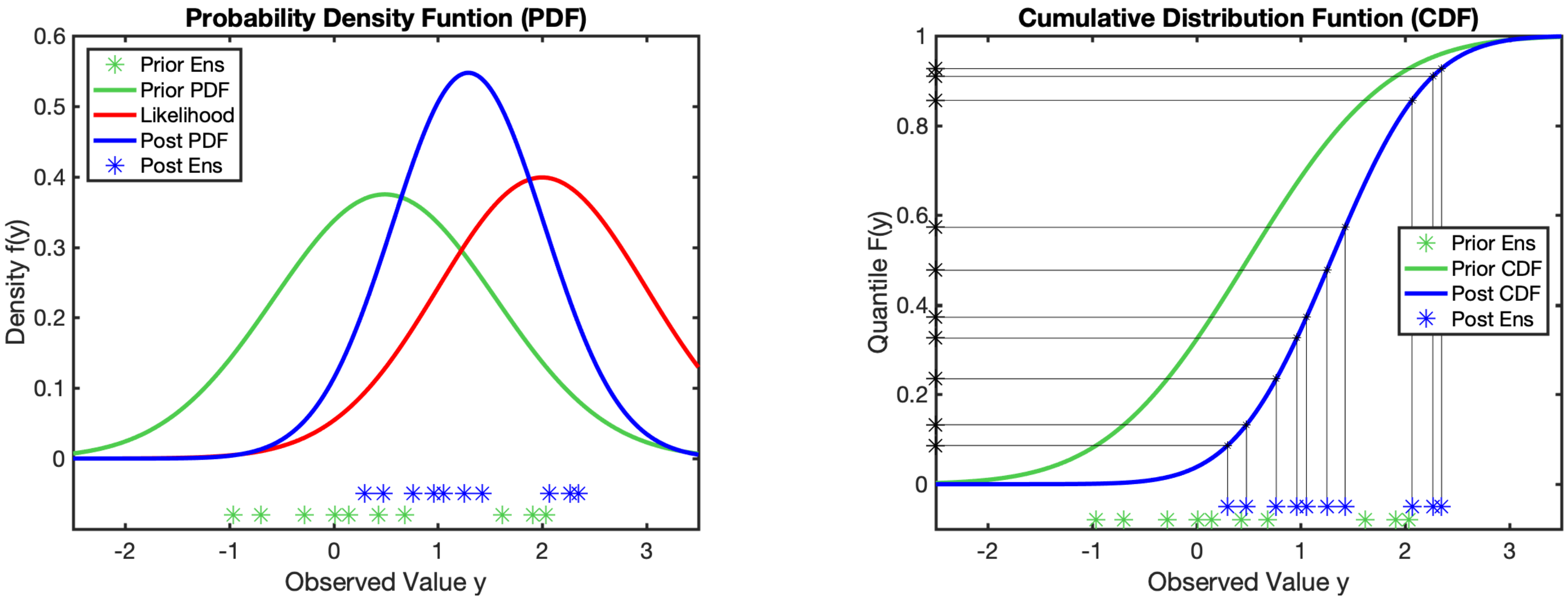




This example uses a normal PDF

Application 1: Bayesian filtering for an observed variable

For normal prior and likelihood, this is identical to existing deterministic Ensemble Adjustment Kalman Filter (EAKF) described in Section 1.



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	Notes
Normal	Normal	Normal	EAKF
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	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods

Prior	Likelihood	Posterior	Notes
Normal	Normal	Normal	EAKF
Lognormal	Lognormal	Lognormal	Trans. EAKF
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	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods

Prior	Likelihood	Posterior	Notes
Normal	Normal	Normal	EAKF
Lognormal	Lognormal	Lognormal	Trans. EAKF
Gamma	Gamma	Gamma	Bishop
Inverse Gamma	Inverse Gamma	Inverse Gamma	Bishop
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	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods

Prior	Likelihood	Posterior	Notes
Normal	Normal	Normal	EAKF
Lognormal	Lognormal	Lognormal	Trans. EAKF
Gamma	Gamma	Gamma	Bishop
Inverse Gamma	Inverse Gamma	Inverse Gamma	Bishop
Beta	Beta	Beta	Doubly
Beta prime	Beta prime	Beta prime	bounded
Exponential	Exponential	Exponential	
Pareto	Pareto	Pareto	
Genl. Gamma given p	Genl. Gamma given p	Genl. Gamma given p	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods

Prior	Likelihood	Posterior	Notes
Normal	Normal	Normal	EAKF
Lognormal	Lognormal	Lognormal	Trans. EAKF
Gamma	Gamma	Gamma	Bishop
Inverse Gamma	Inverse Gamma	Inverse Gamma	Bishop
Beta	Beta	Beta	Doubly
Beta prime	Beta prime	Beta prime	bounded
Exponential	Exponential	Exponential	Applications?
Pareto	Pareto	Pareto	
Genl. Gamma given p	Genl. Gamma given p	Genl. Gamma given p	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods

Prior	Likelihood	Posterior	Notes
Normal	Normal	Normal	EAKF
Lognormal	Lognormal	Lognormal	Trans. EAKF
Gamma	Gamma	Gamma	Bishop
Inverse Gamma	Inverse Gamma	Inverse Gamma	Bishop
Beta	Beta	Beta	Doubly
Beta prime	Beta prime	Beta prime	bounded
Exponential	Exponential	Exponential	Applications?
Pareto	Pareto	Pareto	
Genl. Gamma given p	Genl. Gamma given p	Genl. Gamma given p	
Gamma	Poisson	Gamma	
Skew normal	Normal	Skew normal	Hodyss & Campbell
Truncated normal	Normal	Trunc. normal	

Useful families for continuous priors and likelihoods (2)

Prior	Likelihood	Posterior	Notes
Bounded Normal Rank Histogram	Any	Bounded Normal Rank Histogram (except tails)	Nearly non-parametric
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	
Delta function	Any	'Weighted' delta function	
Any	Piecewise constant	Piecewise weighted	

Useful families for continuous priors and likelihoods (2)

Prior	Likelihood	Posterior	Notes
Bounded Normal Rank Histogram	Any	Bounded Normal Rank Histogram (except tails)	Nearly non-parametric
Huber	Huber	Piecewise normal and exponential	Outliers
Weighted sum of two normals	Normal	Weighted sum of two normals	(also Chan)
Sum of N normals same variance	Normal	Weighted sum of N normals same variance	
Delta function	Any	'Weighted' delta function	
Any	Piecewise constant	Piecewise weighted	

Useful families for continuous priors and likelihoods (2)

Prior	Likelihood	Posterior	Notes
Bounded Normal Rank Histogram	Any	Bounded Normal Rank Histogram (except tails)	Nearly non-parametric
Huber	Huber	Piecewise normal and exponential	Outliers
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	Kernel filter, A&A 1999
Delta function	Any	'Weighted' delta function	
Any	Piecewise constant	Piecewise weighted	

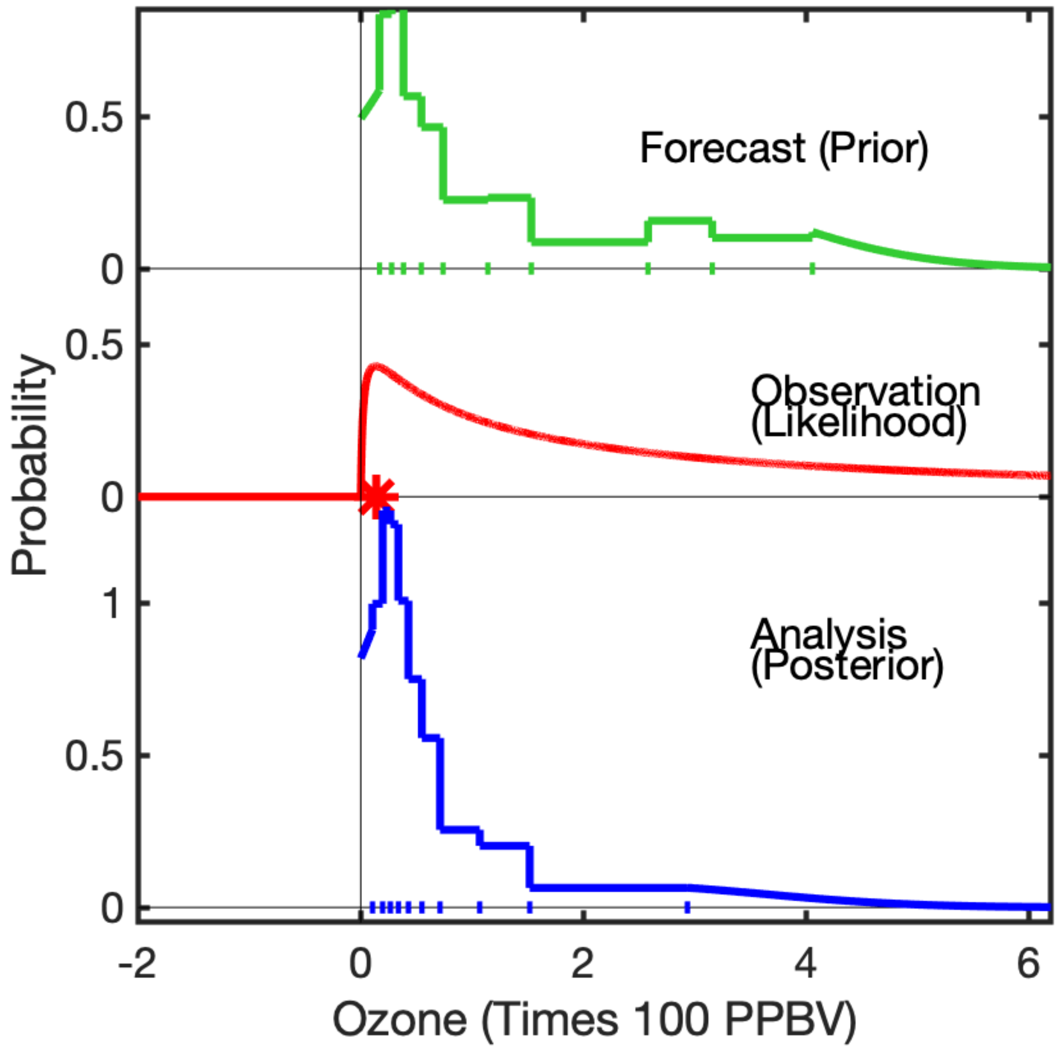
Useful families for continuous priors and likelihoods (2)

Prior	Likelihood	Posterior	Notes
Bounded Normal Rank Histogram	Any	Bounded Normal Rank Histogram (except tails)	Nearly non-parametric
Huber	Huber	Piecewise normal and exponential	Outliers
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	Kernel filter, A&A 1999
Delta function	Any	'Weighted' delta function	'Deterministic' particle filter
Any	Piecewise constant	Piecewise weighted	

Useful families for continuous priors and likelihoods (2)

Prior	Likelihood	Posterior	Notes
Bounded Normal Rank Histogram	Any	Bounded Normal Rank Histogram (except tails)	Nearly non-parametric
Huber	Huber	Piecewise normal and exponential	Outliers
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	Kernel filter, A&A 1999
Delta function	Any	'Weighted' delta function	'Deterministic' particle filter
Any	Piecewise constant	Piecewise weighted	Anything you want by quadrature!!!

What if I Don't Know the Right Distribution?

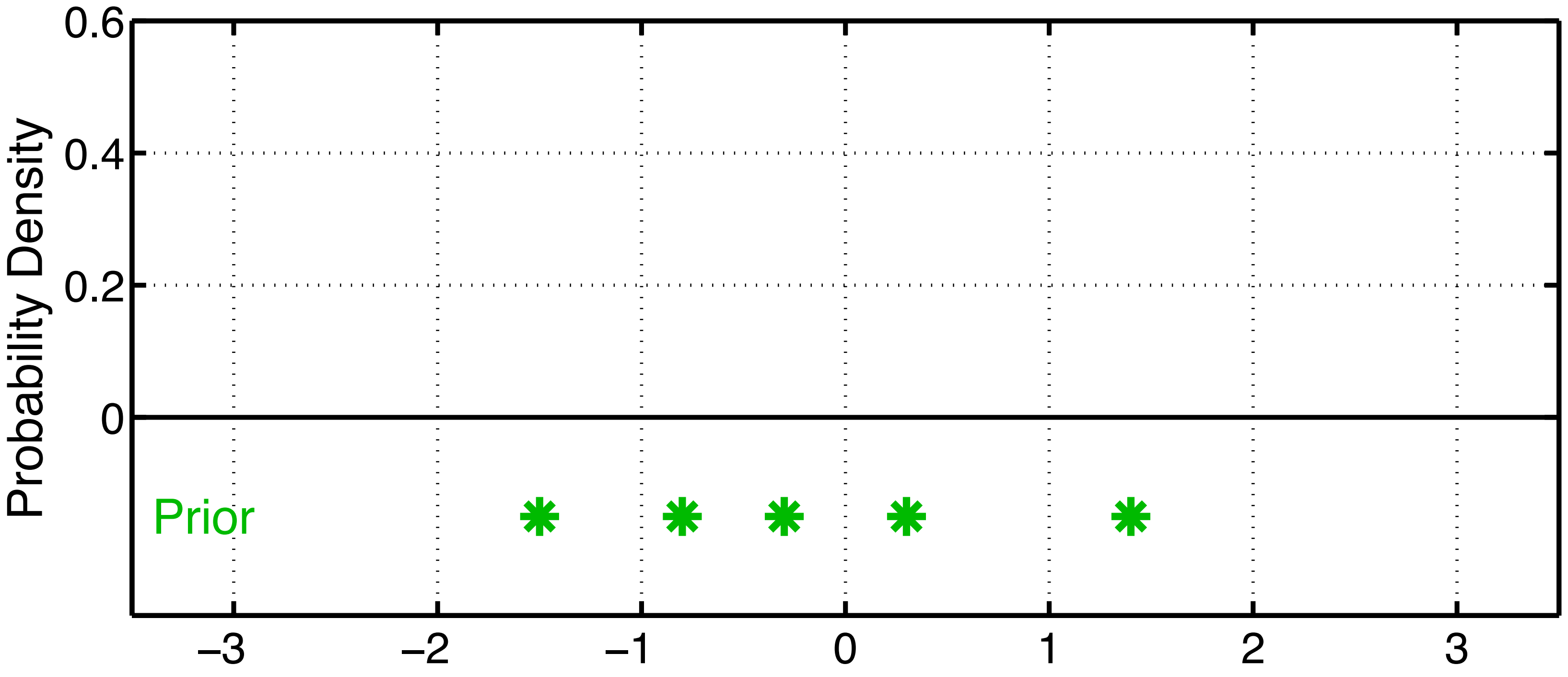


The Bounded Normal Rank Histogram Distribution works well for almost all cases.

Non-parametric. It builds a distribution from the ensemble.

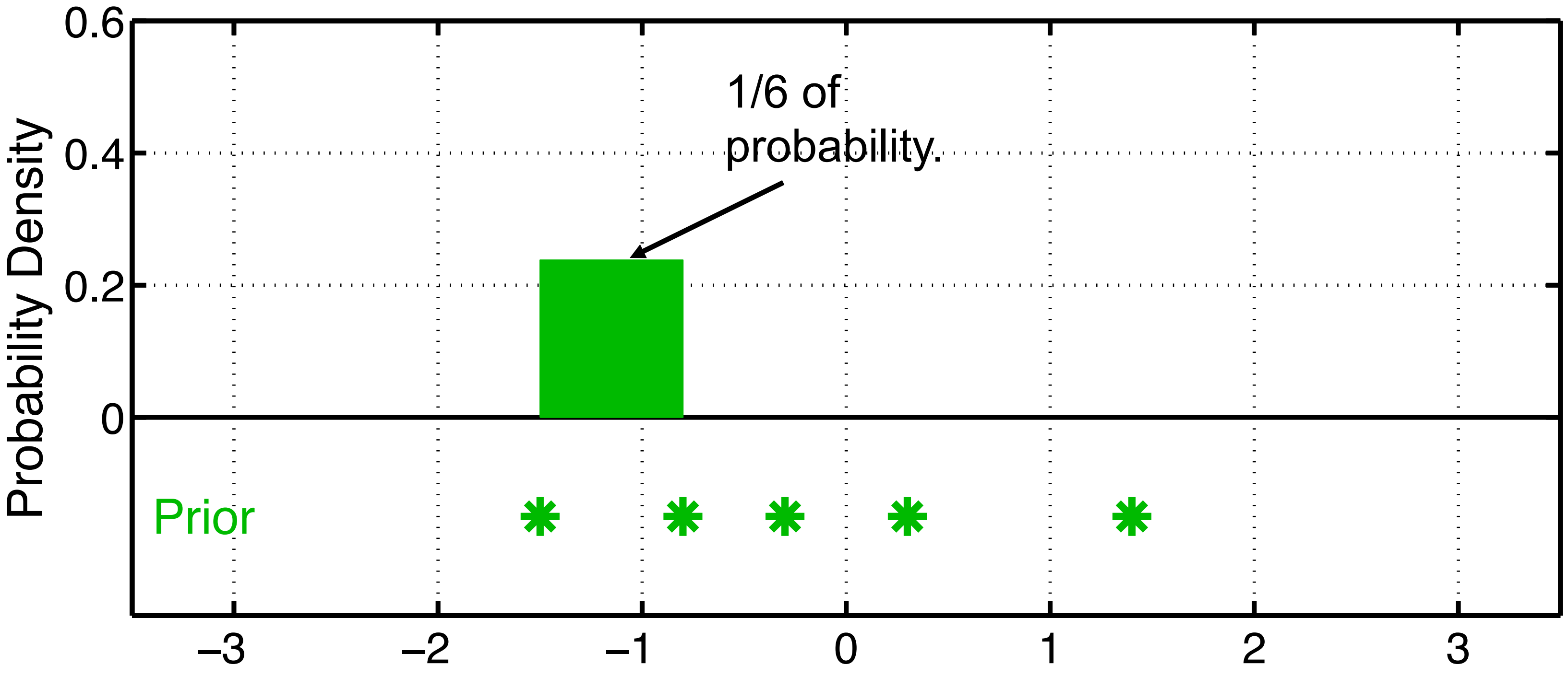
All examples here use this distribution.

Bounded Normal Rank Histogram Continuous Prior



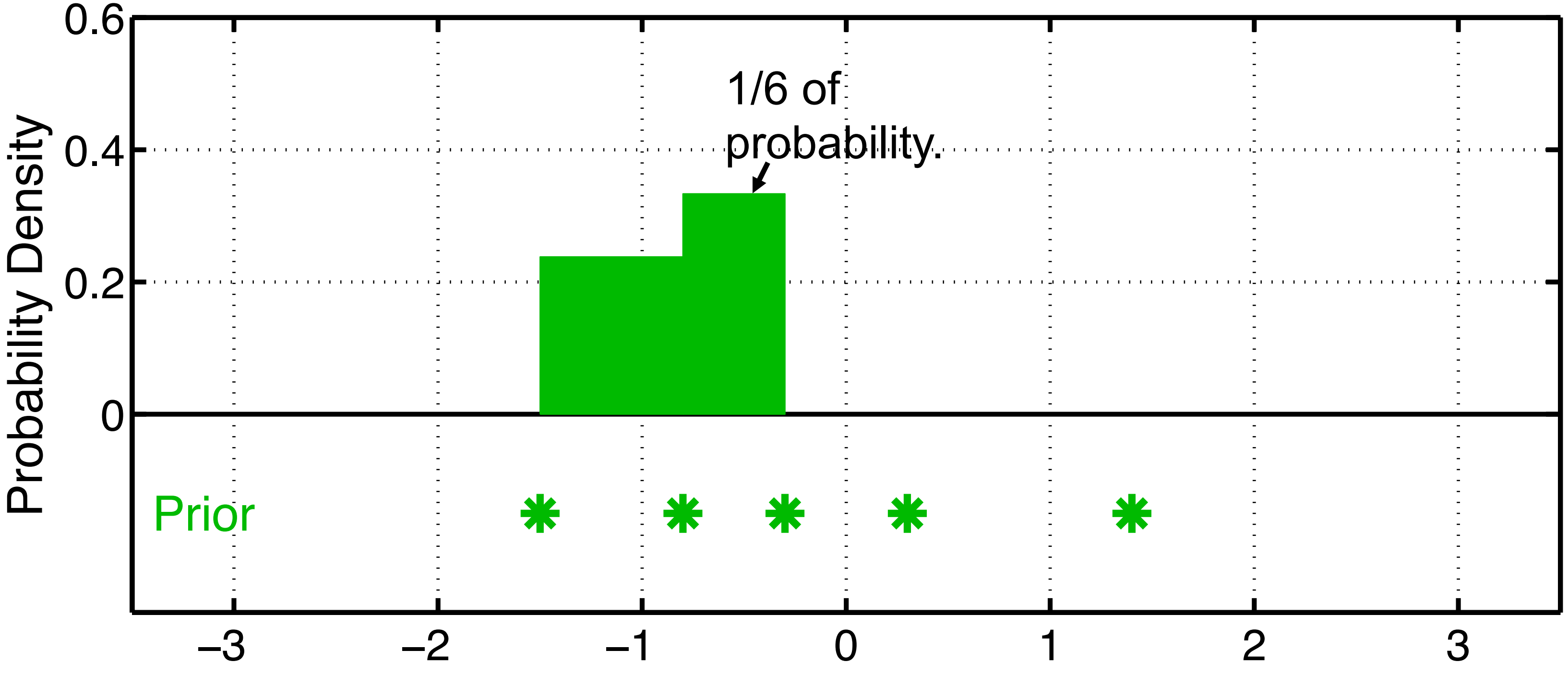
Have a prior ensemble for a state variable (like wind).

Bounded Normal Rank Histogram Continuous Prior



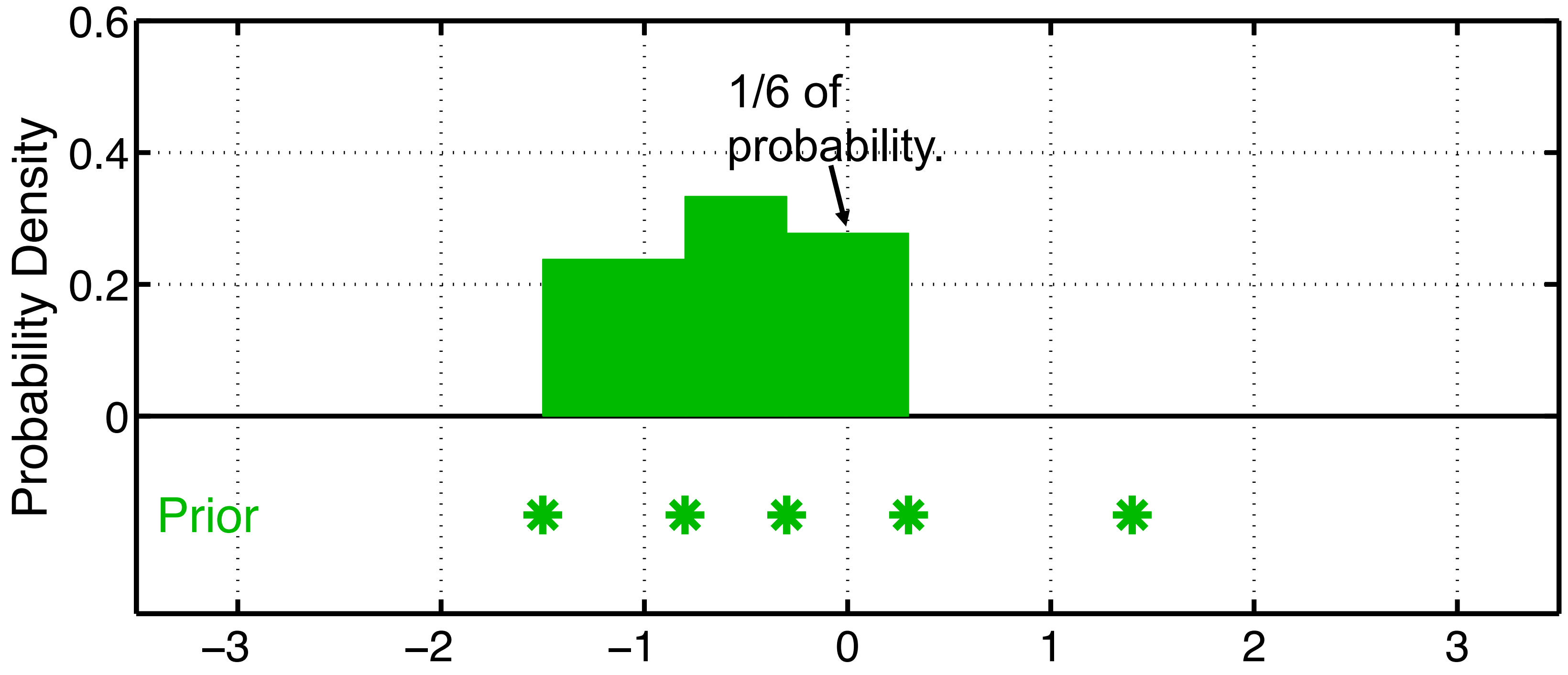
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Bounded Normal Rank Histogram Continuous Prior



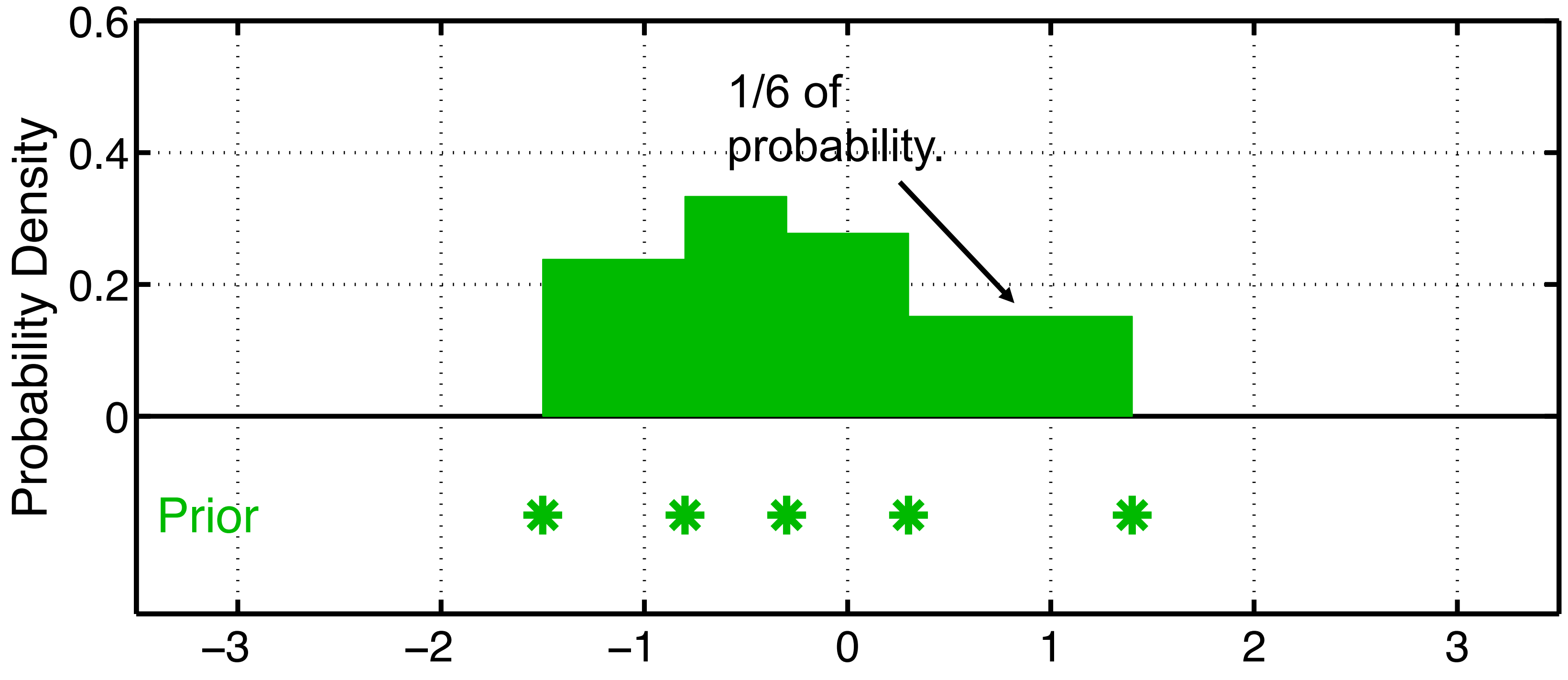
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Bounded Normal Rank Histogram Continuous Prior



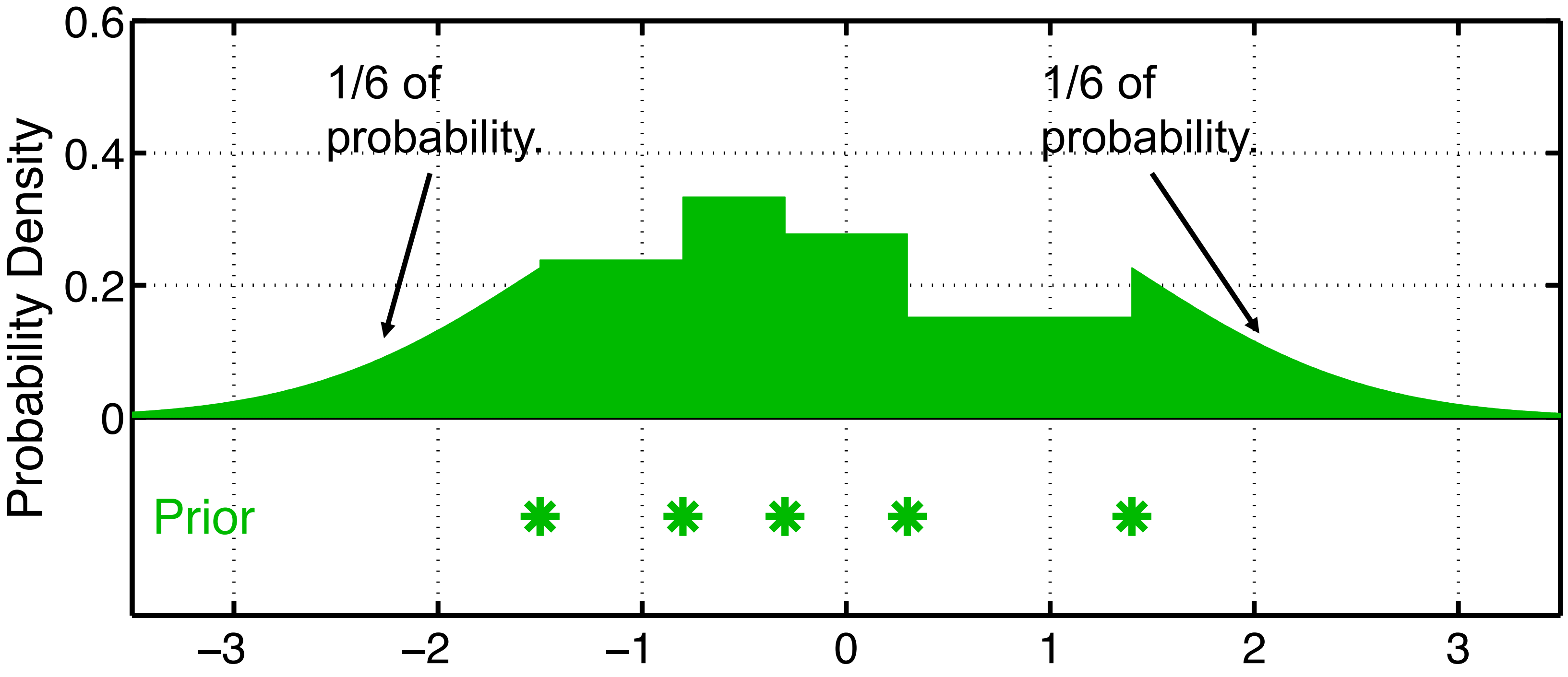
- Place $(\text{ens\_size} + 1)^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Bounded Normal Rank Histogram Continuous Prior



- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

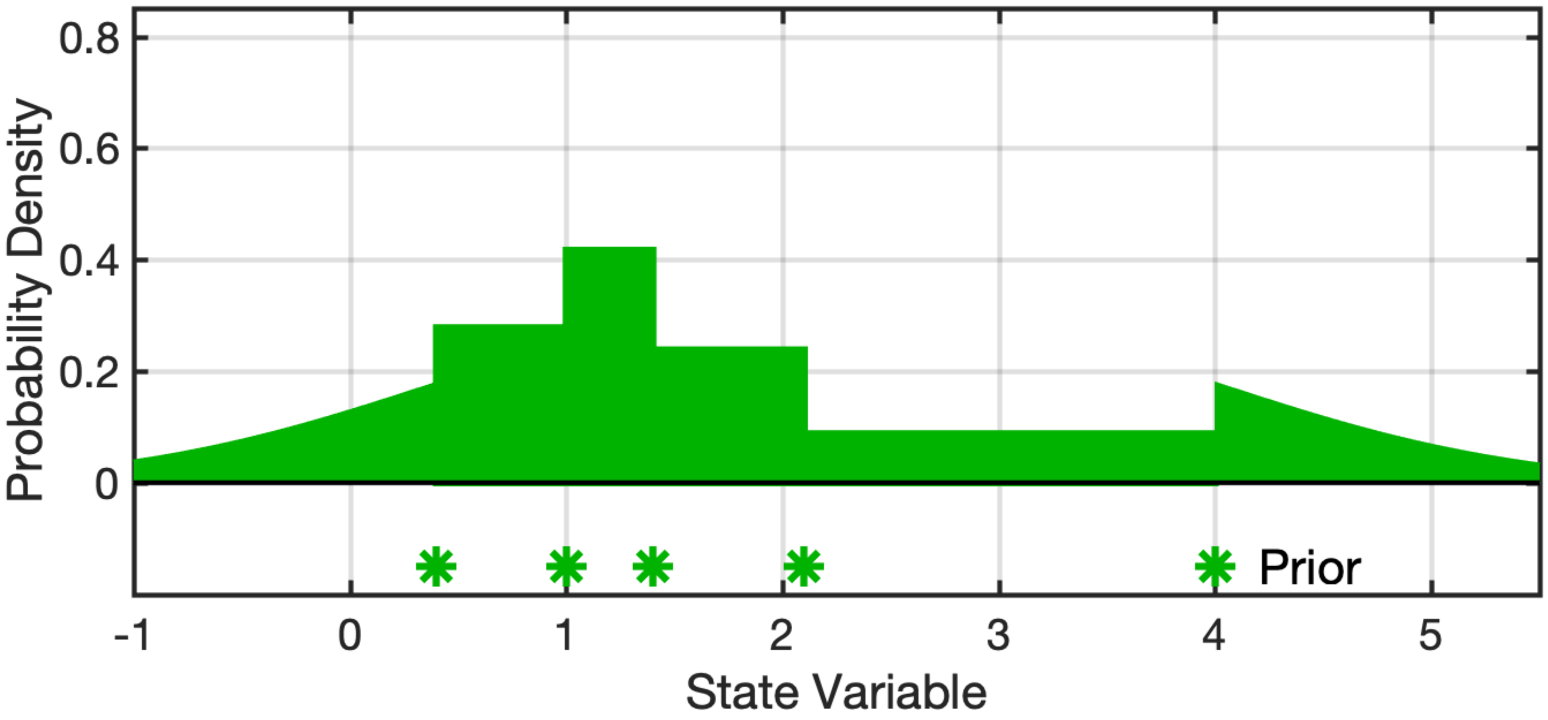
Bounded Normal Rank Histogram Continuous Prior



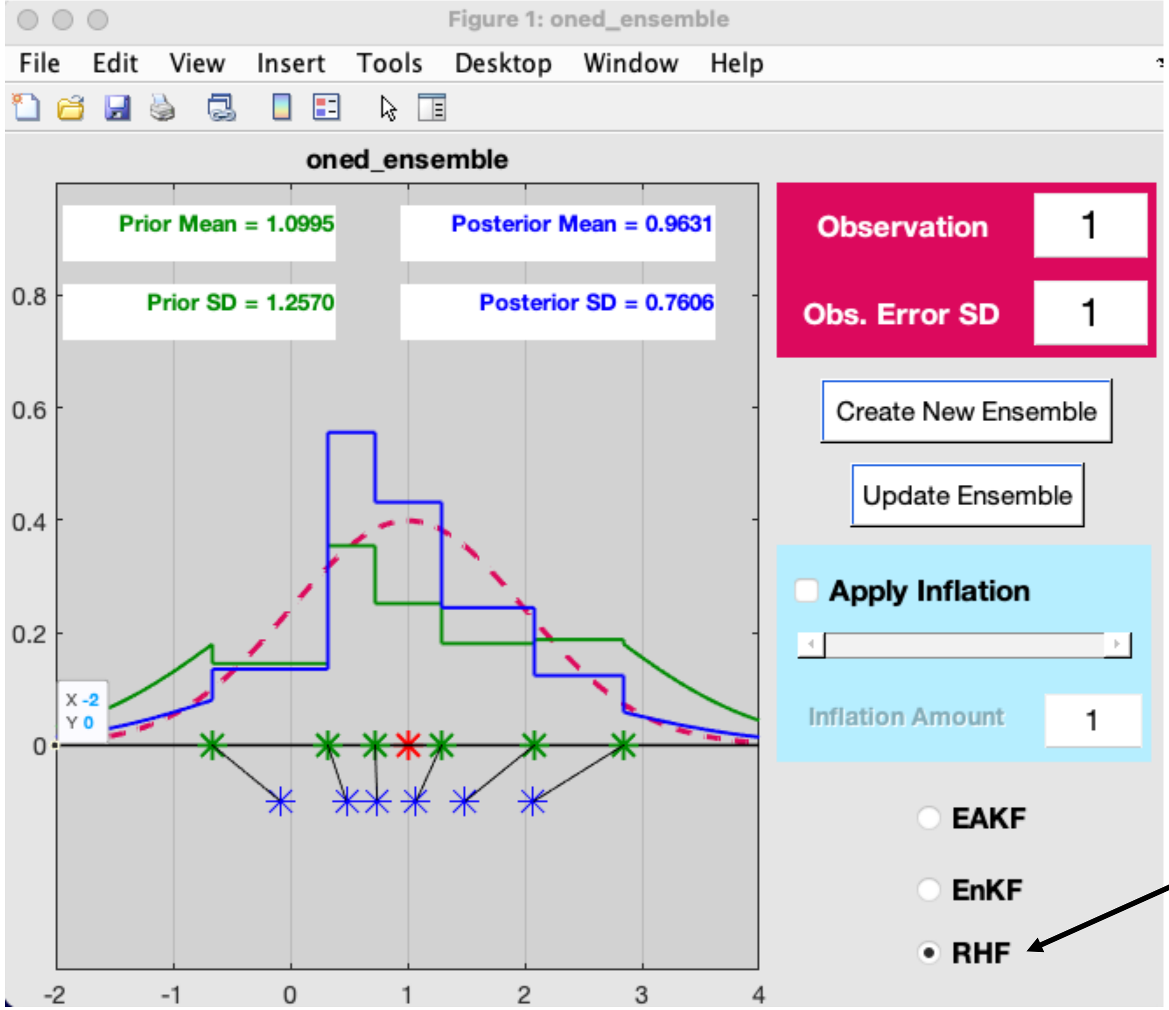
- Partial gaussian kernels on tails, N(*tail_mean*, *ens_sd*).
- *tail_mean selected so that (ens_size + 1)*$^{-1}$ *mass is in tail.*

Bounded Normal Rank Histogram Continuous Prior

Unbounded has normal tails.
Quantiles are exactly uniform, U(0, 1), by construction.

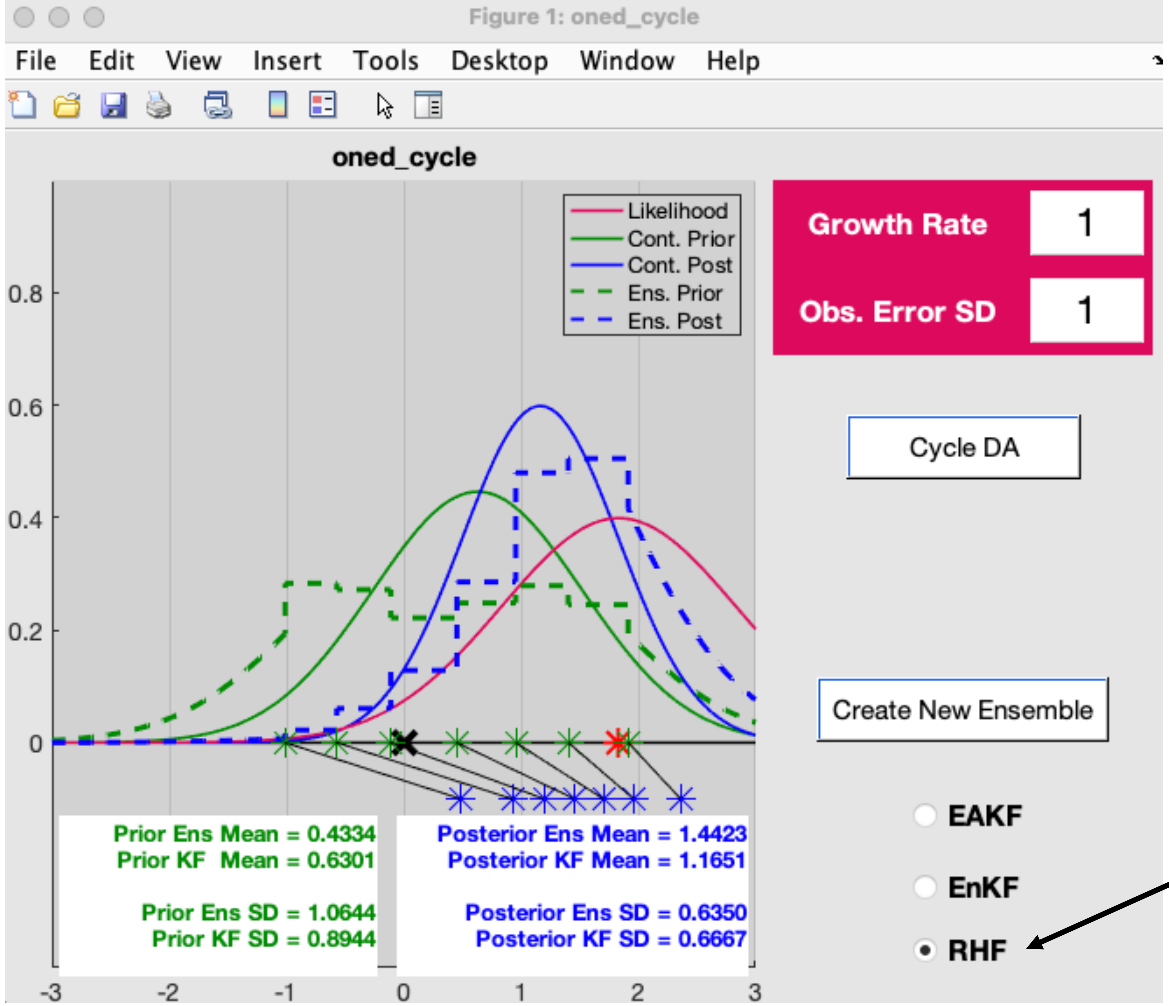


Matlab Hands-on: oned_ensemble



Select RHF to explore what priors and posteriors look like.

Matlab Hands-on: oned_cycle



Select RHF to explore what priors and posteriors look like.

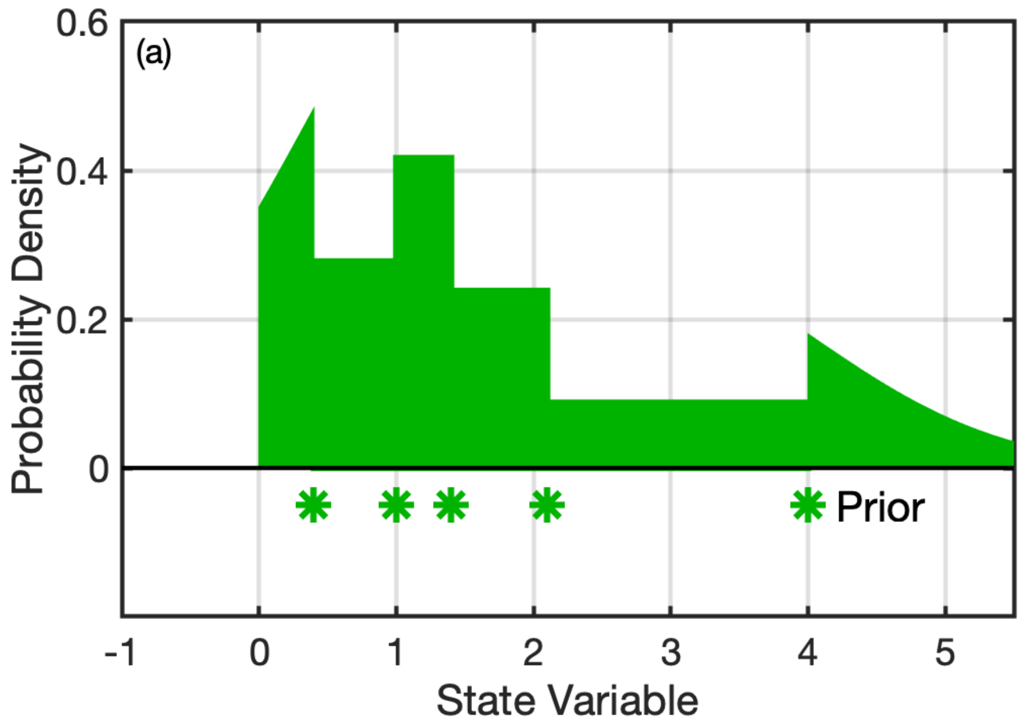
Matlab Hands-on: More RHF example

The RHF can also be explored in:

oned_model
oned_model_inf
twod_ensemble
run_lorenz_63
run_lorenz_96
run_lorenz_96_inf

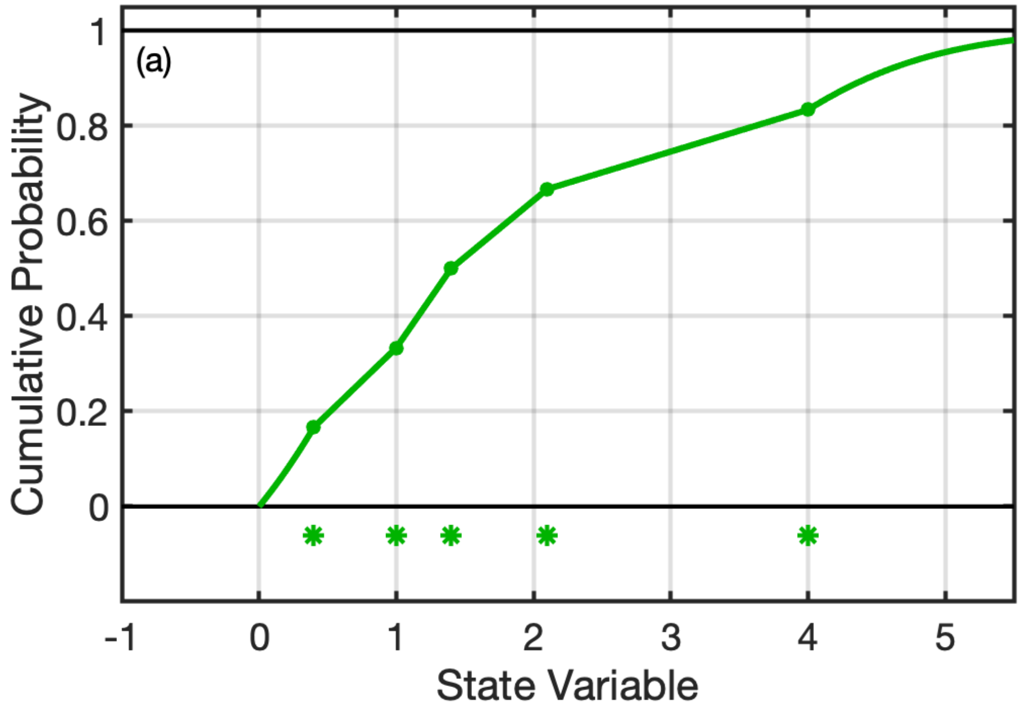
Bounded Normal Rank Histogram Continuous Prior

Bounded has truncated tail, that is part of a normal.
Quantiles are exactly U(0, 1) by construction.



Bounded Normal Rank Histogram Continuous Prior

Bounded has truncated tail, that is part of a normal.
Quantiles are exactly U(0, 1) by construction.
This is the corresponding CDF, doesn't look so weird.



Mixed Distributions: A Challenge for Tracers and Sources

Mixed Distributions: Have both discrete and continuous probability distribution parts.

Precipitation forecast is an example:

Discrete probability of zero rain (50%),
Continuous distribution for all non-zero amounts,
(zero probability of exactly any given amount).

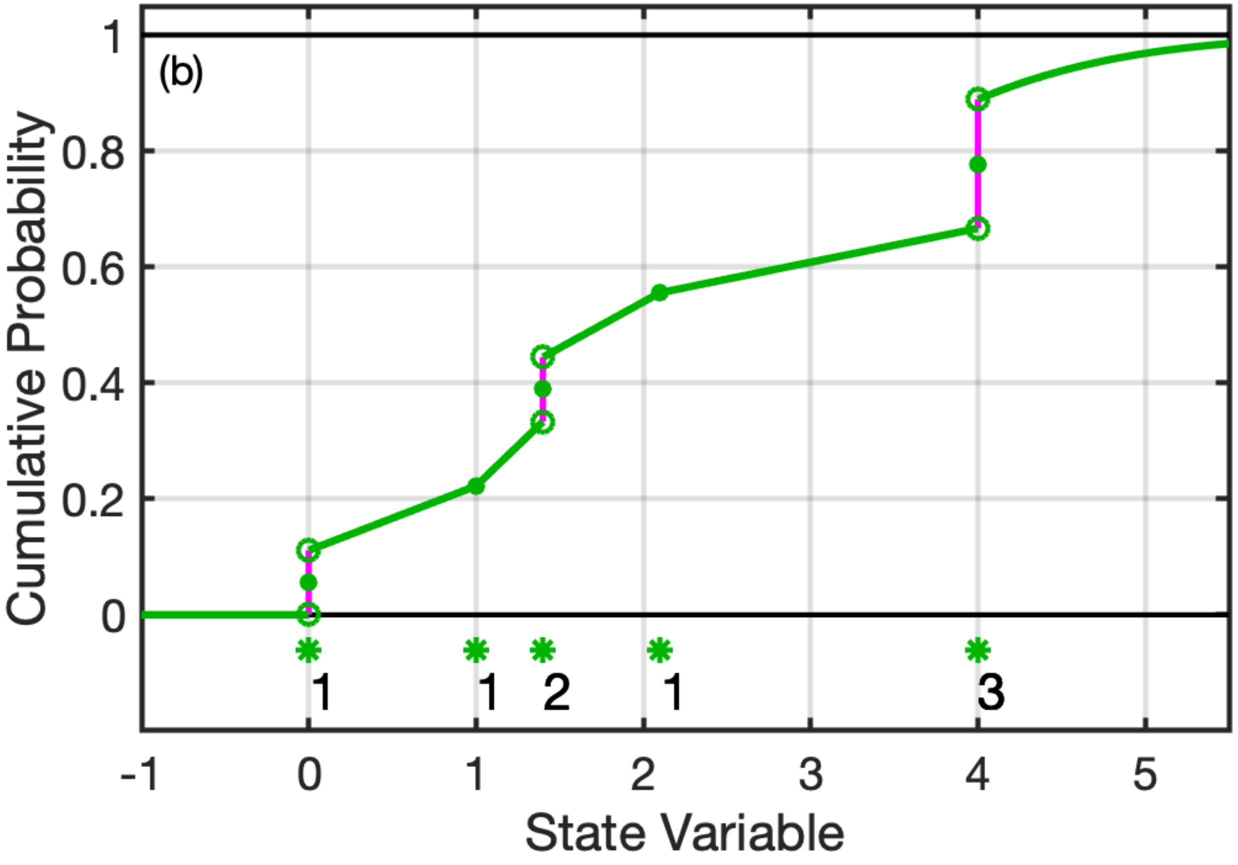
Important for some tracers.

Important for many sources (anthropogenic sources, wildfires, …).

(Anderson et al., 2024, MWR 152, 2111-2127)

Mixed Distributions: A Challenge for Tracers and Sources

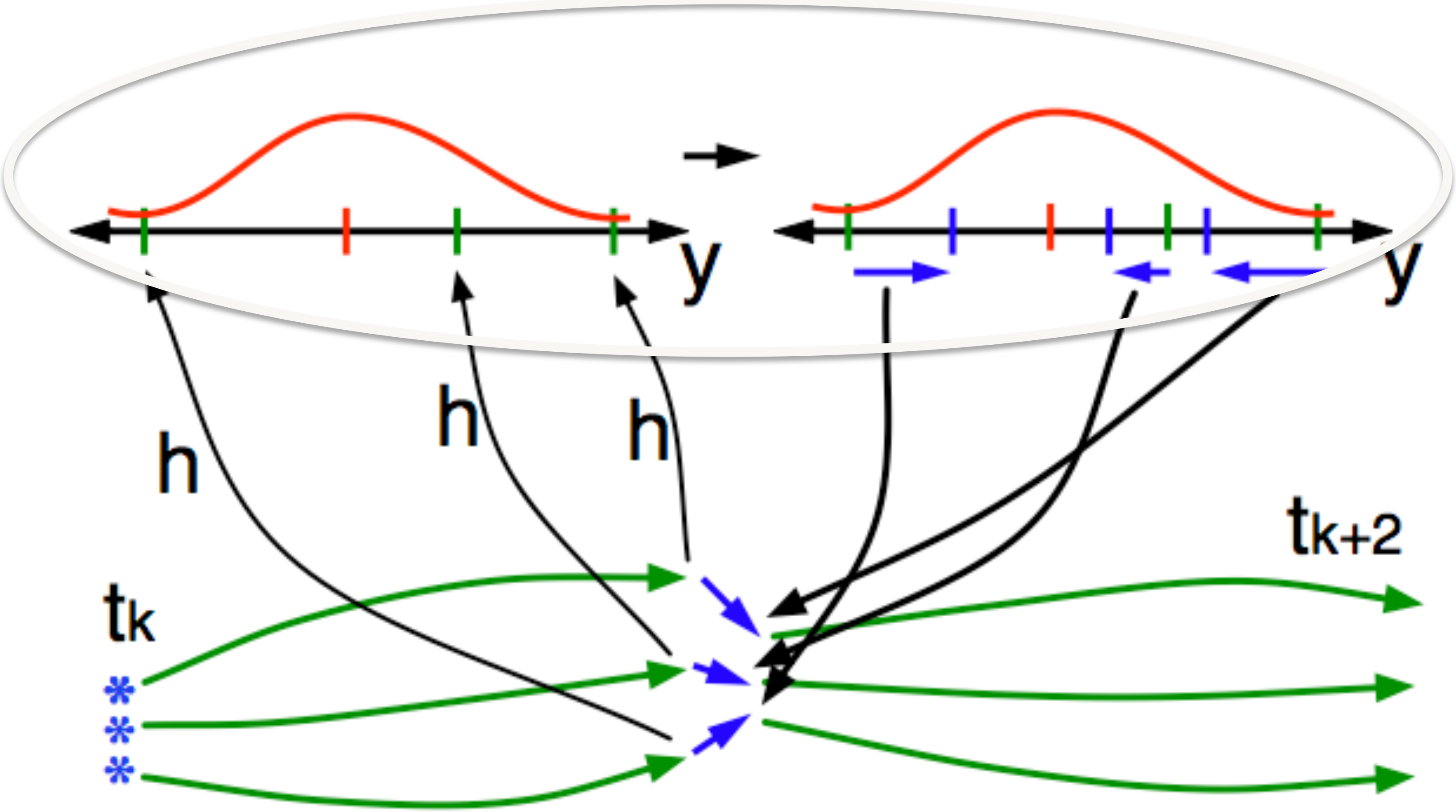
Need to be able to handle duplicate ensemble members.
Want $x = F^{-1}(F(x))$.



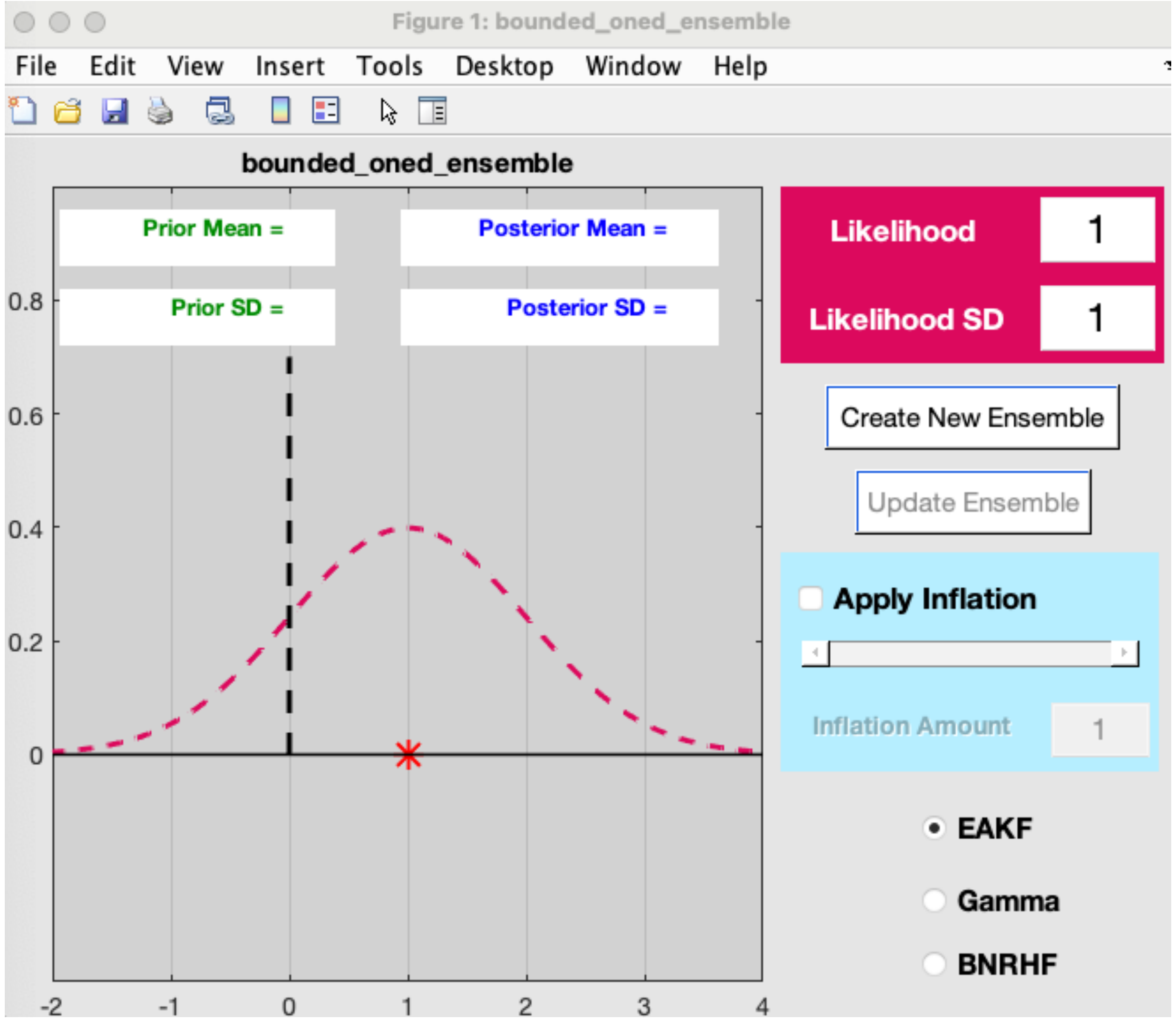
Number by asterisk indicates number of ensemble members with this value.

Quantile Conserving Ensemble Filters in Observation Space

DART now provides nearly general solutions for this step:
(Anderson, 2022, MWR *150*, 1061-1074).



Matlab Hands-on: bounded_oned_ensemble



Controls are similar to oned_ensemble but the quantity is non-negative.

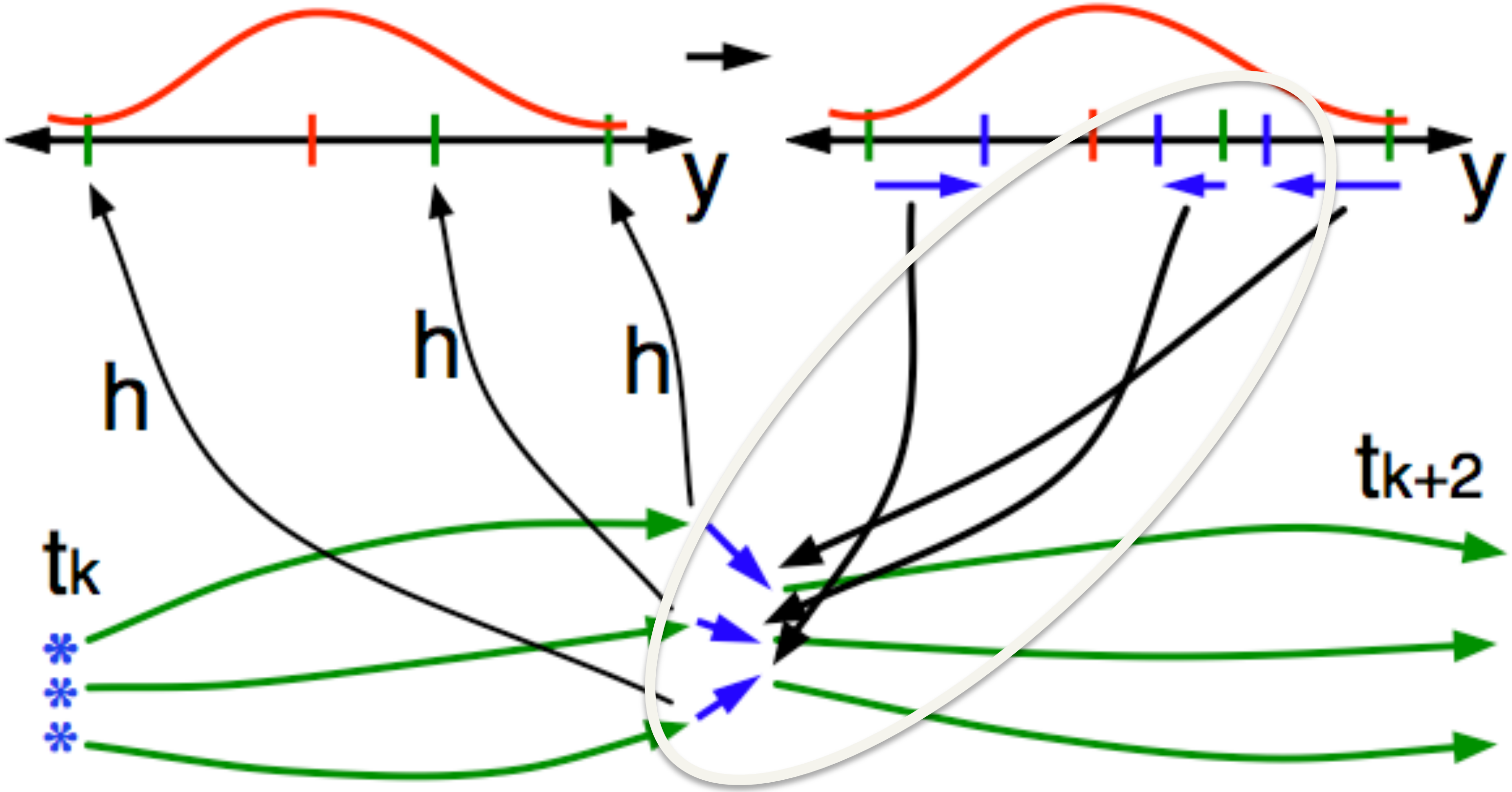
The EAKF (normal prior) doesn't know about this bound. Try to generate negative posterior ensemble members.

The Gamma distribution does know about the bounds.

The bounded normal rank histogram prior also knows about the bounds.

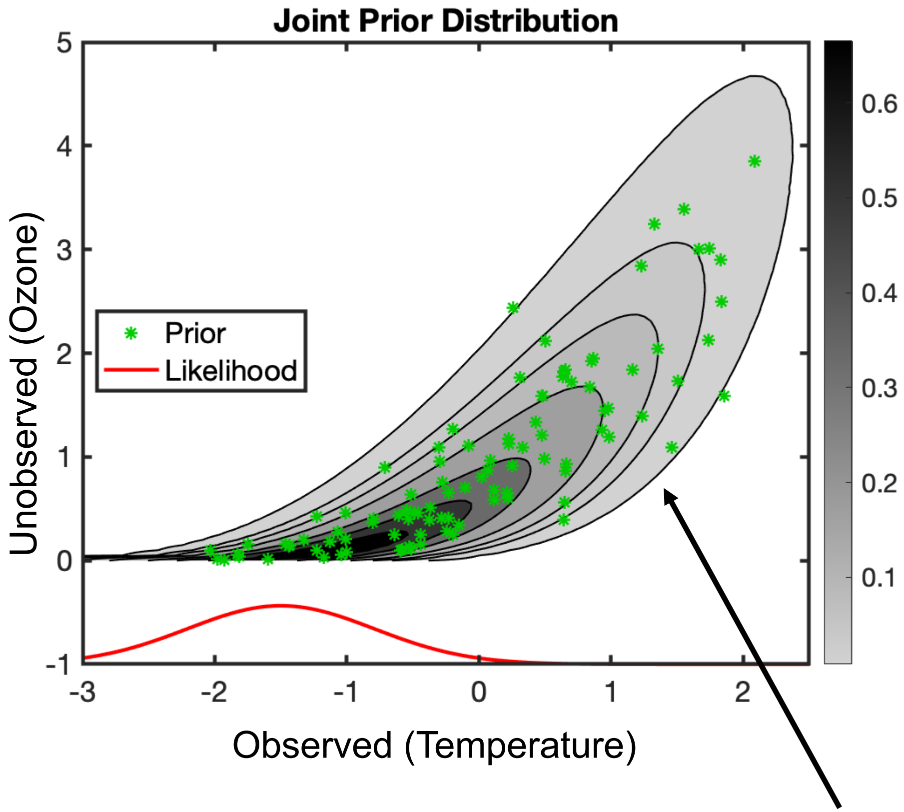
Linear Regression can Wreck Things

Linear regression can destroy benefits of new observation method.



Standard EAKF: Challenged by Non-Gaussian and Nonlinear Relations

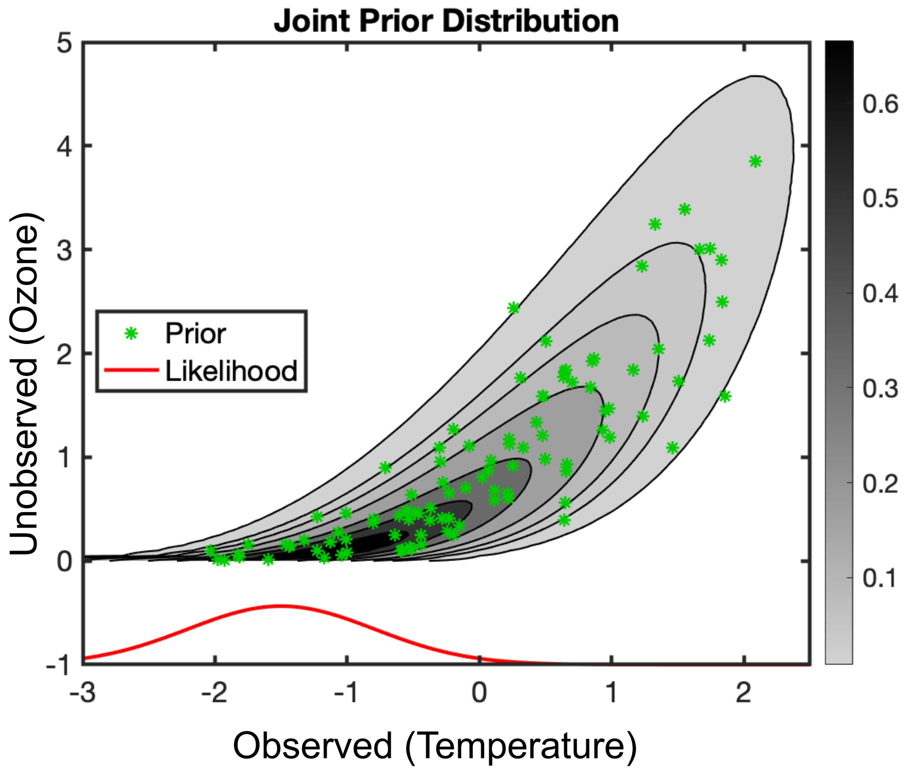
Prior for normal-gamma distribution with 100 member ensemble.



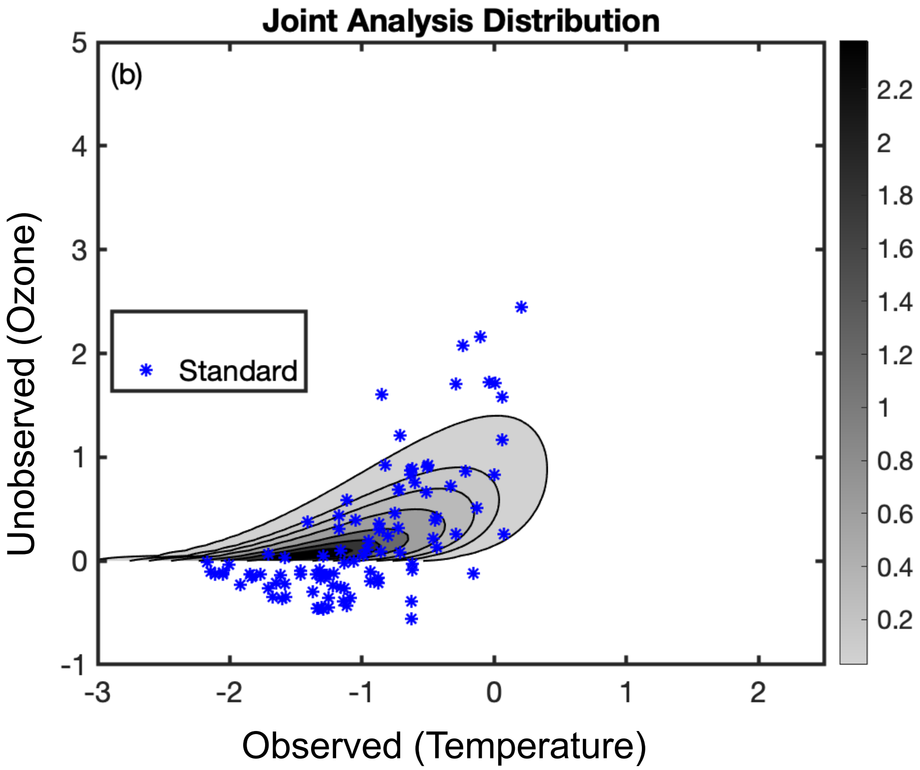
Contours of the correct distribution are 1, 5, 10, 20, 40, 60, 80% of max for all figures.

Standard EAKF: Challenged by Non-Gaussian and Nonlinear Relations

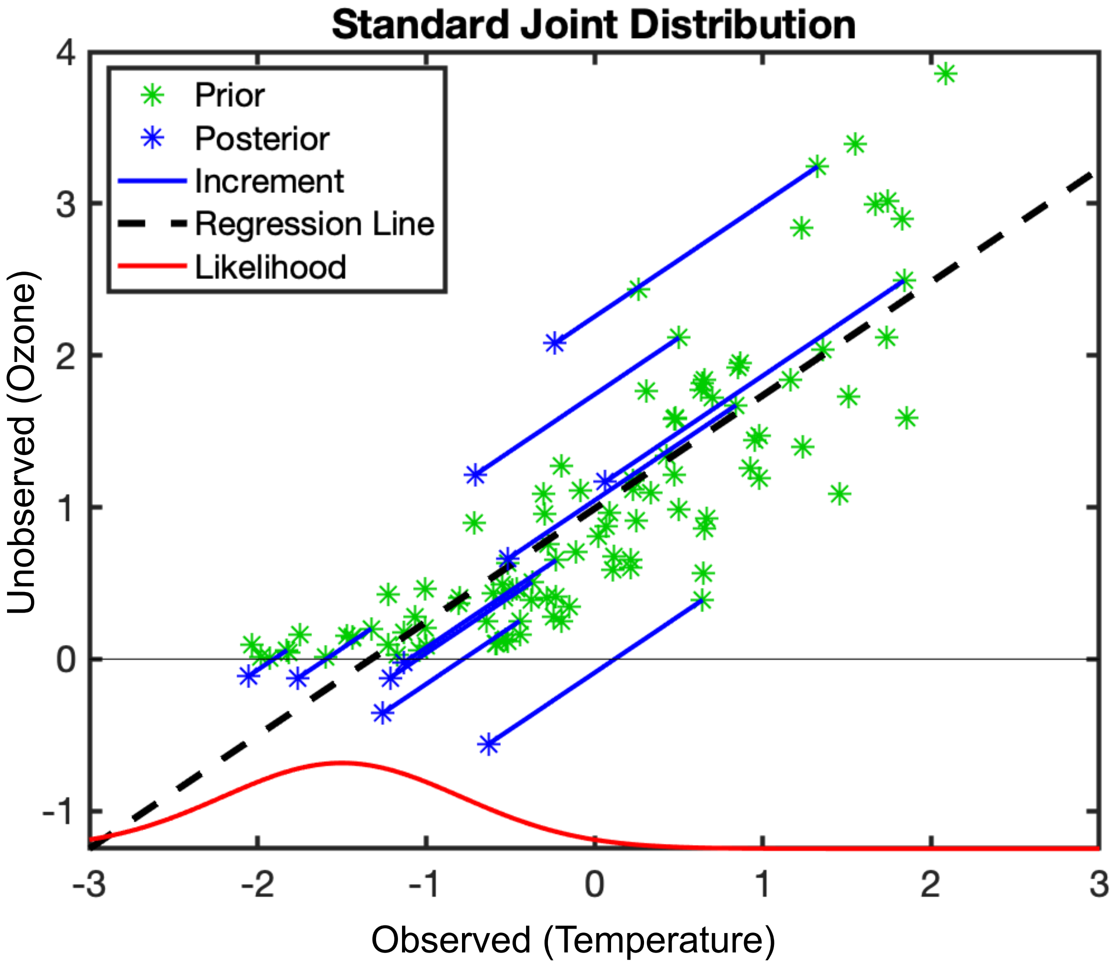
Prior for normal-gamma distribution with 100 member ensemble.



Posterior ensemble has problems.



Standard EAKF: Challenged by Non-Gaussian and Nonlinear Relations

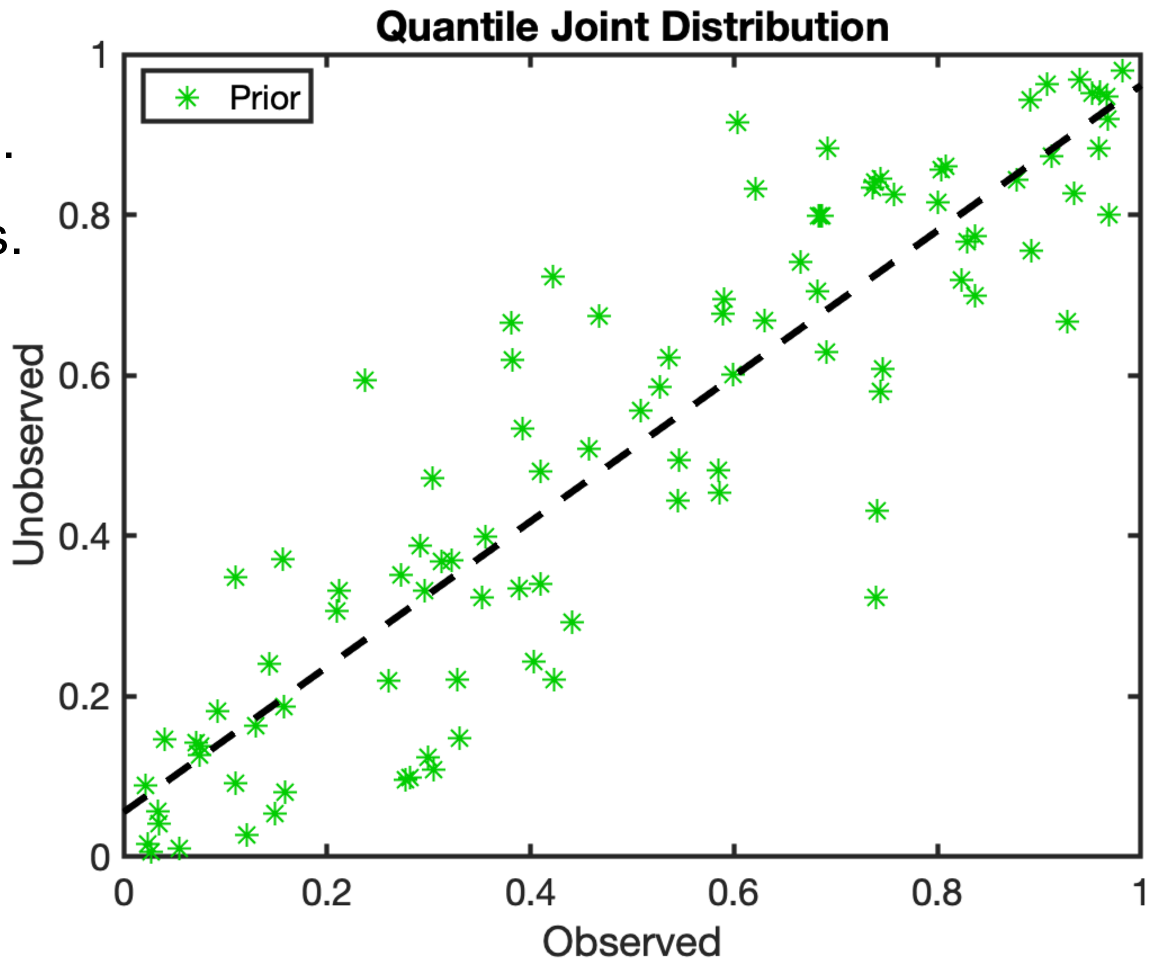


Example regression increment vectors:
Don’t respect bounds,
Struggle with nonlinearity.

Solution, Transform Marginals: Step 1: Compute Quantiles

Apply the probability integral transform:

- Pick an appropriate continuous prior distribution.
- Compute CDF for each member to get quantiles.
- Quantiles are U(0, 1) for appropriate prior.



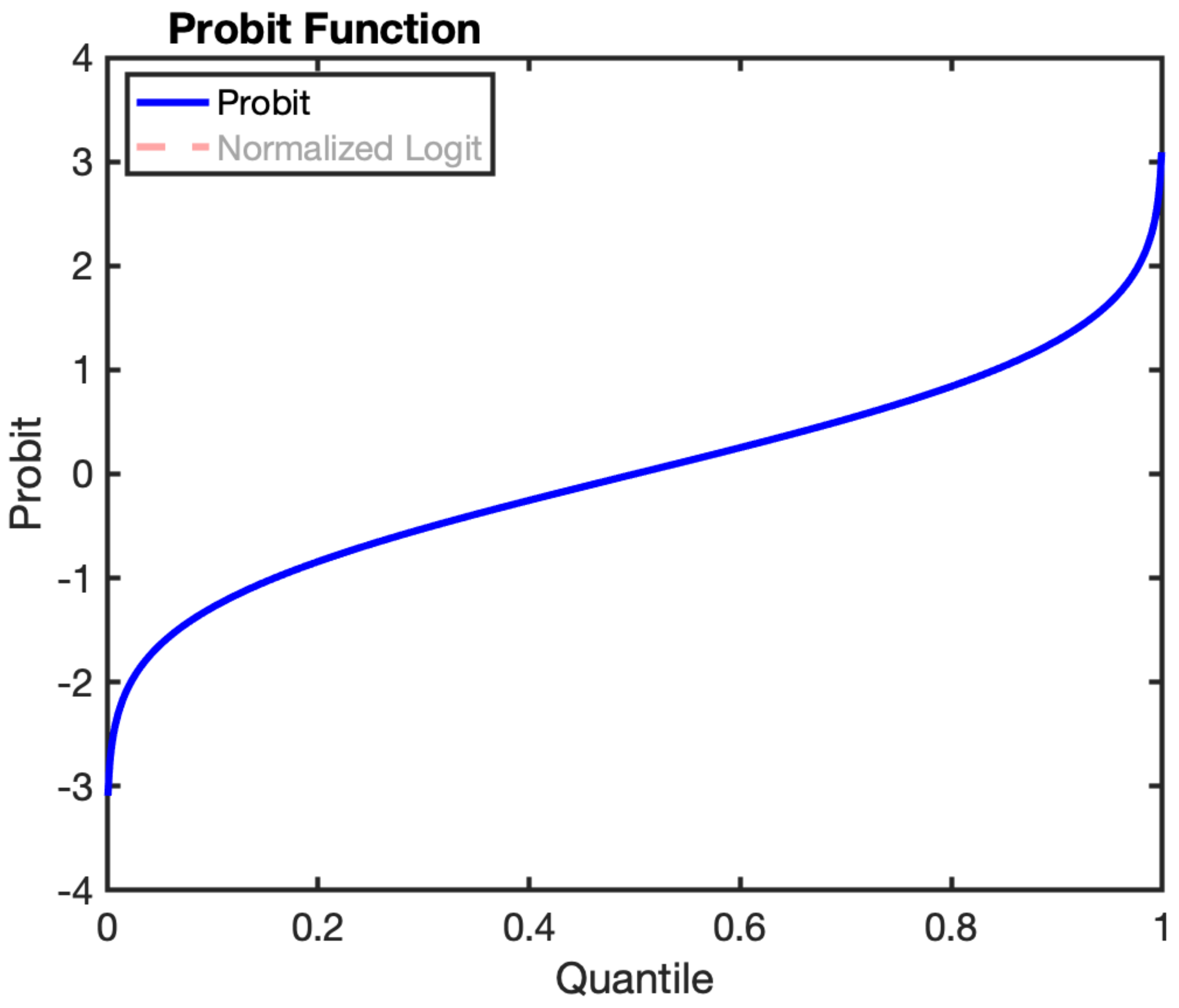
Solution, Transform Marginals: Step 2: Probit Transform of Quantiles

The ‘quantile function’ is the inverse of the CDF for a distribution.

The quantile function for the standard Normal is the probit function (plotted here).

Transforms U(0, 1) to Normal(0, 1).

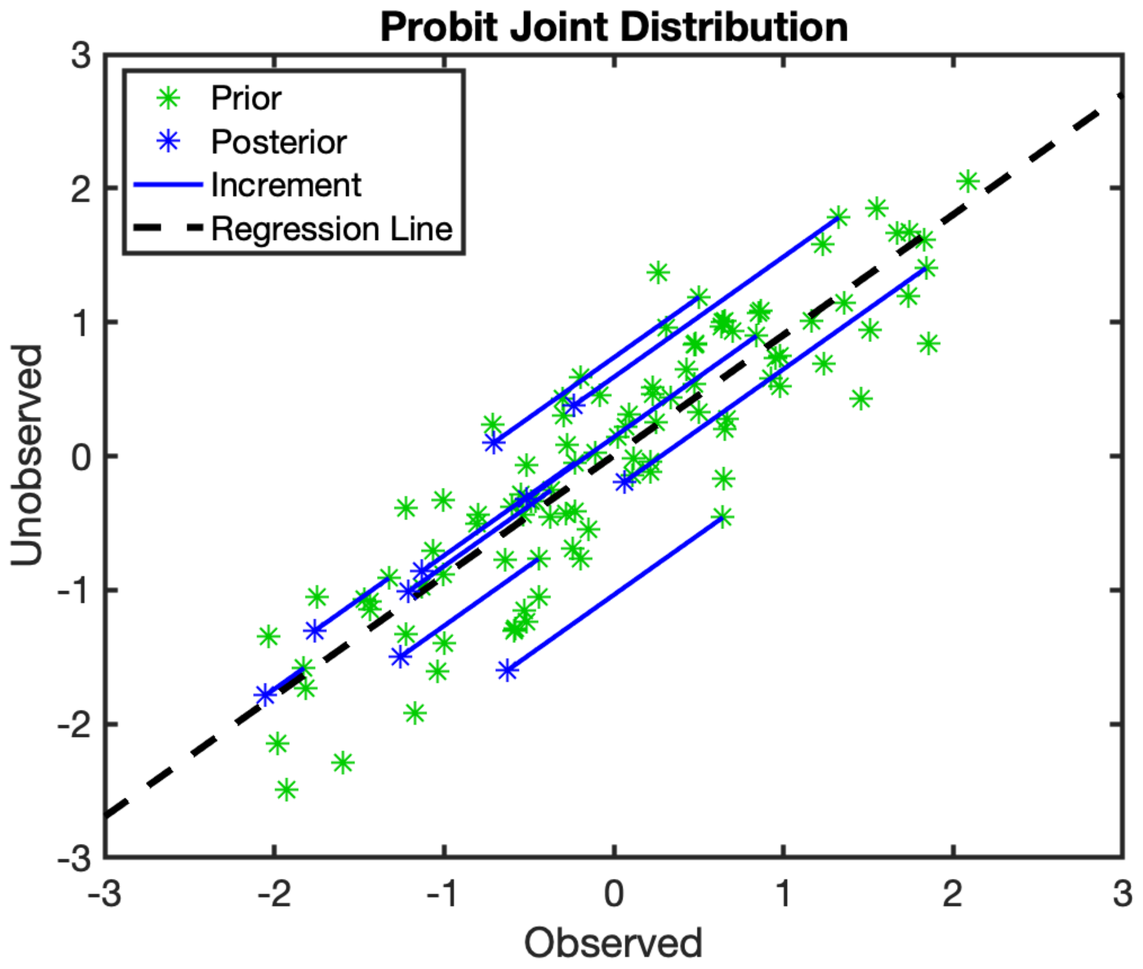
Marginal distributions should be N(0, 1).



Regression in Probit-Transformed Quantile Space

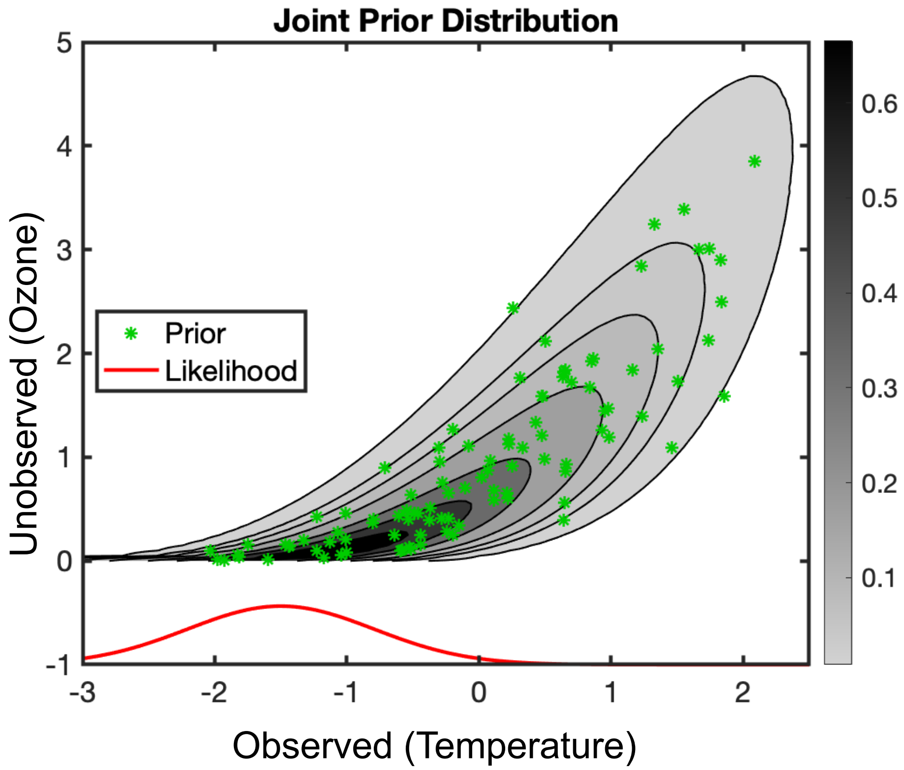
Do the regression of the observed probit increments onto the unobserved probit ensemble.

Linear regression is best unbiased linear estimator (BLUE) in this space.

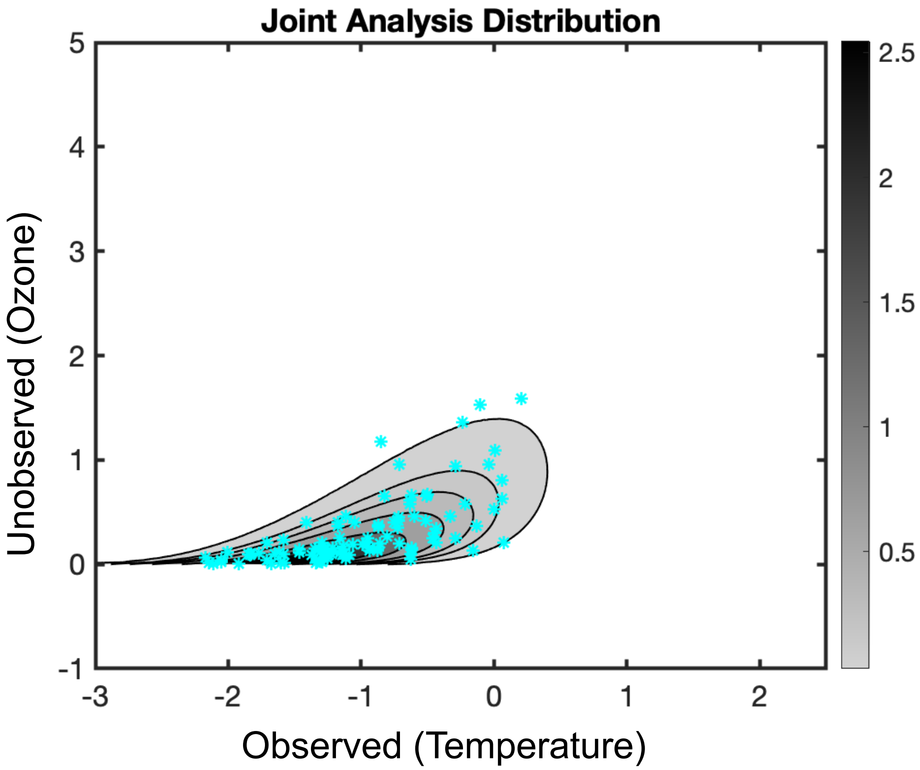


DART: Novel, General Solutions for Nonlinear, Non-Gaussian Problems

Prior for normal-gamma distribution with 100 member ensemble.

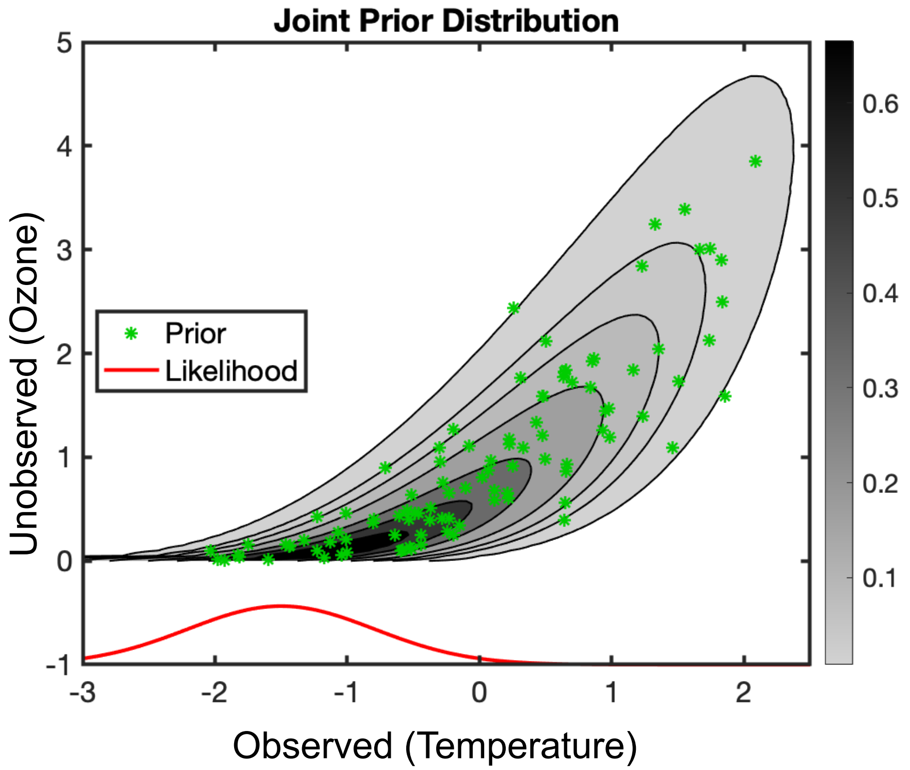


Bounds enforced. Nonlinear aspect respected.

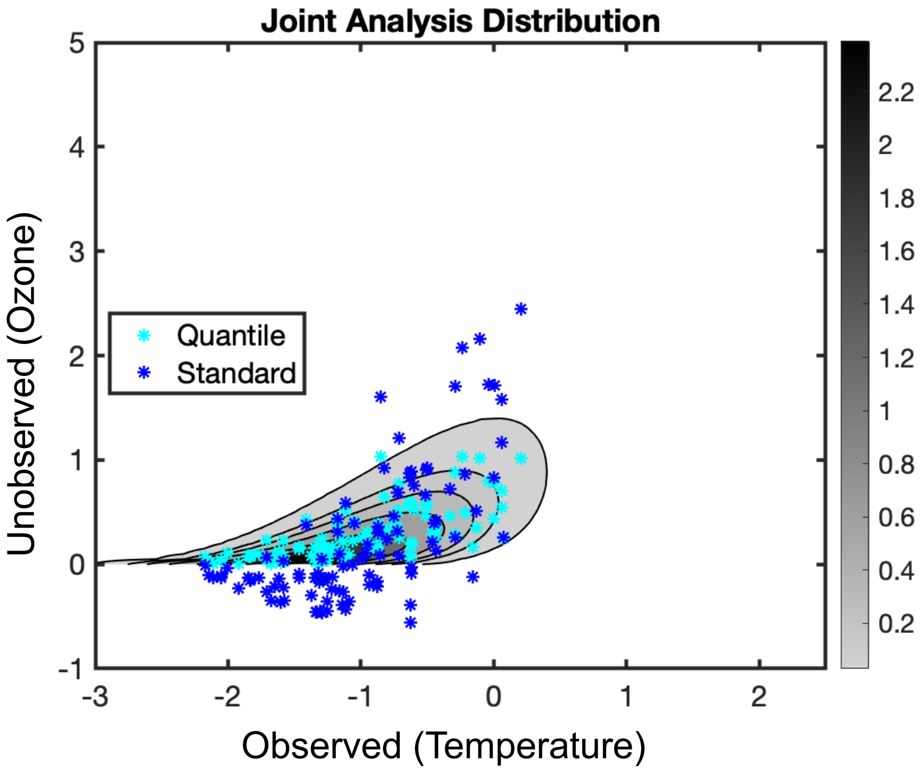


DART: Novel, General Solutions for Nonlinear, Non-Gaussian Problems

Prior for normal-gamma distribution with 100 member ensemble.



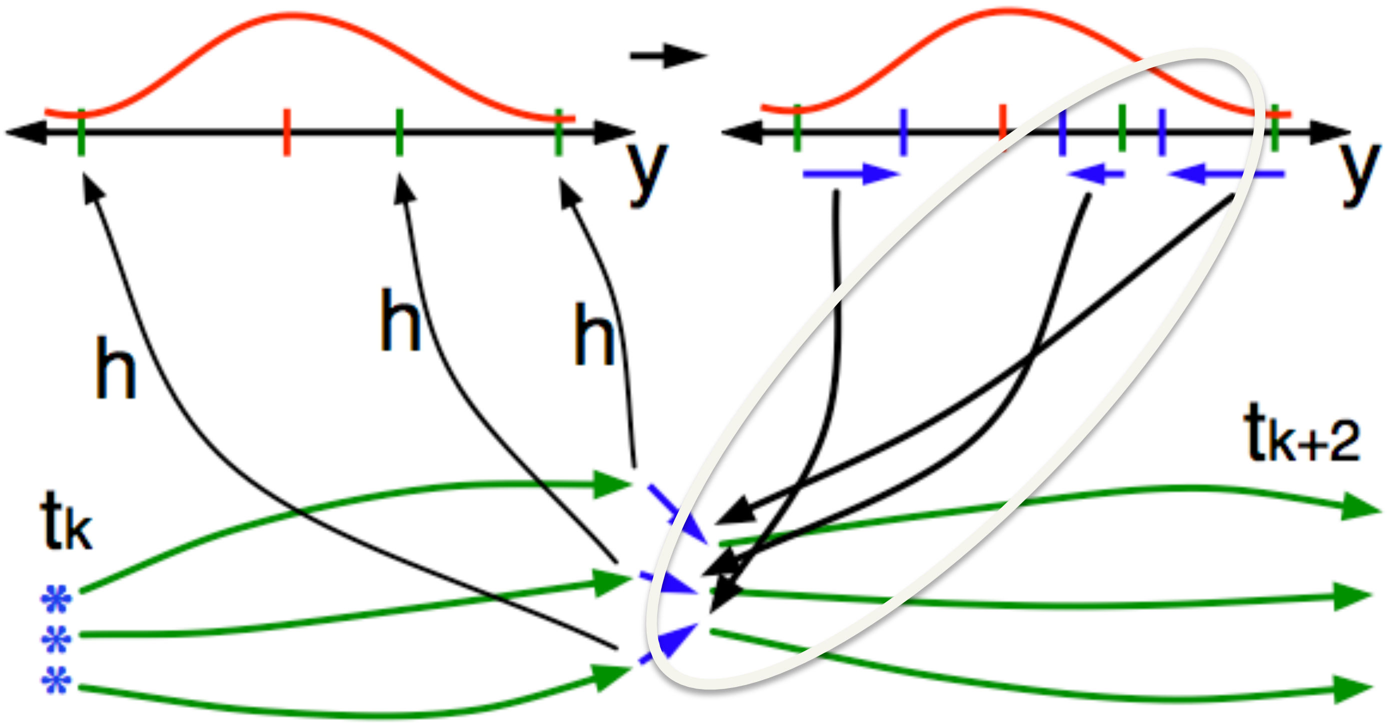
Bounds enforced. Nonlinear aspect respected.



DART Now Implements Regression in a Transformed Space

Can update unobserved variables with regression in a transformed space for each state variable.

(Anderson, 2023, MWR *151*, 2759-2777)



Regression in Probit-Transformed Quantile Space

y_n^p , y_n^a, x_n^p, *n=1, …N* are prior and posterior (analysis) ensembles of observed variable y and unobserved variable x.

F_x^p and F_y^p are continuous CDFs appropriate for x and y.

$\Phi(z)$ is the CDF of the standard normal, $\Phi^{-1}(p)$ is the probit function.

$\tilde{x}_n^p = \Phi^{-1}\left[F_x^p\left(x_n^p\right)\right]$, $\tilde{y}_n^p = \Phi^{-1}\left[F_y^p\left(y_n^p\right)\right]$ and $\tilde{y}_n^a = \Phi^{-1}\left[F_y^p\left(y_n^a\right)\right]$ are probit space.

$\Delta\tilde{y}_n = \tilde{y}_n^a - \tilde{y}_n^p$ is probit space observation increment.

$\Delta\tilde{x}_n = \frac{\tilde{\sigma}_{x,y}}{\tilde{\sigma}_{y,y}}\Delta\tilde{y}_n$ regress increments in probit space (eq. 5 Anderson 2003).

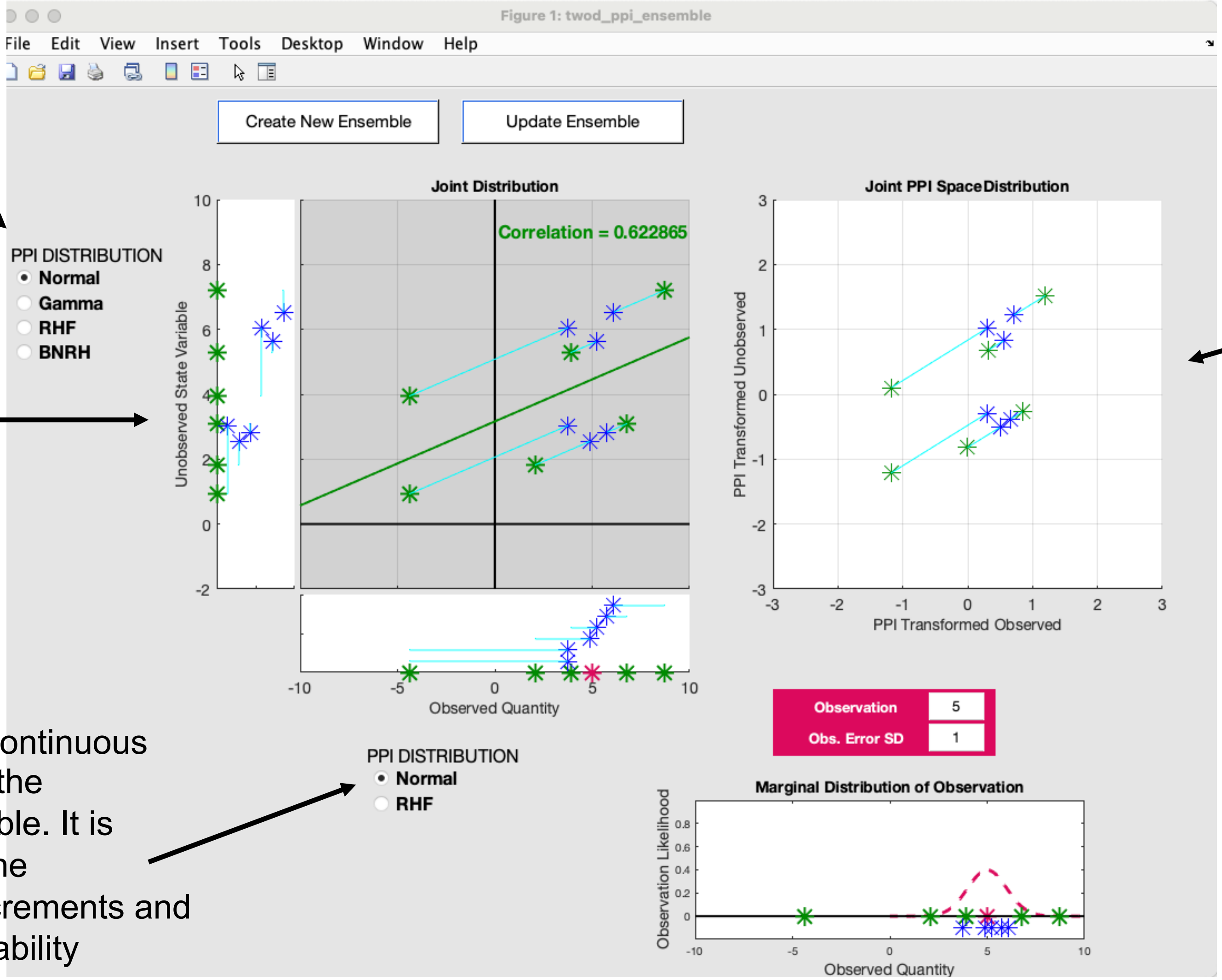
$\tilde{x}_n^a = \tilde{x}_n^p + \Delta\tilde{x}_n$ is posterior ensemble in probit space.

$x_n^a = \left(F_x^p\right)^{-1}\left[\Phi\left(\tilde{x}_n^a\right)\right]$ is posterior ensemble.

Matlab Hands-on: twod_ppi_ensemble

For the unobserved variable, select the distribution for the probit probability integral transform.

This plot is similar to twod_ensemble except that the unobserved variable is nonnegative.



This panel shows the prior and posterior in the probit probability integral transformed space.

Select from a continuous distribution for the observed variable. It is used both for the observation increments and the probit probability integral transform.

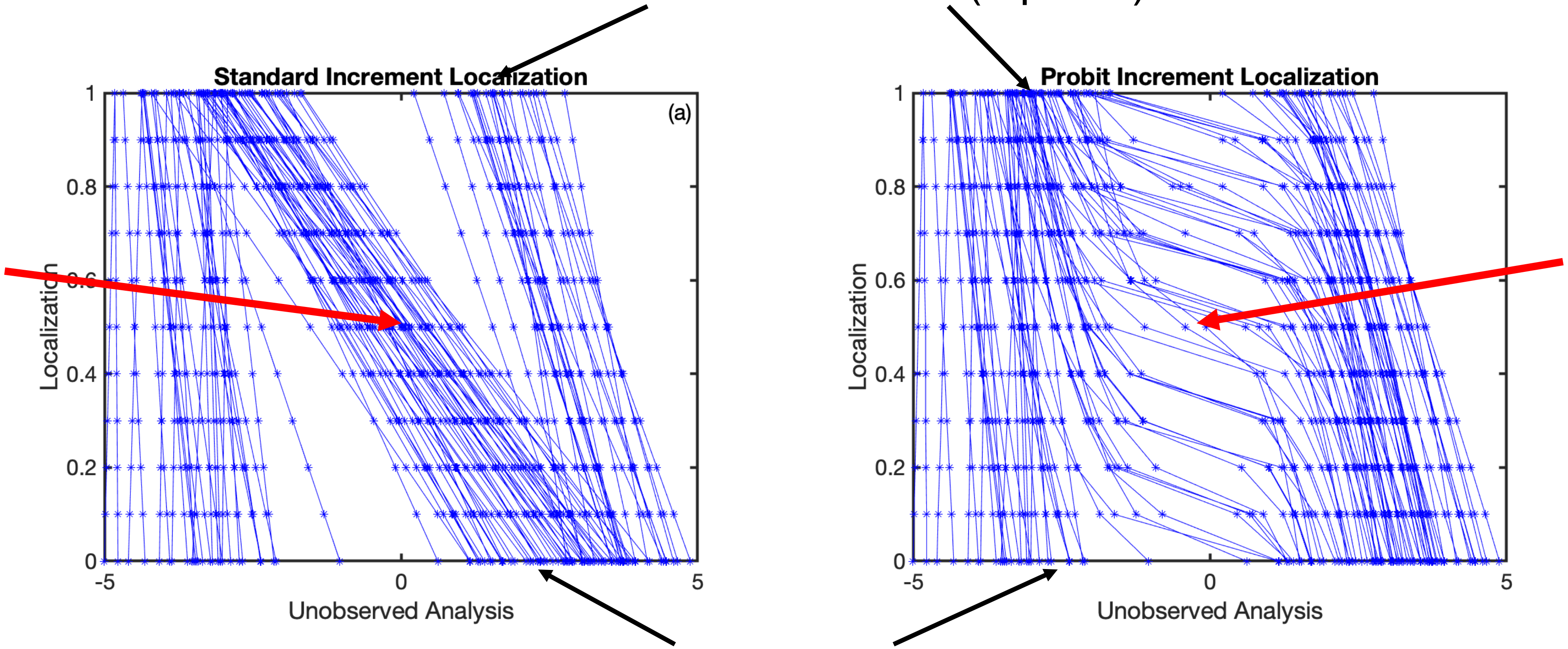
Try to create a case with negative posterior members for the unobserved variable.

Localization of Probit Increments: Normal-binormal example

Standard increment localization may ignore prior constraints (values around zero are very unlikely).
Probit increment localization 'knows' prior was binormal.

Unobserved Posterior (top row)

Lots of members with values near zero, inconsistent with prior.

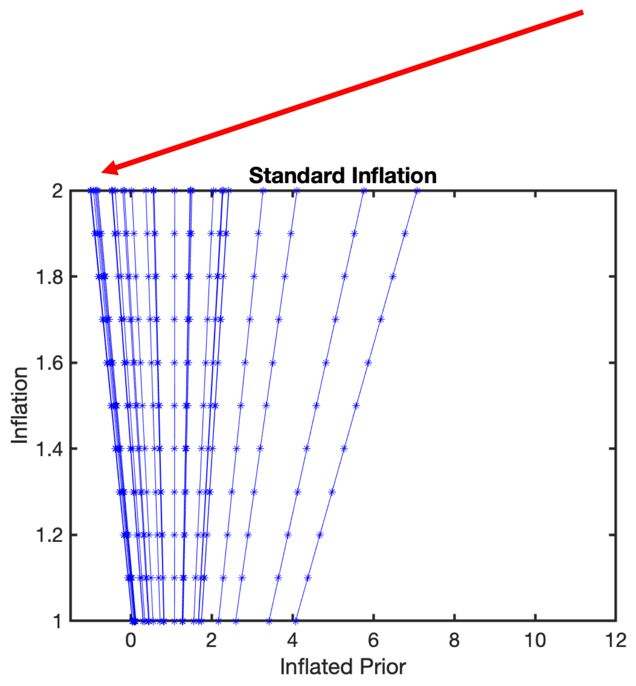


Few members with values near zero, consistent with prior.

Unobserved Prior (bottom row)

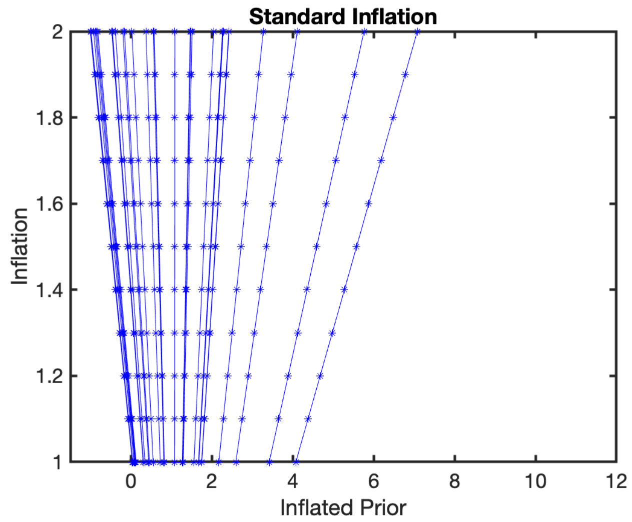
Inflation in Probit Space: Gamma example

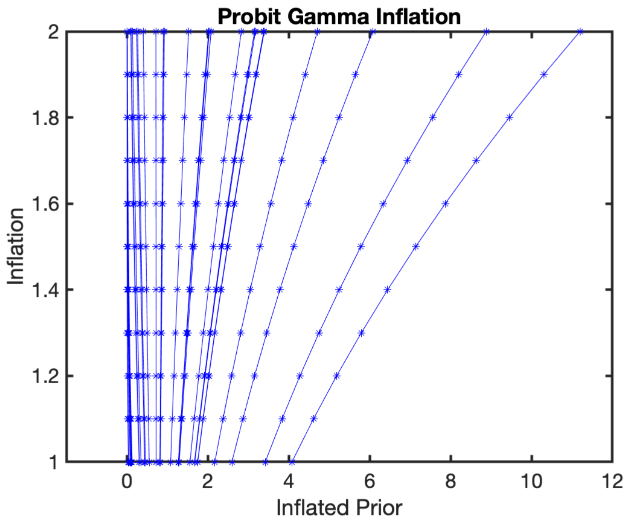
Standard inflation may violate prior constraints.



Inflation in Probit Space: Gamma example

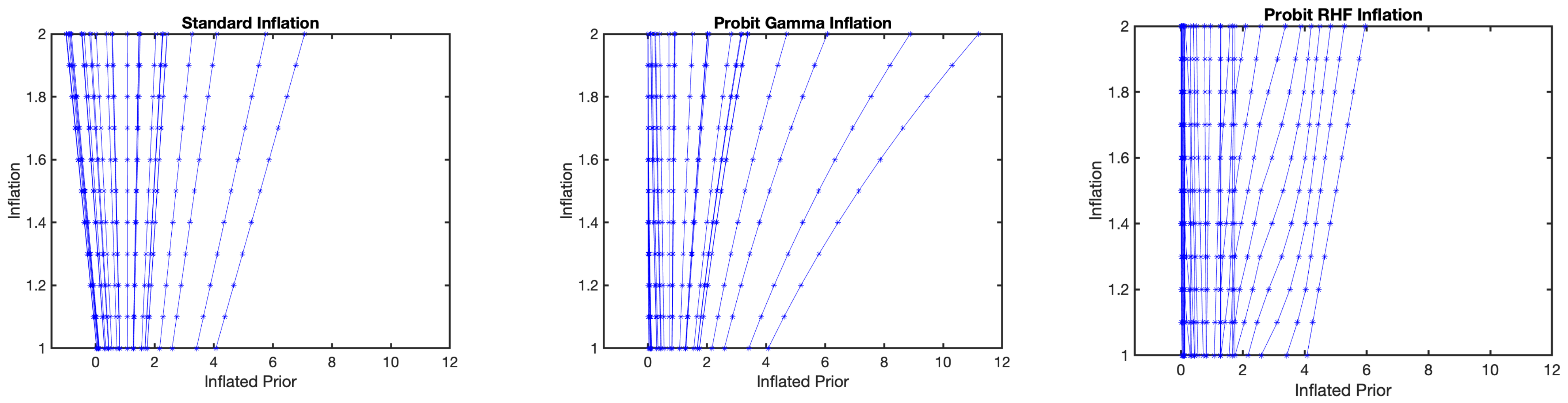
Standard inflation may violate prior constraints.
Inflation can be done in probit space.



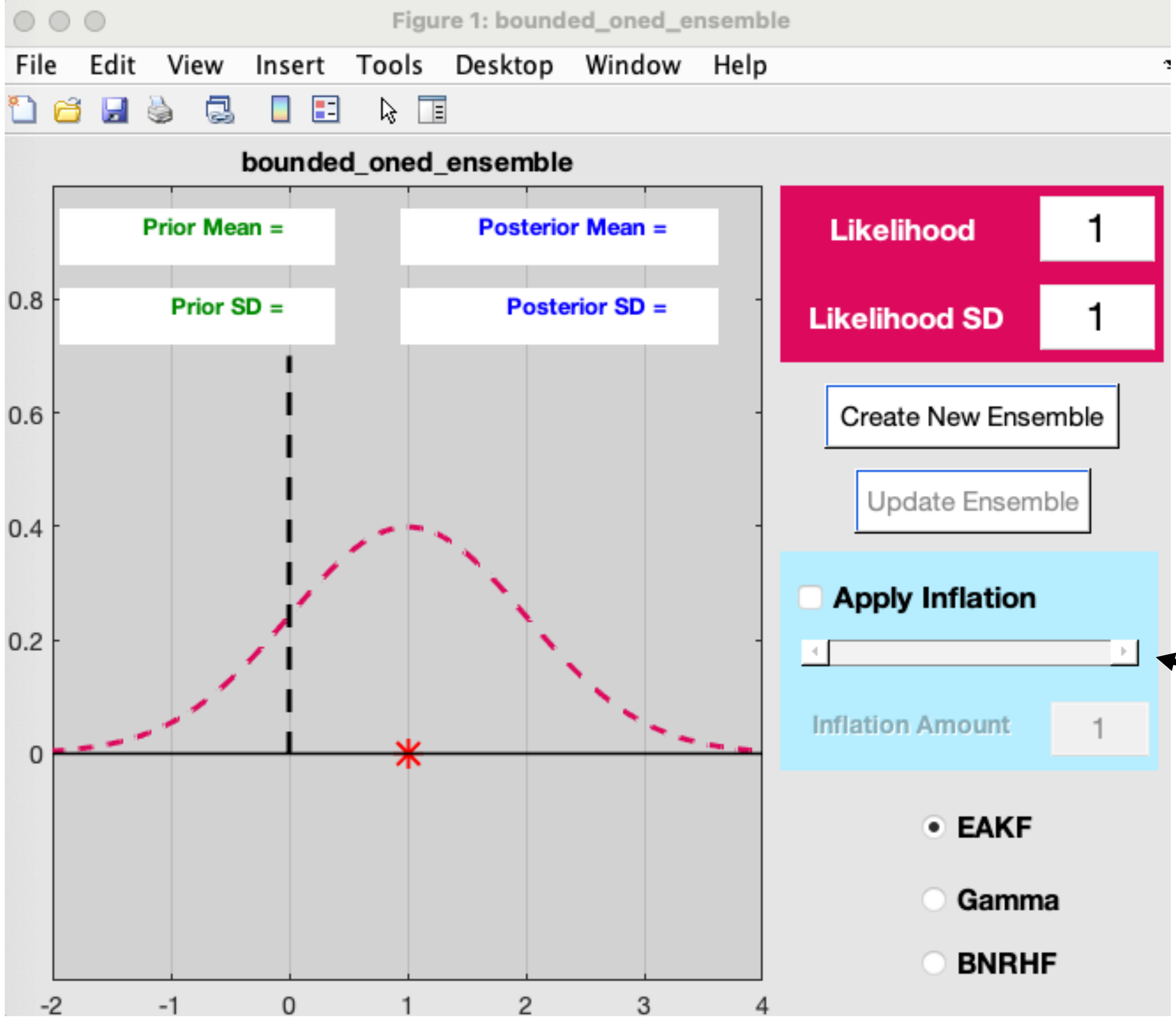


Inflation in Probit Space: Gamma example

Standard inflation may violate prior constraints.
Inflation can be done in probit space. And with BNRH!



Matlab Hands-on: bounded_oned_ensemble (2)



Try applying inflation with the different continuous priors.

What about the normal-normal case?

Computing increments in regular space is equivalent to computing increments in probit space.

Recall that the QCEFF normal filter in observation space is equivalent to the traditional EAKF in observation space.

Similarly, the method here is identical to the EAKF for unobserved updates.

The EAKF is equivalent to the Kalman Filter for normal/normal cases.

The QCEFF normal combined with probit space regression here is an ensemble generalization of the EAKF and the Kalman filter.