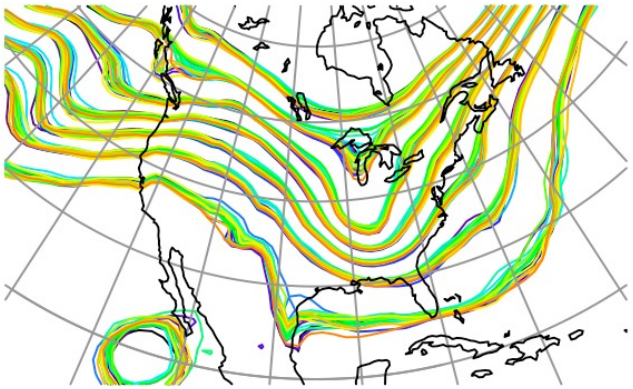
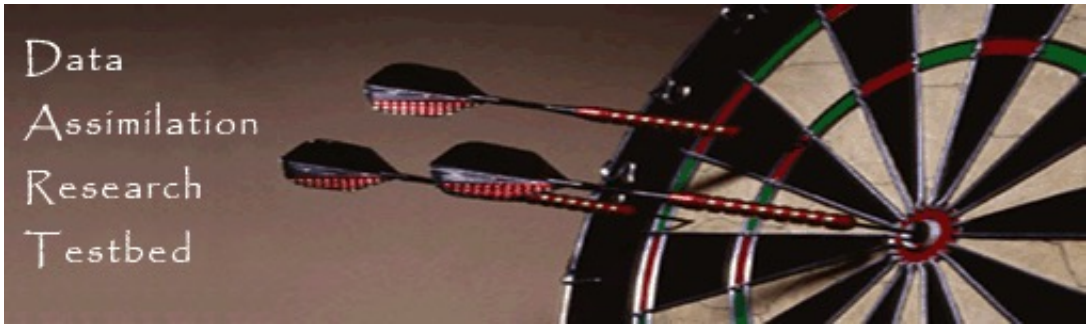


A Quantile Conserving Ensemble Filtering Framework for Non-Gaussian, Nonlinear Data Assimilation

Jeff Anderson, NCAR Data Assimilation Research Section

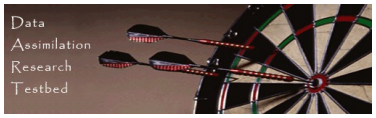


Schematic of a Sequential Ensemble Filter

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

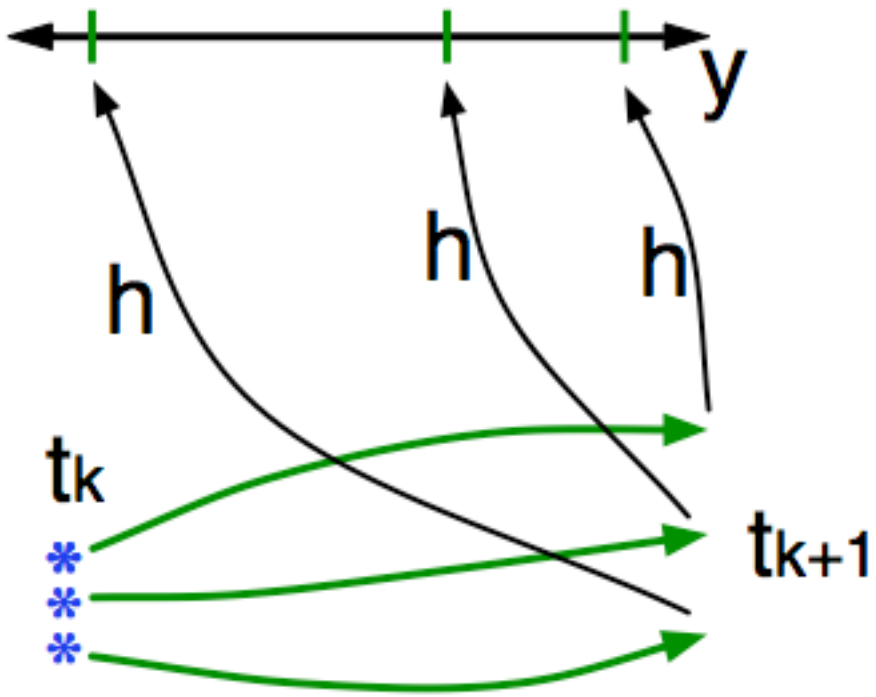
Ensemble state estimate after using previous observation (analysis)

Ensemble state at time of next observation (prior)



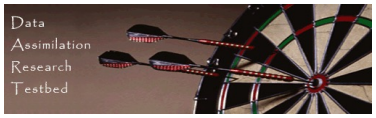
Schematic of a Sequential Ensemble Filter

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



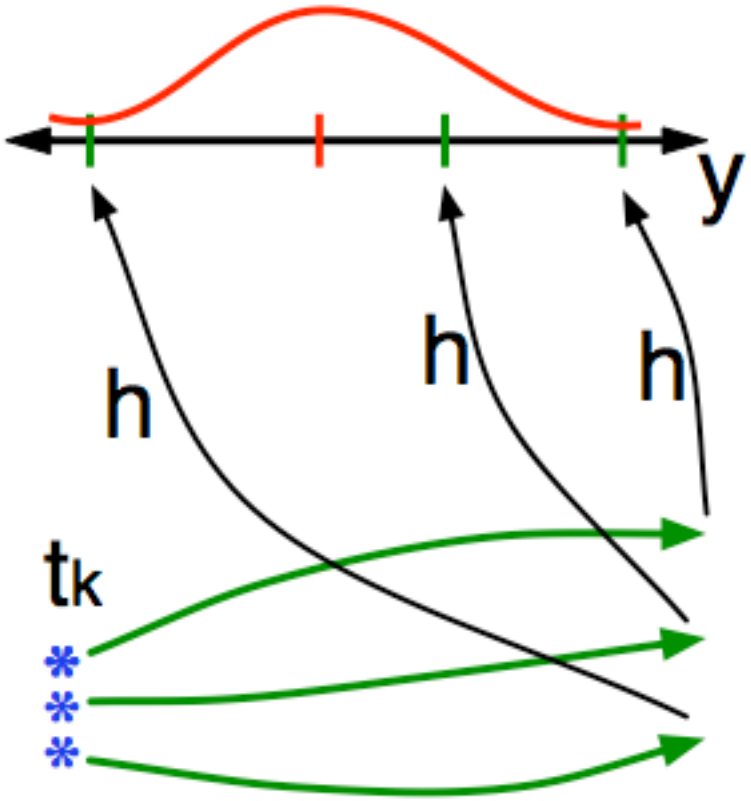
Theory: observations from instruments with uncorrelated errors can be done sequentially.

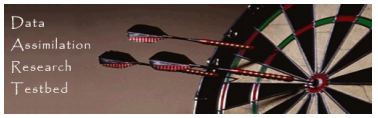
Can think about single observation without (too much) loss of generality.



Schematic of a Sequential Ensemble Filter

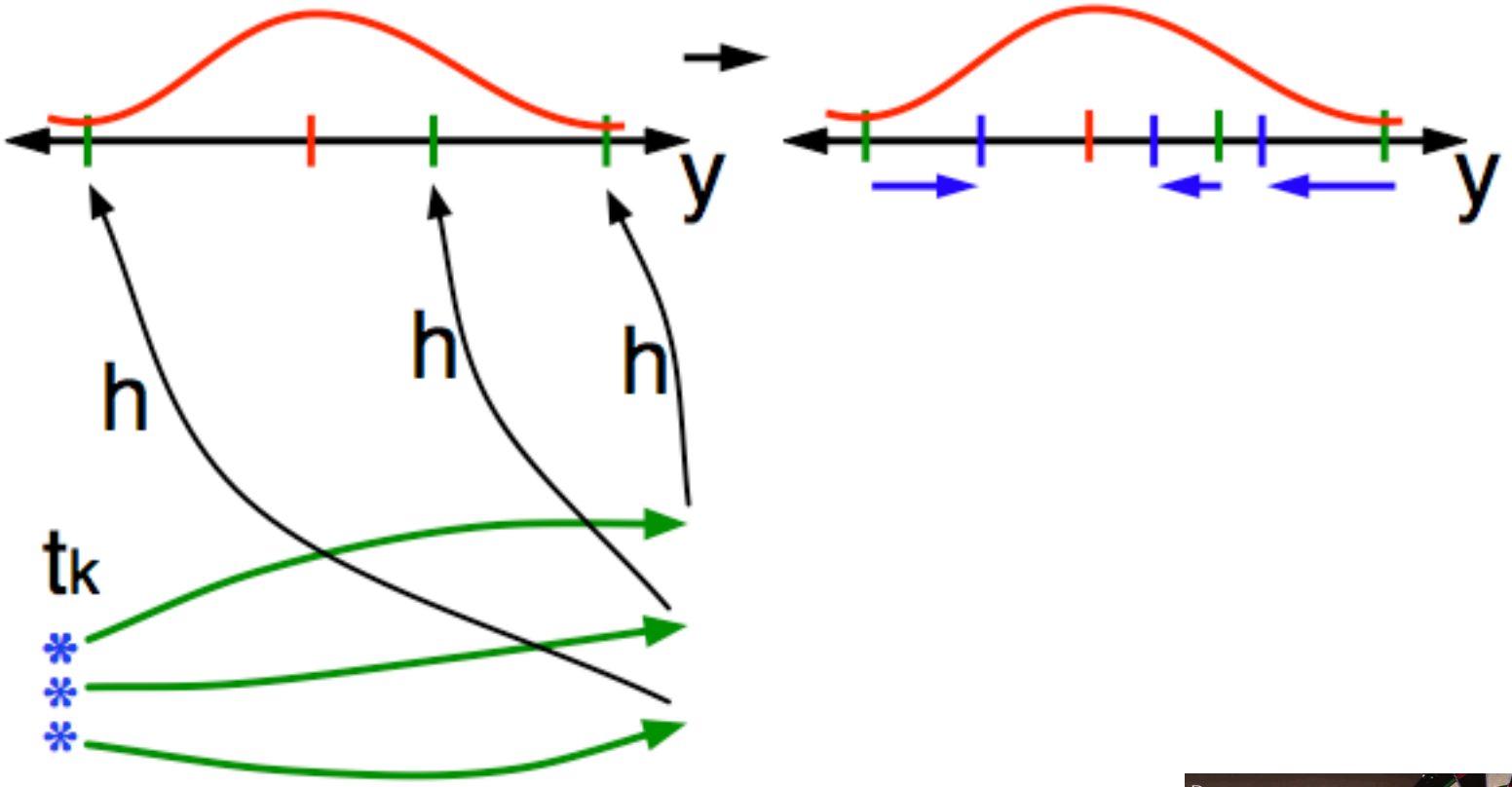
3. Get observed value and observational error distribution from observing system.

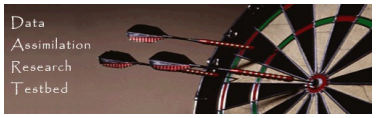




Schematic of a Sequential Ensemble Filter

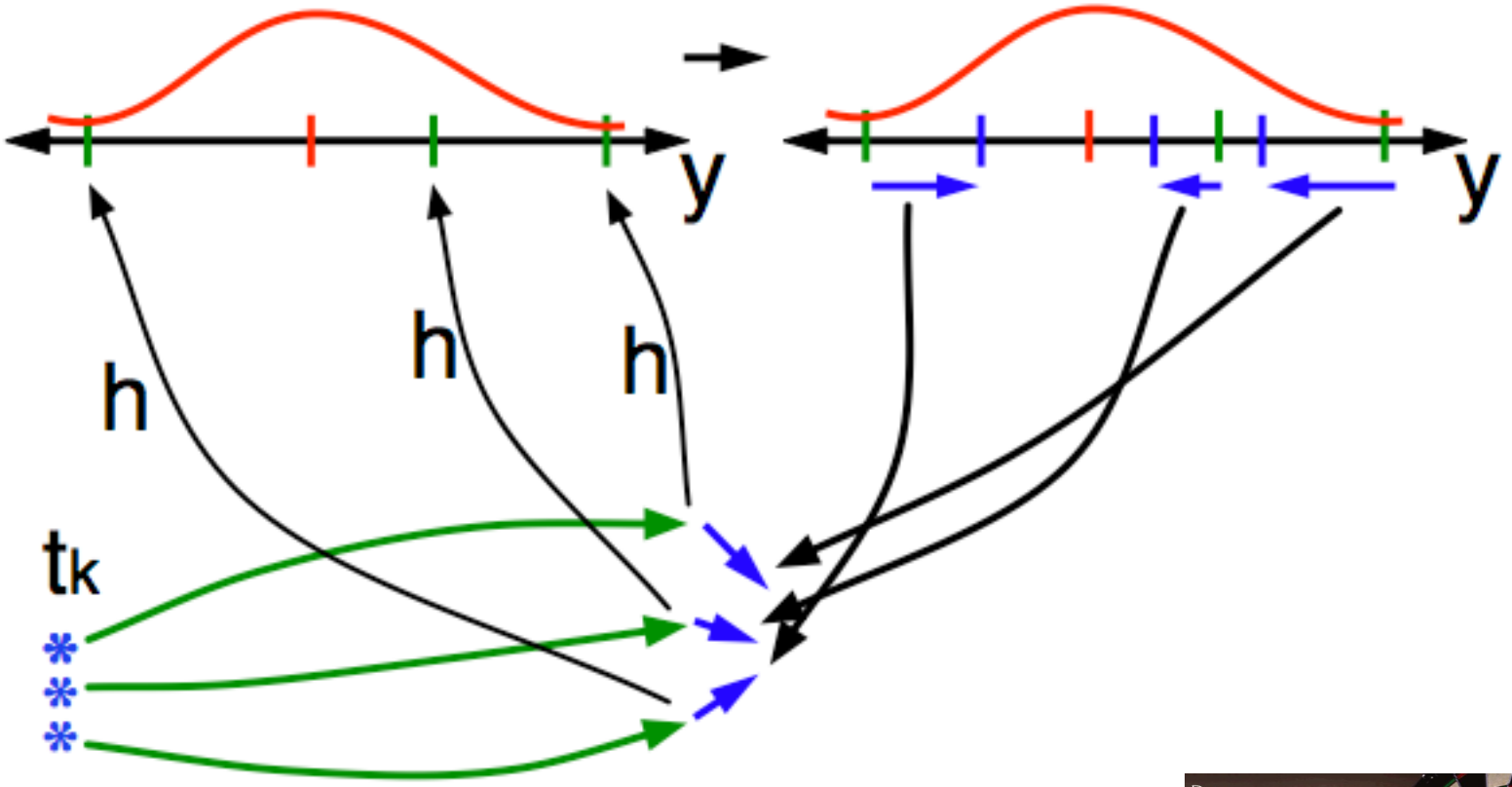
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).

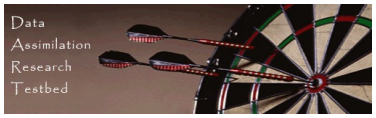




Schematic of a Sequential Ensemble Filter

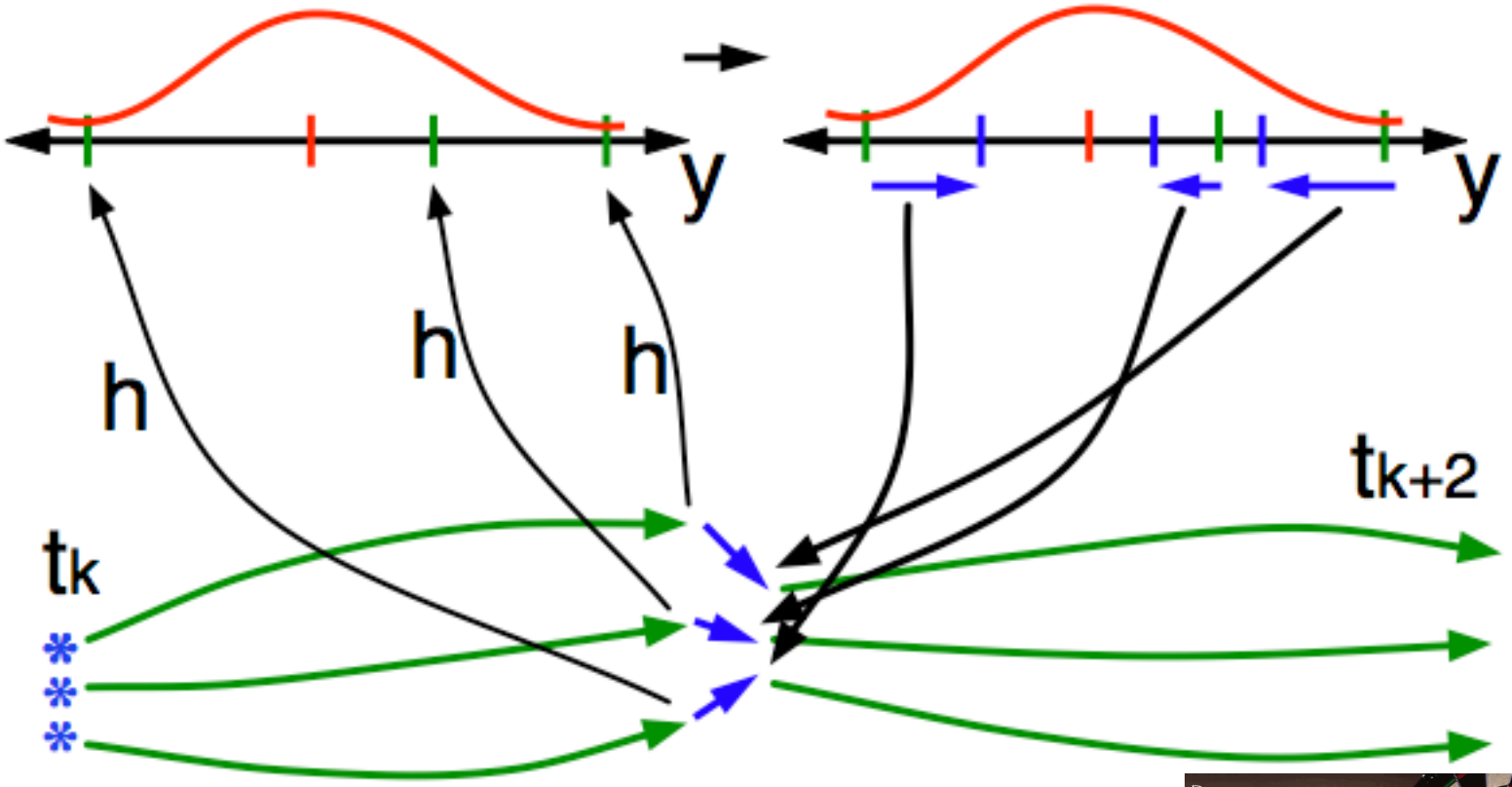
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.

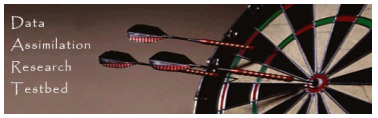




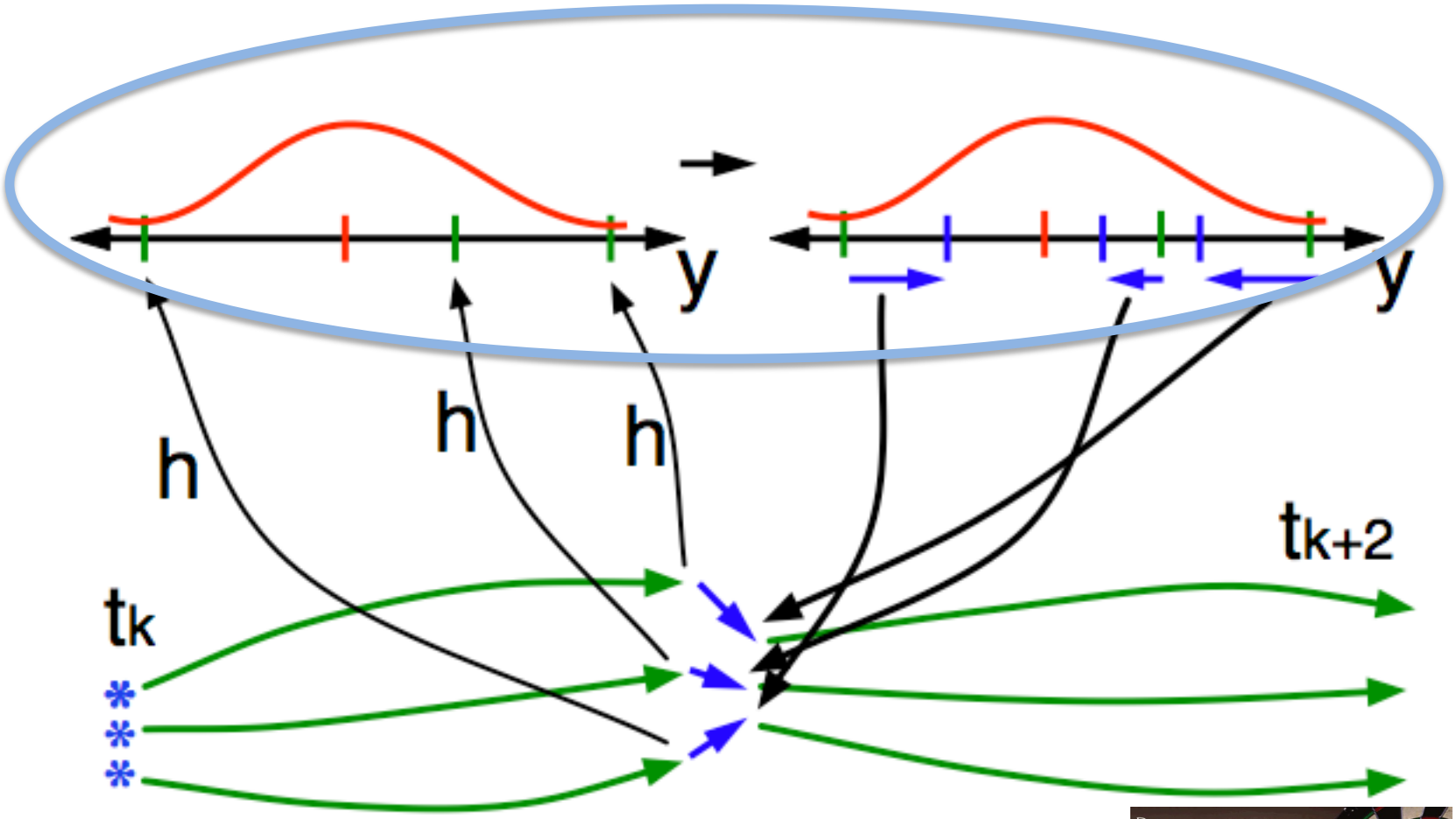
Schematic of a Sequential Ensemble Filter

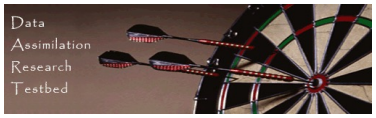
6. When all ensemble members for each state variable are updated, integrate to time of next observation …



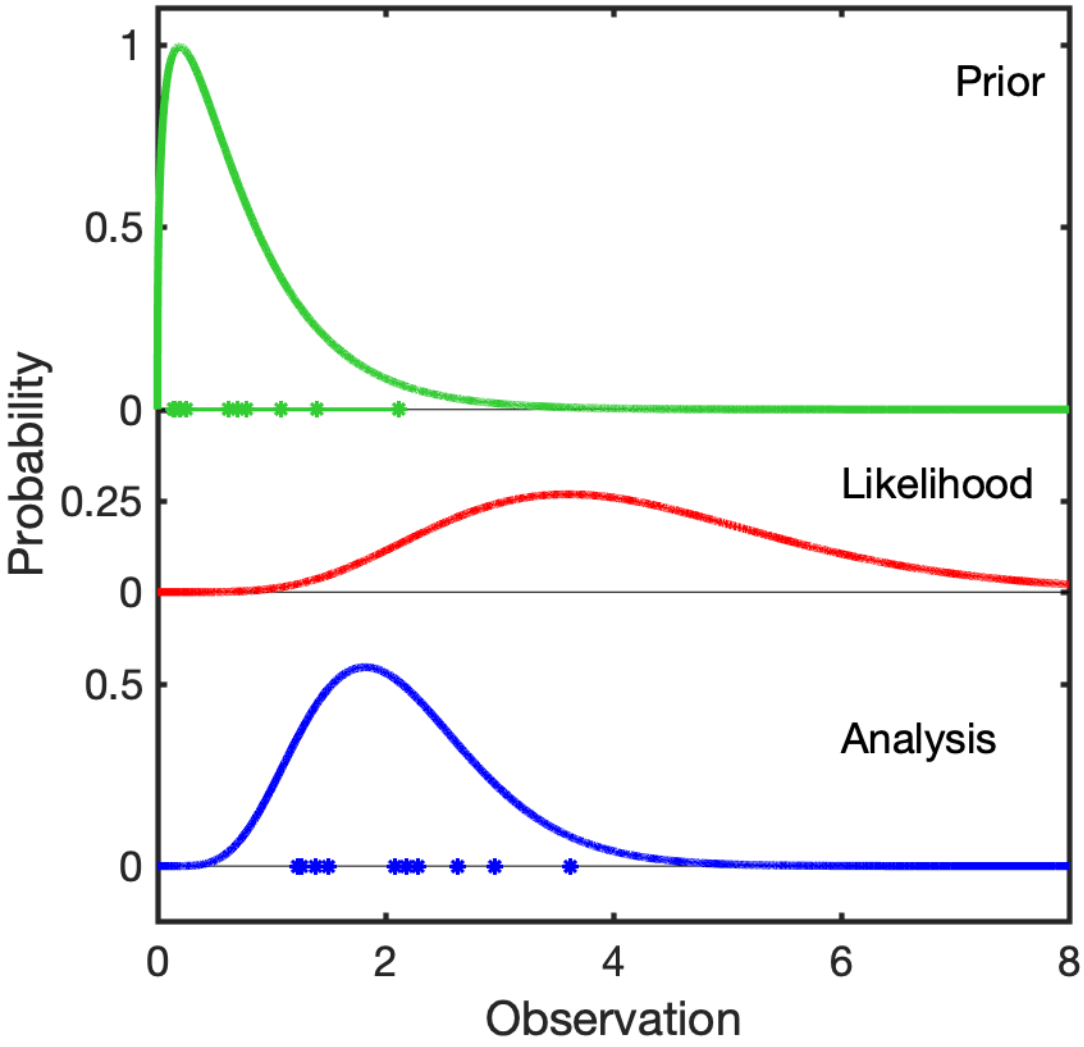


DART now provides nearly general solutions for this step: (Anderson, 2022, MWR*150*, 1061-1074).





Example: Gamma prior, Gamma Likelihood



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

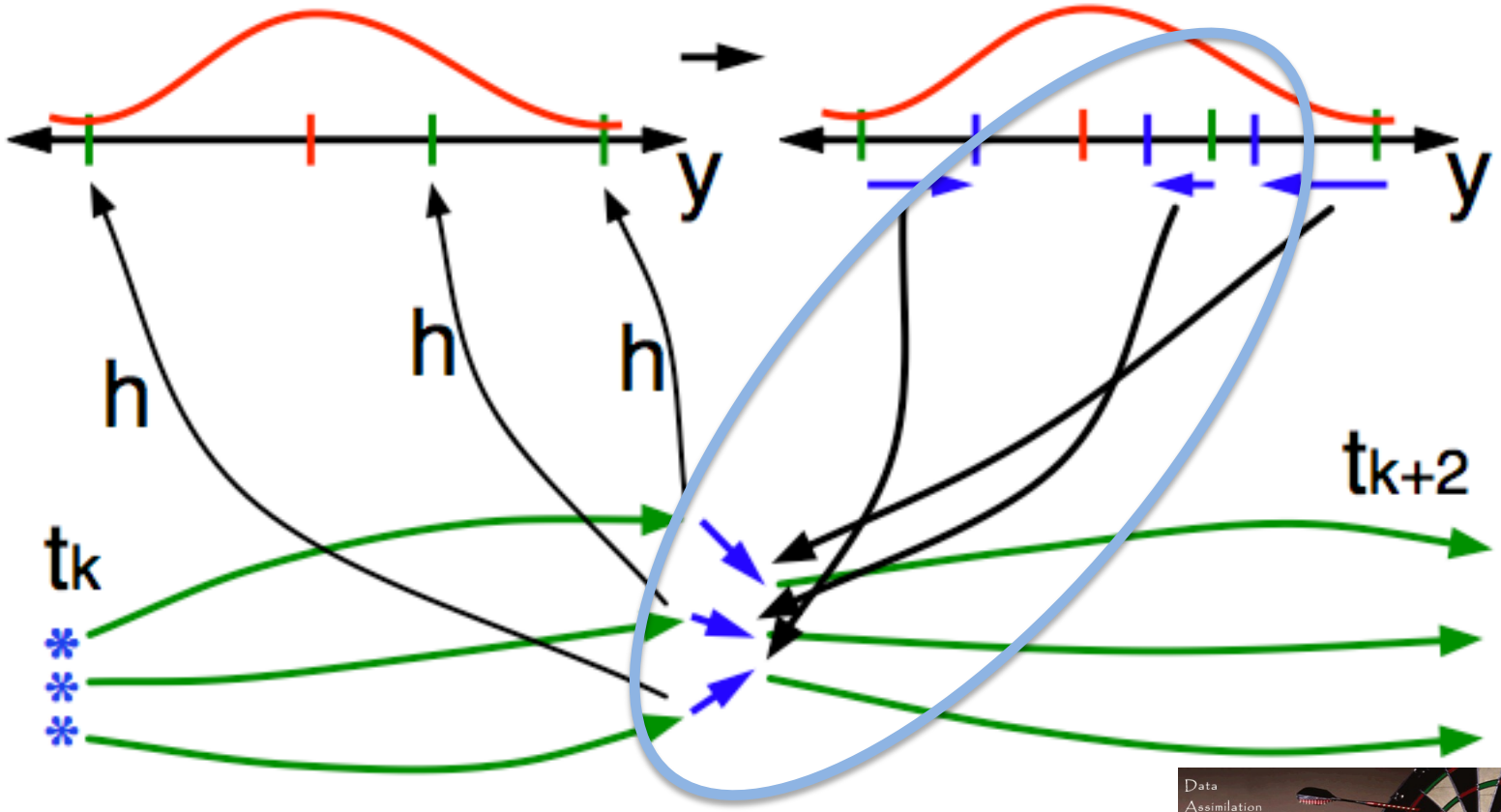
Gamma prior enforces non-negativity.

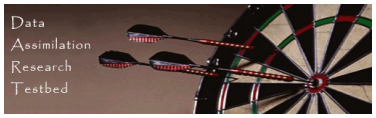
Gamma likelihood leads to gamma posterior.

Regression in a Transformed Space

Linear regression can destroy benefits of new observation method.

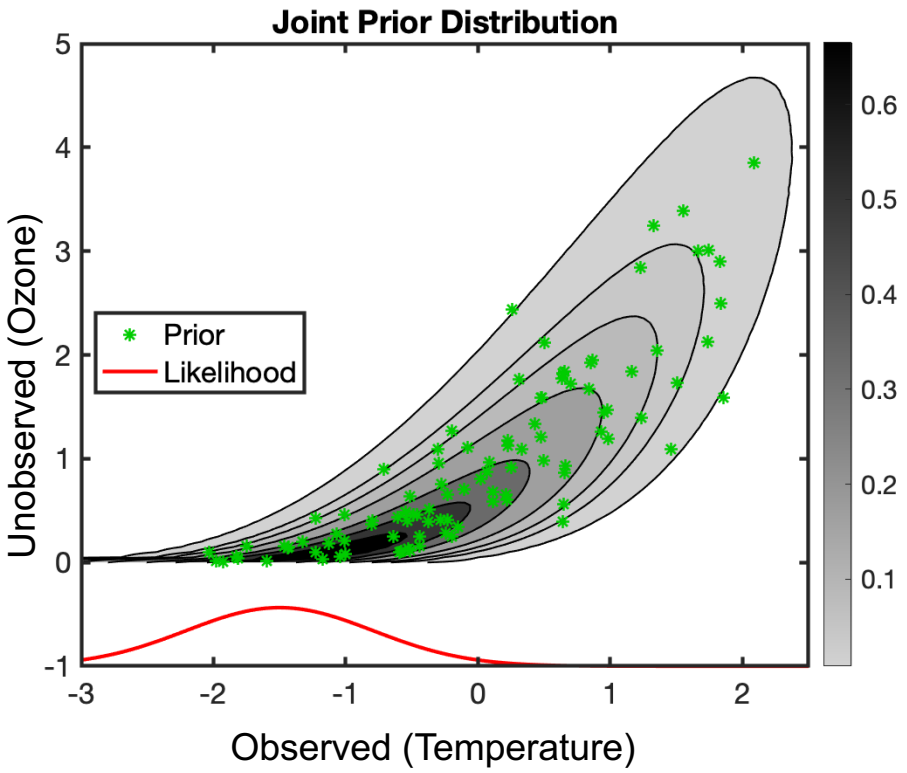
This talk focuses on updating unobserved variables with regression in a transformed space that extends the benefits to all state variables.





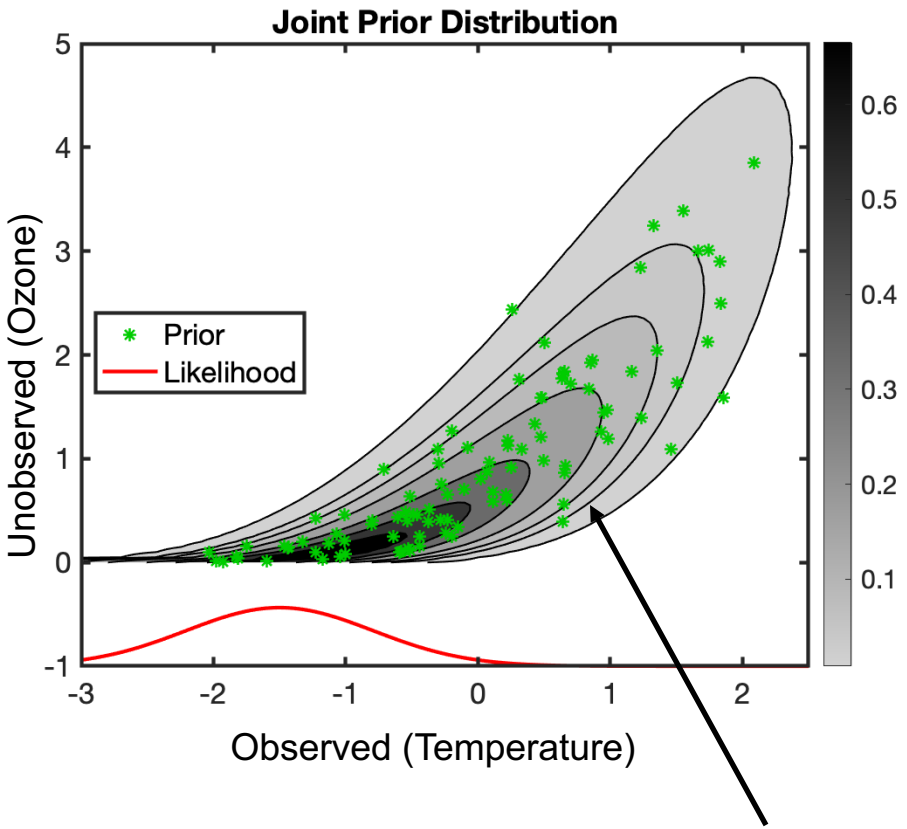
Standard EnKF: Challenged by Non-Gaussian and Nonlinear Relations

Prior for normal-gamma distribution with 100 member ensemble.



Standard EnKF: Challenged by Non-Gaussian and Nonlinear Relations

Prior for normal-gamma distribution with 100 member ensemble.

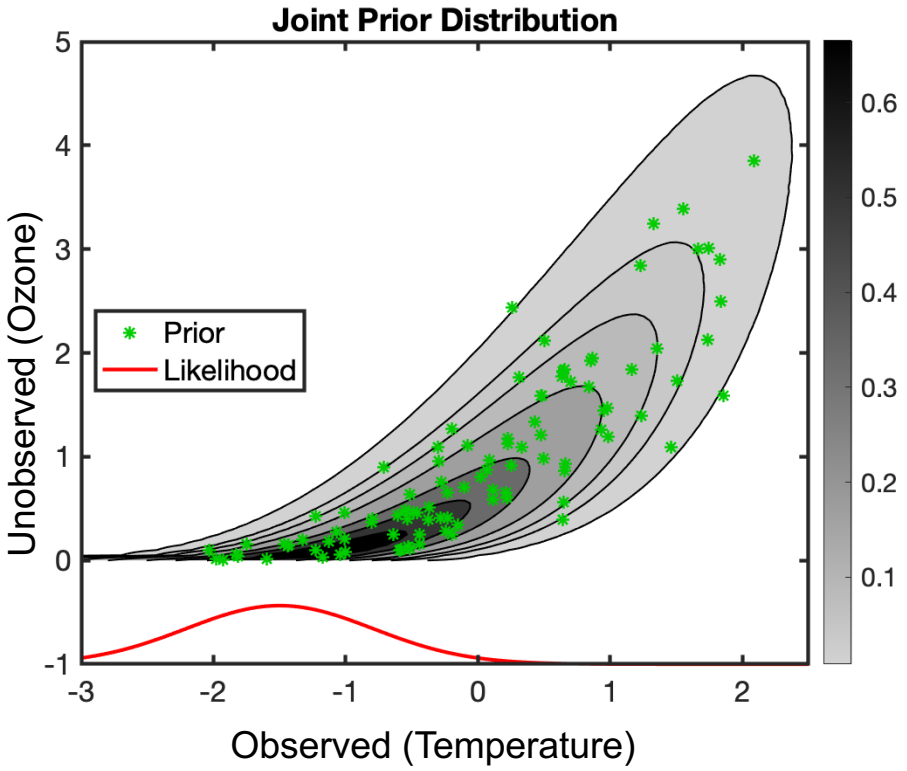


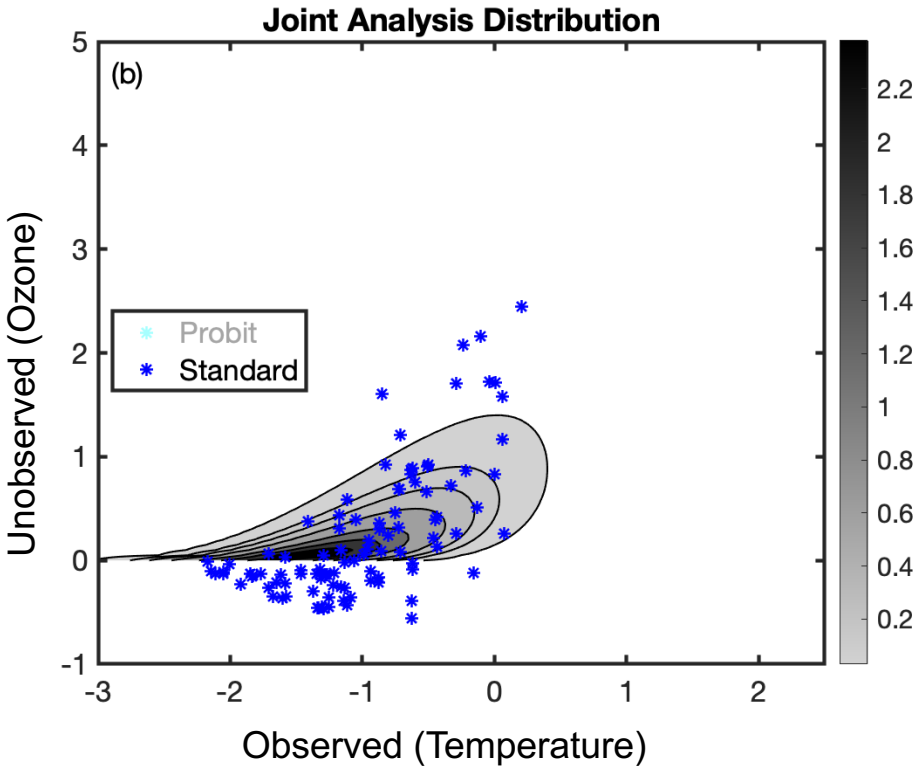
Correct distribution contours are 1, 5, 10, 20, 40, 60, 80% of max for all figures.

Standard EnKF: Challenged by Non-Gaussian and Nonlinear Relations

Prior for normal-gamma distribution with 100 member ensemble.

Posterior ensemble has problems.



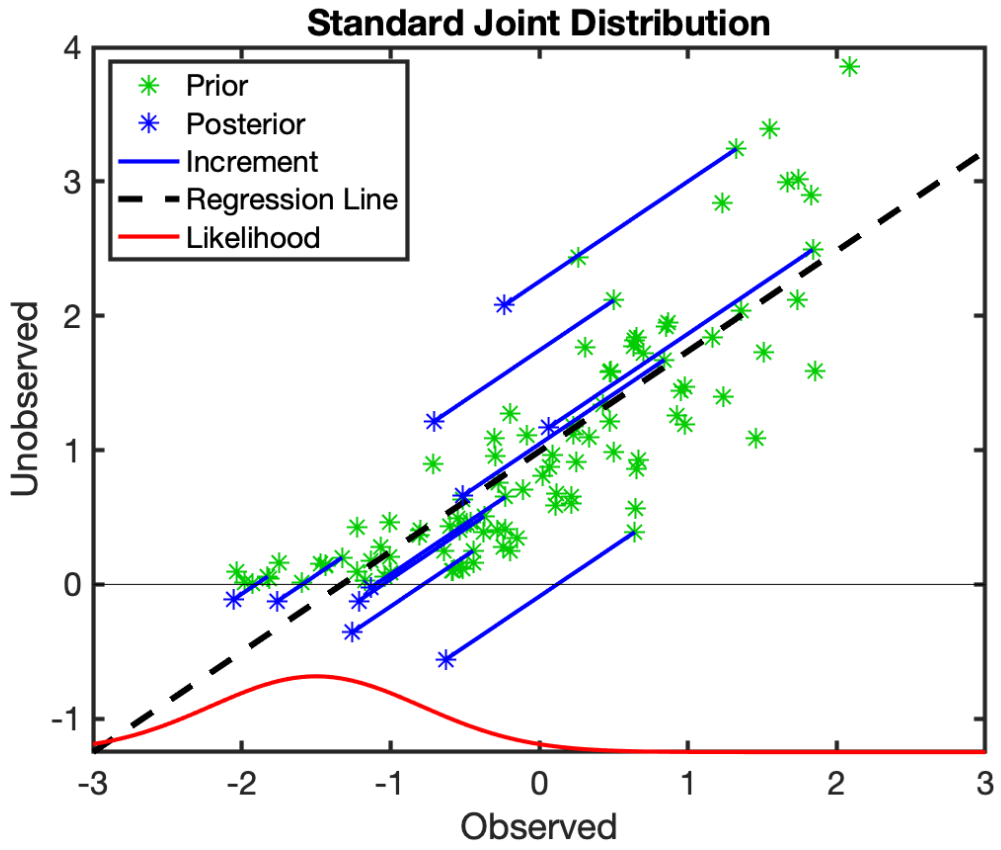


Problems with Linear Regression of Increments

Example regression increment vectors:

Don't respect bounds,

Struggle with nonlinearity.



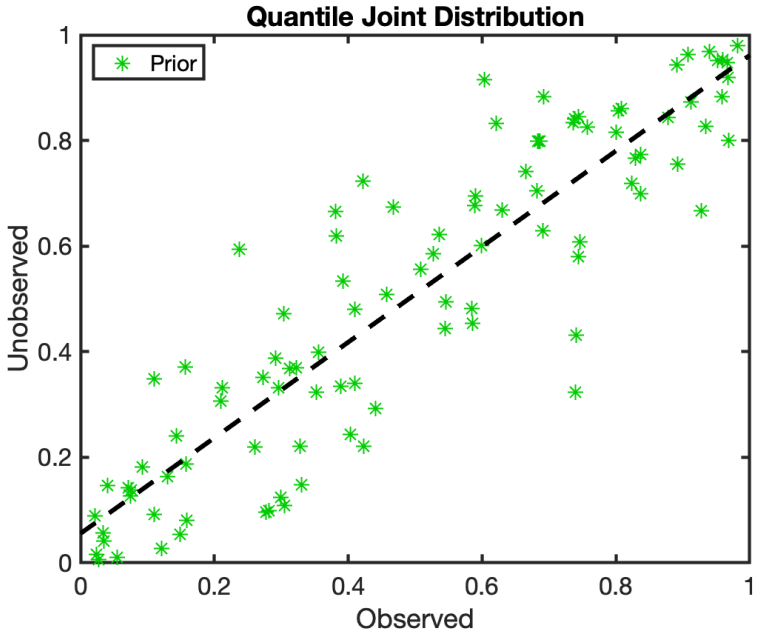
Solution, Transform Marginals: Step 1: Compute Quantiles

Pick an appropriate continuous prior distribution.

Compute CDF for each ensemble member to get quantiles.

Distribution of quantiles is U(0, 1) for appropriate prior.

This is the probability integral transform.

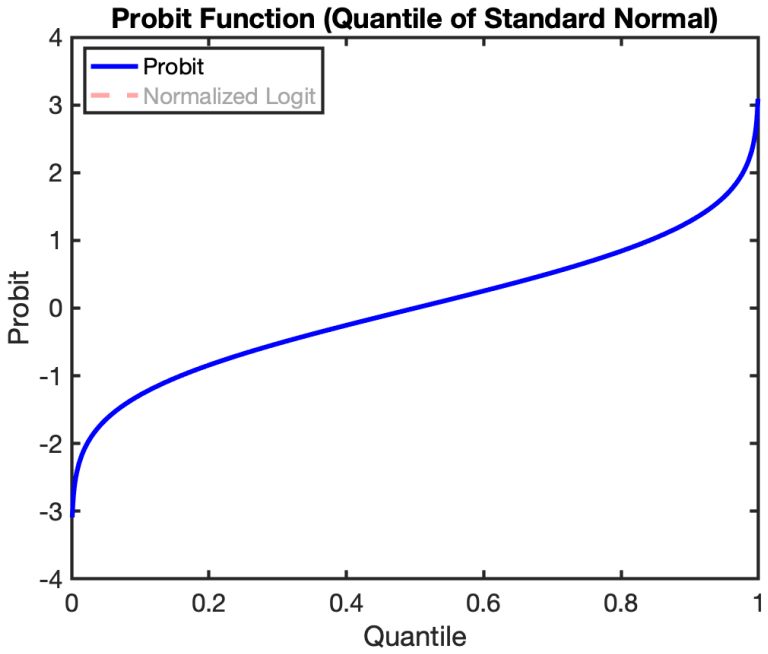


Solution, Transform Marginals: Step 2: Probit Transform of Quantiles

Probit function, the quantile function for the standard Normal.

Transforms U(0, 1) to unbounded.

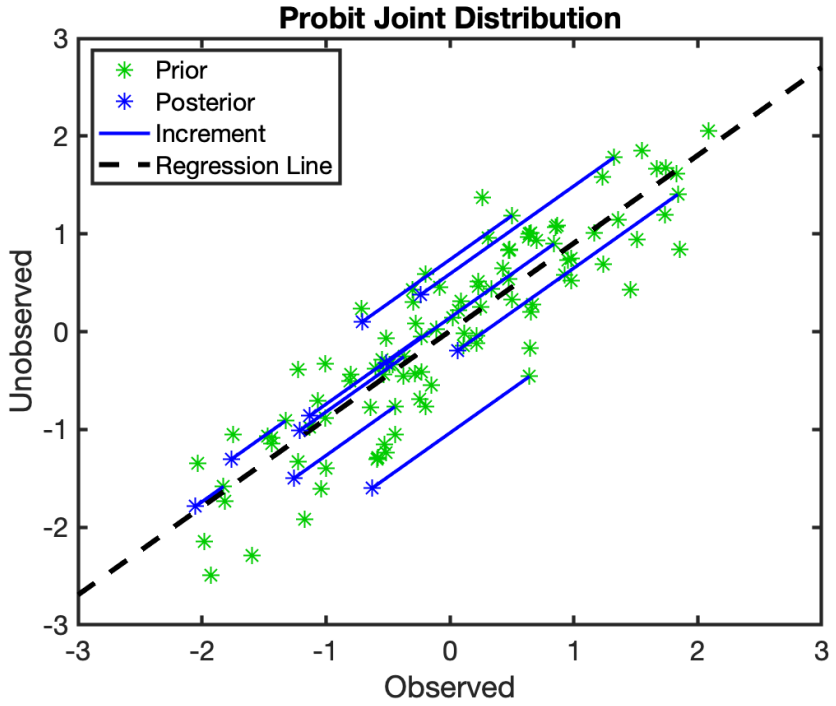
Marginal distributions should be N(0, 1).



Regression in Probit-Transformed Quantile Space

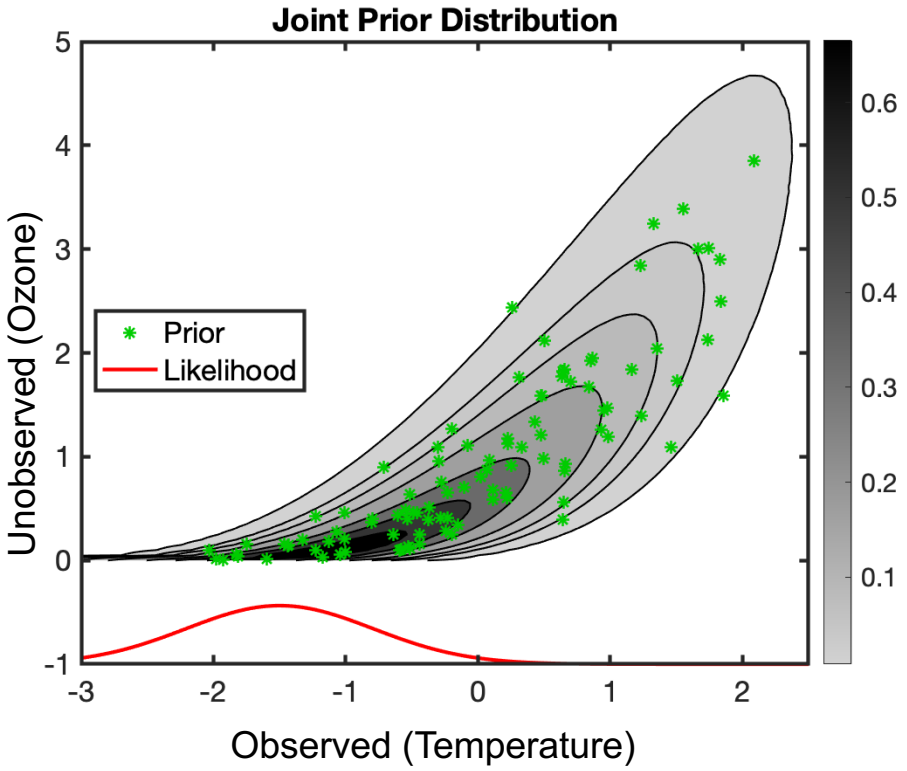
Do the regression of the observed probit increments onto the unobserved probit ensemble.

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

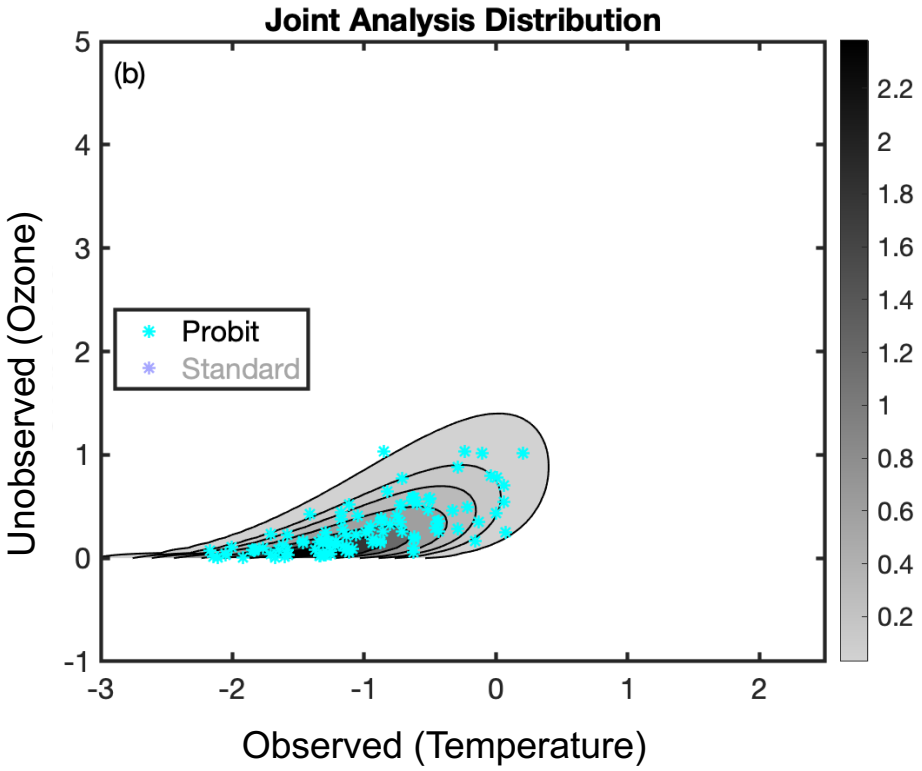


DART: Novel, General Solutions for Nonlinear, Non-Gaussian Problems

Prior for normal-gamma distribution with 100 member ensemble.



Bounds enforced. Nonlinear aspect respected.



Probit-Transformed Quantile Regression Algorithm

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

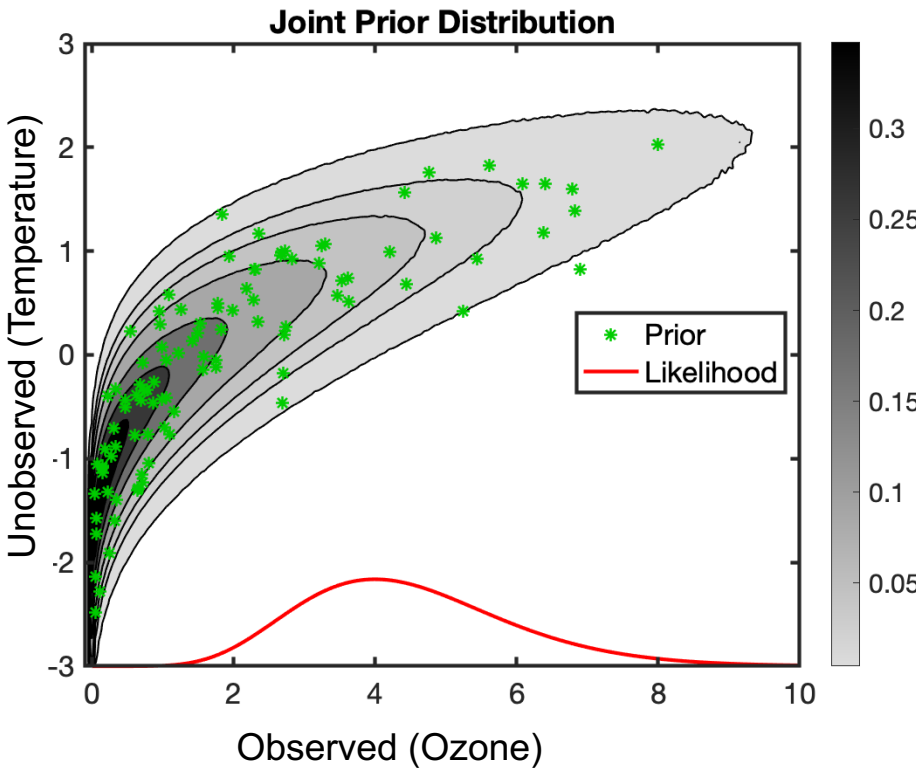
$\Delta\tilde{x}_n = \frac{\tilde{\sigma}_{x,y}}{\tilde{\sigma}_{y,y}}\Delta\tilde{y}_n$ regress increments in probit space (eq. 5 Anderson 2003)

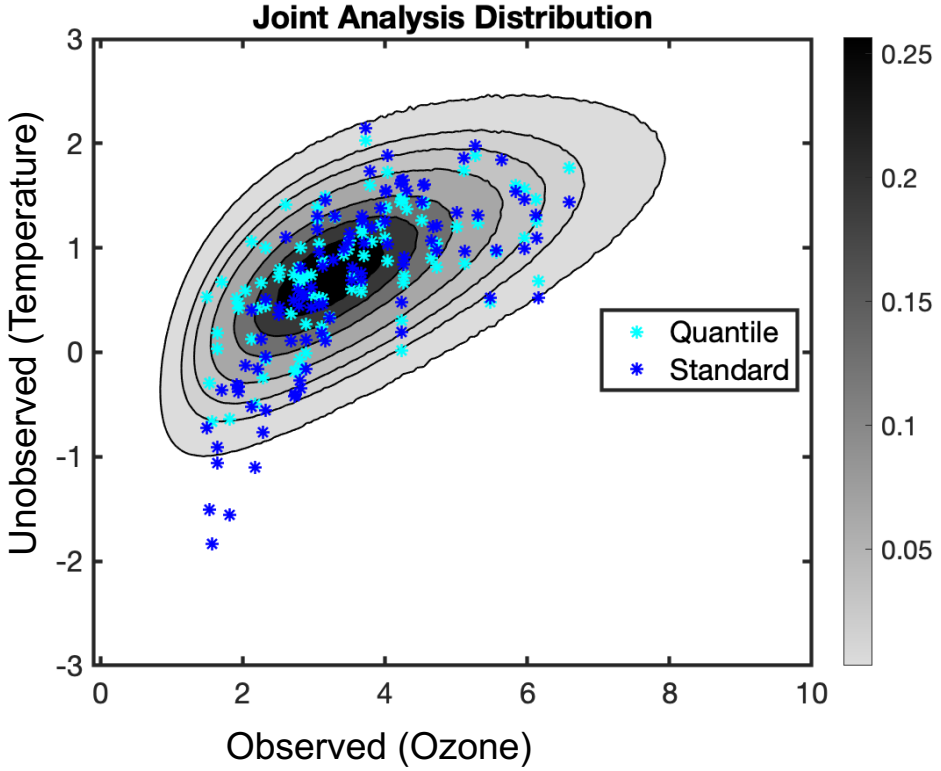
$\tilde{x}_n^a = \tilde{x}_n^p + \Delta\tilde{x}_n$ is posterior ensemble in probit space

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

Example 4: Gamma observed, normal unobserved

No bounds issues, new methods handle curvature.
Could improve impact of tracer obs on meteorology.

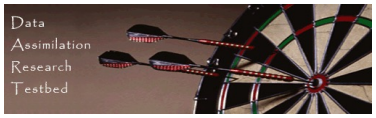




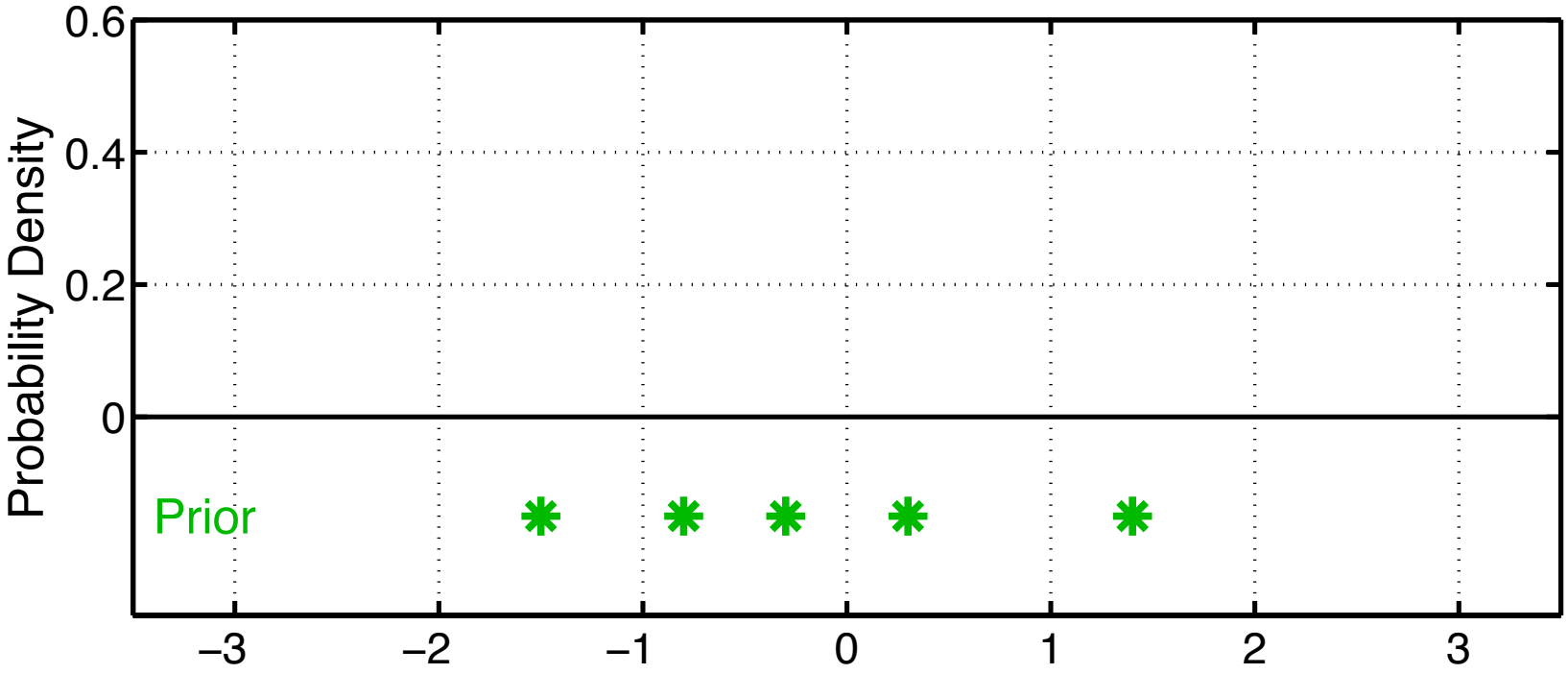
But, we may not know the right distribution family.

Can use a non-parametric continuous prior.

Rank histogram piecewise constant distribution.

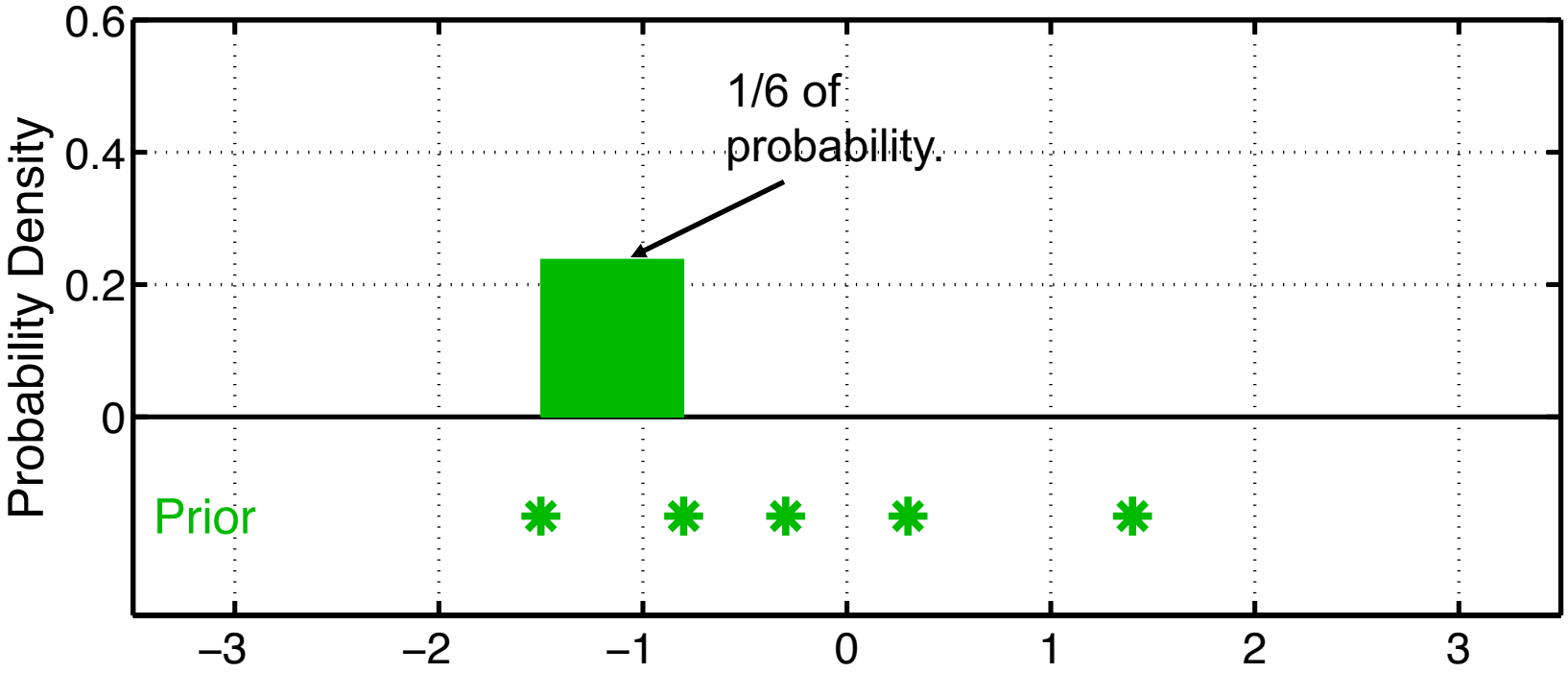


Rank Histogram Continuous Prior



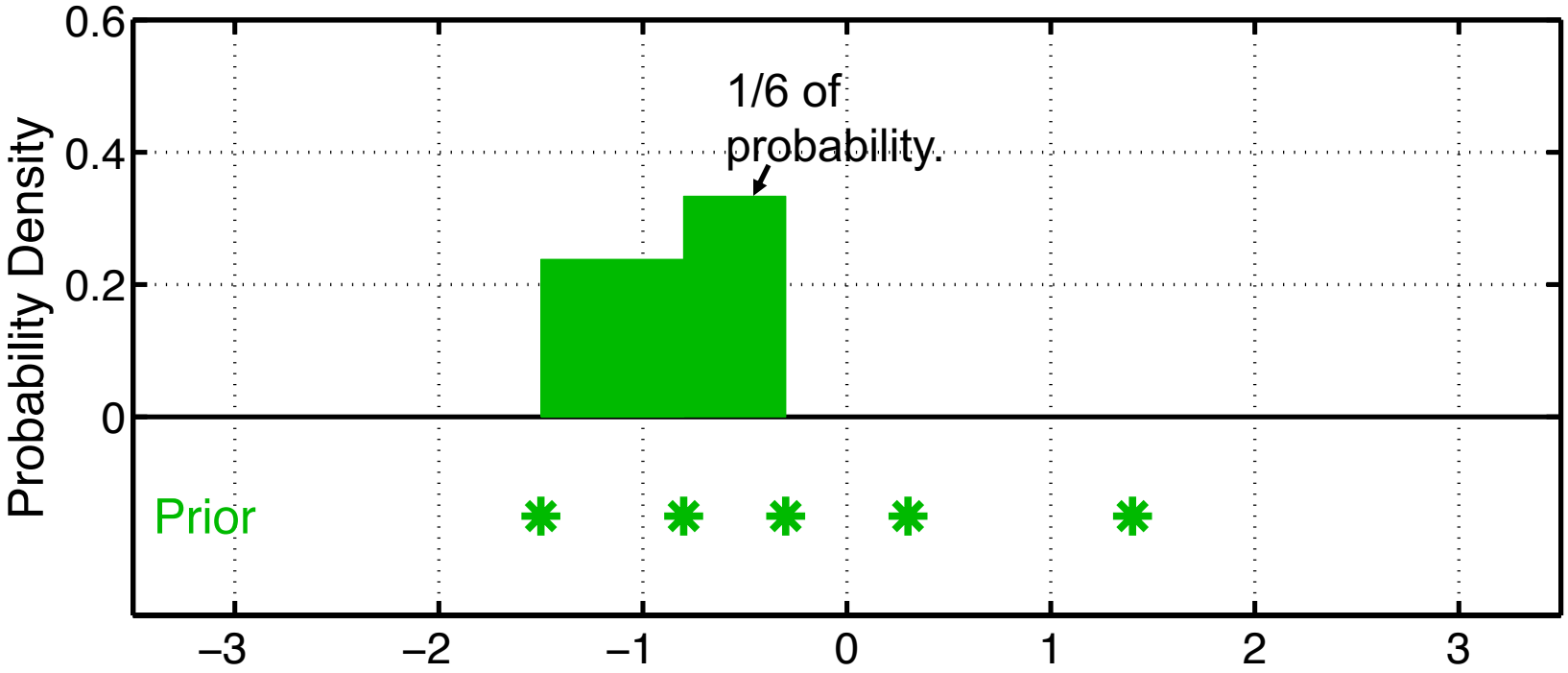
Have a prior ensemble for a state variable (like wind).

Rank Histogram Continuous Prior



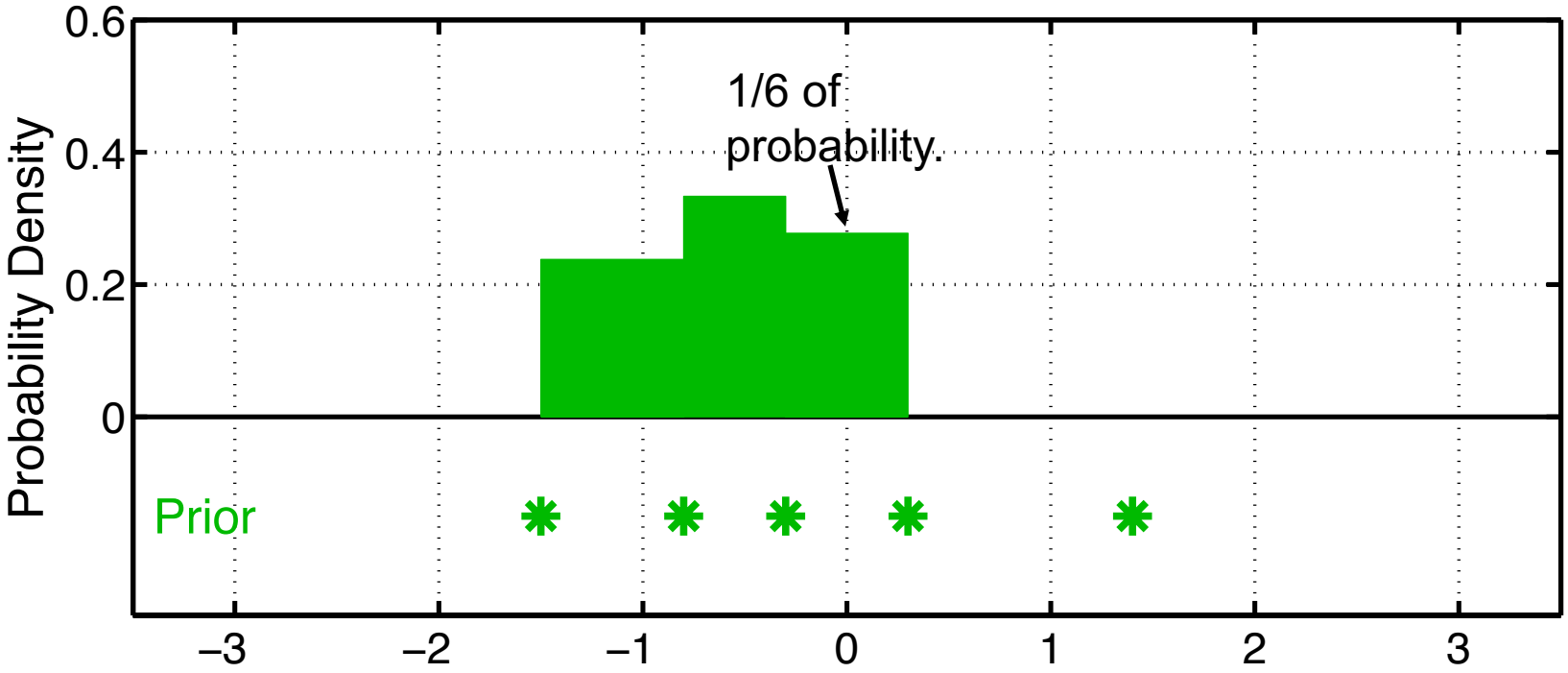
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Rank Histogram Continuous Prior



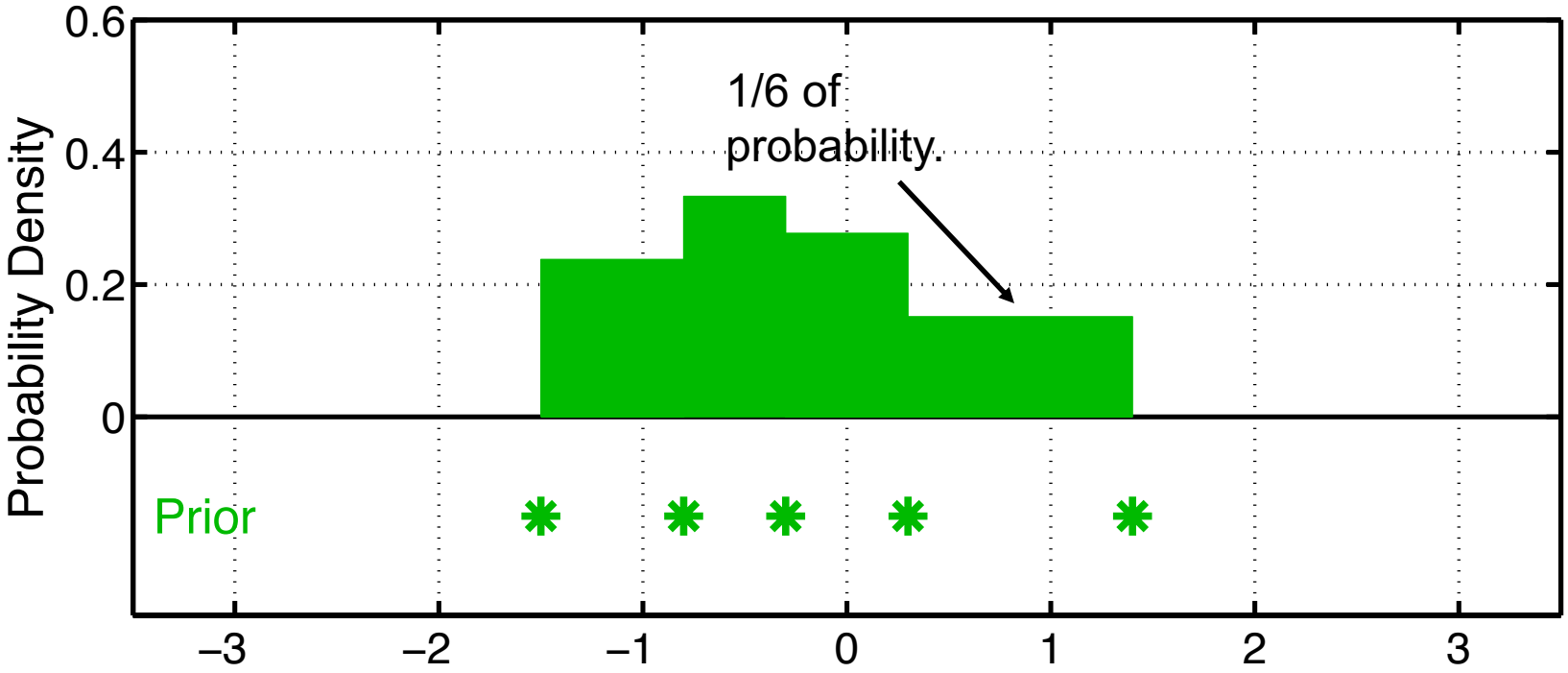
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Rank Histogram Continuous Prior



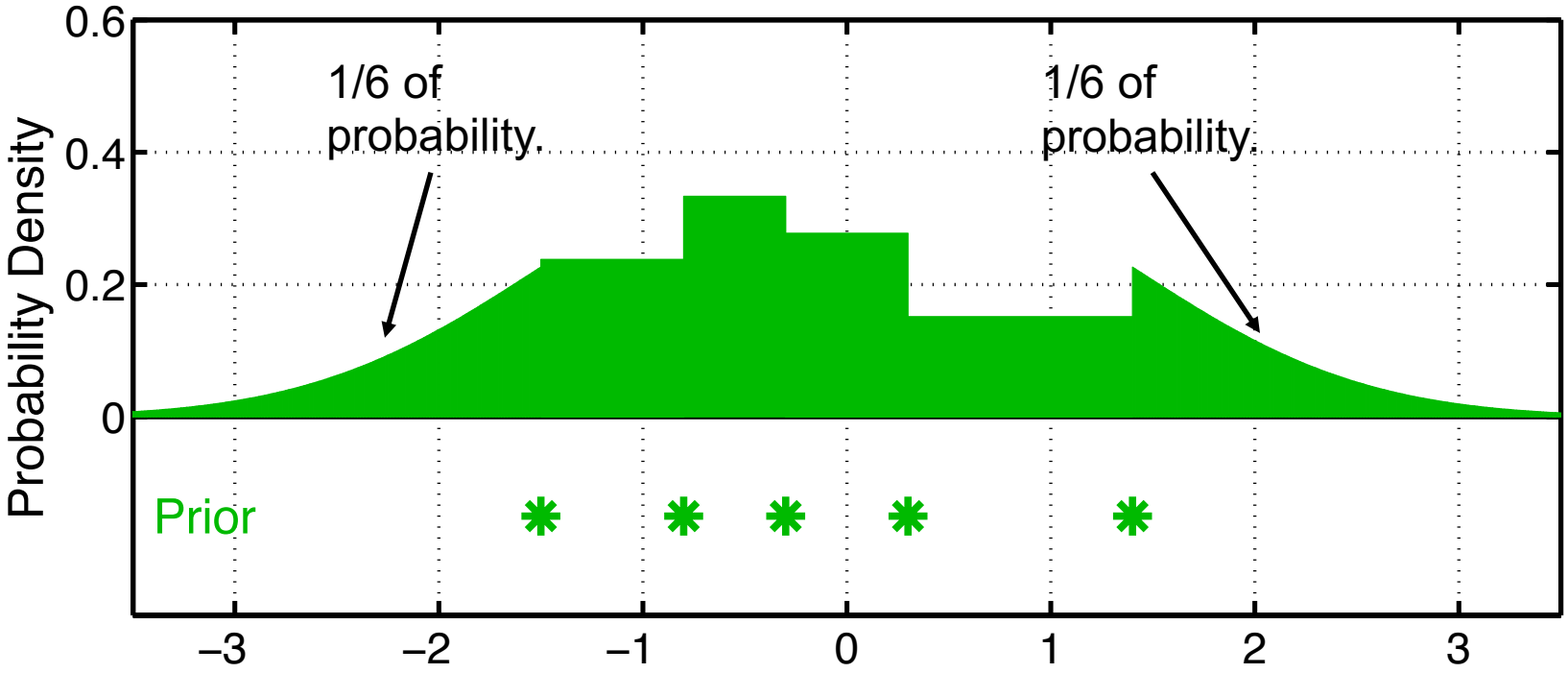
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Rank Histogram Continuous Prior



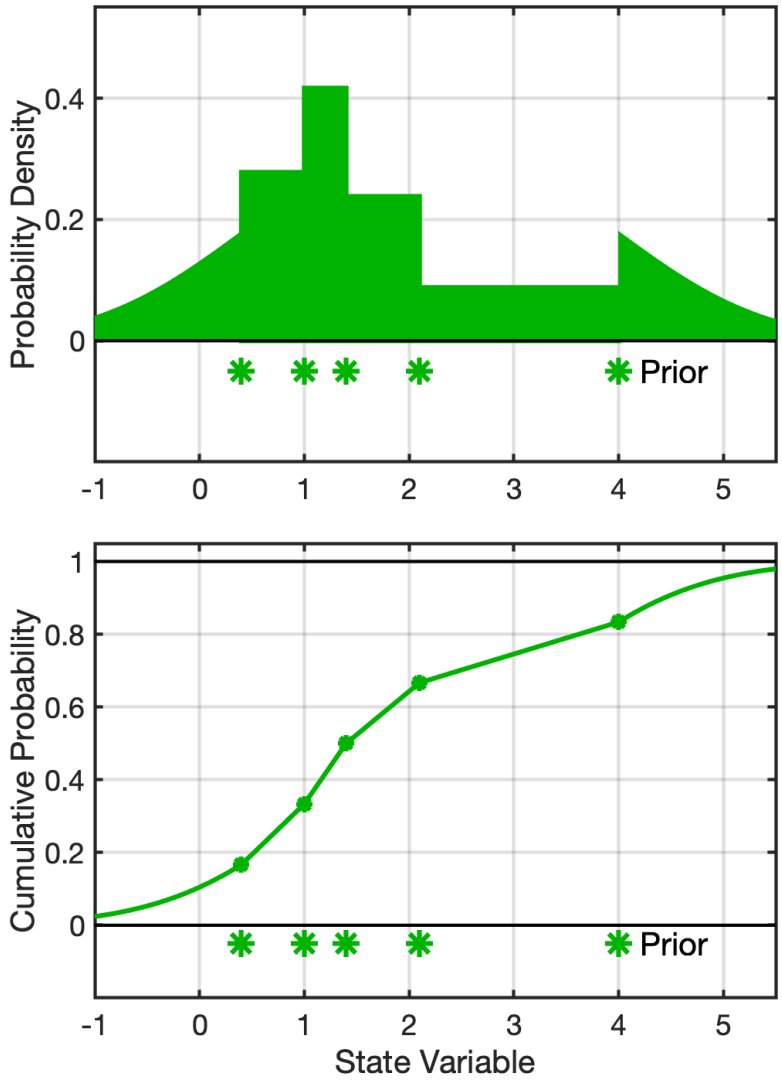
- Place (ens_size + 1)$^{-1}$ mass between adjacent ensemble members.
- Reminiscent of rank histogram evaluation method.

Rank Histogram Continuous Prior



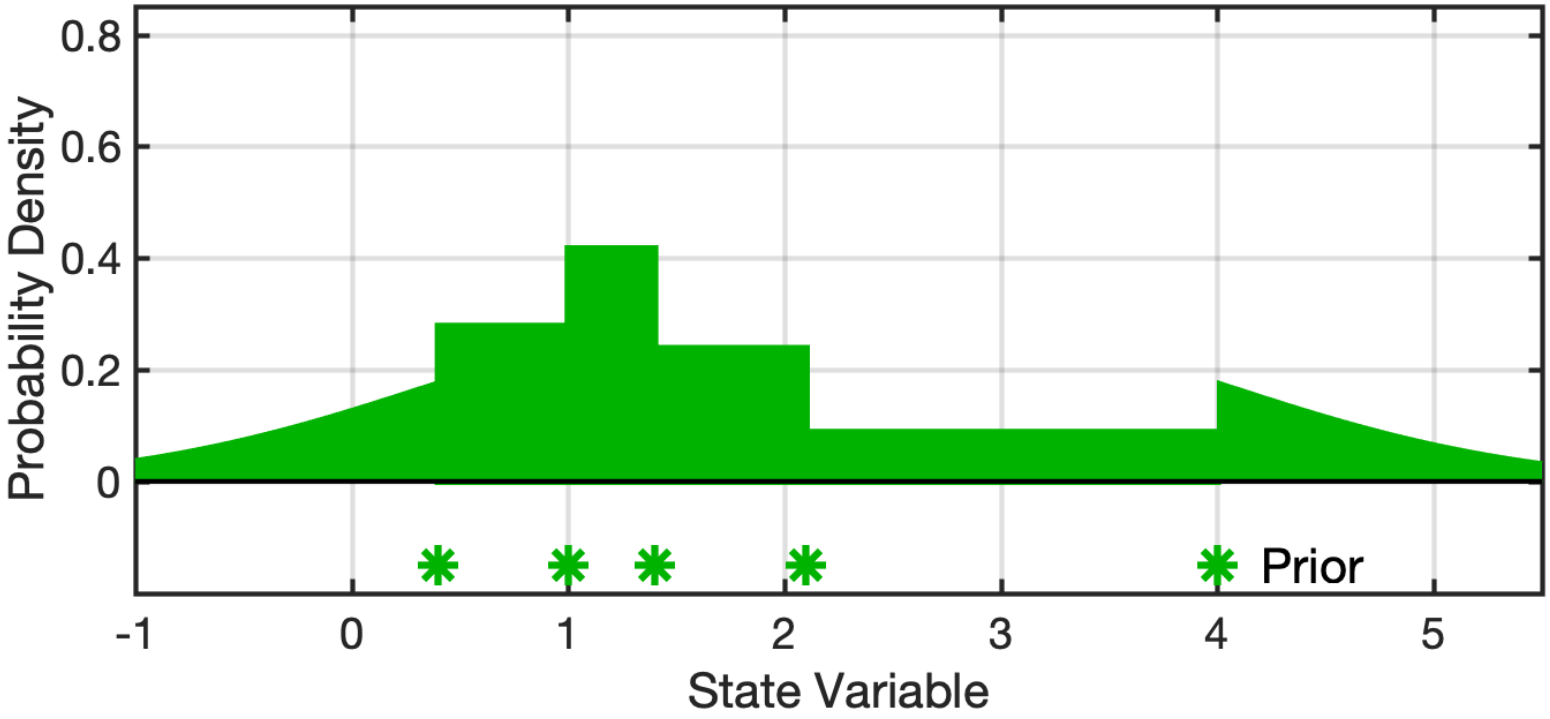
- Partial gaussian kernels on tails, N(*tail_mean*, *ens_sd*).
- *tail_mean selected so that (ens_size + 1)$^{-1}$ mass is in tail*.

Rank Histogram Prior PDF and CDF



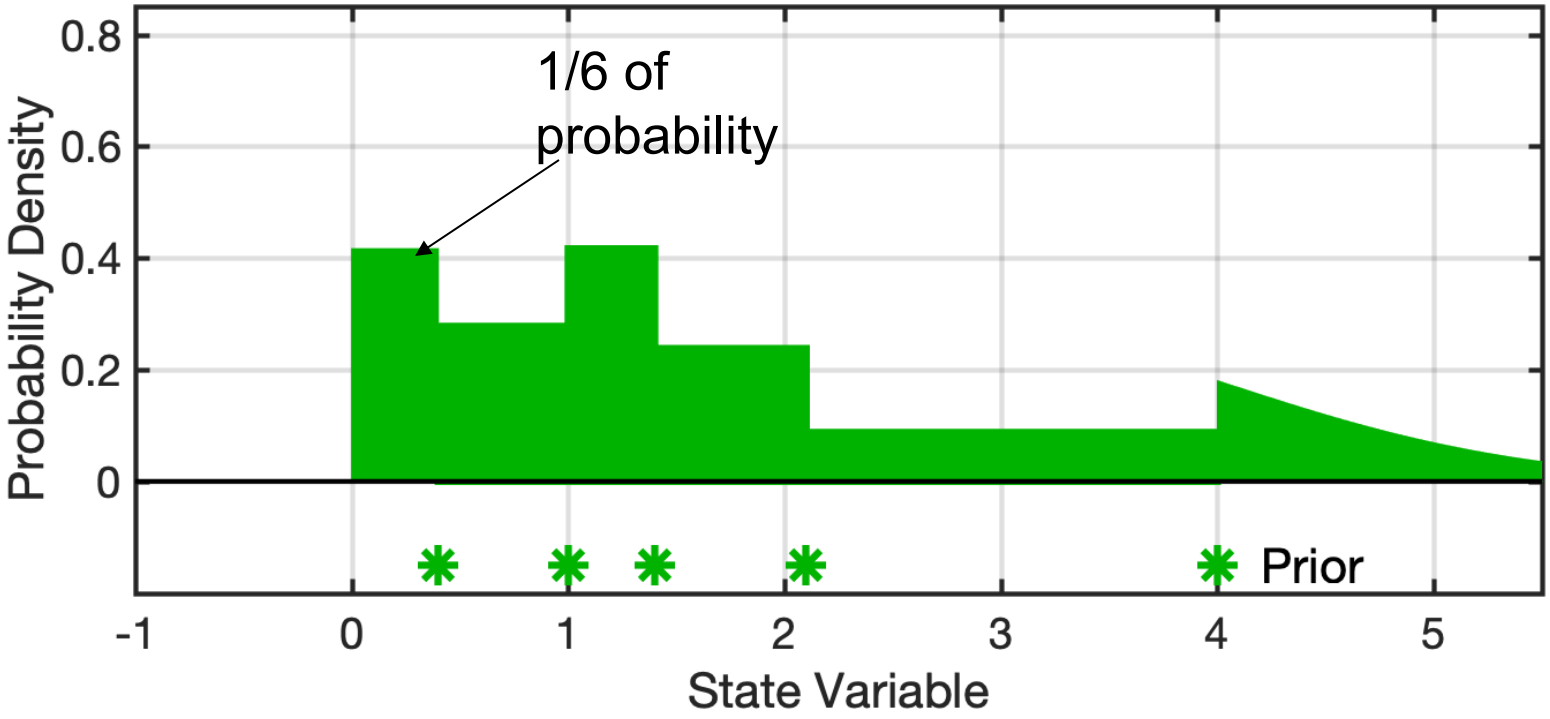
Rank Histogram Continuous Prior

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

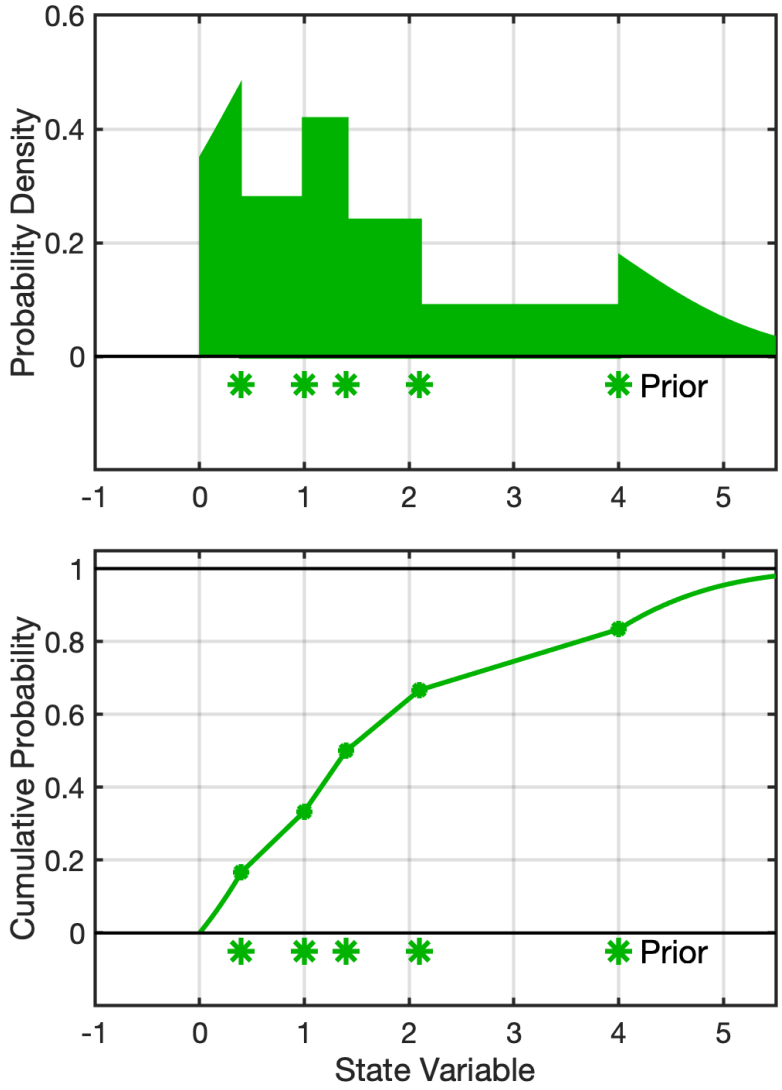


Bounded Rank Histogram Continuous Prior

Bounded has truncated tail.
Quantiles are exactly U(0, 1) by construction.

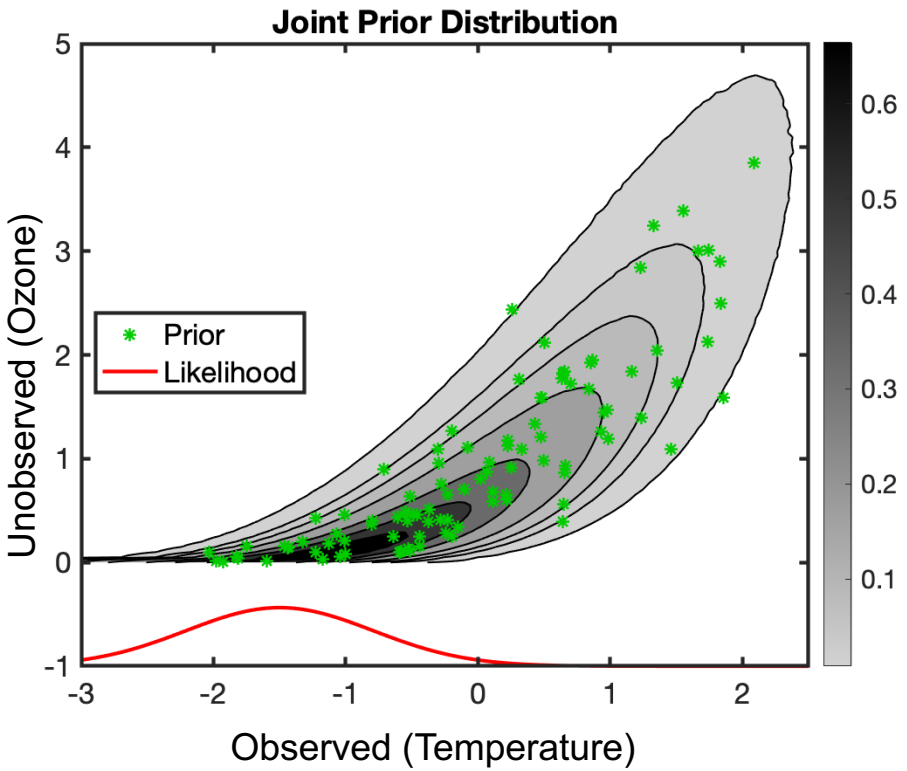


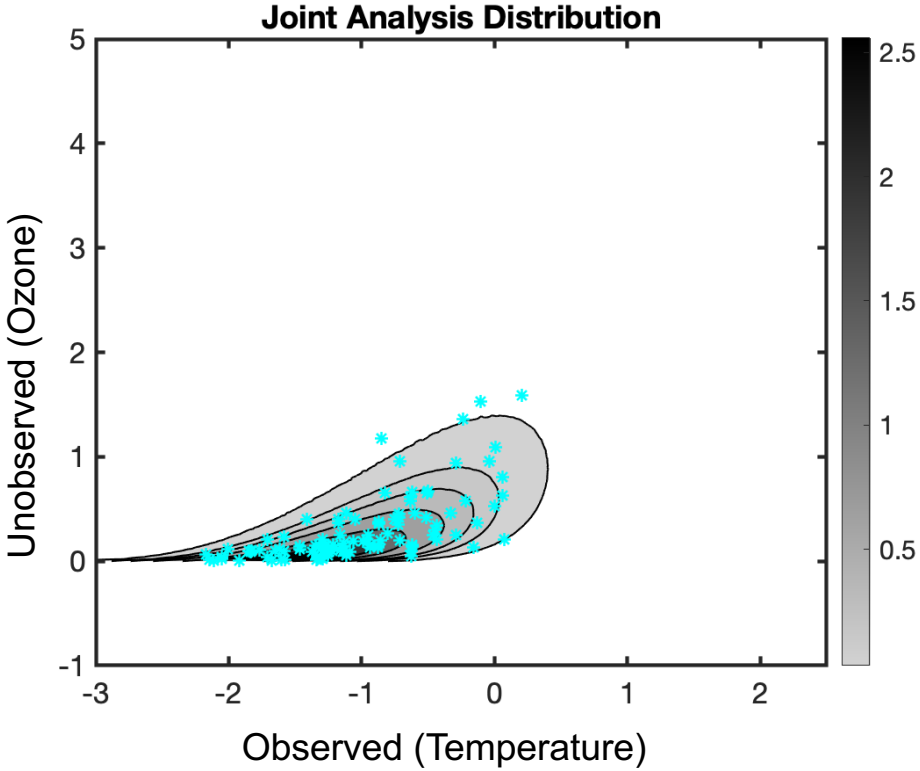
Rank Histogram Prior PDF and CDF with Lower Bound



Example 5: Normal observed, gamma unobserved with bounded RH

No need to classify prior distribution.
RH continuous prior distribution finds it.
This is very similar to case with explicit gamma prior.

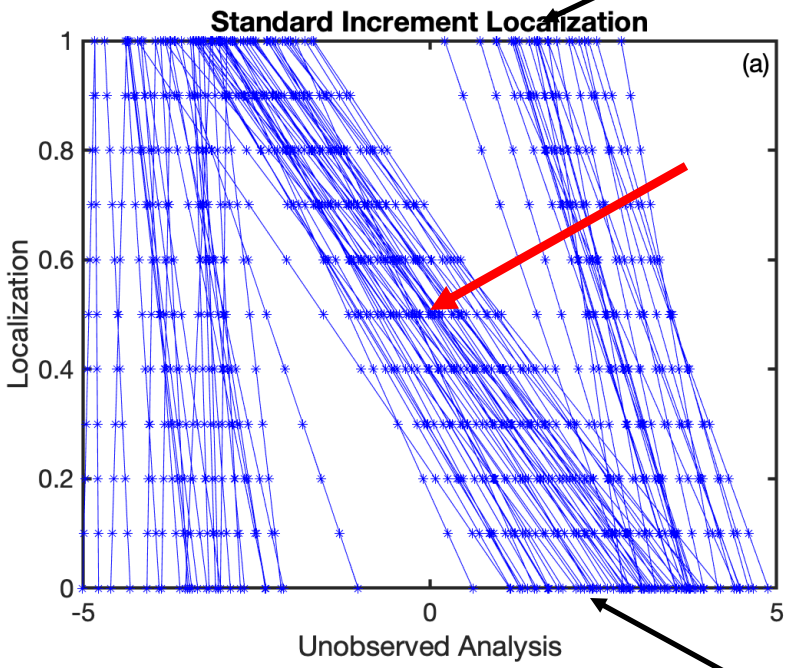


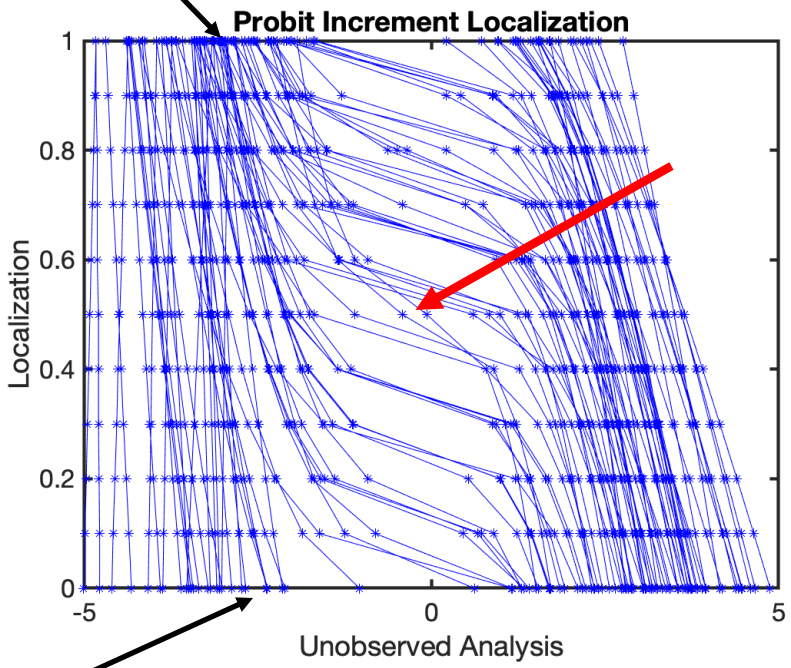


Localization of Probit Increments: Normal-binormal example

Standard increment localization may ignore prior constraints.
Probit increment localization 'knows' prior was binormal.

Unobserved Posterior (top row)

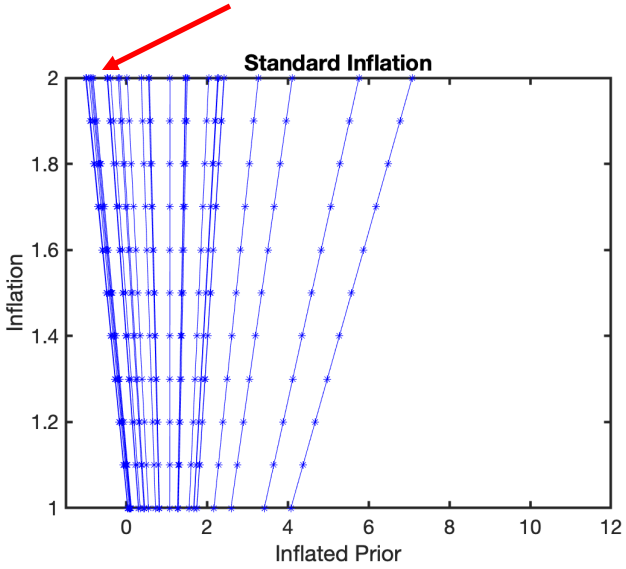




Unobserved Prior (bottom row)

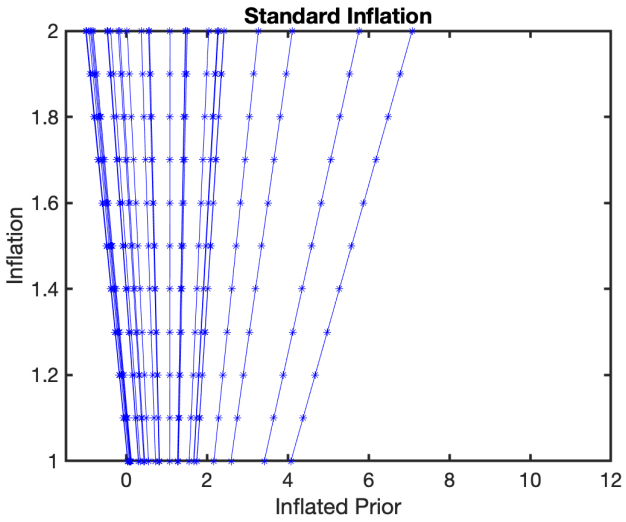
Inflation in Probit Space: Gamma example

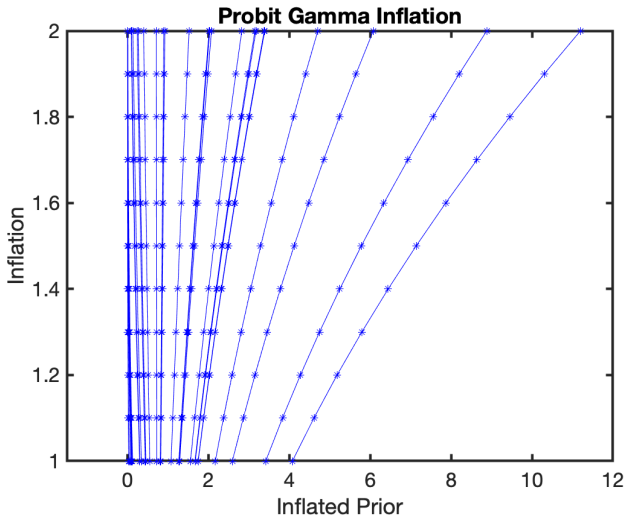
Standard inflation may violate prior constraints.



Inflation in Probit Space: Gamma example

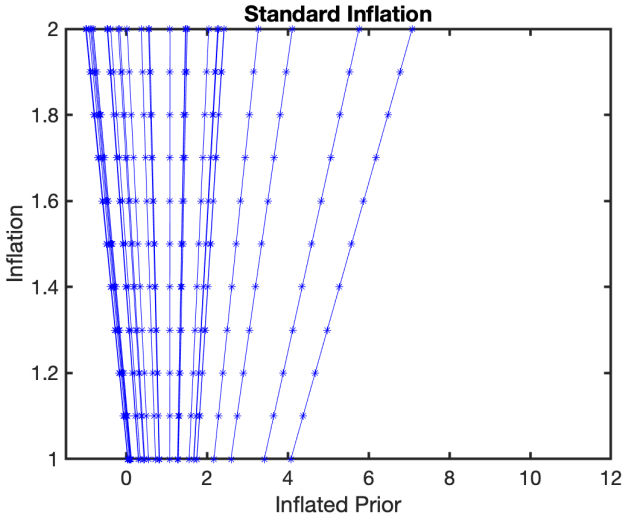
Standard inflation may violate prior constraints.
Inflation can be done in probit space.

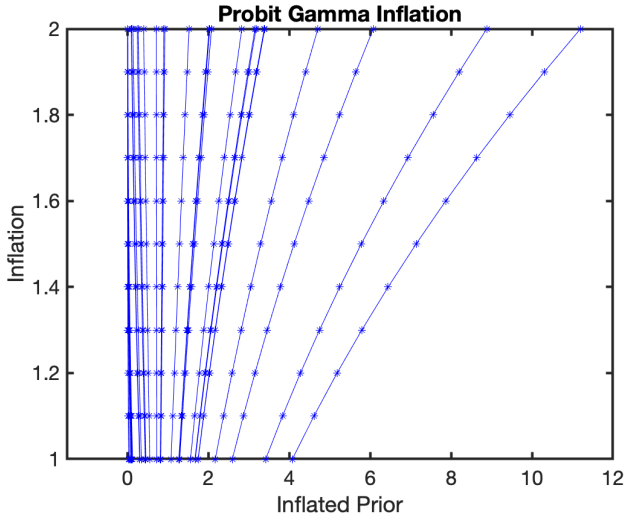


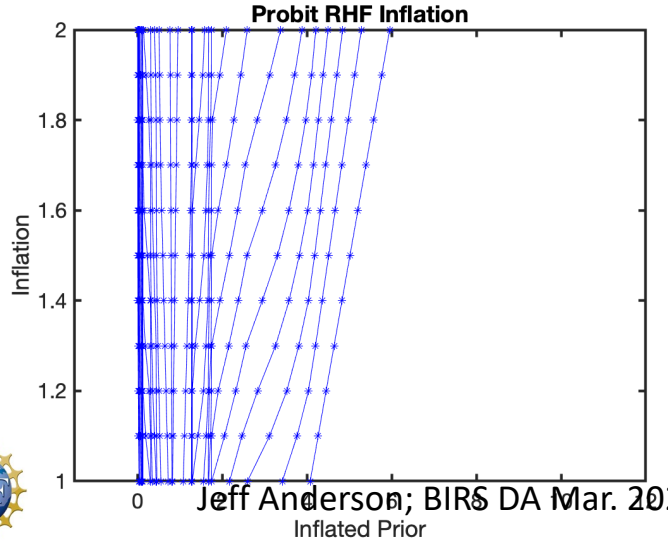


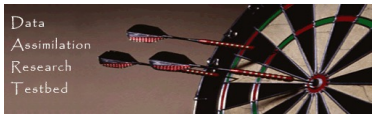
Inflation in Probit Space: Gamma example

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



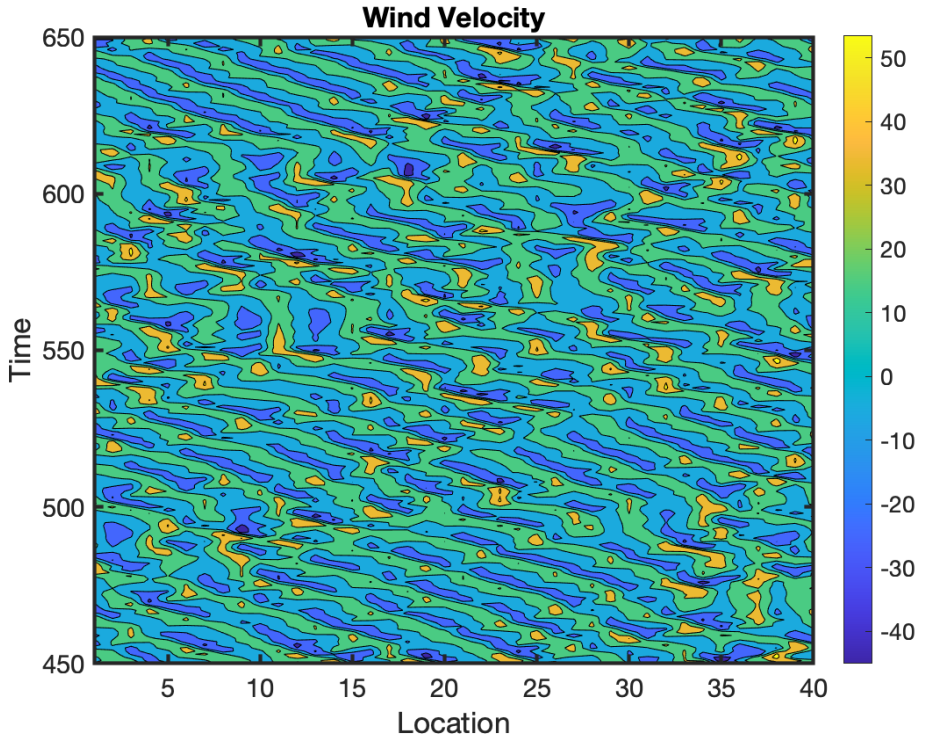


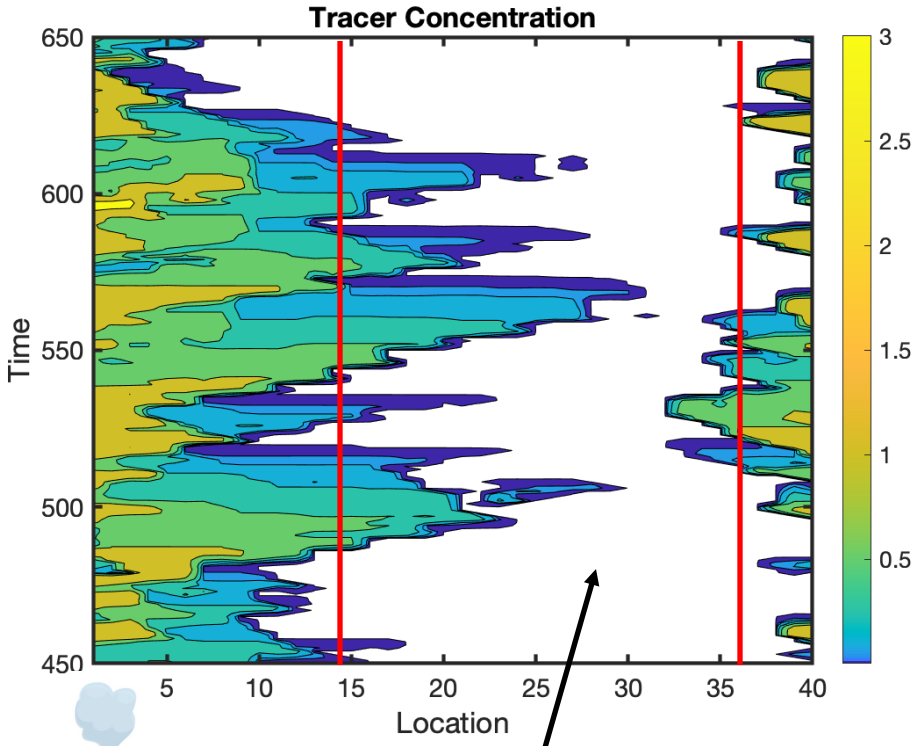




Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





Constant Source.

Concentration can be zero far from source.

Low-Order Tracer Advection Model Example

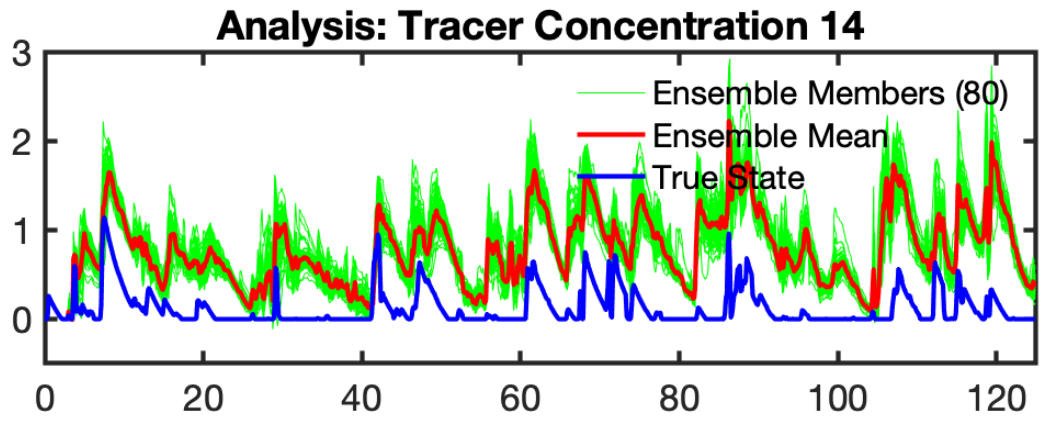
Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.

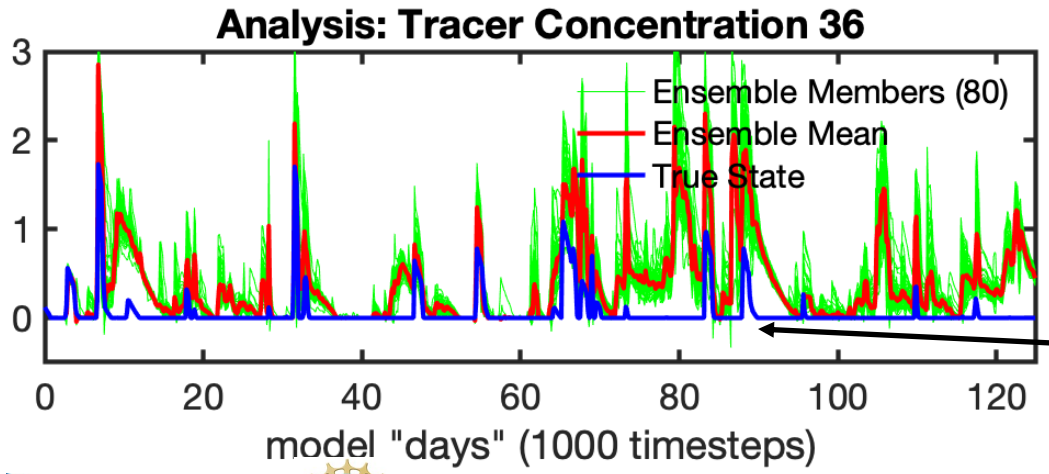
Observe state and concentration infrequently at each point.

Concentration error is truncated normal.

Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





Observe state and concentration infrequently at each point.

Concentration error is truncated normal.

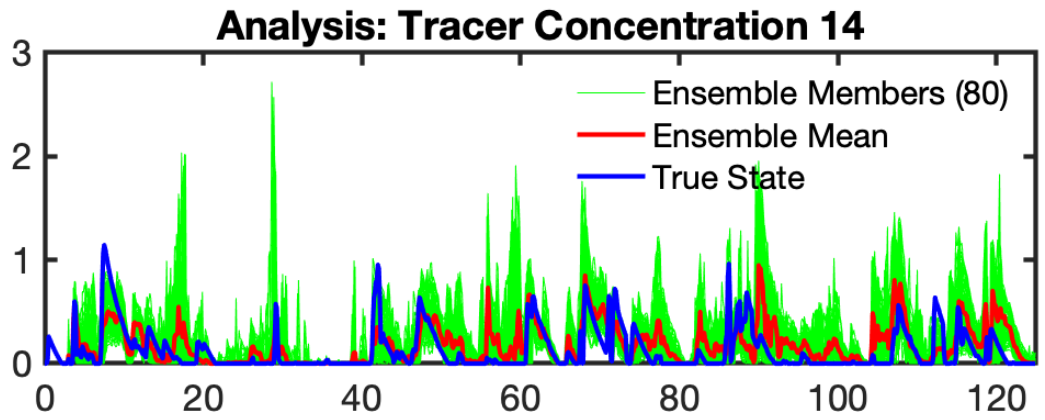
EAKF has large bias for tracers.

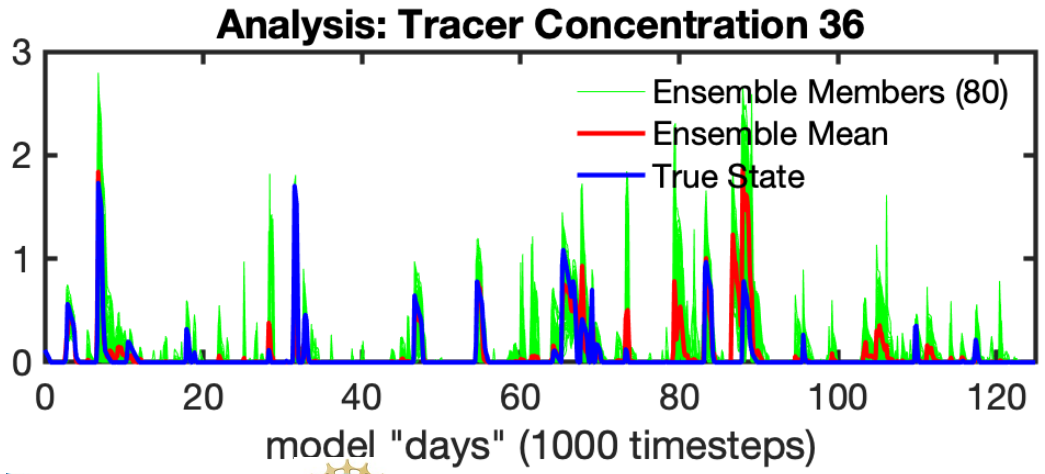
Can't go to all zeros.

Some negative values.

Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





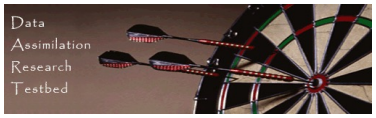
Observe state and concentration infrequently at each point.

Concentration error is truncated normal.

RHF with RH quantile regression is unbiased.

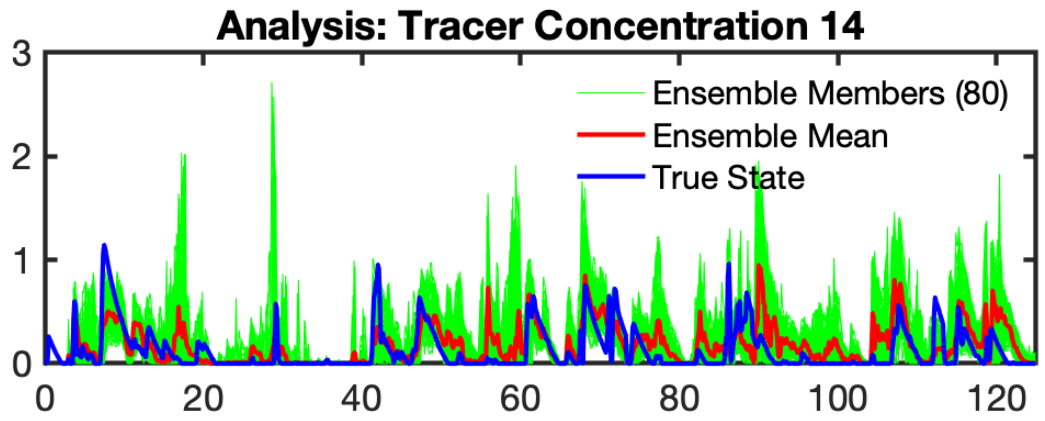
Can go to all zeros.

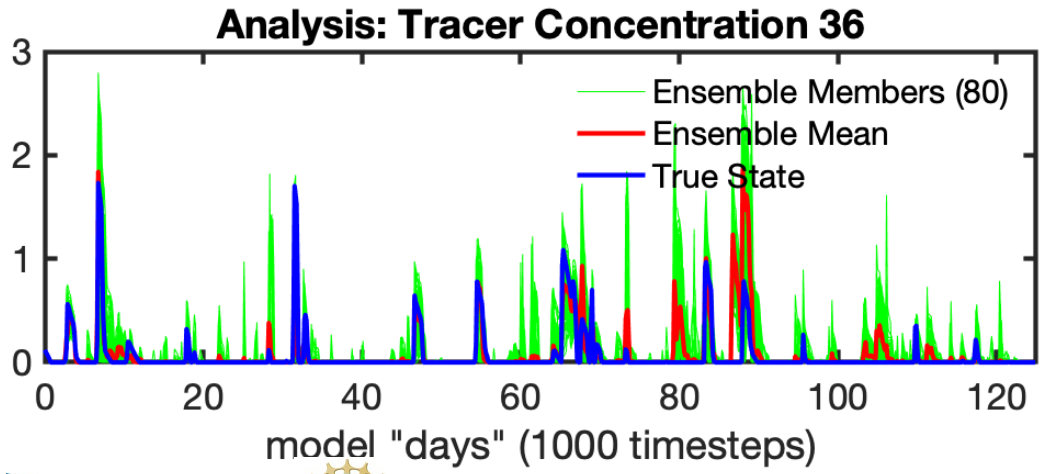
No negative values.



Low-Order Tracer Advection Model Example

Each grid point has Lorenz-96 state, tracer concentration, tracer source/sink.
Multiple of state treated as wind, conservatively advects tracer.
Example: single time constant source at grid point 1.





Gamma, inverse gamma, and log-normal distributions can't do this.

RHF with RH quantile regression unbiased.

Can go to all zeros.

No negative values.

Implementing in DART

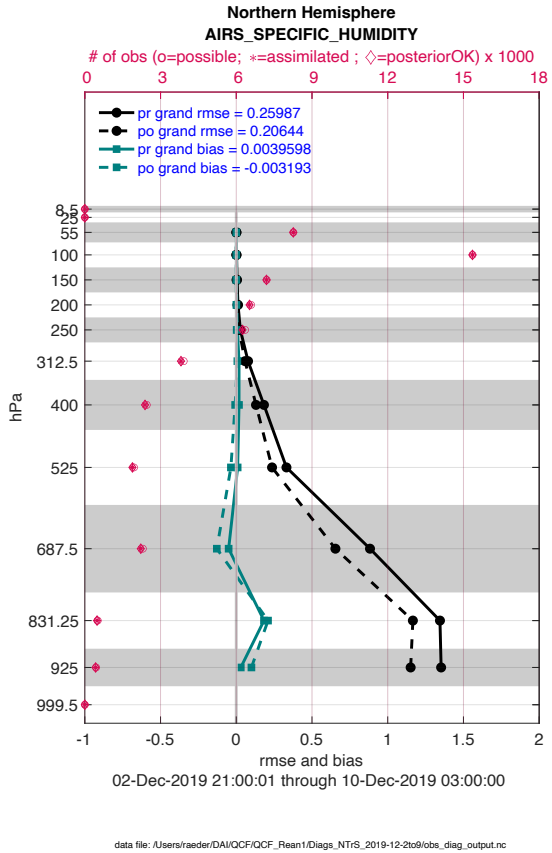
- Implementation in DART is trivial:
  - Convert all joint space state variables to probit space.
  - Apply standard algorithms.
- Adaptive inflation, localization, sampling error correction, all work.
- Parallelization is unchanged.
- Supported release in the next few months.

Large-scale tests and timing

Initial Tests for NWP in 1 degree CAM global atmosphere.
Using (bounded) rank histogram for all observations and regressions.

Results slightly better!

Time penalty around 1% of assimilation cost!

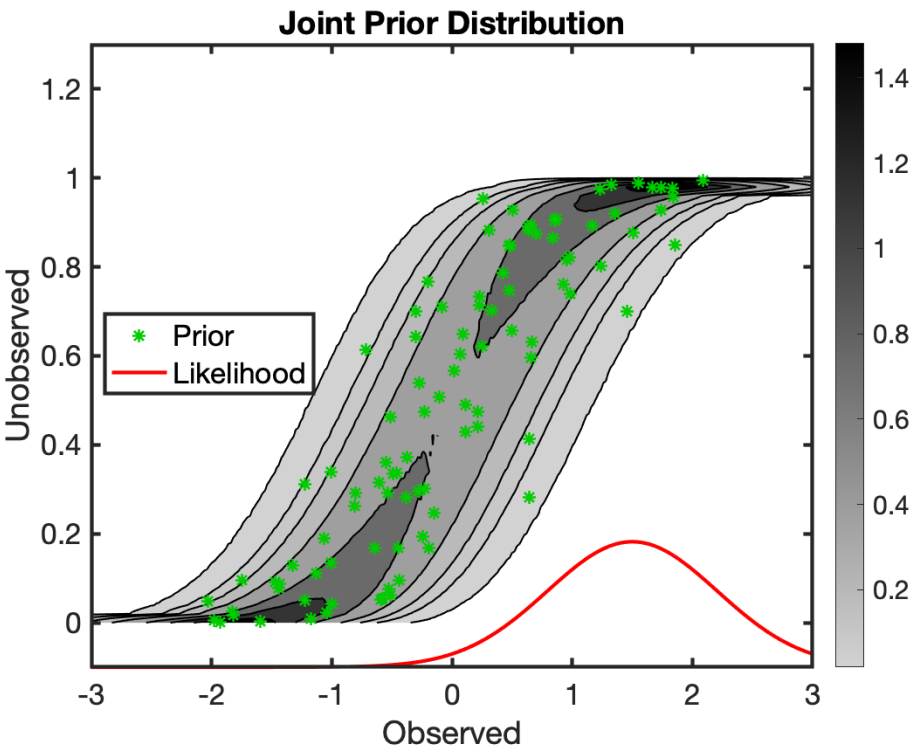


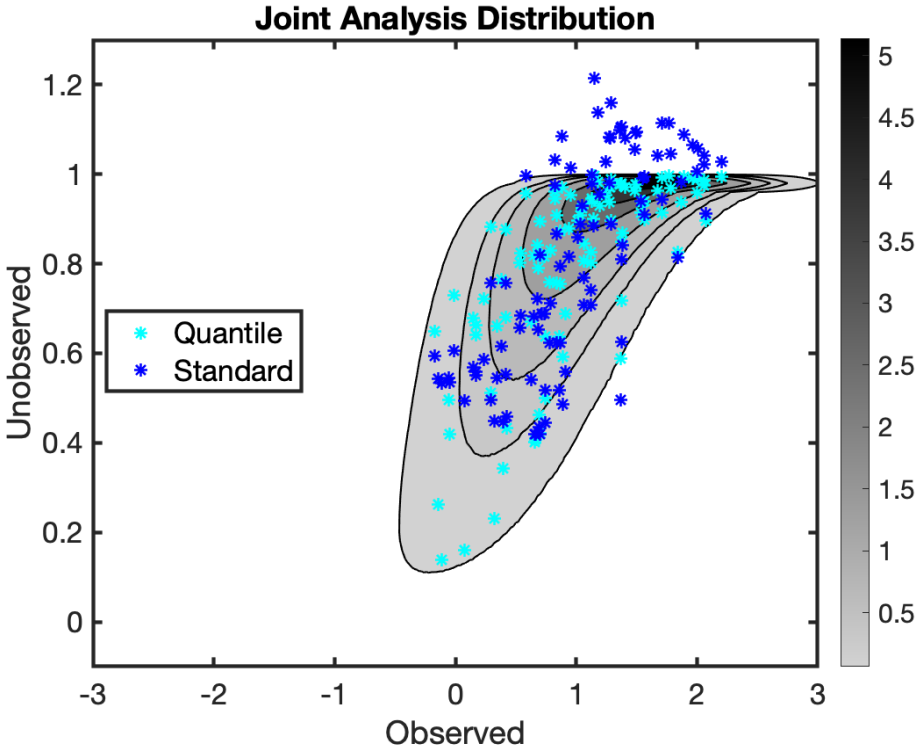
A DART release, that includes the algorithms described here and the idealized tracer model, is now available at https://github.com/NCAR/DART/releases/tag/v11.0.0-alpha (QR code below) or see the DART website: dart.ucar.edu

Thanks to:
Chris Riedel, Helen Kershaw,
Marlee Smith, Molly Wieringa
and the rest of the DAReS team.

Example 2: Normal observed, beta unobserved

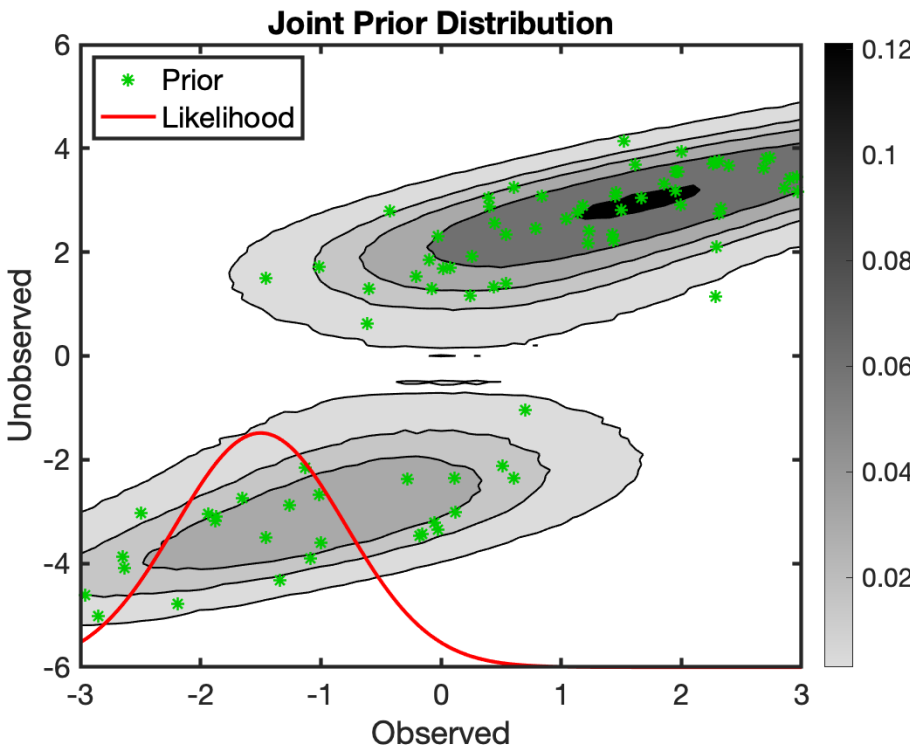
Probit respects bounds.

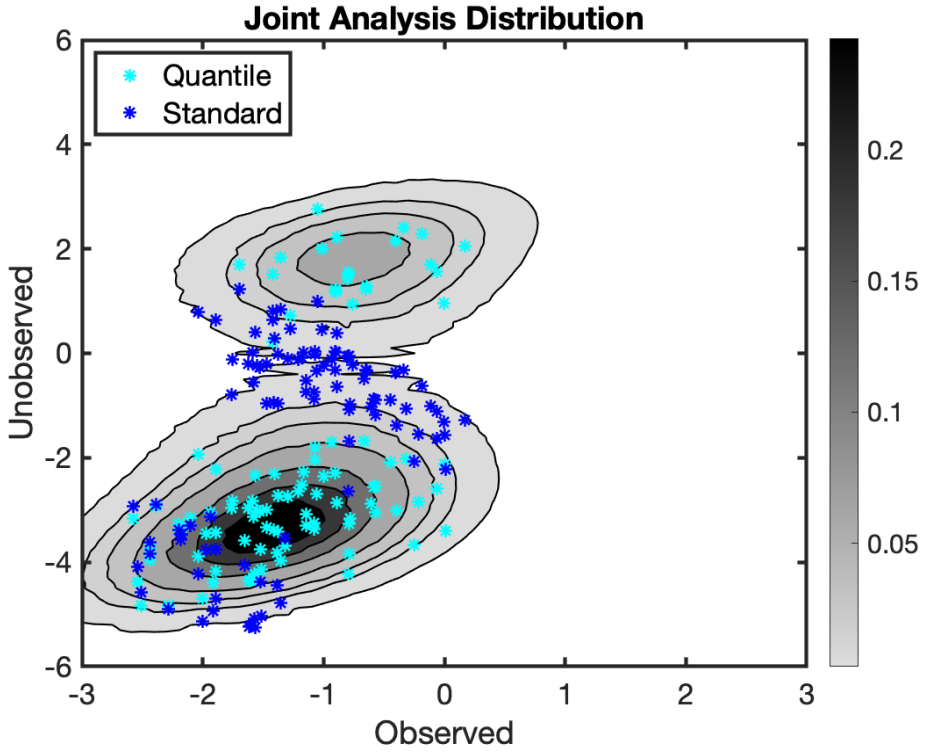




Example 3: Normal observed, binormal unobserved

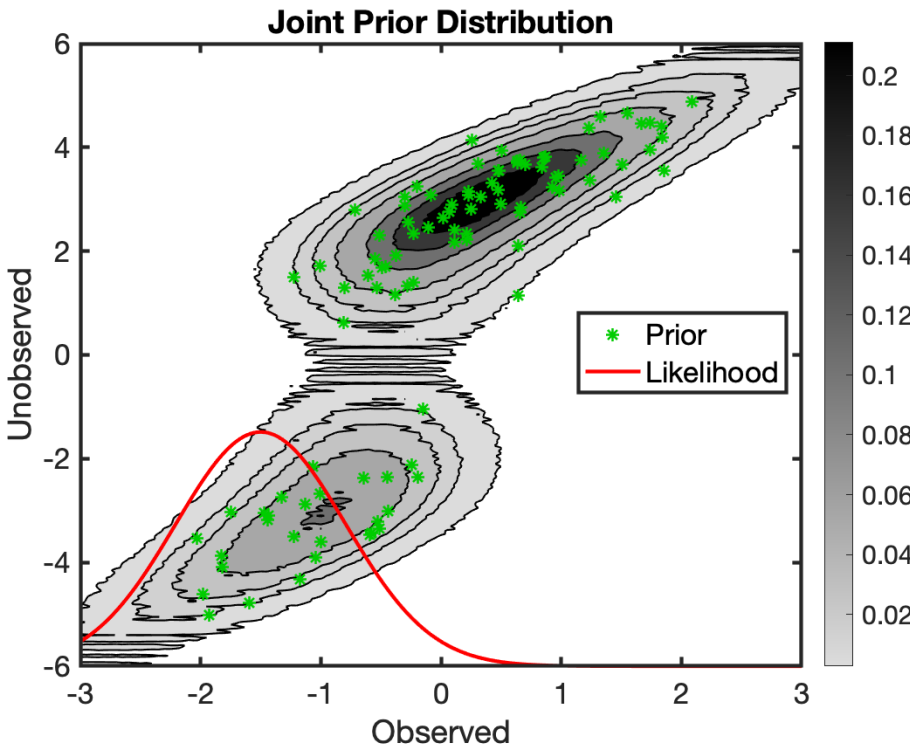
Probit moves members from one mode to the other.
Also adjusts ensemble in the modes.

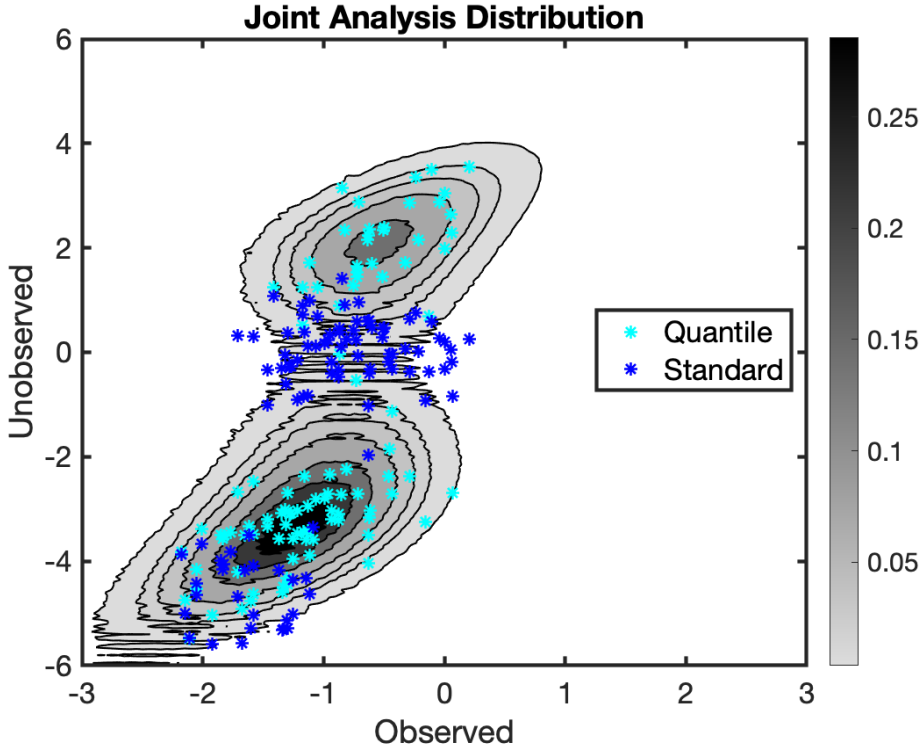




Example 6: Normal observed, binormal unobserved with RH

No need to classify prior distribution, RH finds it.
Avoids clustering for prior, much cheaper to invert CDF.





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.