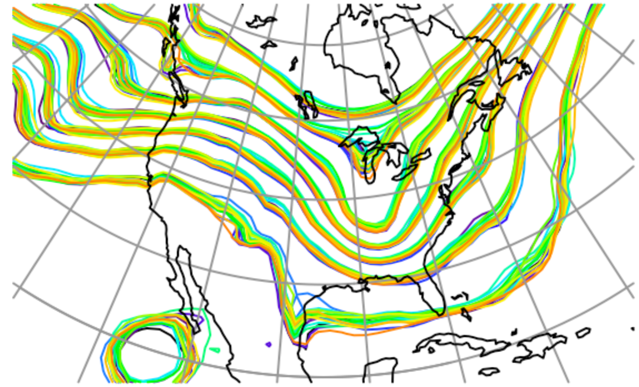


Non-Gaussian, Nonlinear Extensions for Ensemble Filter Data Assimilation with a Marginal Correction Rank Histogram Filter

Jeff Anderson, NCAR Data Assimilation Research Section

Outline

Many ensemble assimilation methods are Gaussian & linear.

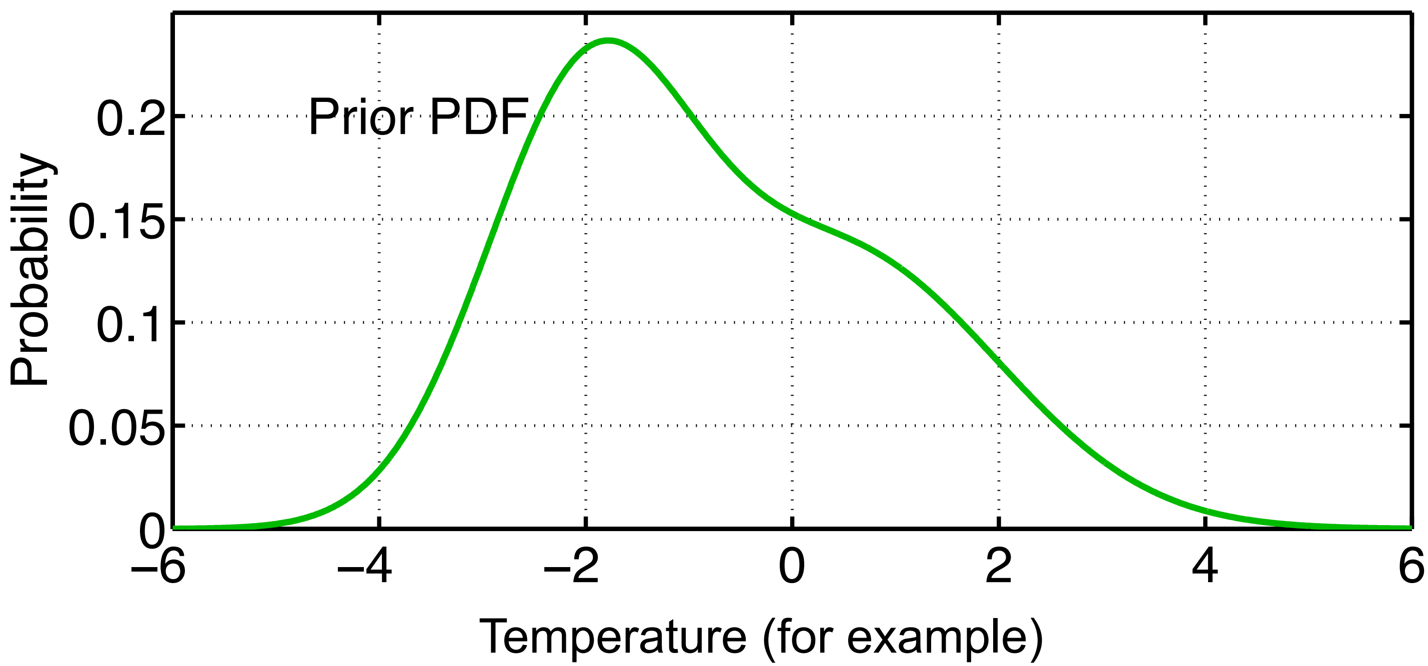
Tracer applications are neither.

Describe methods to deal with this.

Bayes Rule (1D example)

Bayes rule is the key to ensemble data assimilation.

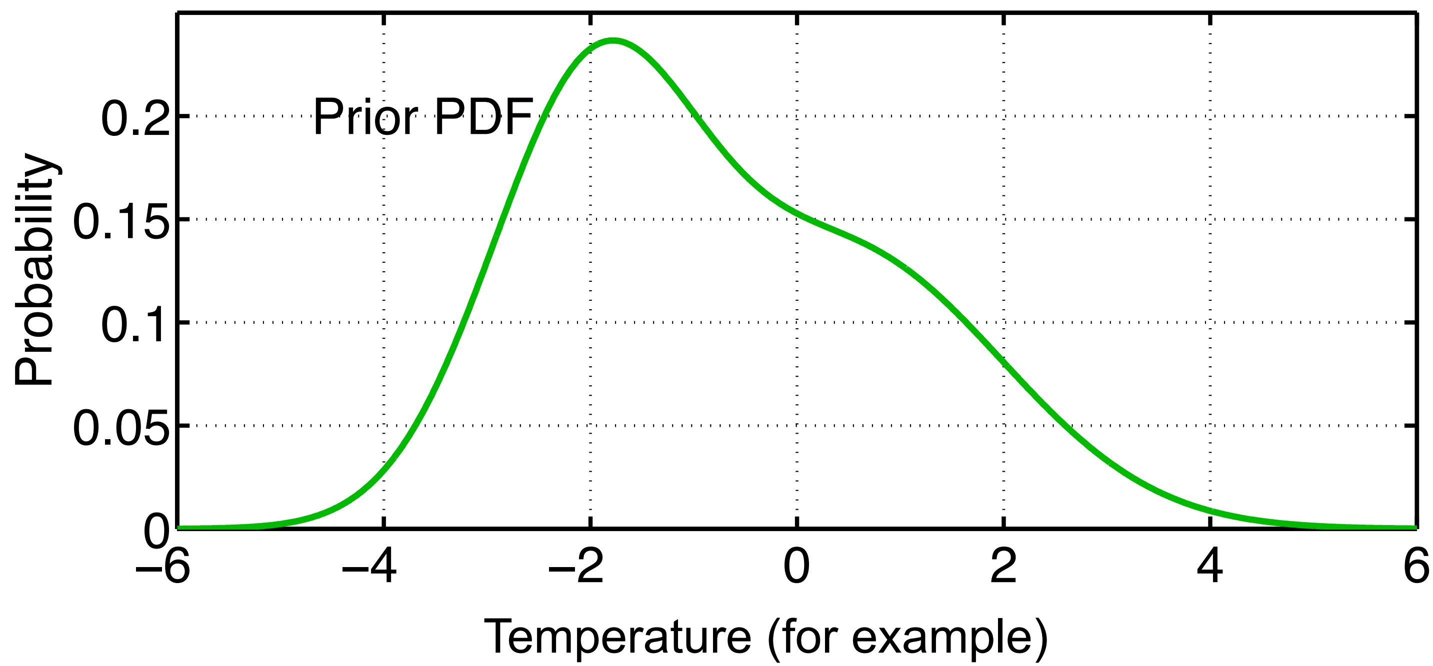
$$P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k | \mathbf{x}) P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Bayes Rule (1D example)

Prior: from model forecast.

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

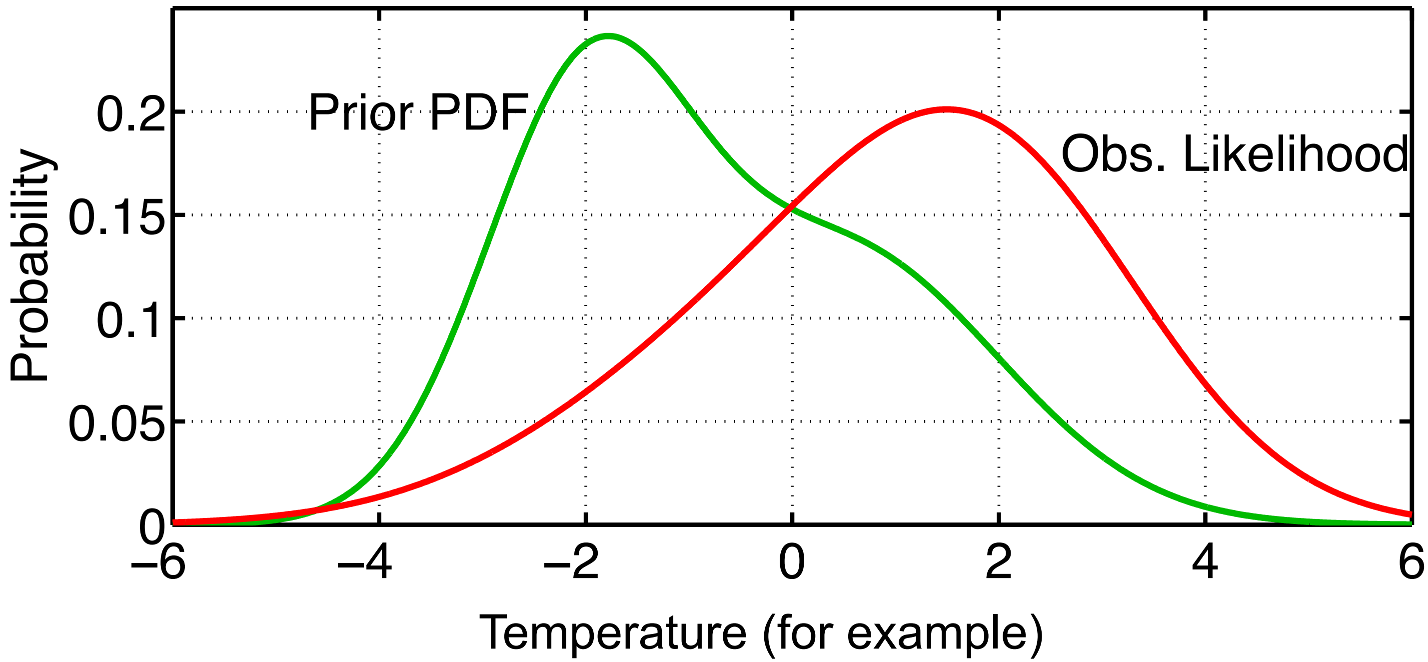


Data Assimilation Research Testbed

Bayes Rule (1D example)

Likelihood:
from instrument.

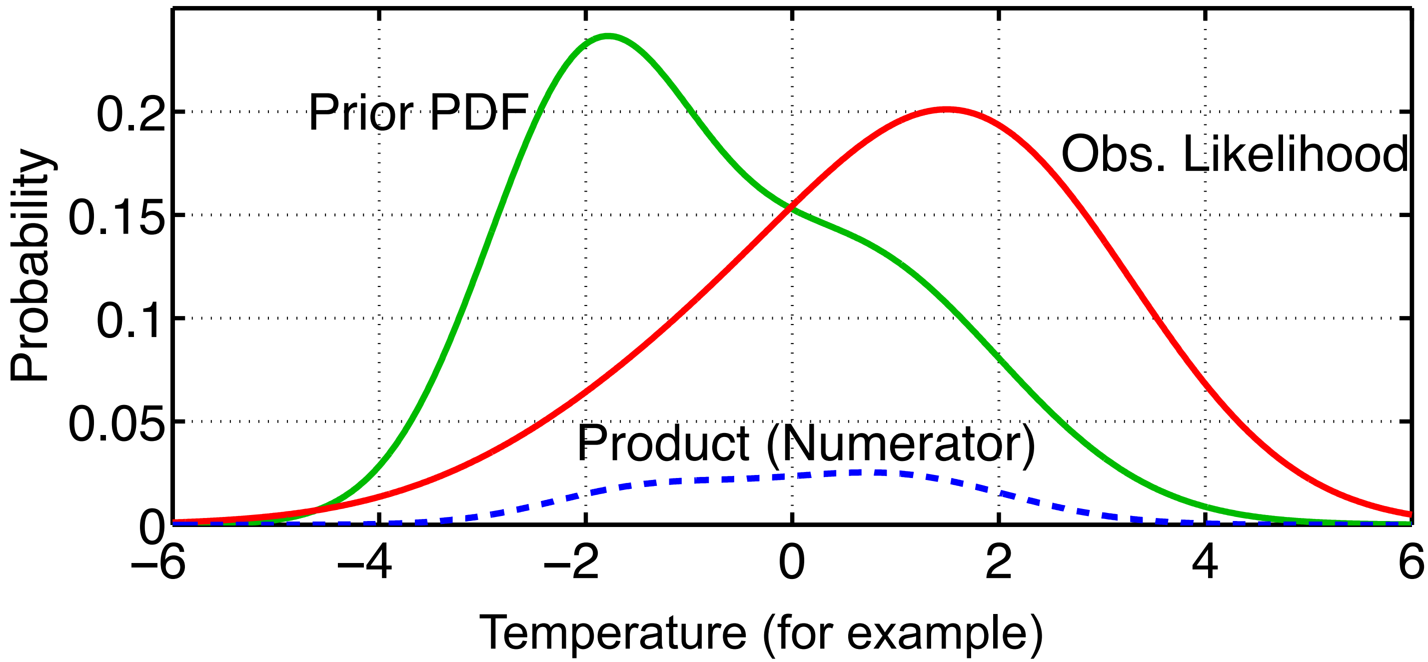
$$P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Data
Assimilation
Research
Testbed

Bayes Rule (1D example)

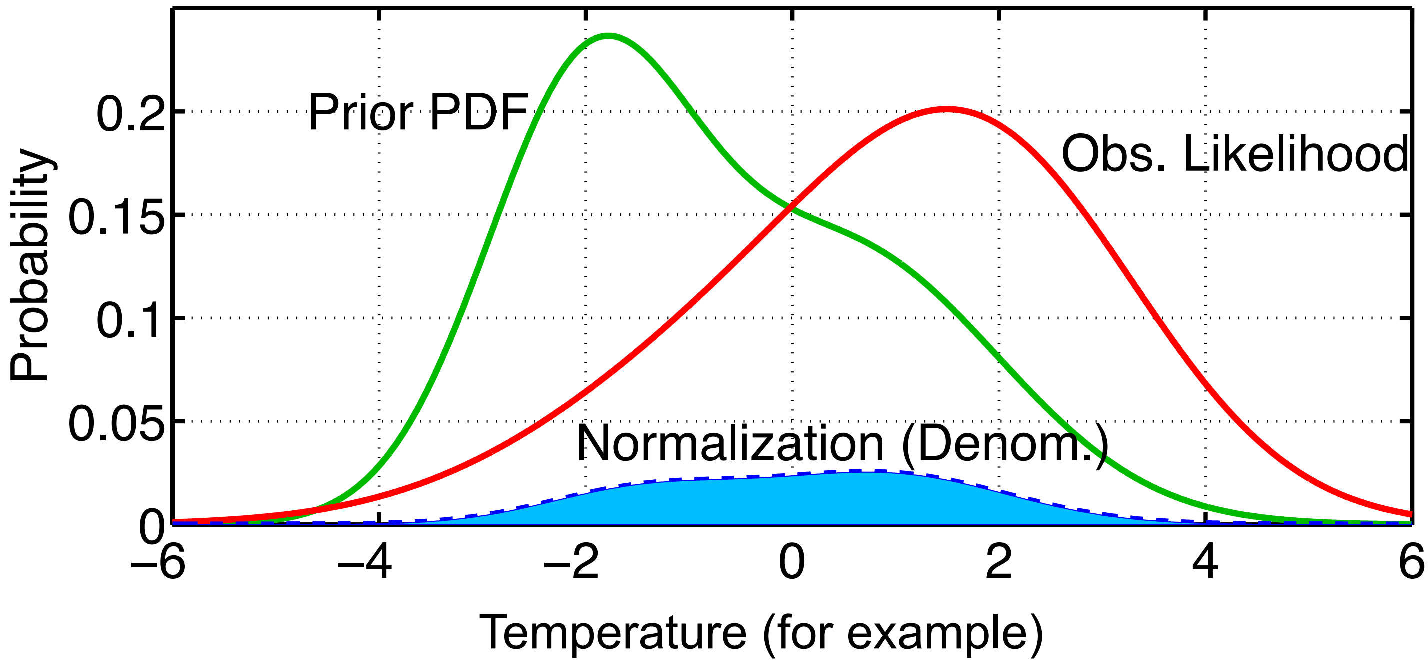
$$P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k | \mathbf{x}) P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Data Assimilation Research Testbed

Bayes Rule (1D example)

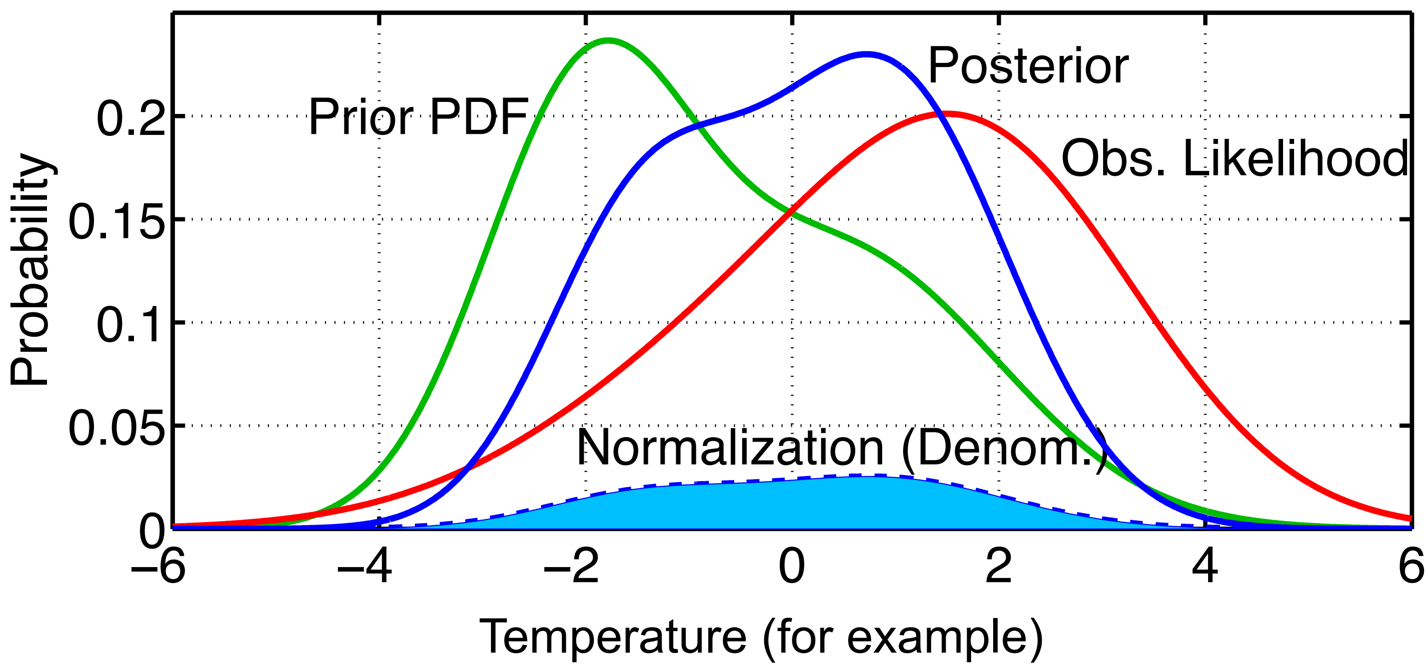
$$P\left(\mathbf{x}_{t_k} \middle| \mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k} \middle| \mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Bayes Rule (1D example)

Posterior:
(analysis).

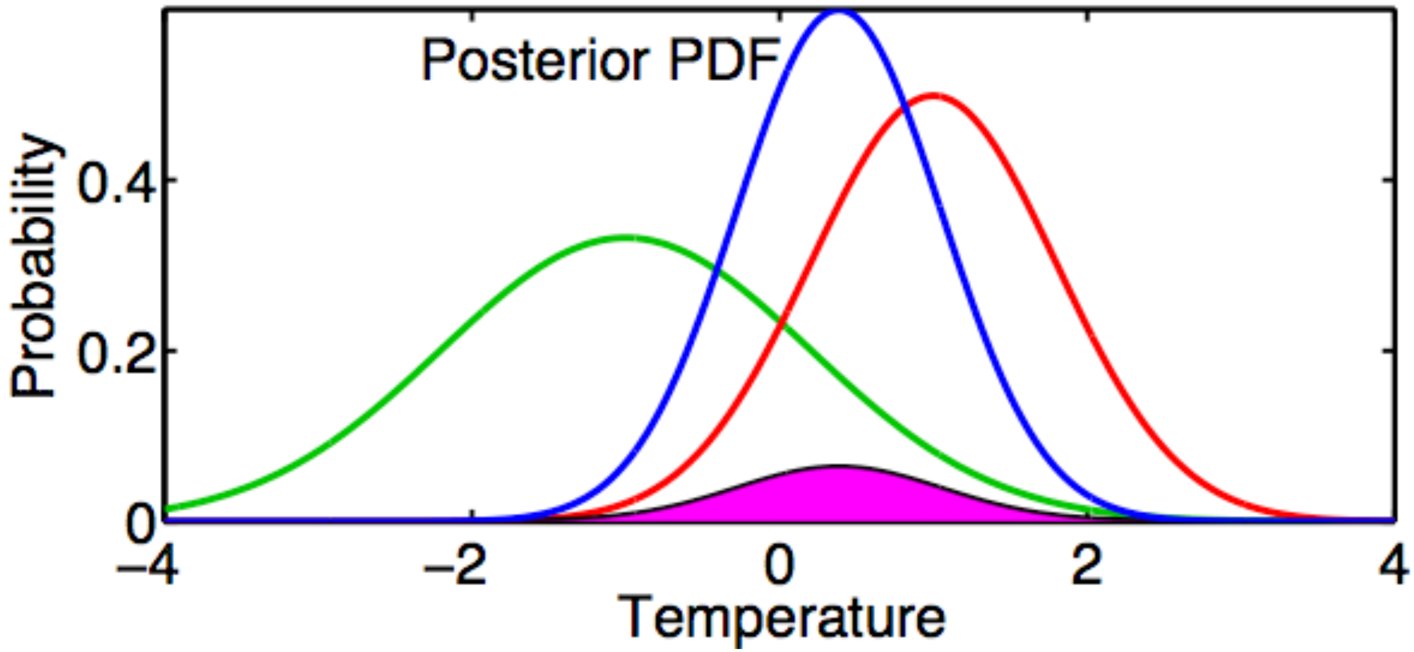
$$P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k|\mathbf{x})P\left(\mathbf{x}_{t_k}|\mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Bayes Rule (1D example)

Most ensemble assimilation algorithms assume Gaussians.
May be okay for quantity like temperature.

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

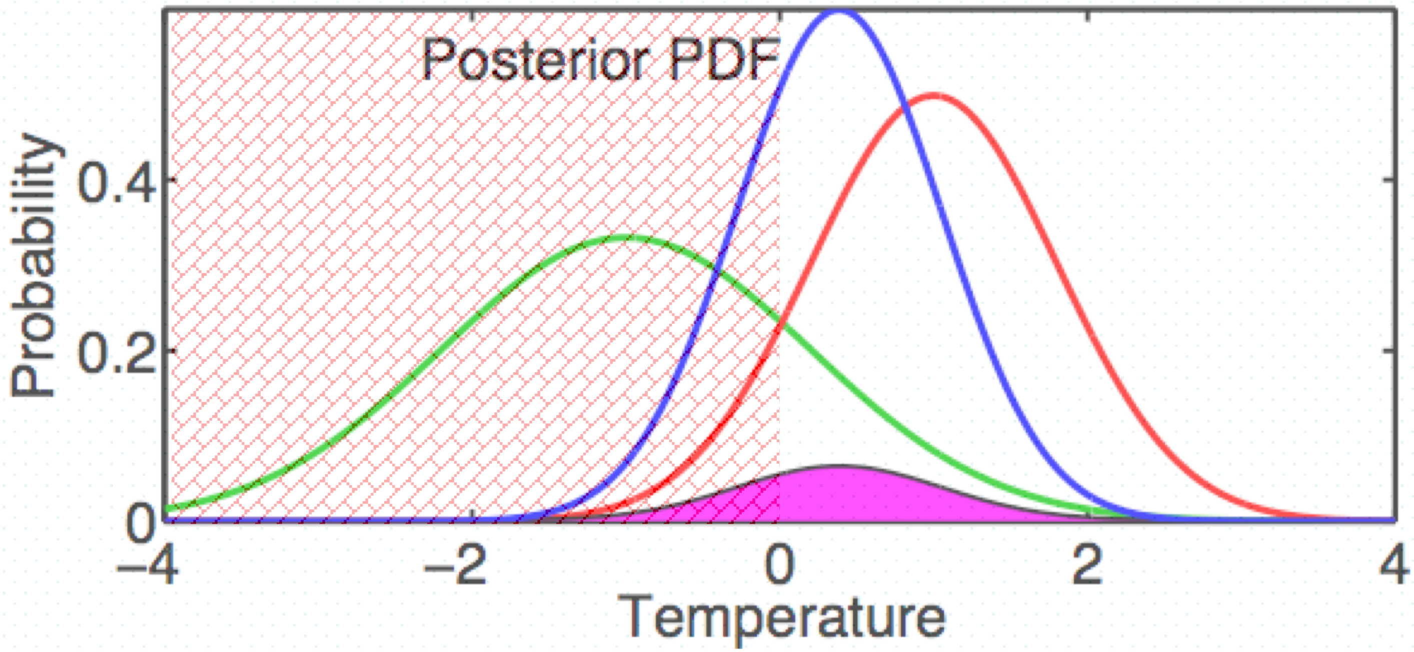


Bayes Rule (1D example)

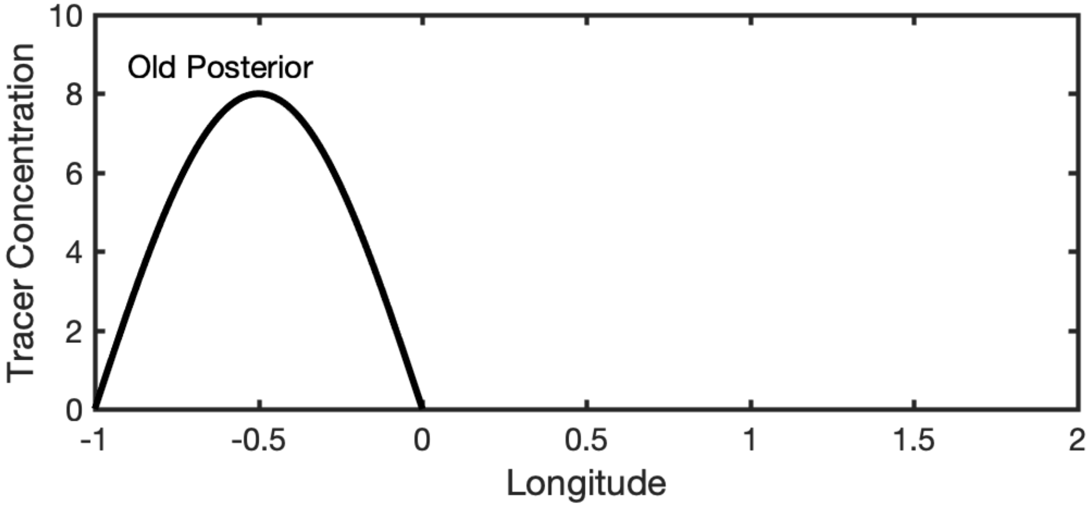
Most ensemble assimilation algorithms assume Gaussians.

Tracer concentration is bounded. Gaussian a poor choice.

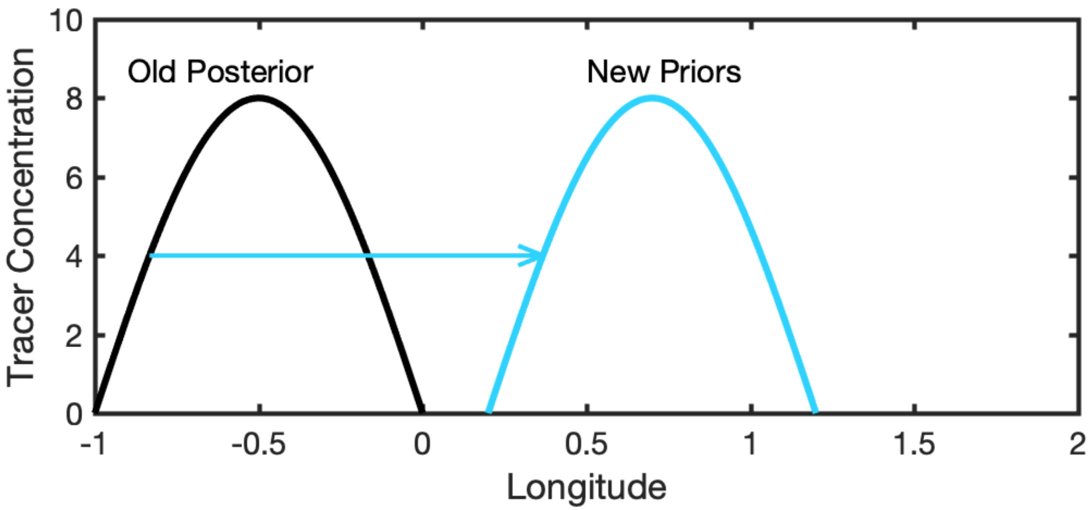
$$P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_k}\right) = \frac{P(\mathbf{y}_k | \mathbf{x}) P\left(\mathbf{x}_{t_k} | \mathbf{Y}_{t_{k-1}}\right)}{Normalization}$$



Advection of Tracer -> Nonlinear Prior for Concentration & Wind

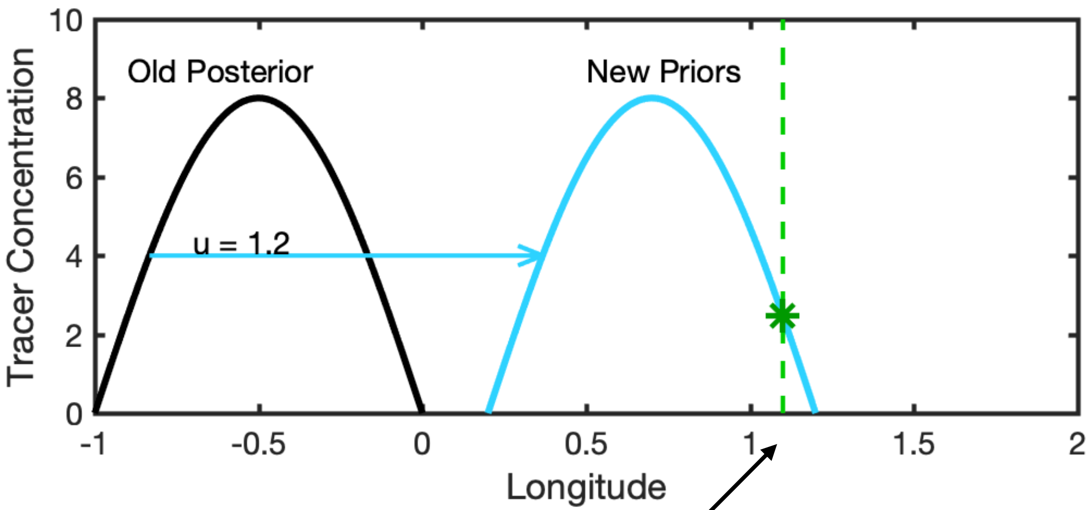


Advection of Tracer -> Nonlinear Prior for Concentration & Wind



Tracer advected by uncertain winds.

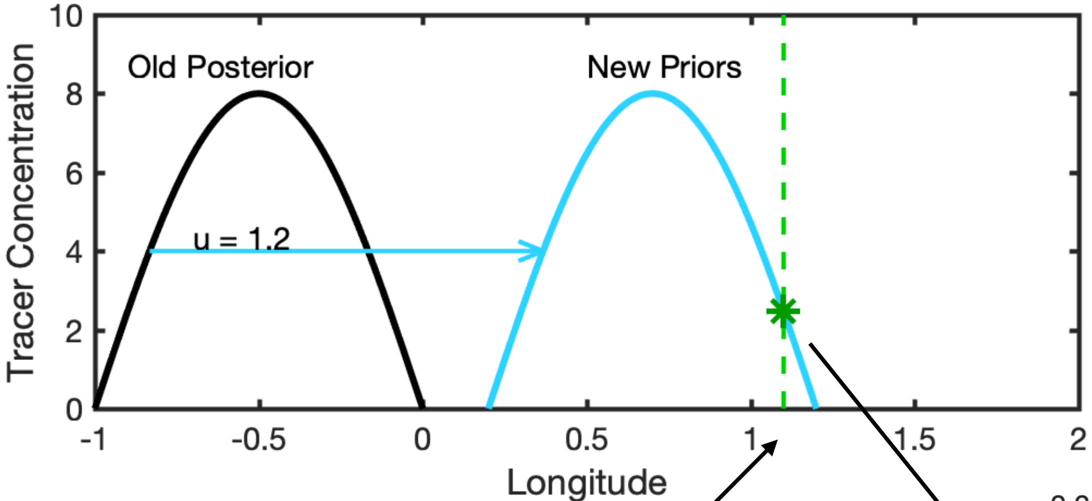
Advection of Tracer -> Nonlinear Prior for Concentration & Wind



Concentration Observed at Longitude 1.1

Tracer advected by uncertain winds.

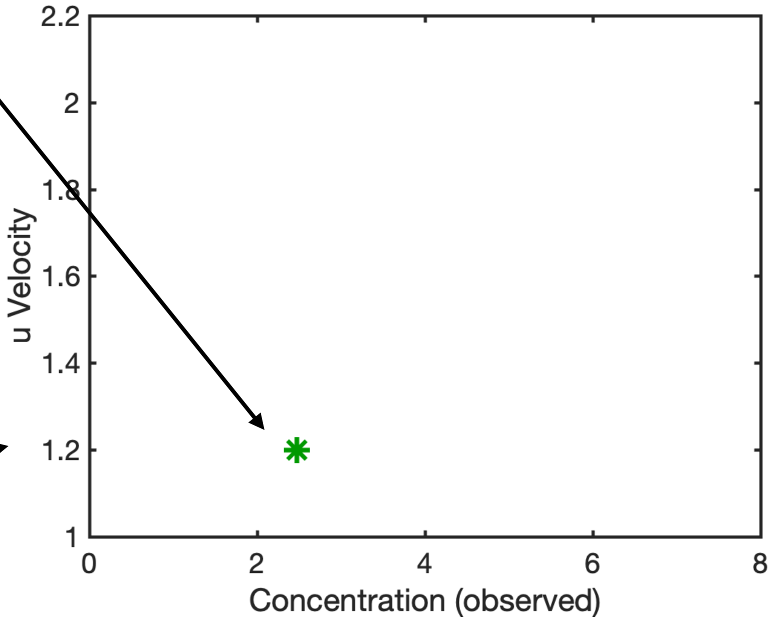
Advection of Tracer -> Nonlinear Prior for Concentration & Wind



Tracer advected by uncertain winds.

Concentration Observed at Longitude 1.1

Bivariate Prior shows how Concentration Observation Impacts Wind.

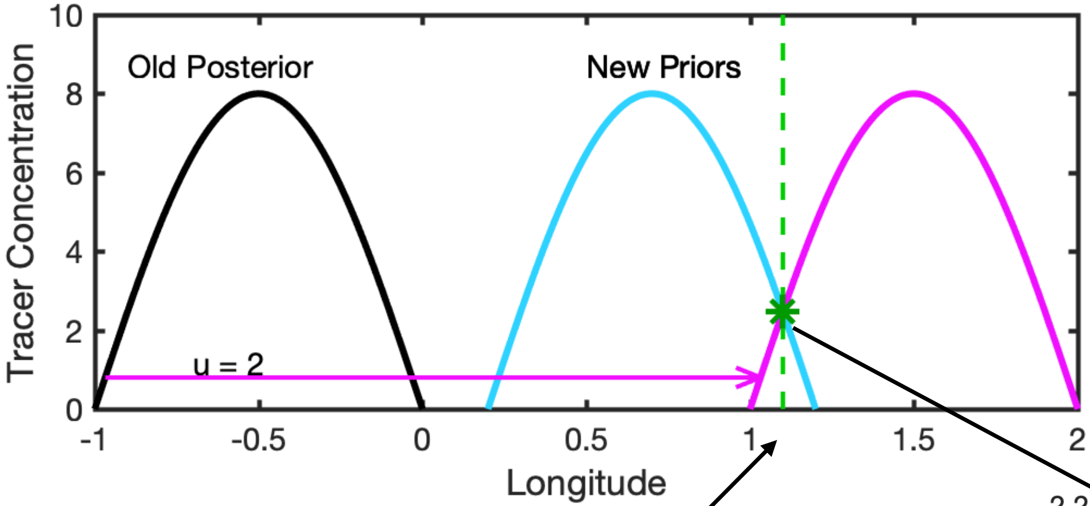


NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

NSF

Data Assimilation Research Testbed

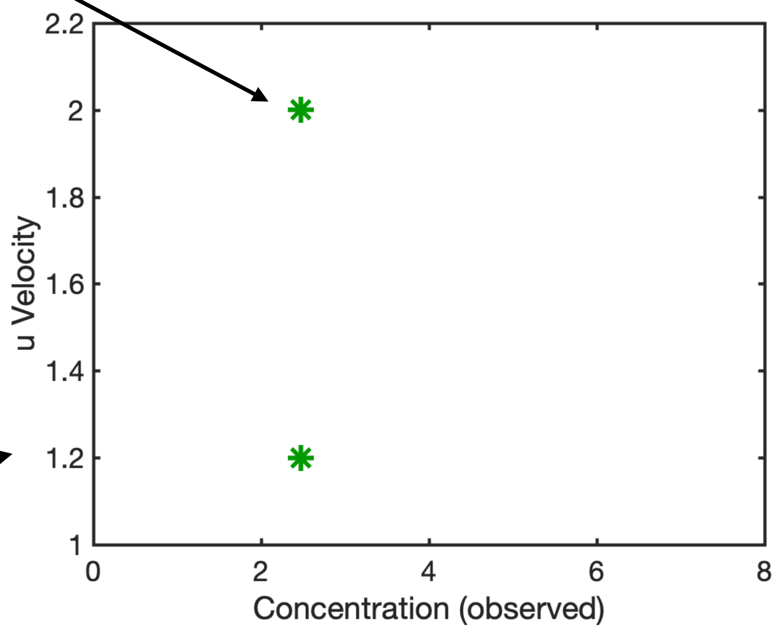
Advection of Tracer -> Nonlinear Prior for Concentration & Wind



Tracer advected by uncertain winds.

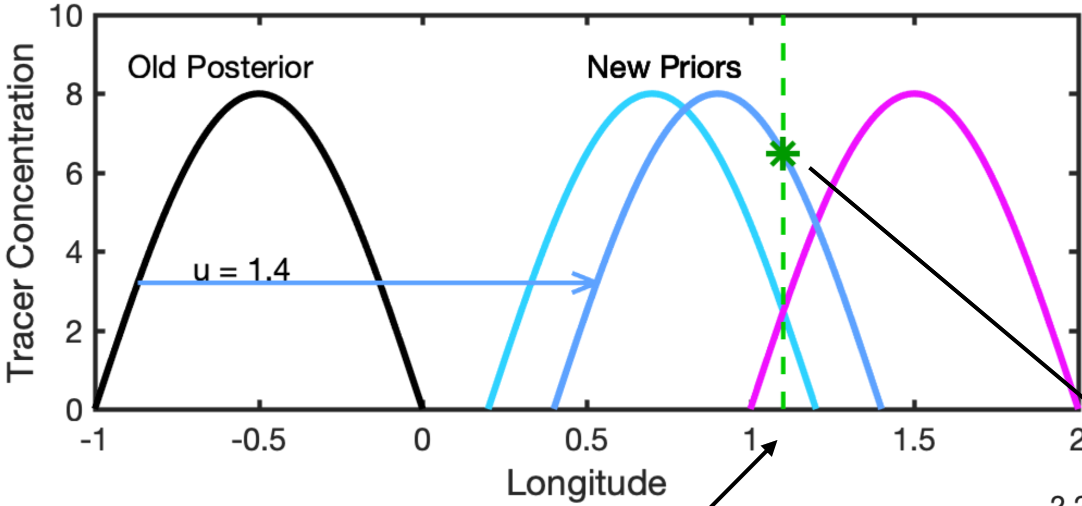
Concentration Observed at Longitude 1.1

Bivariate Prior shows how Concentration Observation Impacts Wind.



Data Assimilation Research Testbed

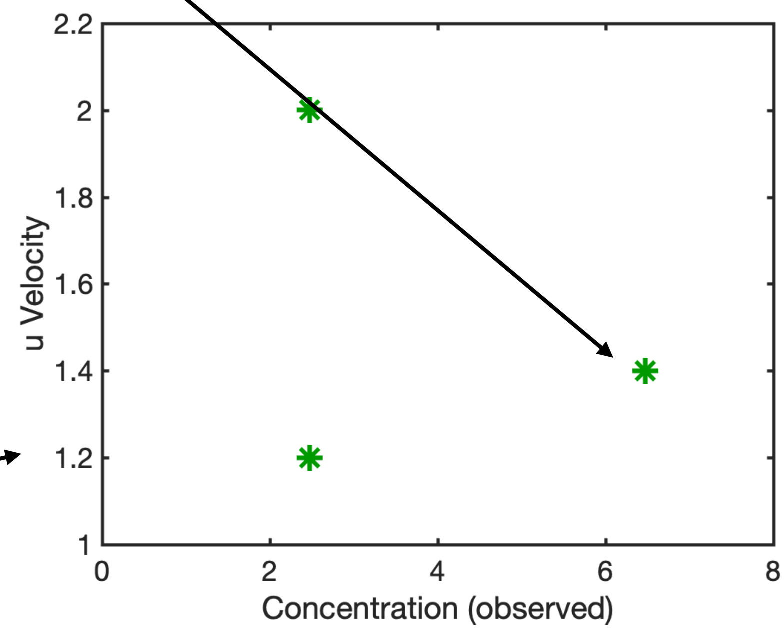
Advection of Tracer -> Nonlinear Prior for Concentration & Wind



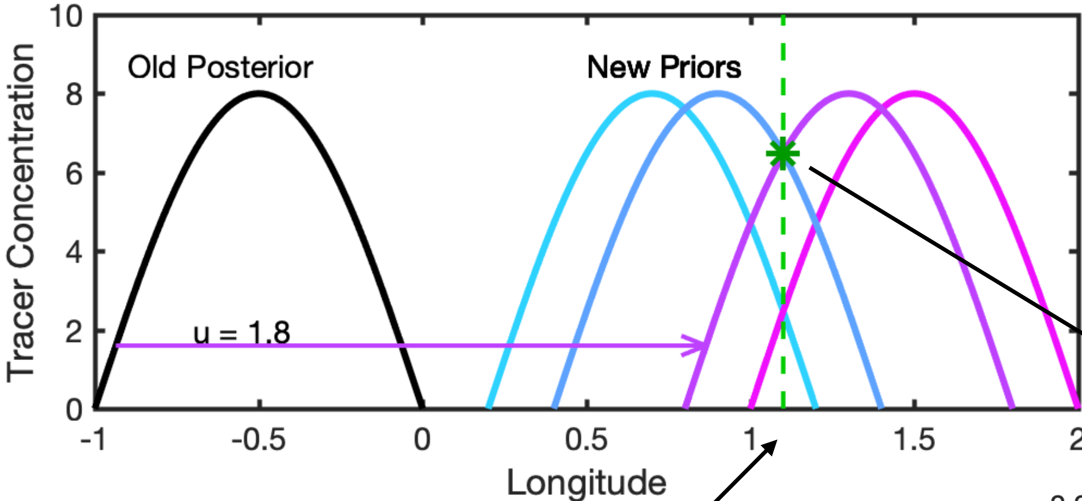
Tracer advected by uncertain winds.

Concentration Observed at Longitude 1.1

Bivariate Prior shows how Concentration Observation Impacts Wind.

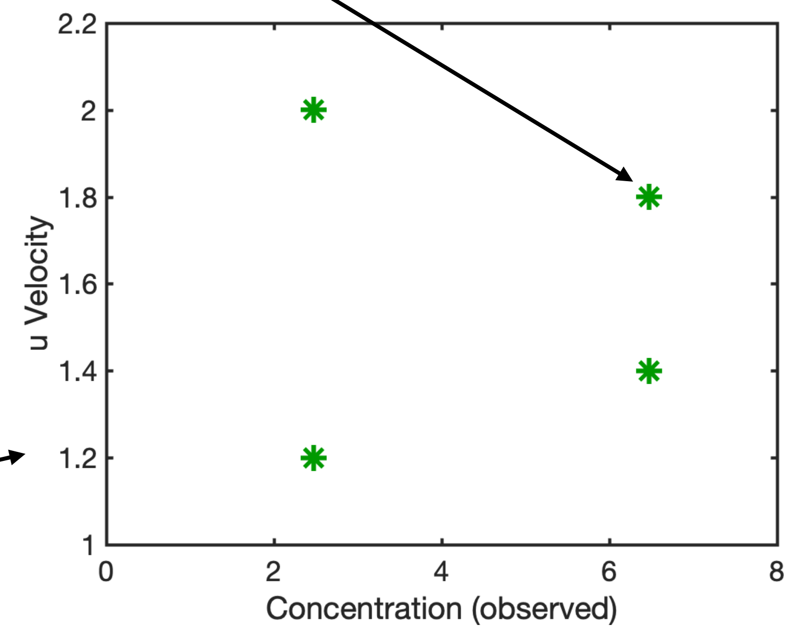


Advection of Tracer -> Nonlinear Prior for Concentration & Wind



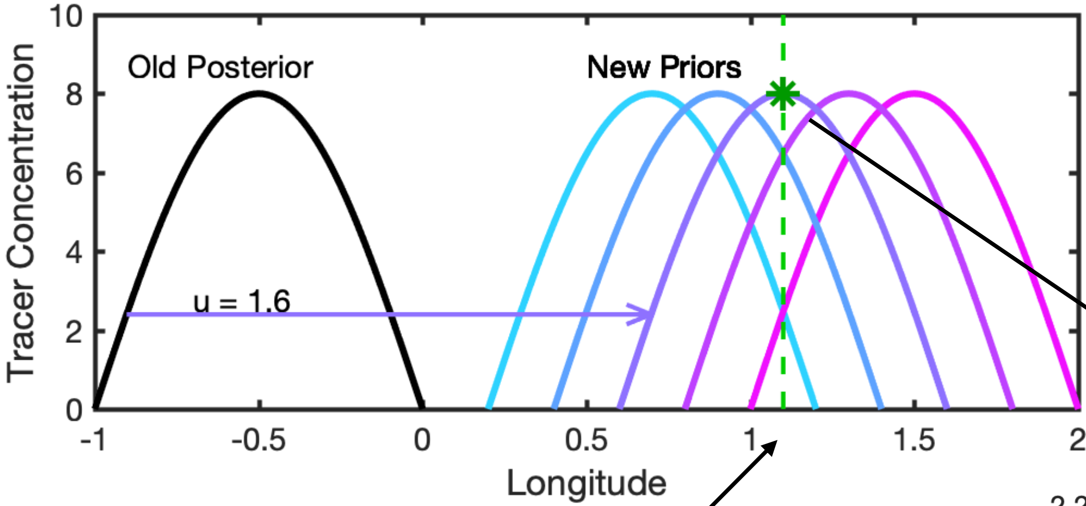
Tracer advected by uncertain winds.

Concentration Observed at Longitude 1.1



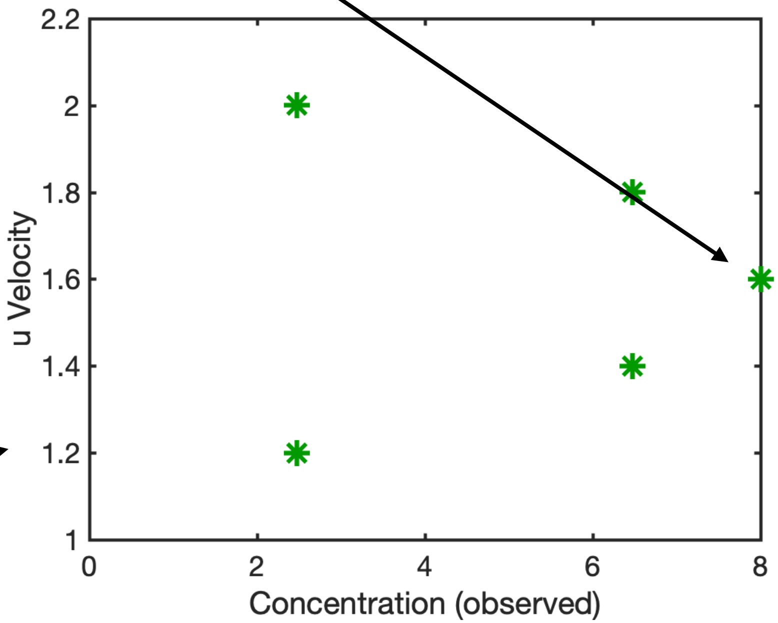
Bivariate Prior shows how Concentration Observation Impacts Wind.

Advection of Tracer -> Nonlinear Prior for Concentration & Wind



Tracer advected by uncertain winds.

Concentration Observed at Longitude 1.1



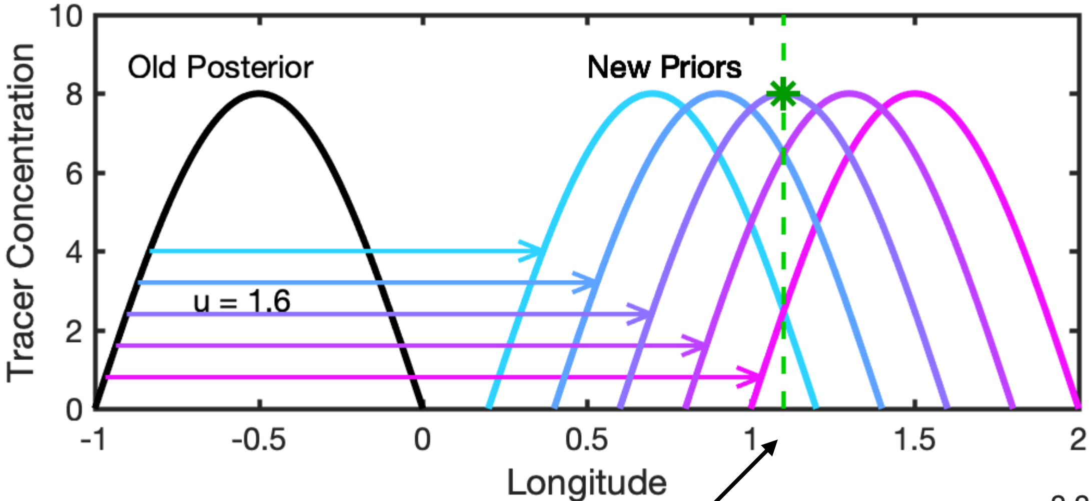
Bivariate Prior shows how Concentration Observation Impacts Wind.

NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

NSF

Data Assimilation Research Testbed

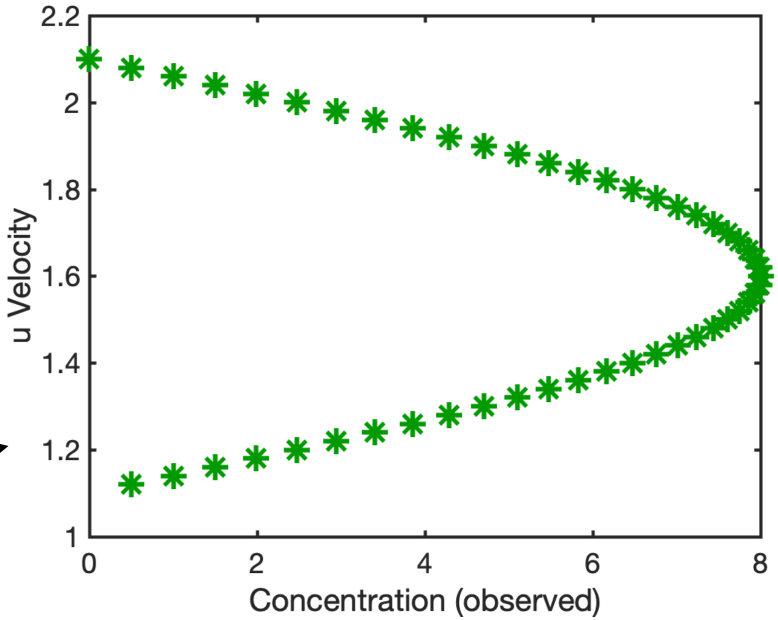
Advection of Tracer -> Nonlinear Prior for Concentration & Wind



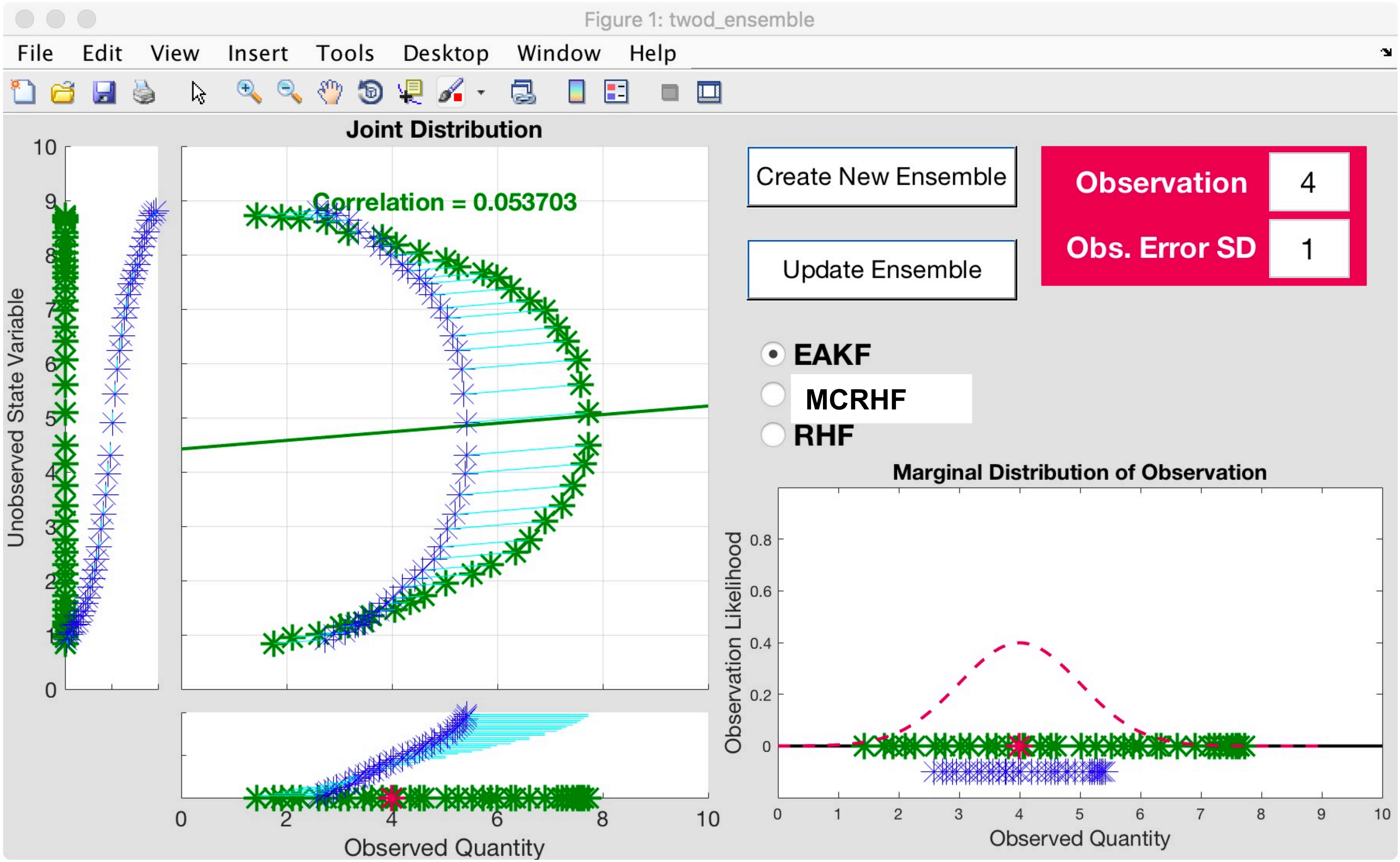
Tracer advected by uncertain winds.

Concentration Observed at Longitude 1.1

Bivariate Prior shows how Concentration Observation Impacts Wind.



Advection of Cosine Tracer: EAKF



Challenges for Tracer Assimilation

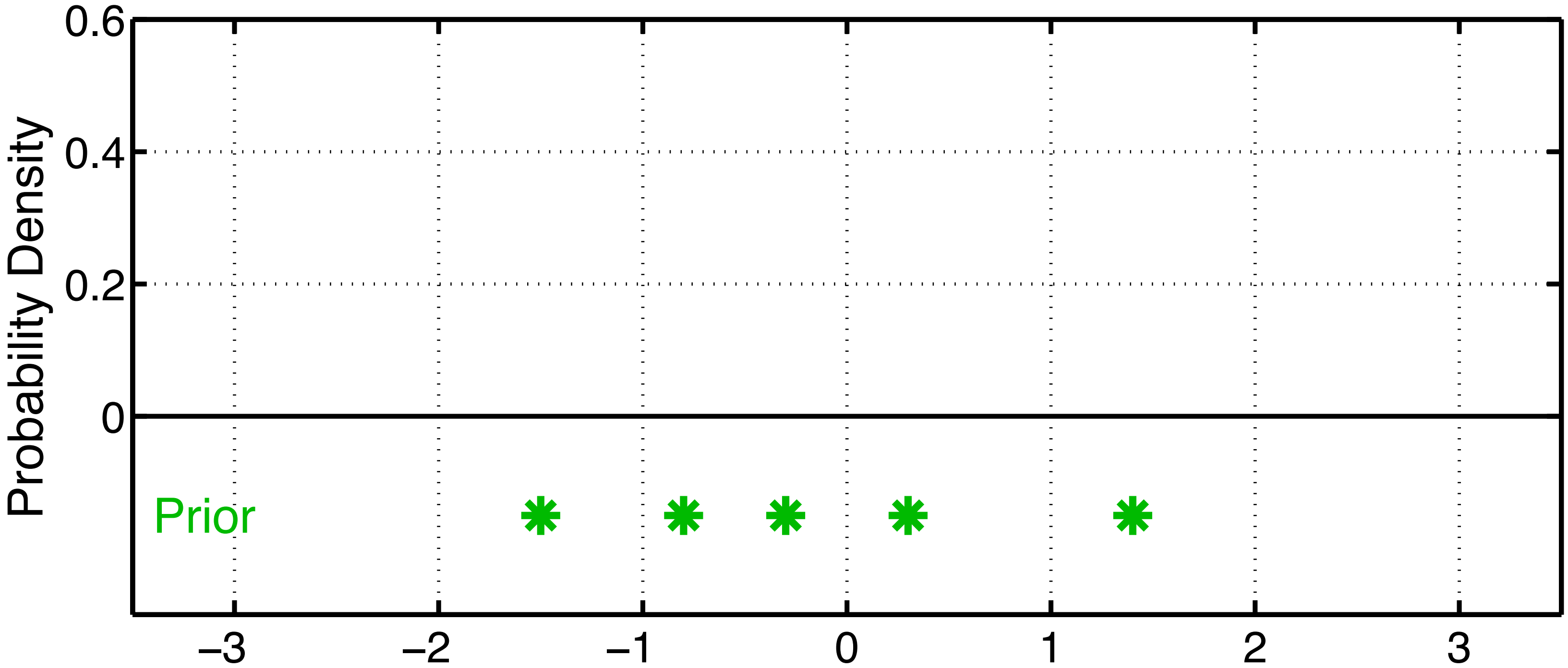
Non-Gaussian bounded priors.

Nonlinear bivariate priors.

Solution: More general representation of priors and likelihoods.

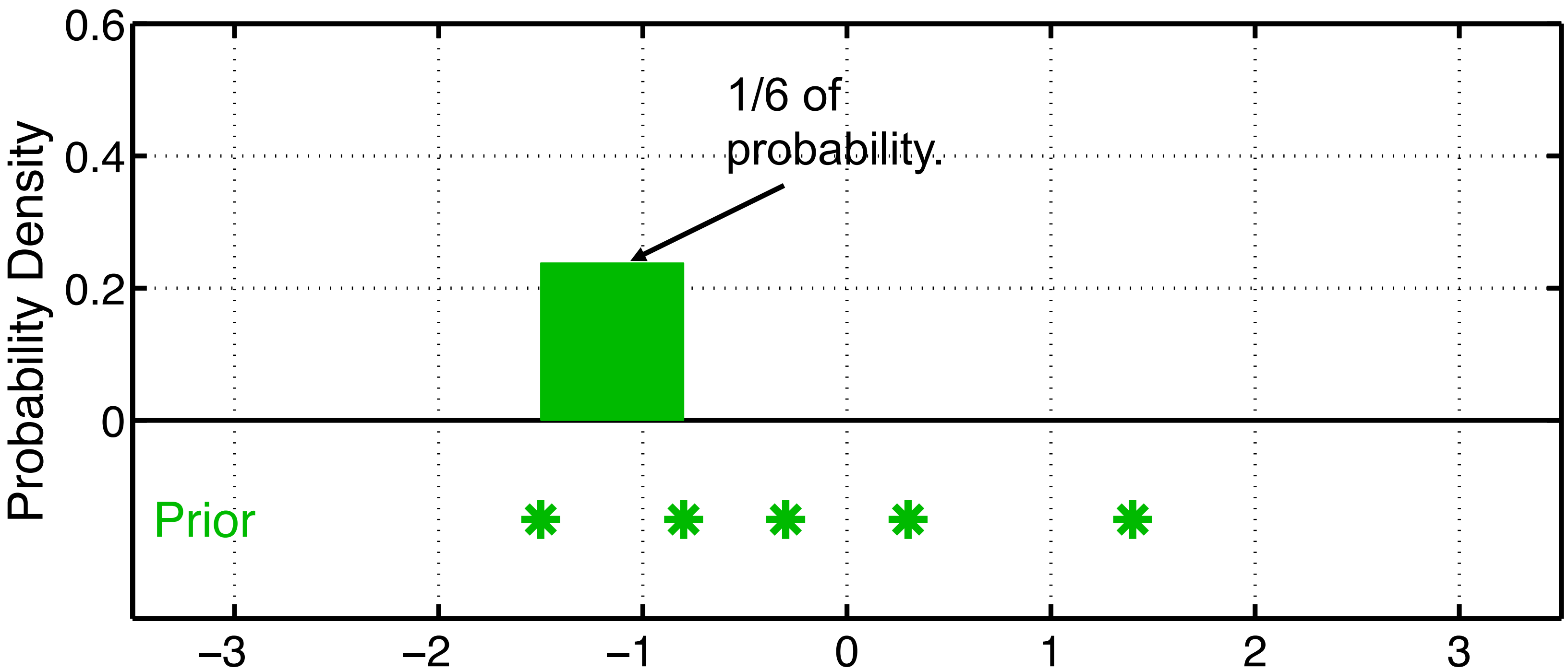
Rank Histogram Filters for State Variables.

Marginal Correction Rank Histogram (MCRHF)



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

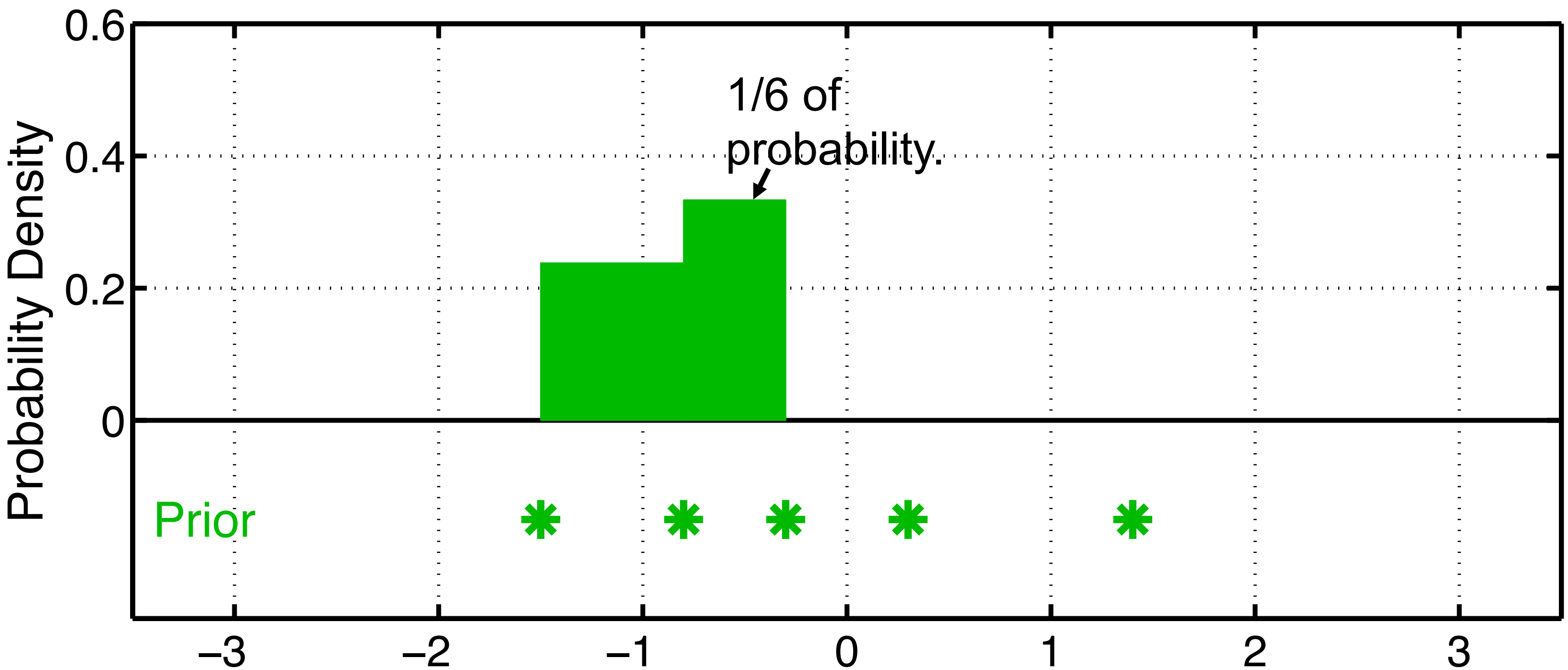
Marginal Correction Rank Histogram (MCRHF)



Step 1: Get continuous prior distribution density.

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

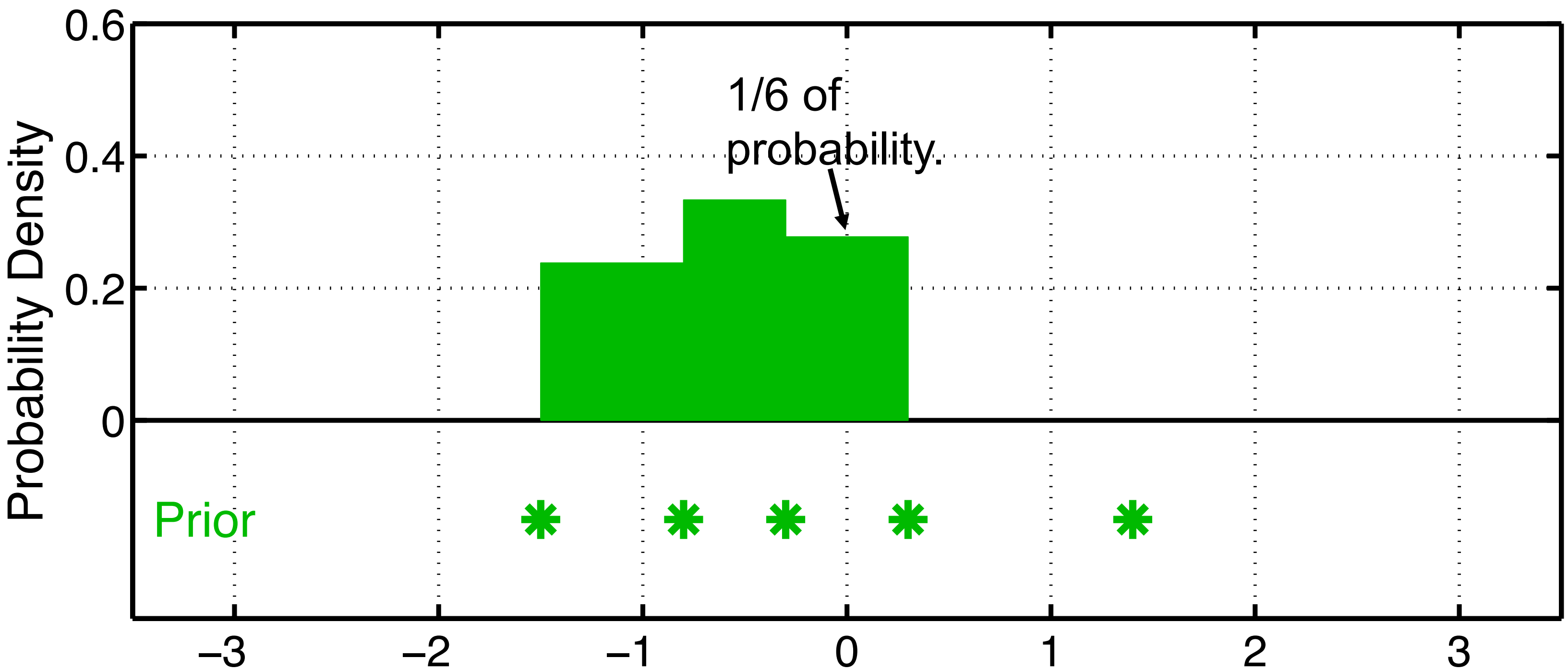
Marginal Correction Rank Histogram (MCRHF)



Step 1: Get continuous prior distribution density.

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

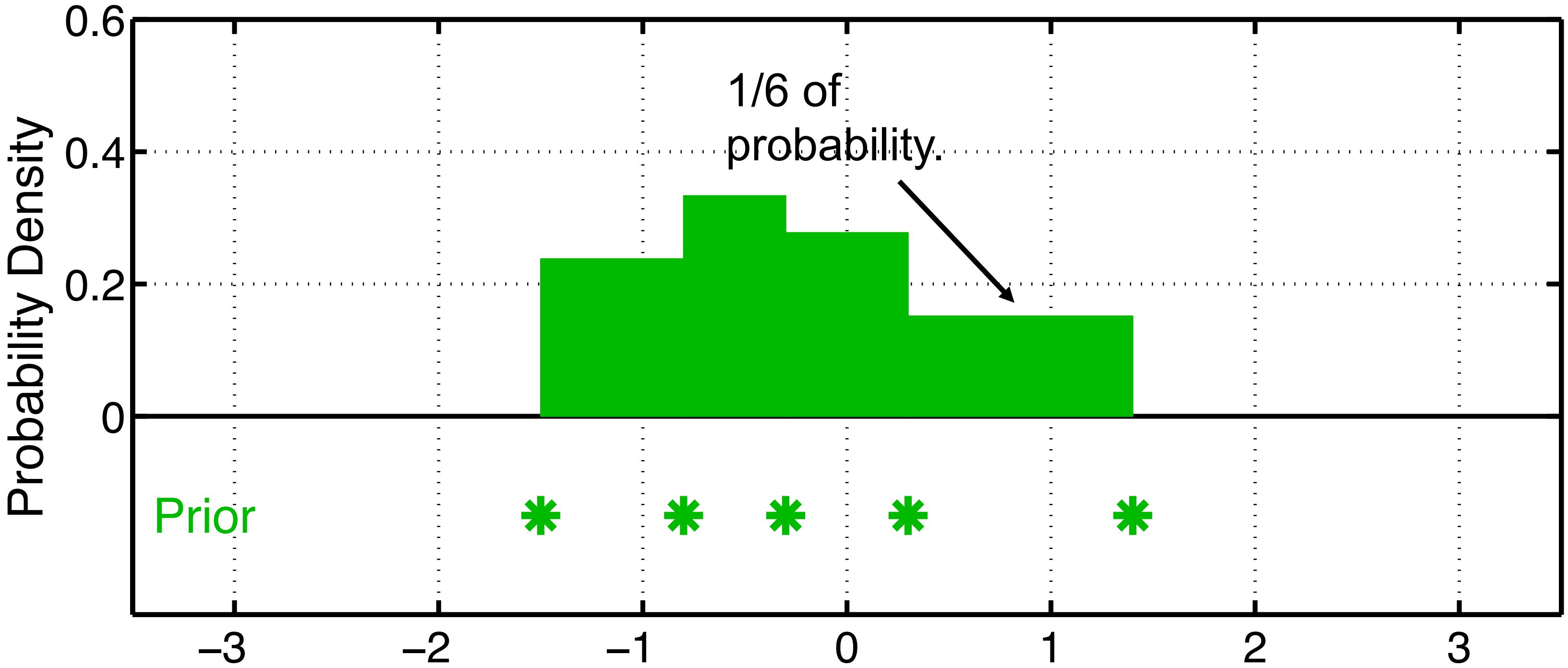
Marginal Correction Rank Histogram (MCRHF)



Step 1: Get continuous prior distribution density.

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

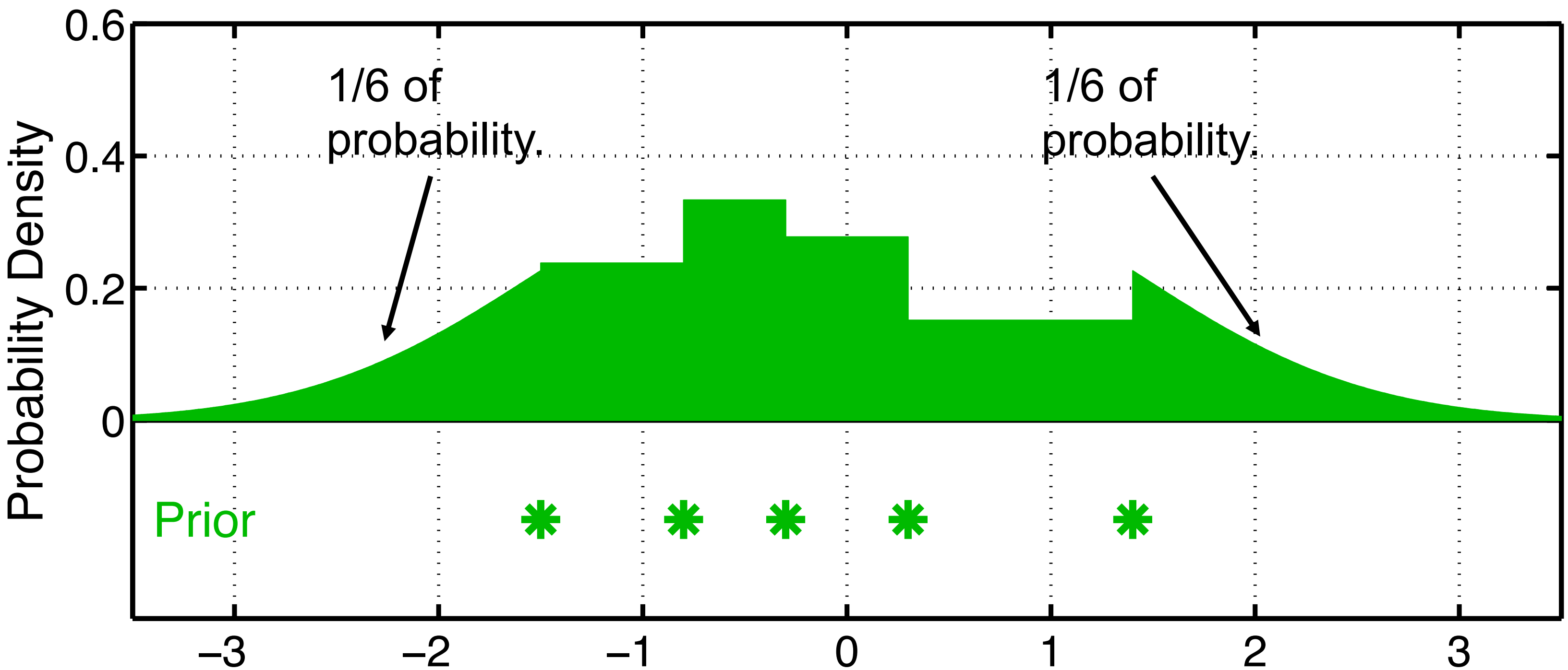
Marginal Correction Rank Histogram (MCRHF)



Step 1: Get continuous prior distribution density.

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

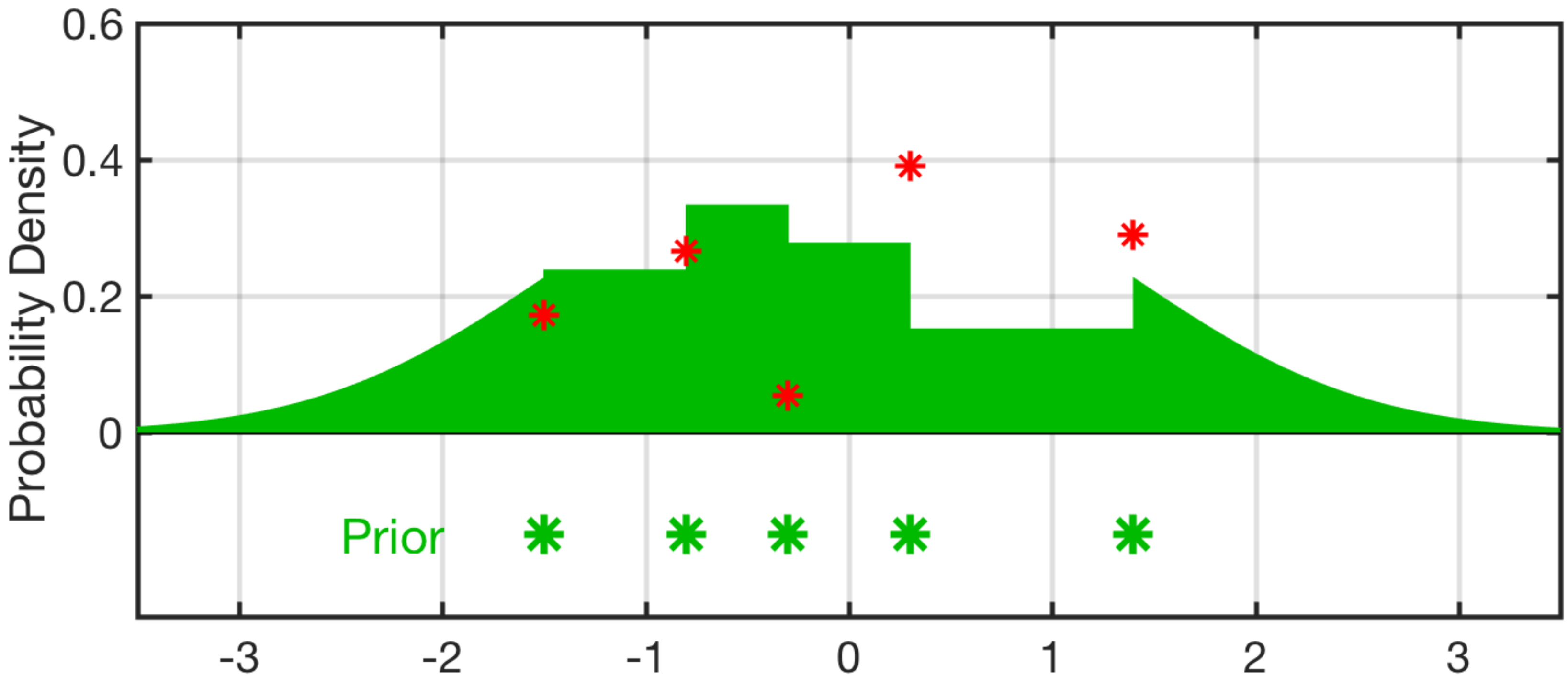
Marginal Correction Rank Histogram (MCRHF)



Step 1: Get continuous prior distribution density.

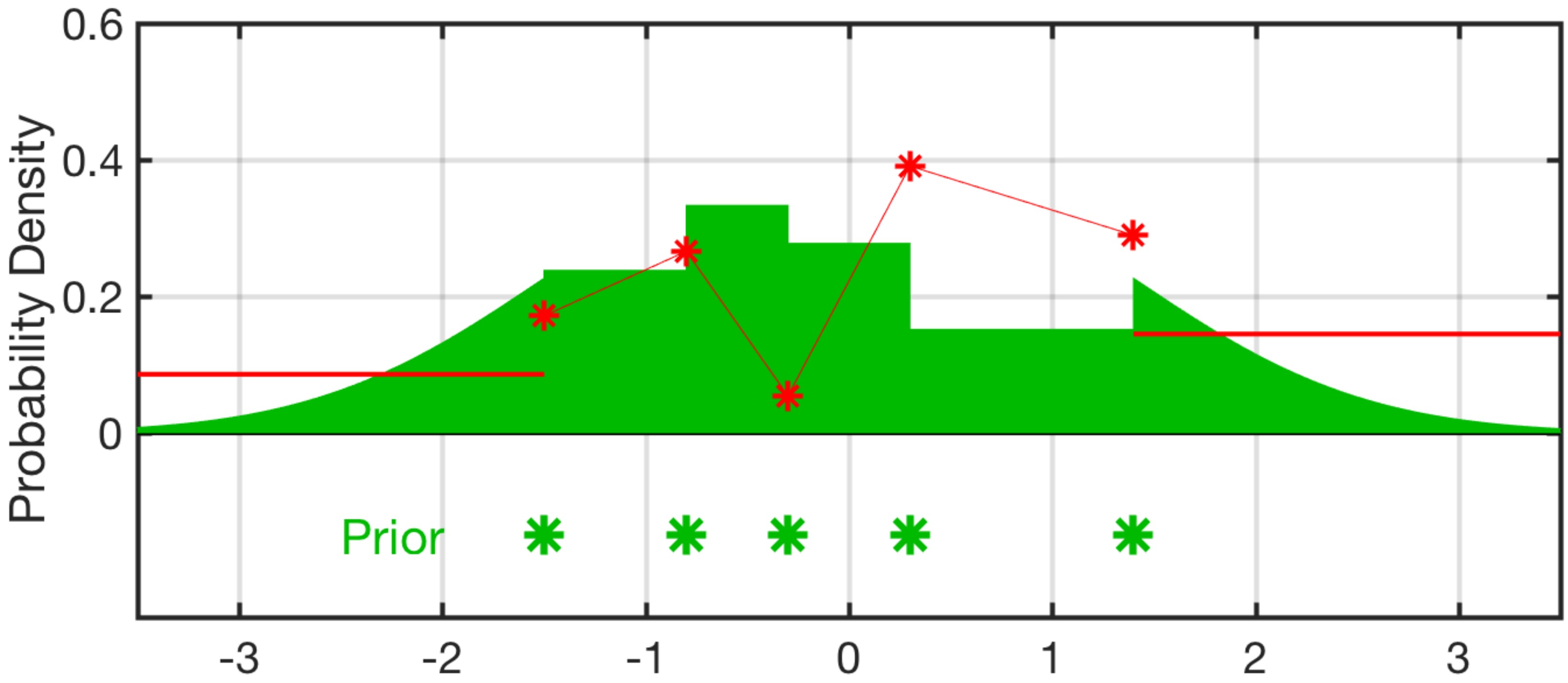
- Partial gaussian kernels on tails, N(*tail_mean*, *ens_sd*).
- *tail_mean selected so that (ens_size + 1)$^{-1}$ mass is in tail.*

Marginal Correction Rank Histogram (MCRHF)



Step 2: Get observation likelihood for each ensemble member.

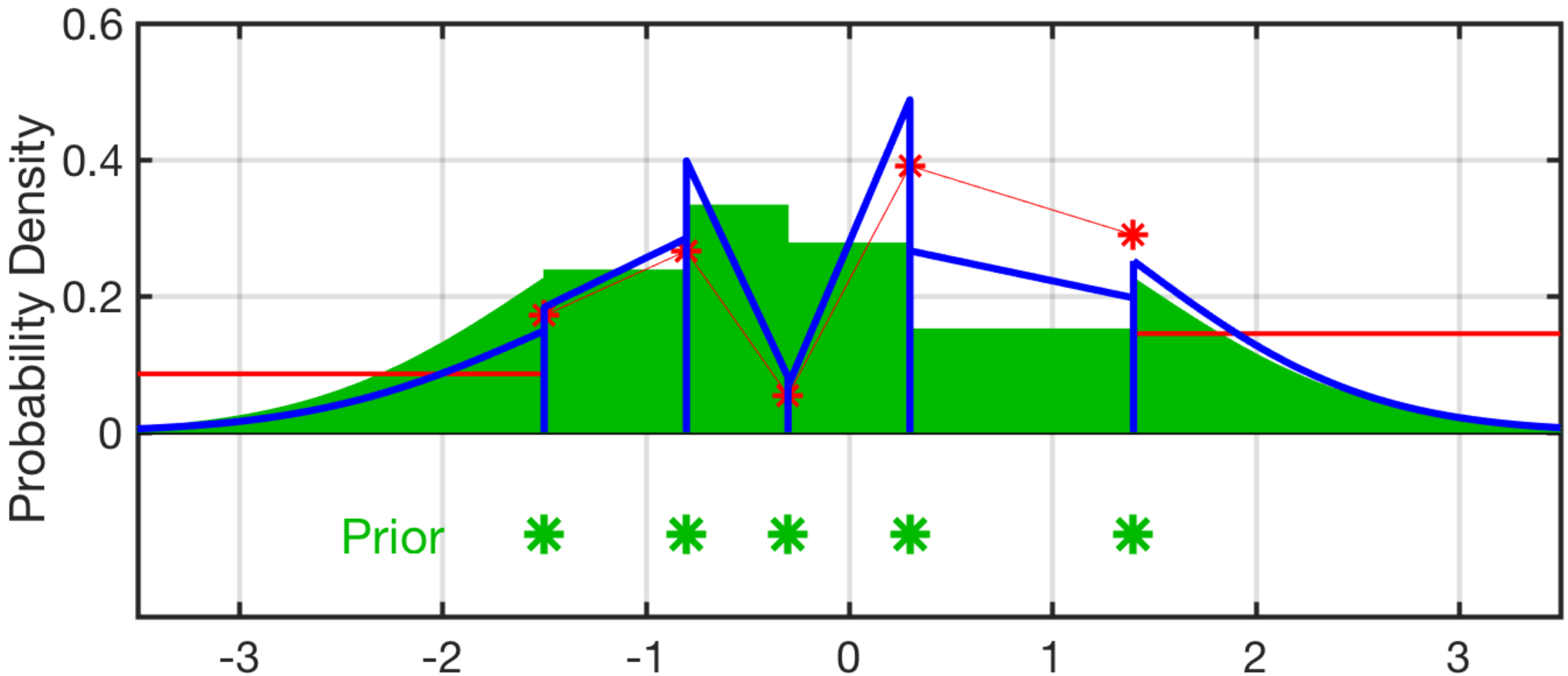
Marginal Correction Rank Histogram (MCRHF)



Step 3: Approximate likelihood with trapezoidal quadrature.

- Use long flat tails.

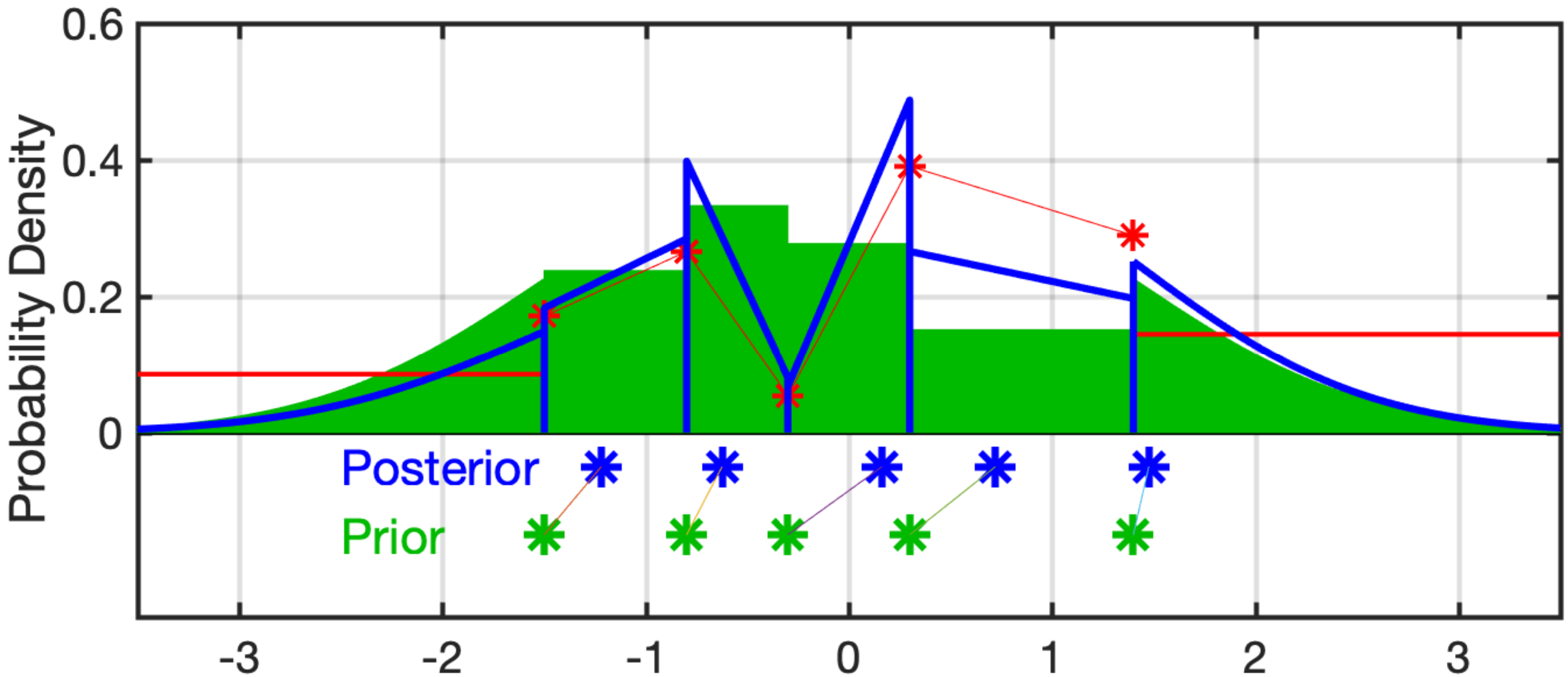
Marginal Correction Rank Histogram (MCRHF)



Step 4: Compute continuous posterior distribution.

- Just Bayes, multiply prior by likelihood and normalize.
- Really simple with uniform likelihood tails.

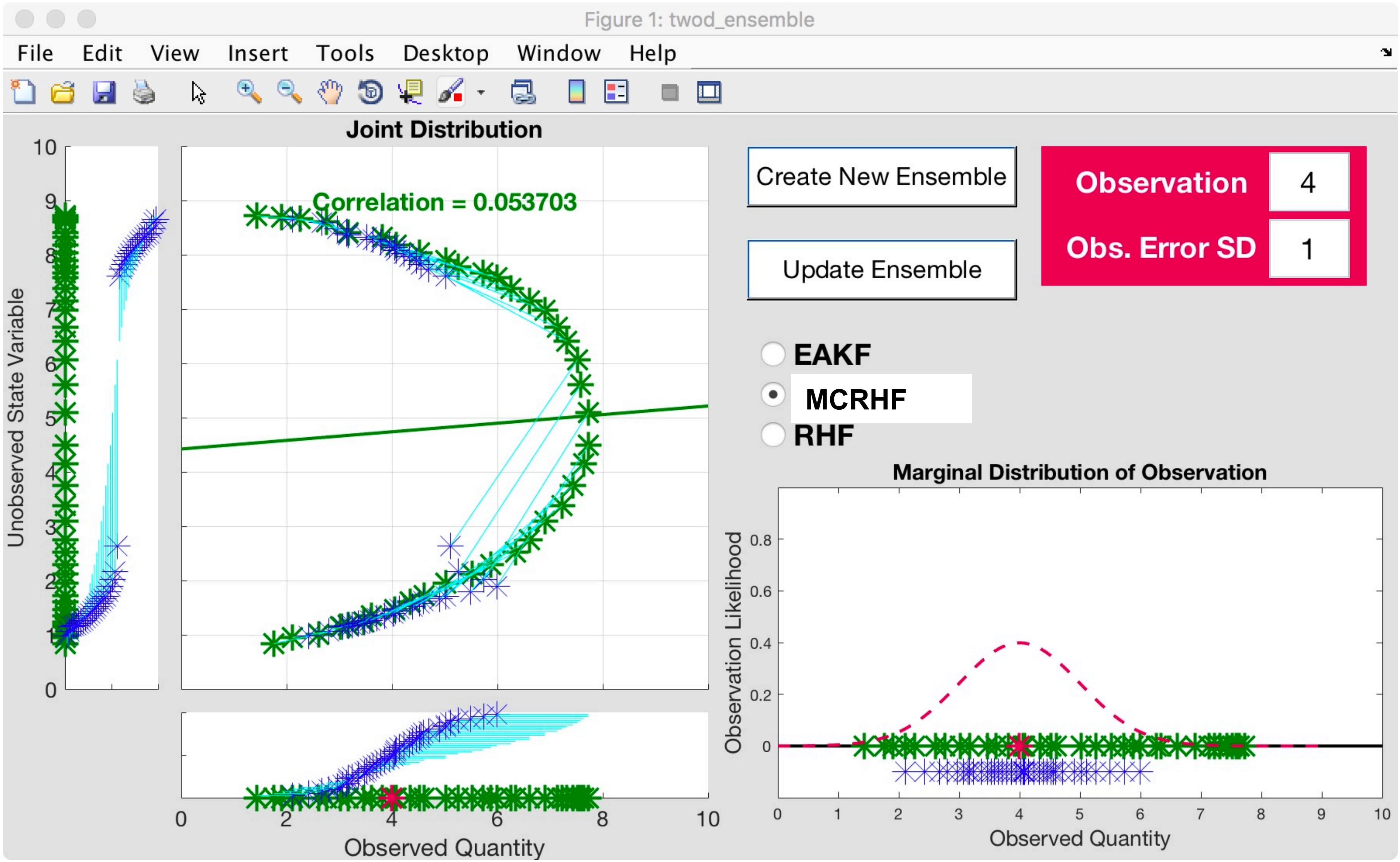
Marginal Correction Rank Histogram (MCRHF)



Step 5: Compute updated ensemble members:

- $(\text{ens\_size} + 1)^{-1}$ of posterior mass between each ensemble pair.
- $(\text{ens\_size} + 1)^{-1}$ in each tail.

Advection of Cosine Tracer: MCRHF



Details for Marginal Correction RHF method (MCRHF)

Do observation RHF with regression for preliminary posterior.

Get RHF State Marginal.

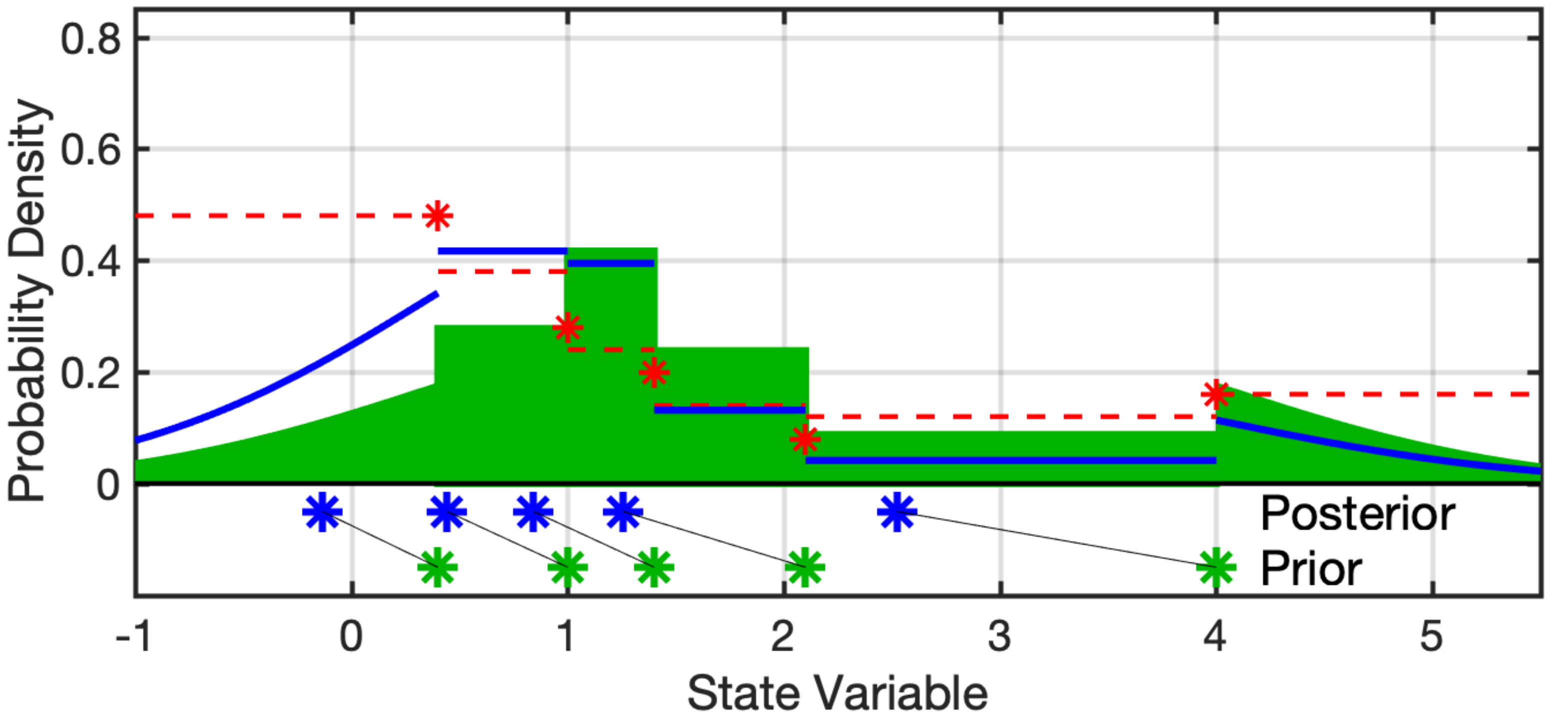
Rank statistics of posterior same as preliminary posterior.
Ensemble member with smallest preliminary posterior value gets smallest posterior value from RHF State Marginal.

Works well for many applications (but more expensive).

MCRHF Capabilities

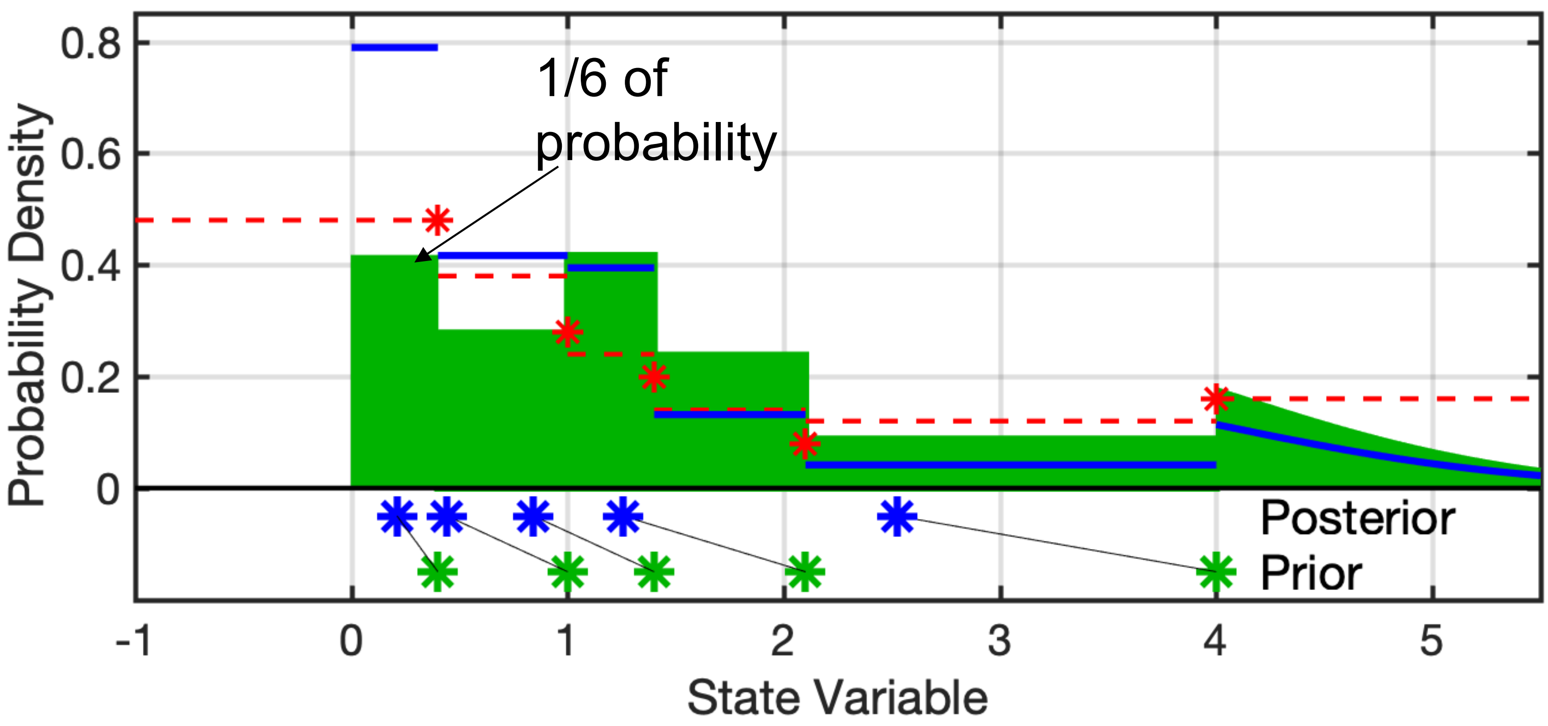
- Enforce additional prior constraints, like boundedness.
- Use arbitrary likelihoods.

Standard MCRHF State Marginal.



MCRHF with Bounded Prior

Bounded State Marginal, same ensemble but positive prior.



Bounded State, Non-Gaussian Likelihoods

Bivariate example.

Log of prior is bivariate Gaussian, so prior is non-negative.

One variable observed.

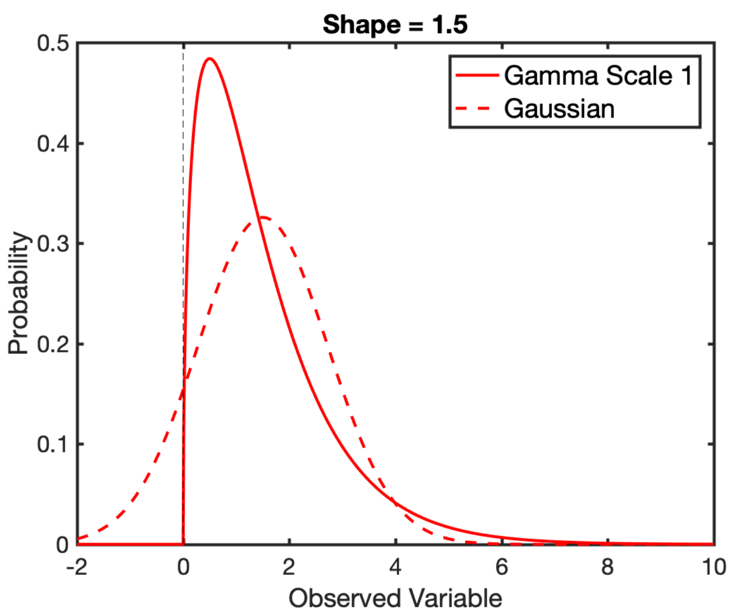
Likelihood is Gamma.

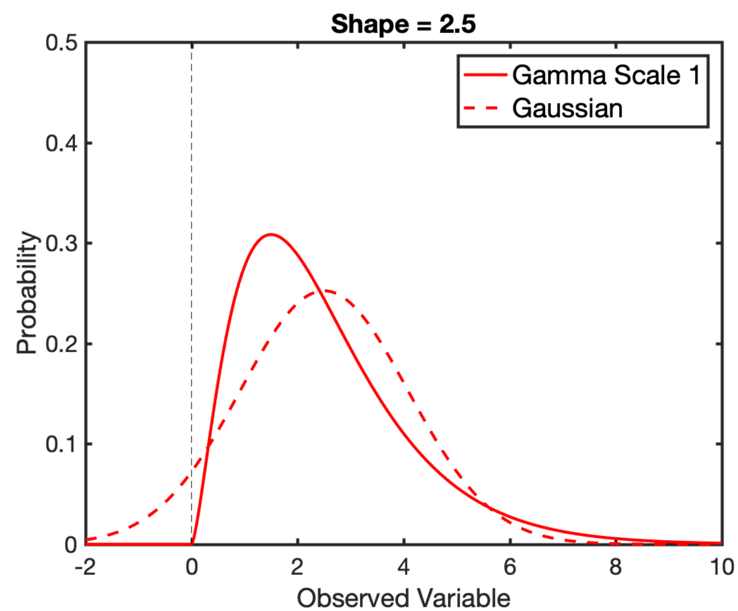
- Shape parameter is same as first prior ensemble.
- Scale parameter is 1.

Assimilate single observation for many random priors.

Bounded State, Non-Gaussian Likelihoods

Compare Gamma likelihood to Gaussian approximation.



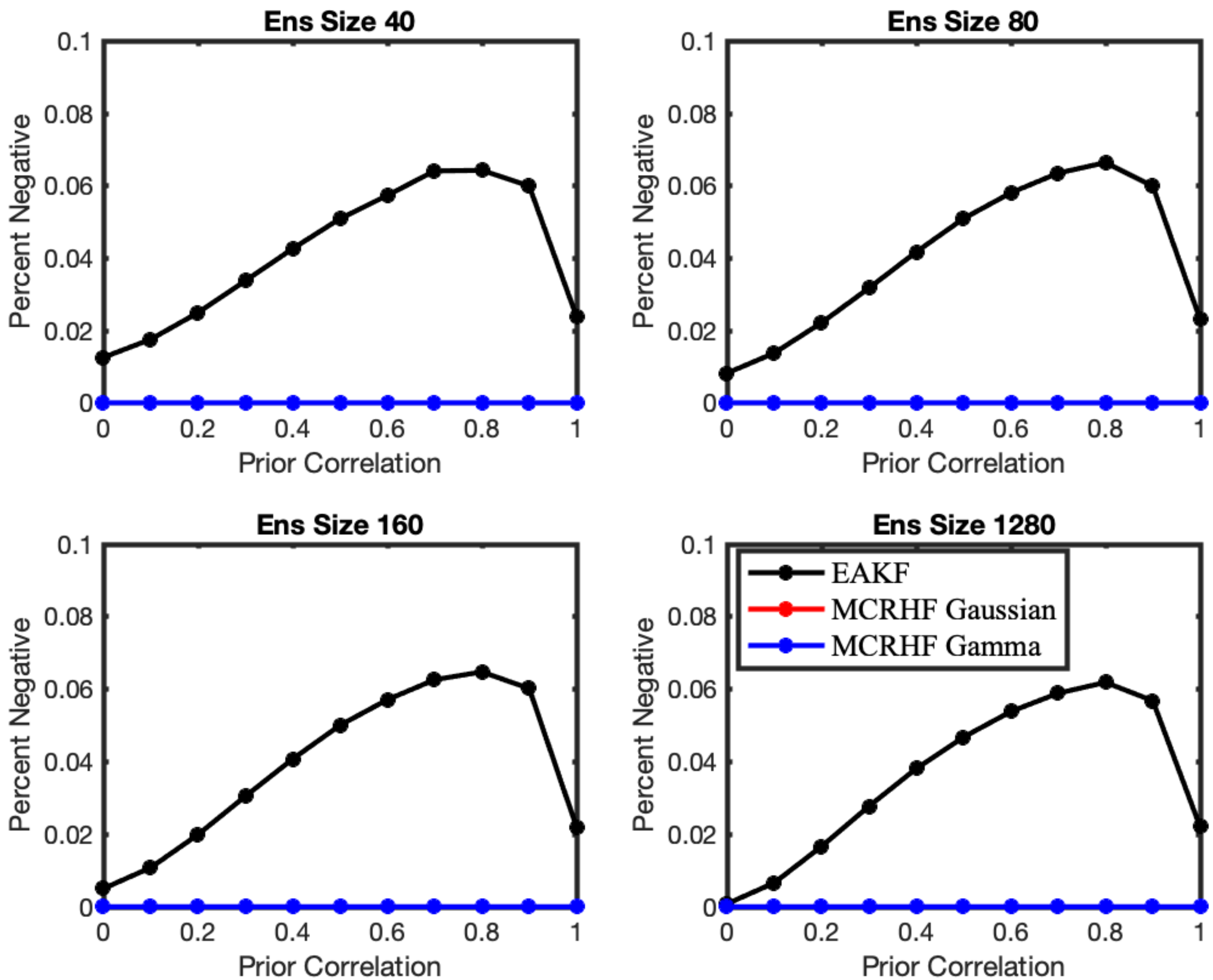


Bounded State, Non-Gaussian Likelihoods

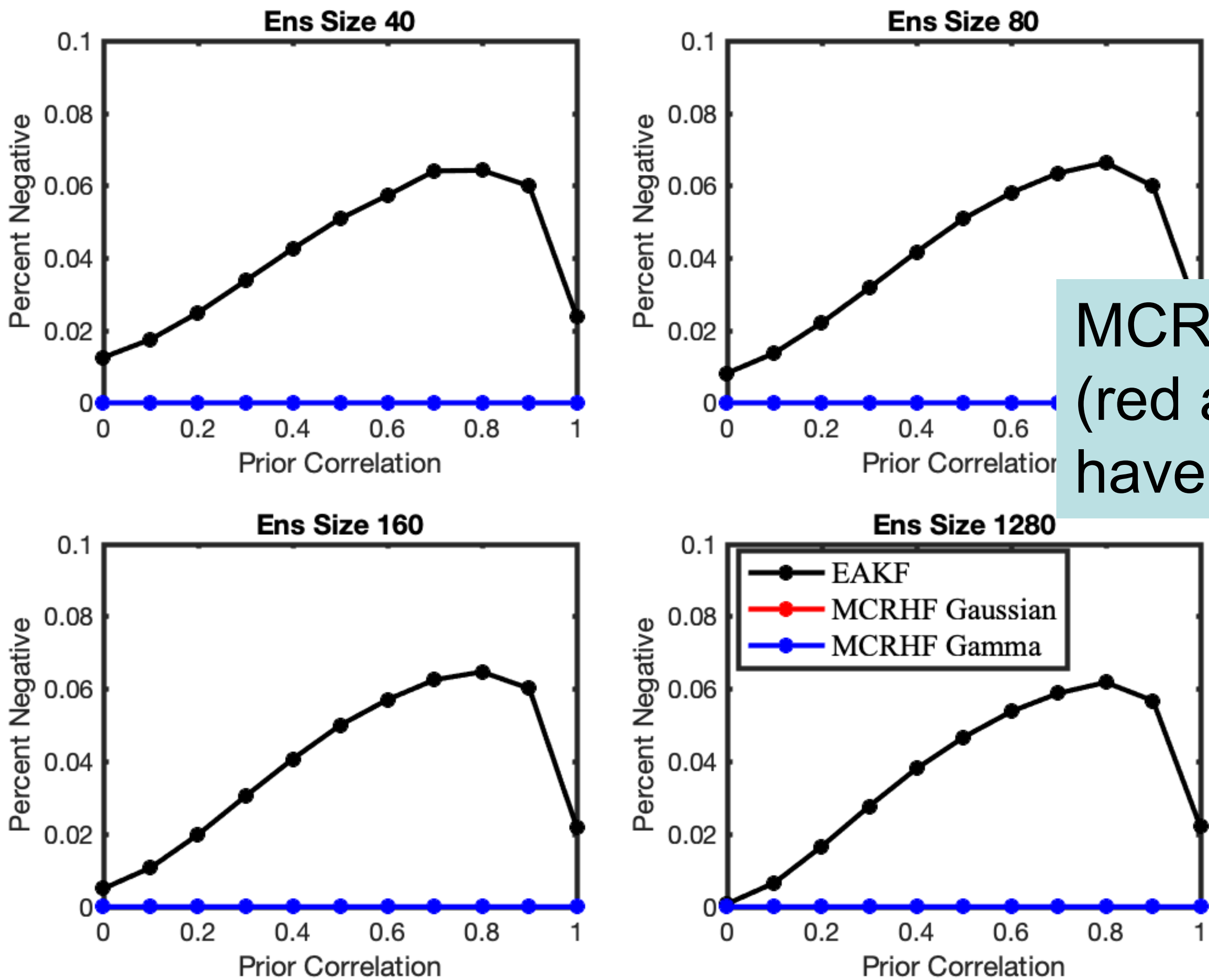
Compare 3 Methods, 4 Ensemble sizes

Observed Var.	Unobserved Var.	Likelihood
EAKF	Regression	Gaussian
RHF	MCRHF	Gaussian
RHF	MCRHF	Gamma

Percent Negative Posterior Members

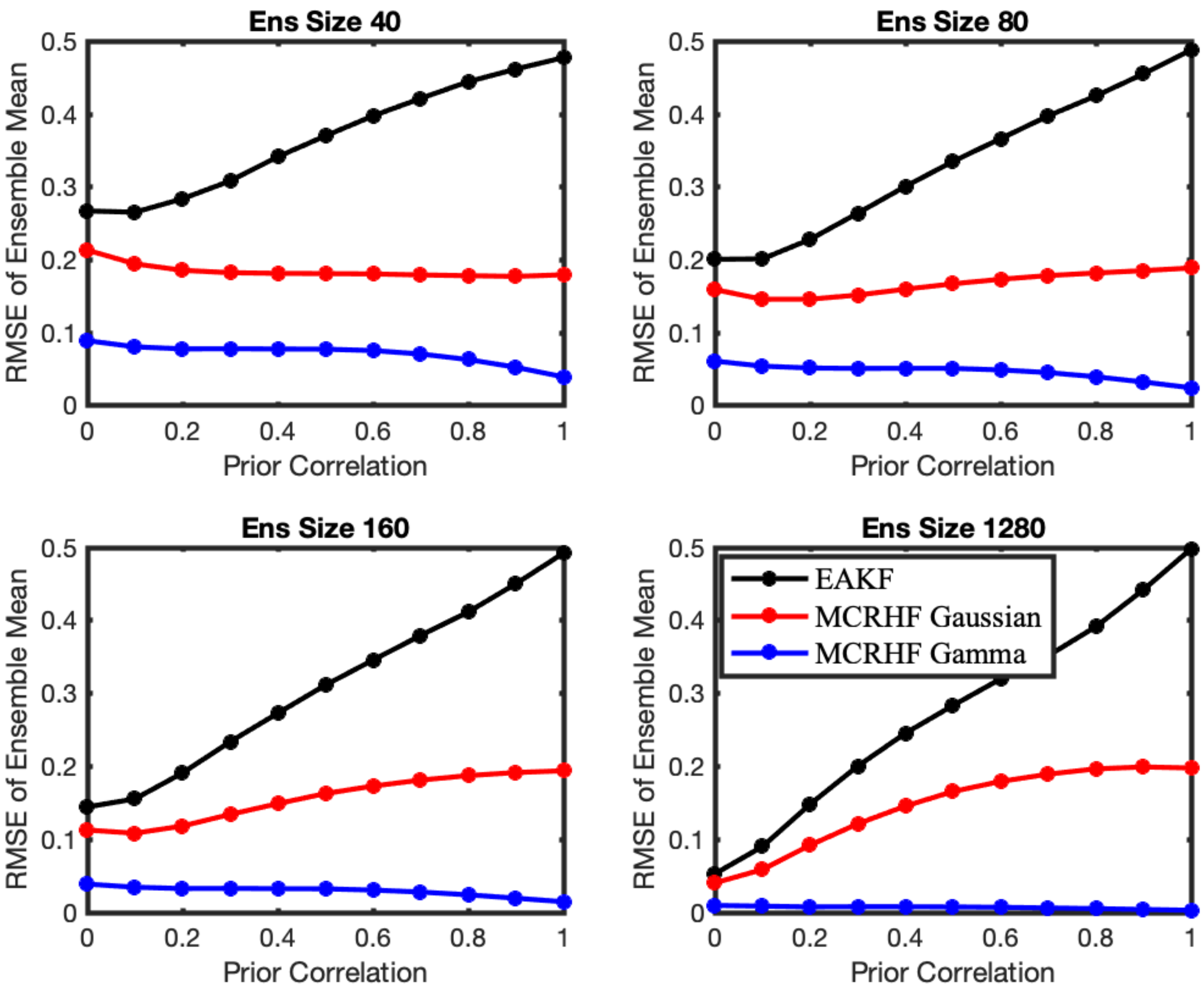


Percent Negative Posterior Members

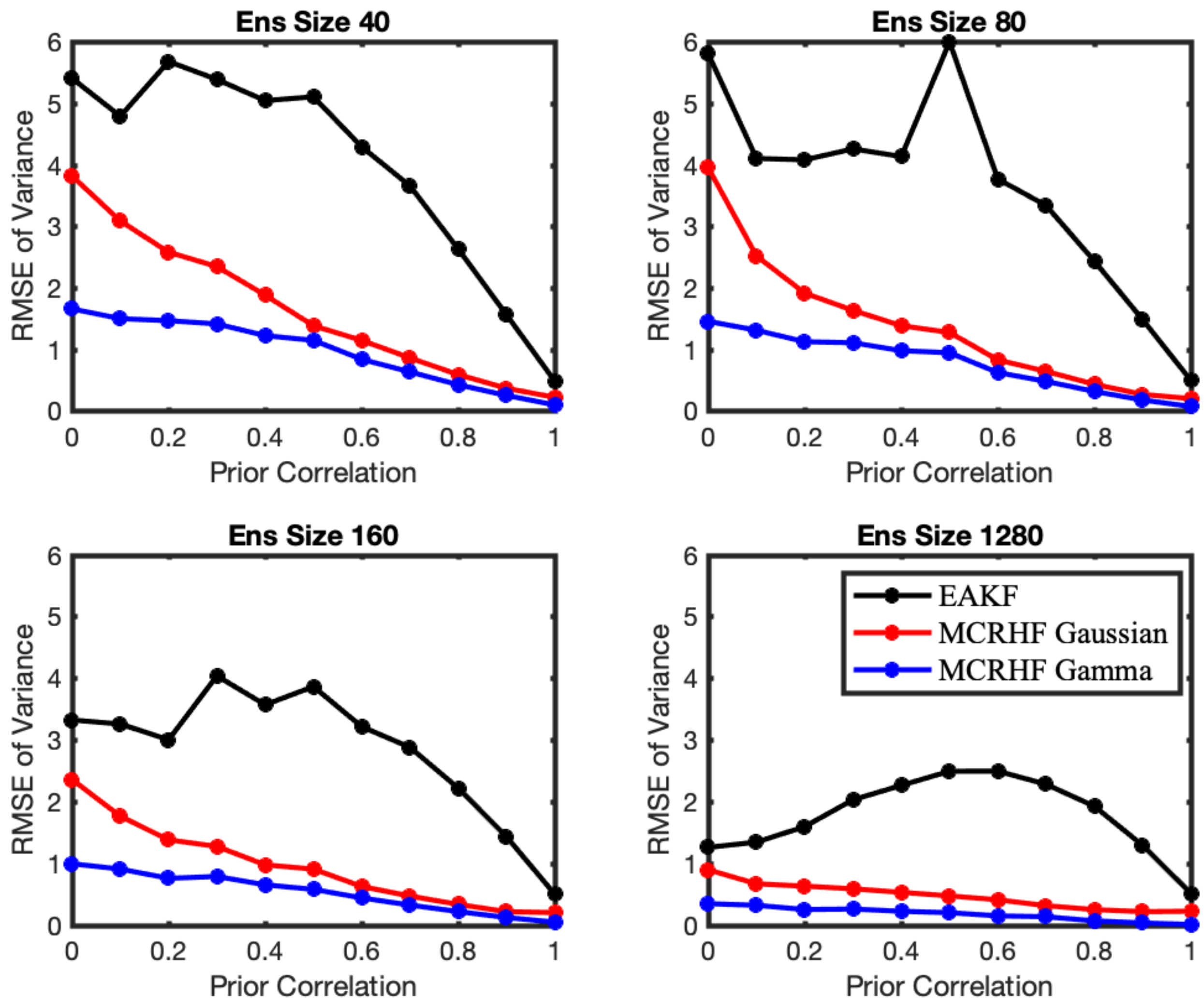


MCRHF both (red and blue) have 0 by design.

RMSE of Posterior Ensemble Mean



RMSE of Posterior Variance



Summary

RHF filters represent non-Gaussian priors, posteriors.

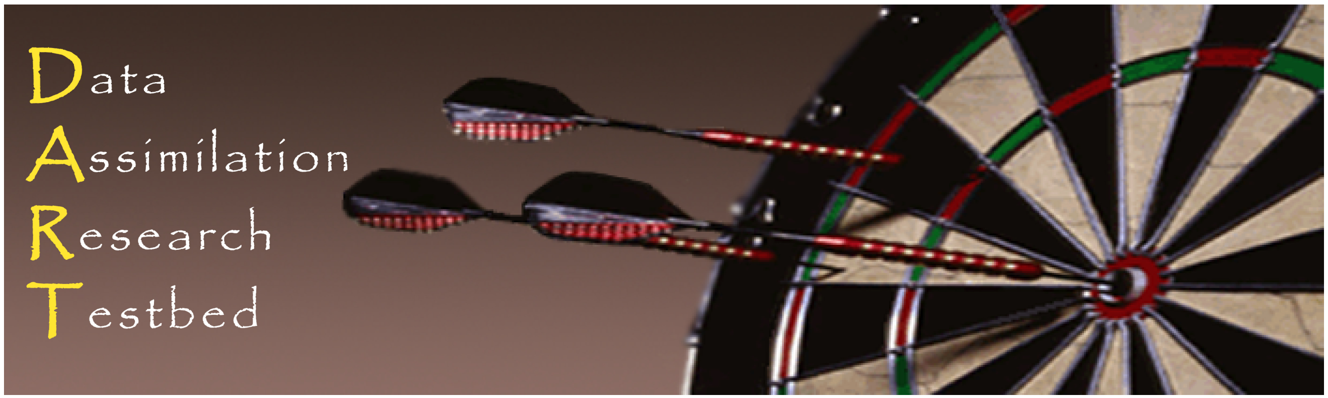
MCRHF allows non-Gaussian, limited non-linearity.

Particularly applicable to bounded quantities like tracers.

MCRHF more expensive, but less than factor of 2.

Ready to test in large applications like tracer transport.
Contact me if you'd like to collaborate.

All results here with DART_LAB tools freely available in DART.



https://dart.ucar.edu/

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