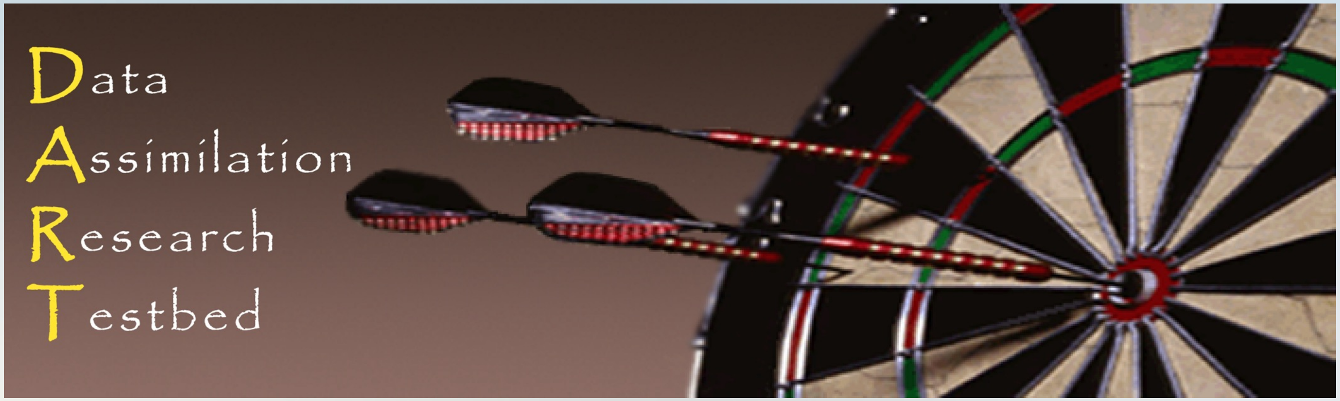


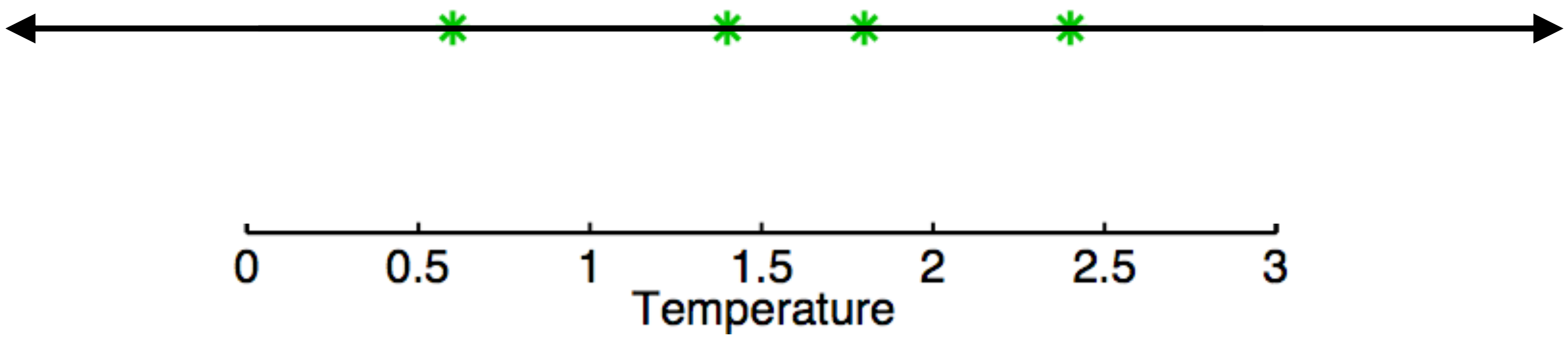
DART_LAB Tutorial Section 3: Inflation and Localization to Improve Performance

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.

The Rank Histogram: Evaluating Filter Performance

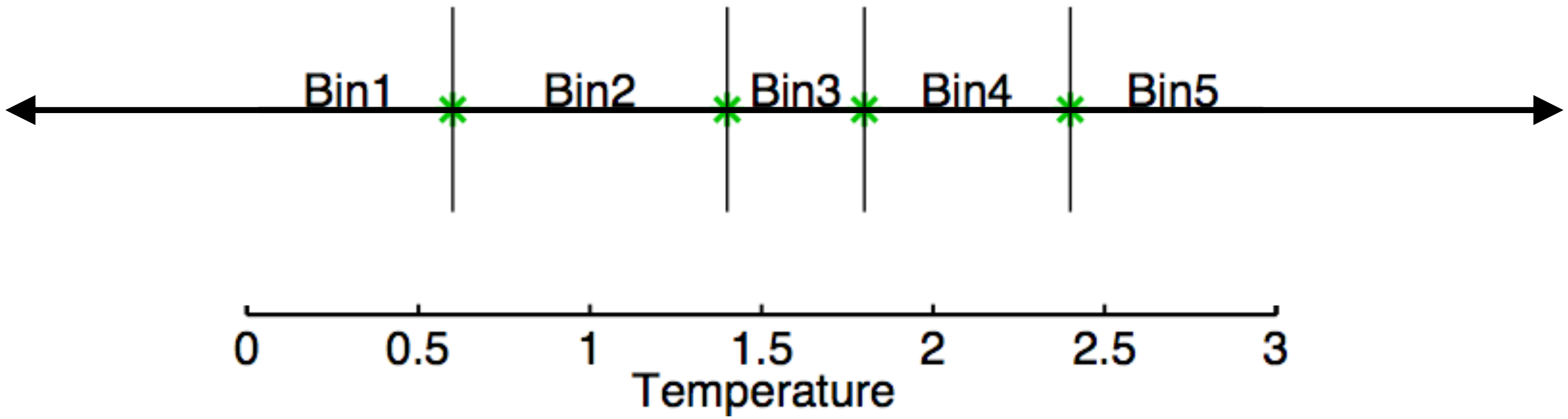
Draw 5 values from a real-valued distribution.

Call the first 4 'ensemble members'.



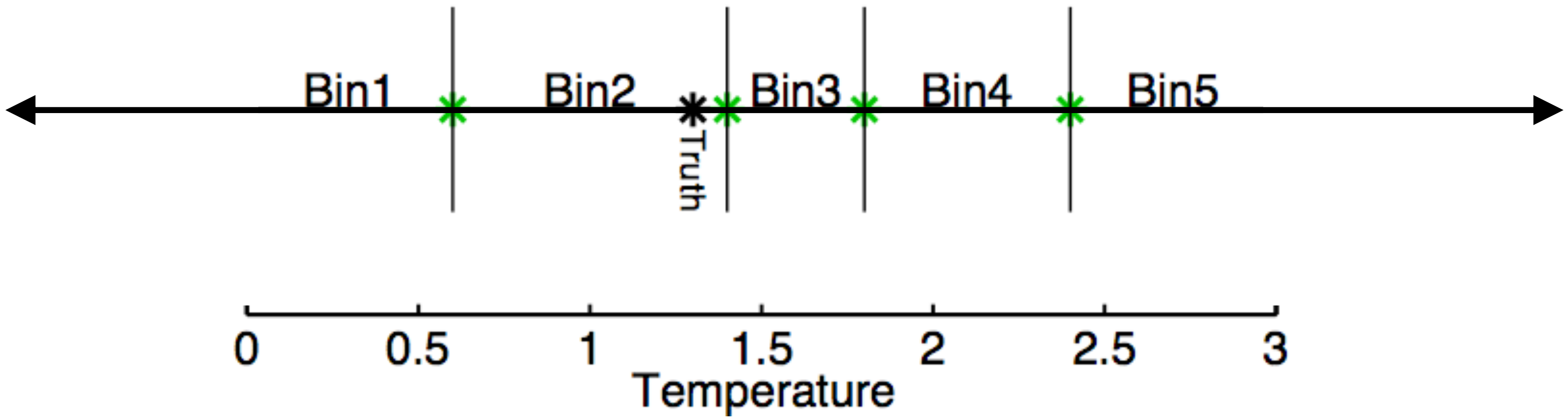
The Rank Histogram: Evaluating Filter Performance

These 4 ‘ensemble members’ partition the real line into 5 bins.



The Rank Histogram: Evaluating Filter Performance

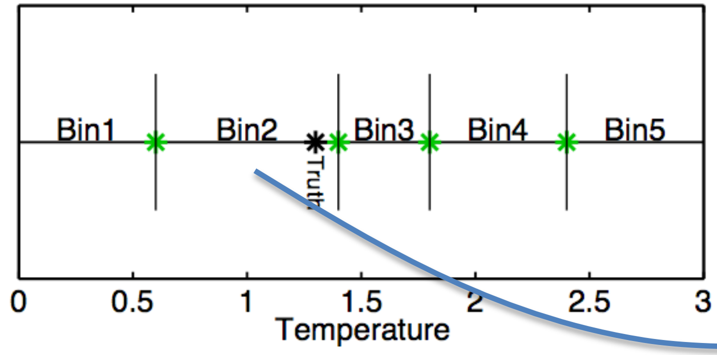
Call the 5th draw the 'truth'.
1/5 chance that this is in any given bin.



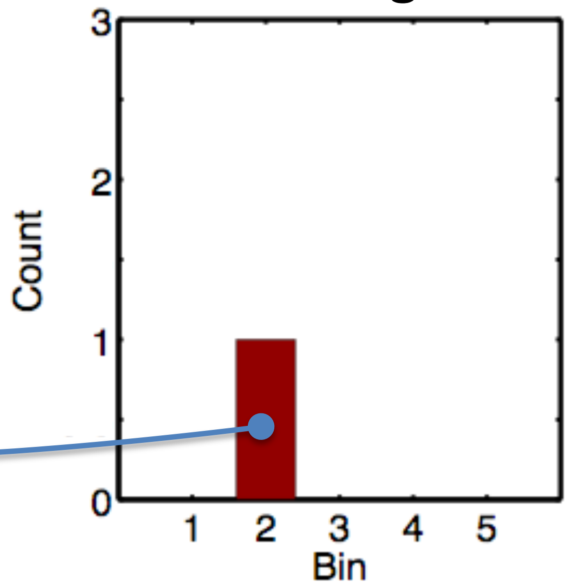
The Rank Histogram: Evaluating Filter Performance

Rank histogram shows the frequency of the truth in each bin over many assimilations.

Same figure as previous slide.



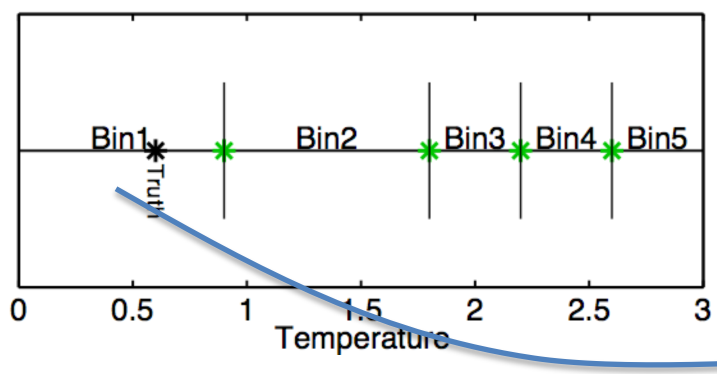
rank histogram



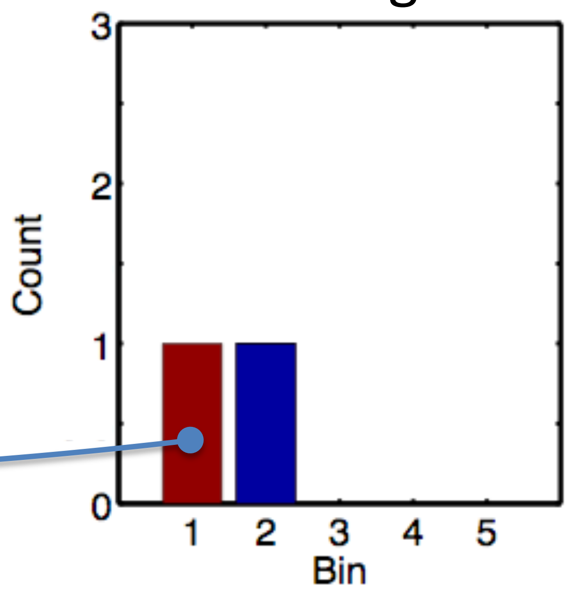
The Rank Histogram: Evaluating Filter Performance

Rank histogram shows the frequency of the truth in each bin over many assimilations.

A new assimilation.

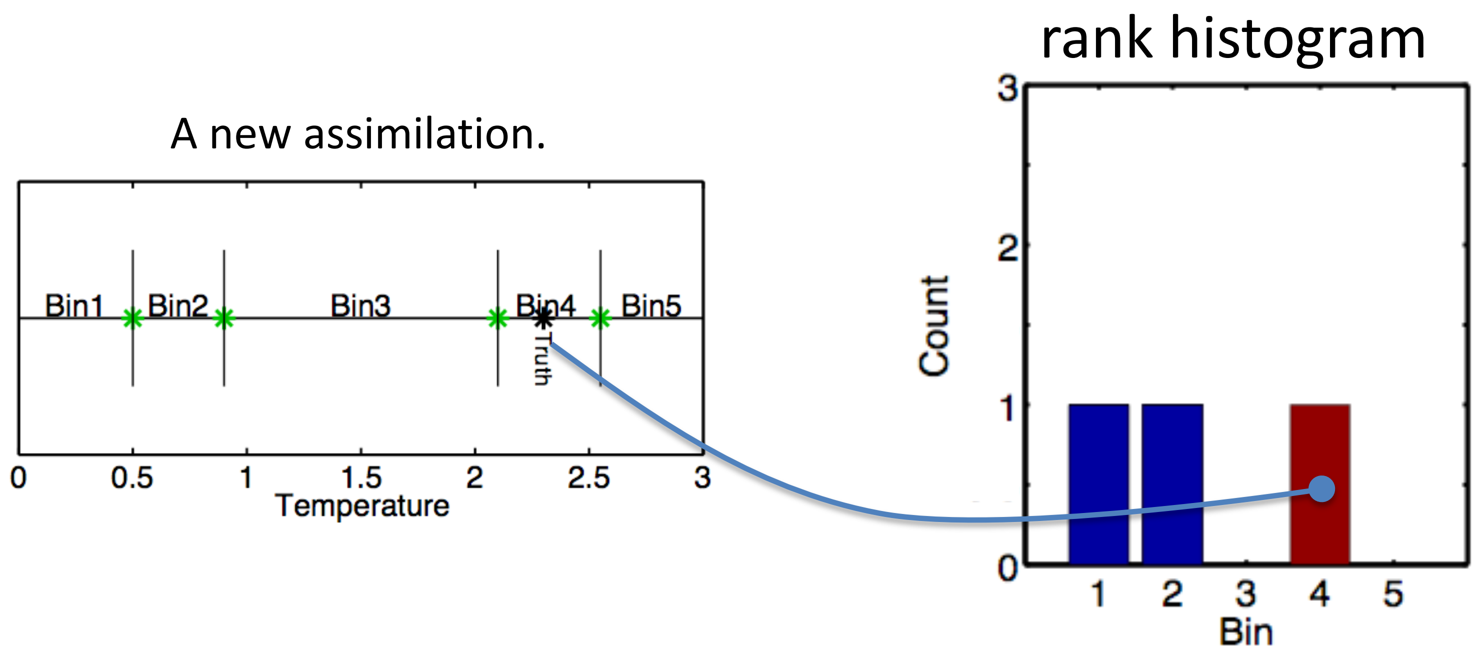


rank histogram



The Rank Histogram: Evaluating Filter Performance

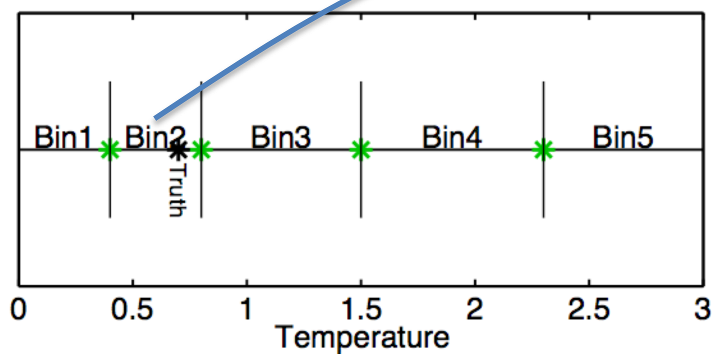
Rank histogram shows the frequency of the truth in each bin over many assimilations.



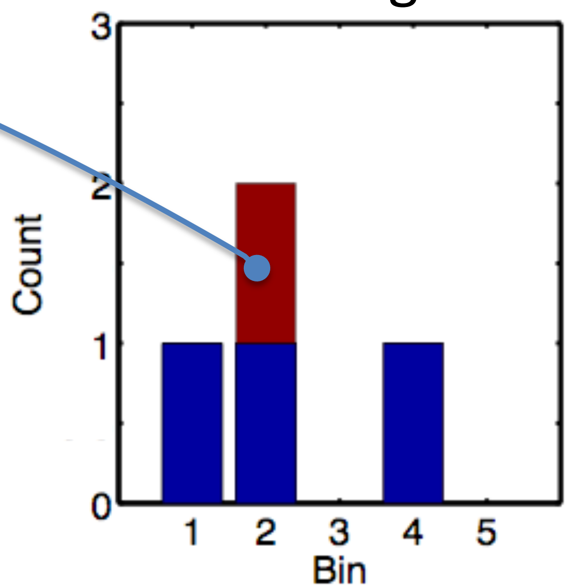
The Rank Histogram: Evaluating Filter Performance

Rank histogram shows the frequency of the truth in each bin over many assimilations.

A new assimilation.

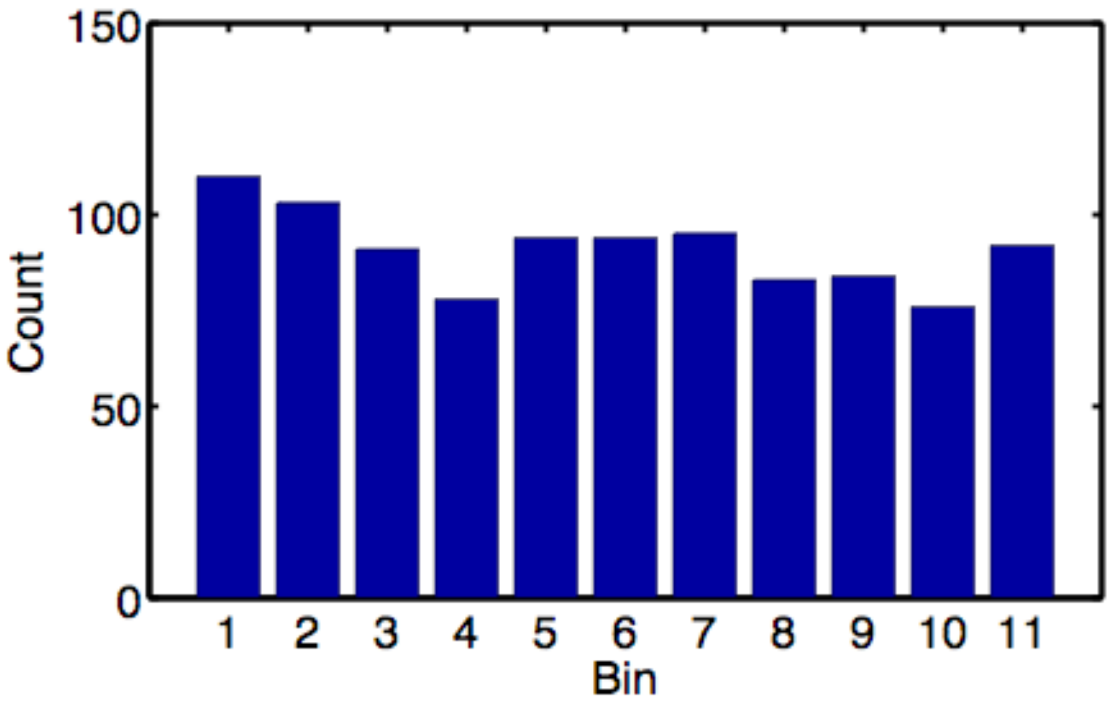


rank histogram



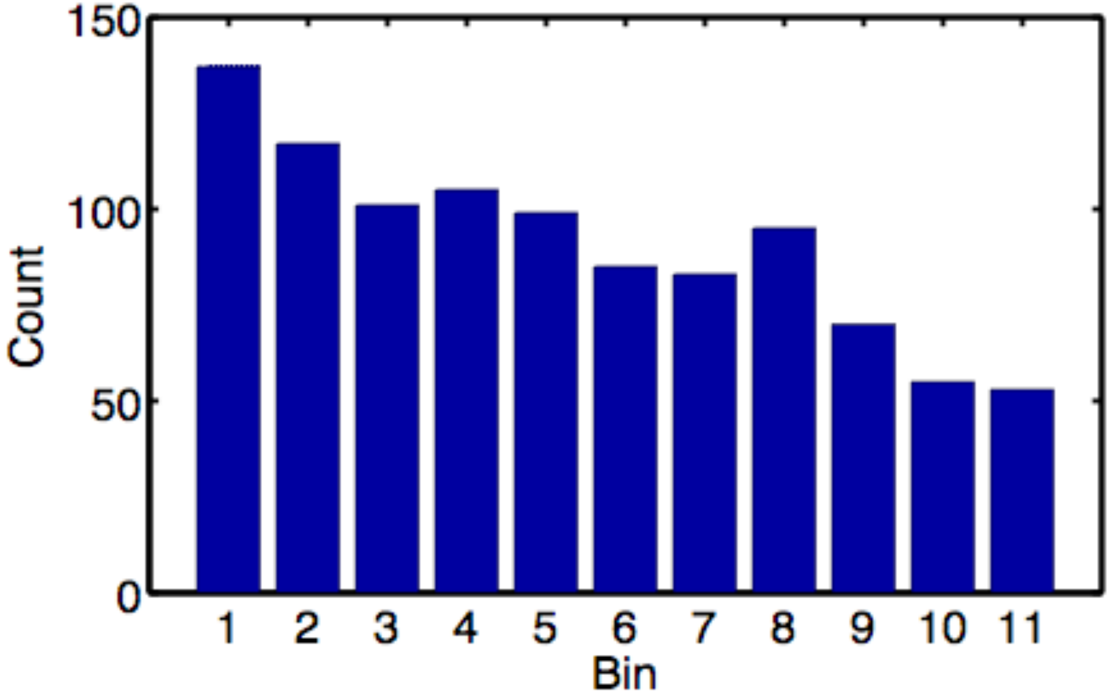
The Rank Histogram: Evaluating Filter Performance

Rank histograms for good (consistent) ensembles should be uniform (caveat sampling noise).
Want truth to look like random draw from ensemble.



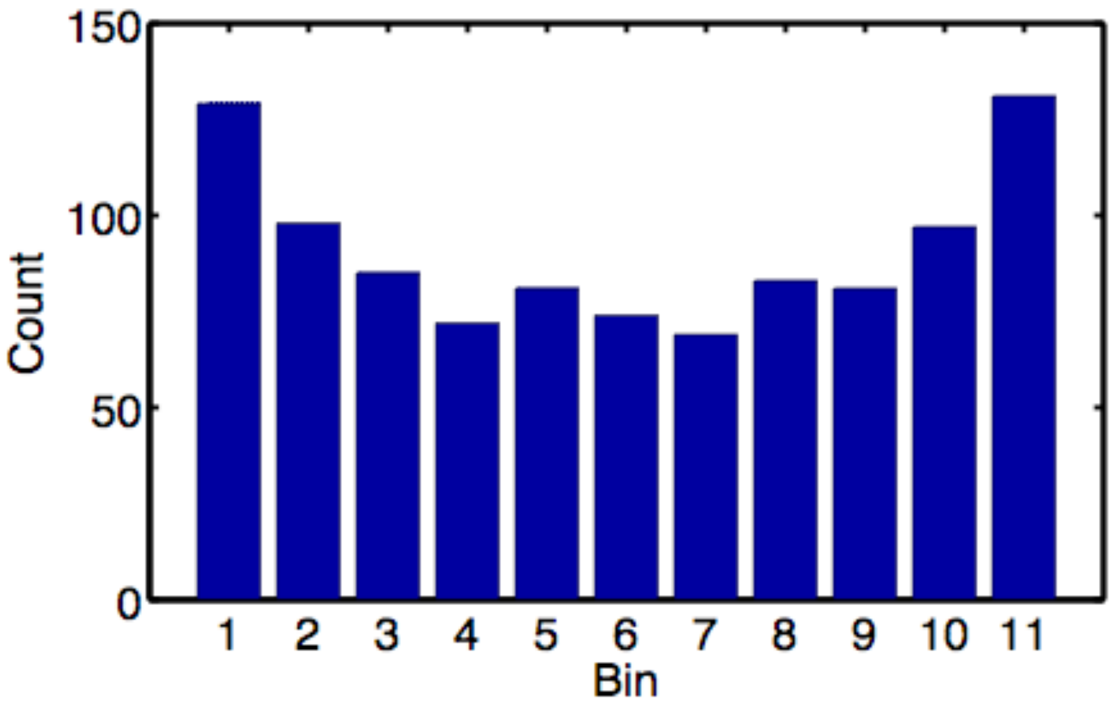
The Rank Histogram: Evaluating Filter Performance

Biased ensembles leads to skewed histograms.



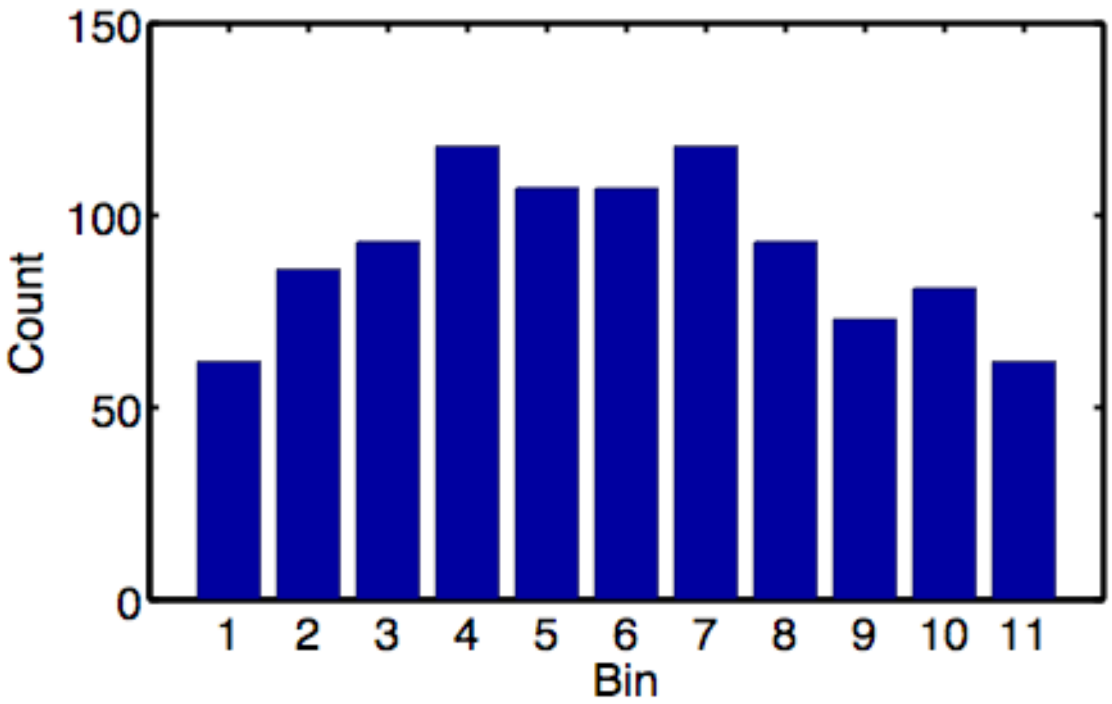
The Rank Histogram: Evaluating Filter Performance

Ensembles with too little spread gives a u-shape.
This is the most common behavior for geophysics.



The Rank Histogram: Evaluating Filter Performance

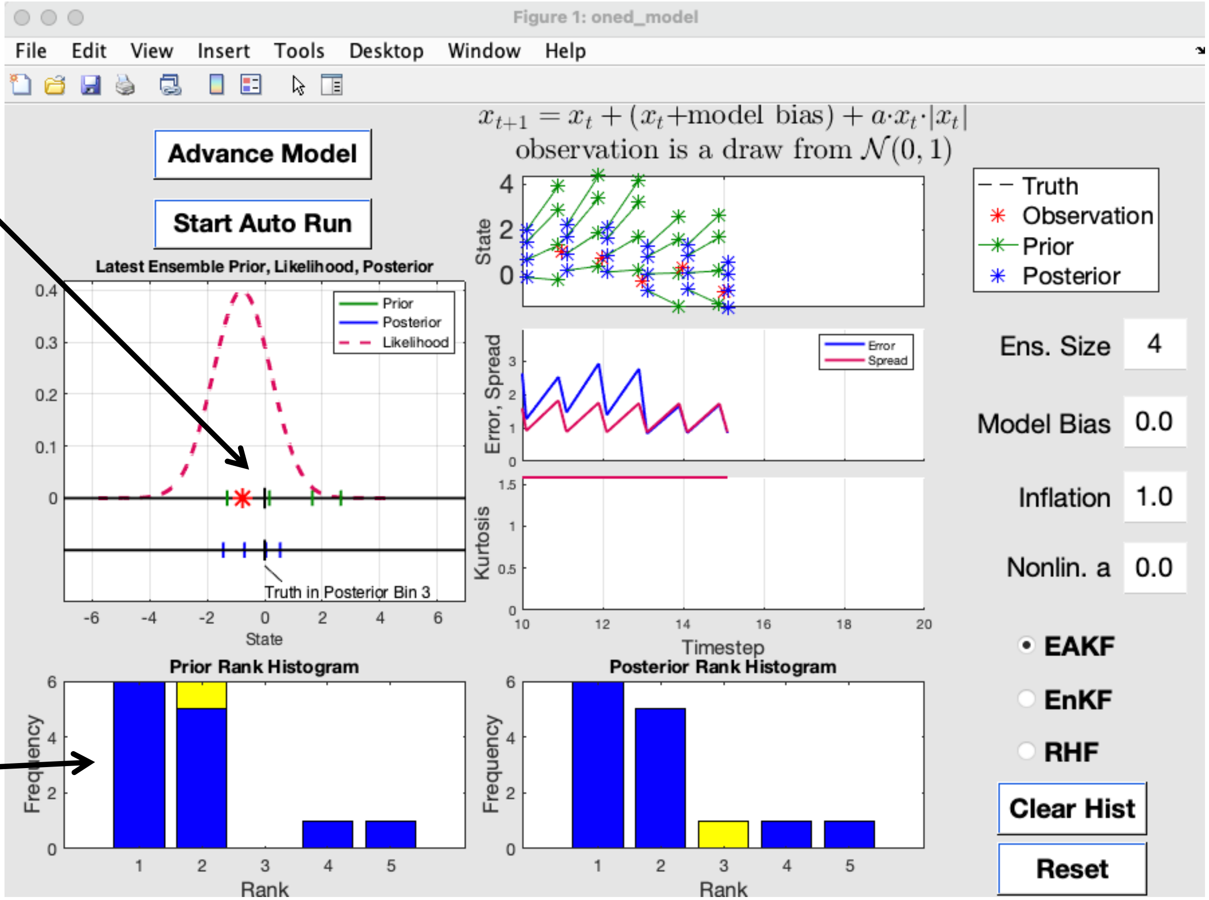
Ensembles with too much spread are peaked in the center.



Matlab Hands-On: oned_model
Understanding the Rank Histogram

Truth is always 0. For this time, truth is between 1st and 2nd prior ensemble members; that's the second bin. It's in the 3rd posterior bin.

Prior (left) and Posterior (right) rank histograms. Yellow is for current time.



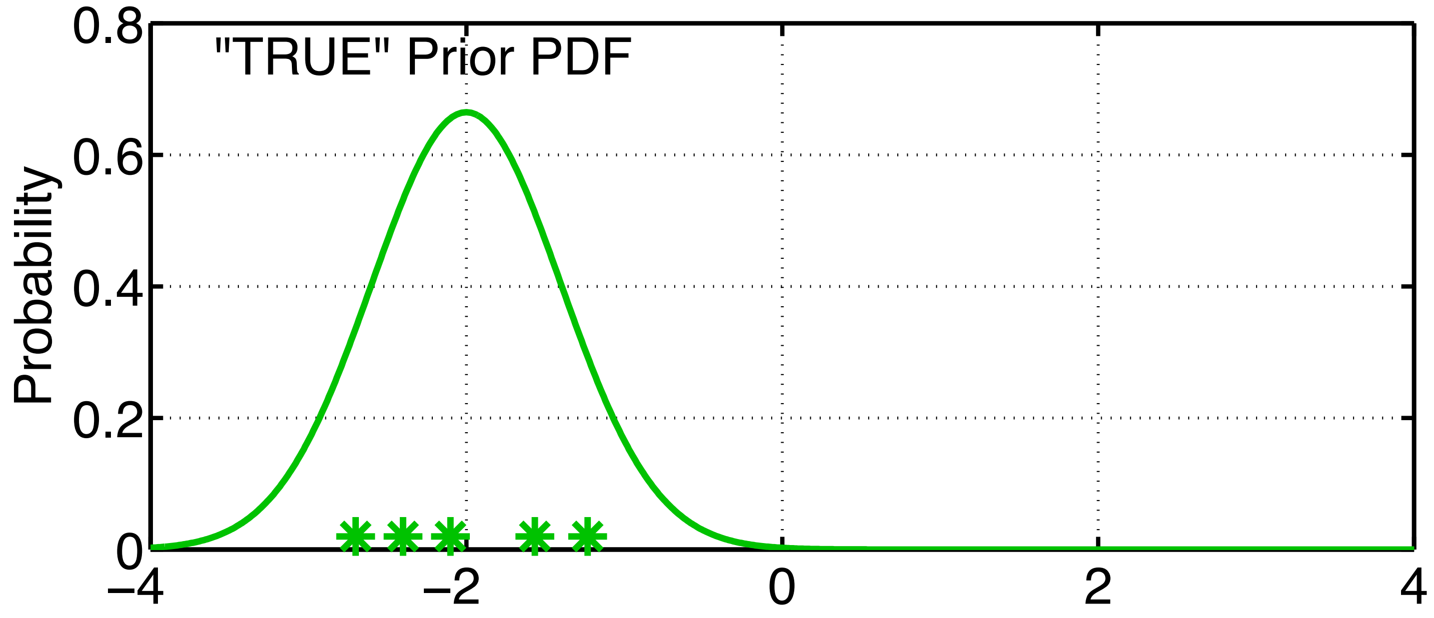
Matlab Hands-On: oned_model
Understanding the Rank Histogram

Explorations:

1. Step through a sequence of advances and assimilations with the top button. Watch the evolution of the rank histogram bins.
2. Add some model bias (less than 1 to start) and see how the filter responds.
3. Add some nonlinearity (a< 1) to the model. How do the different filters respond?
4. Can you break the filter (find setting so that the ensemble moves away from zero) with the options explored so far?

Dealing with systematic error: Variance Inflation

Observations + physical system → 'true' distribution.

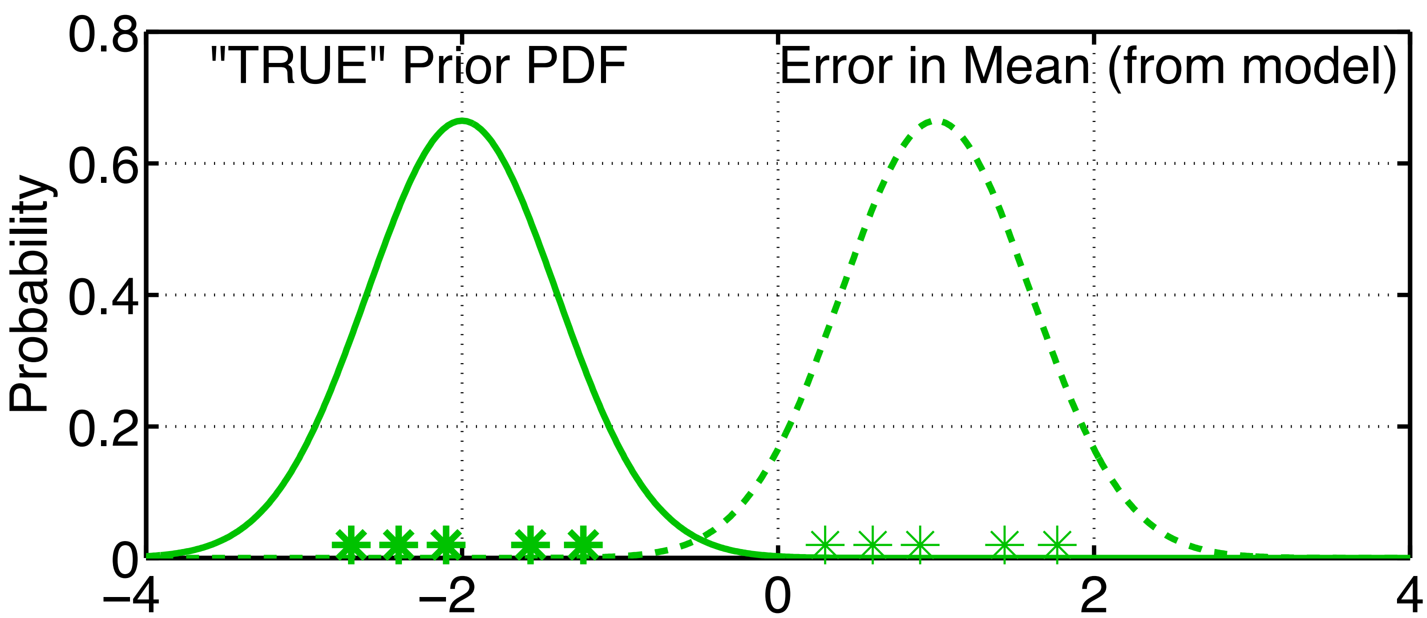


Dealing with systematic error: Variance Inflation

Observations + physical system → 'true' distribution.

Model bias (and other errors) can shift actual prior.

Prior ensemble is too certain (needs more spread).

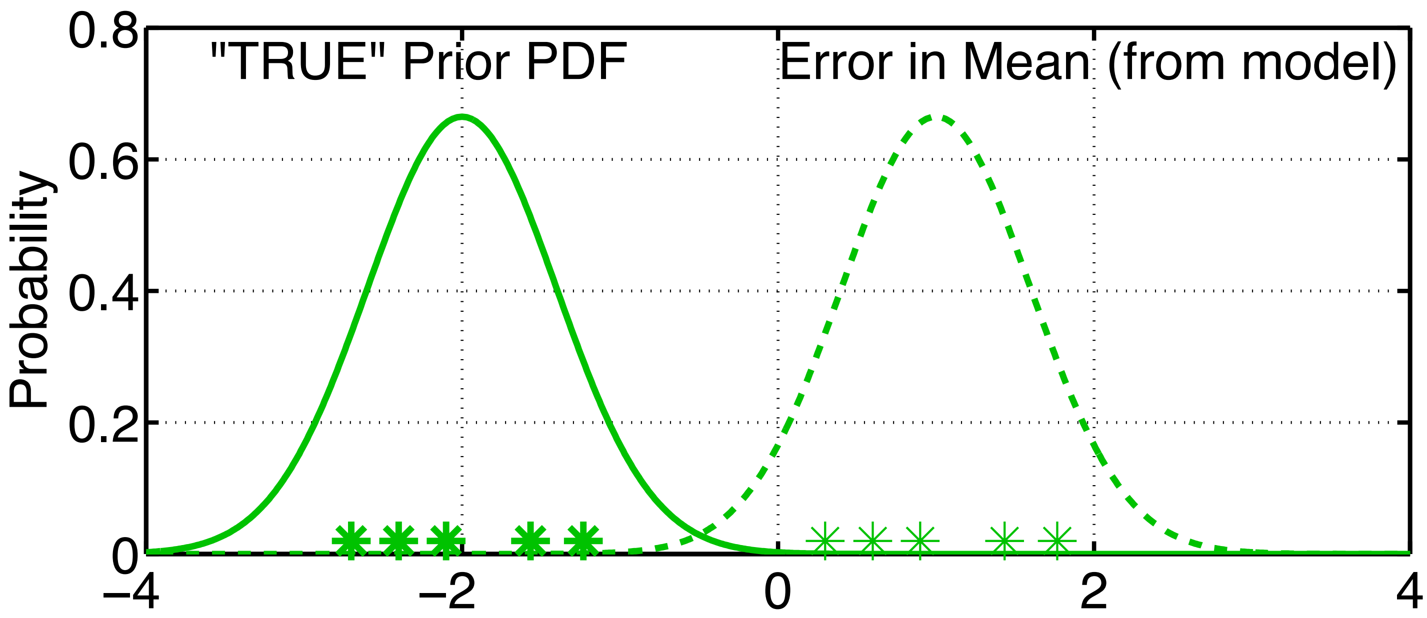


Dealing with systematic error: Variance Inflation

Could correct error if we knew what it was.

With large models, can't know error precisely.

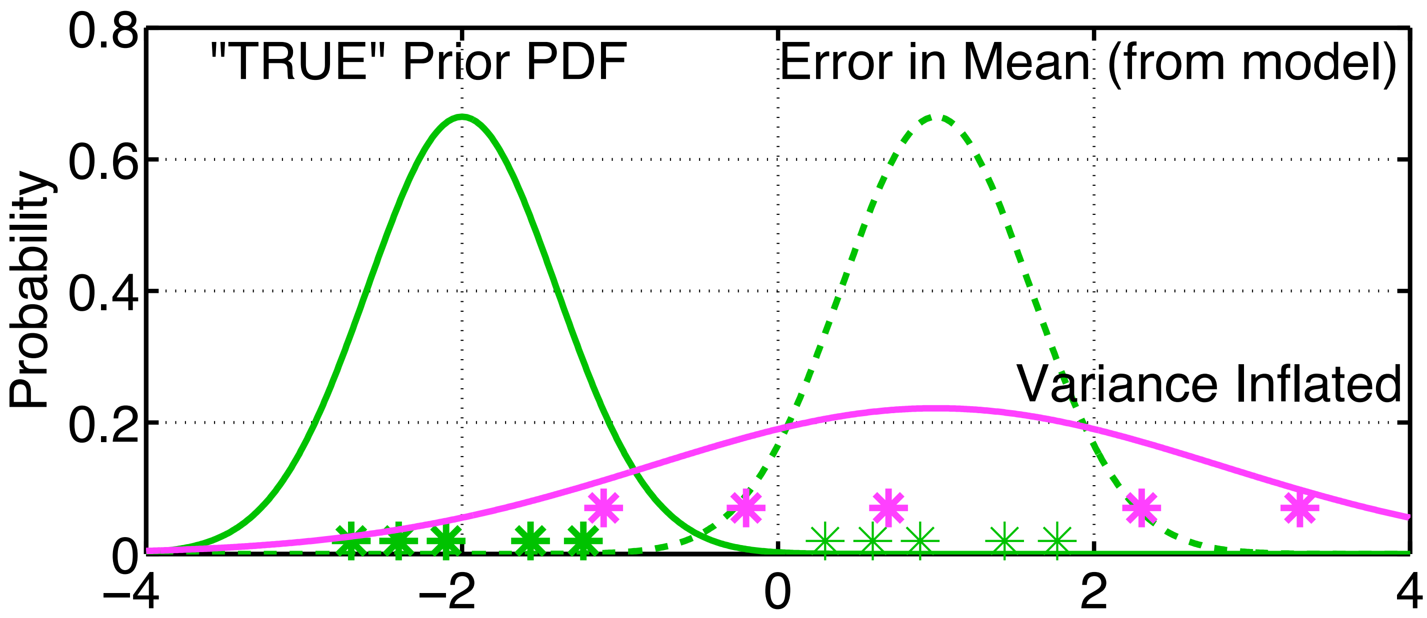
Taking no action can cause observations to be ignored.



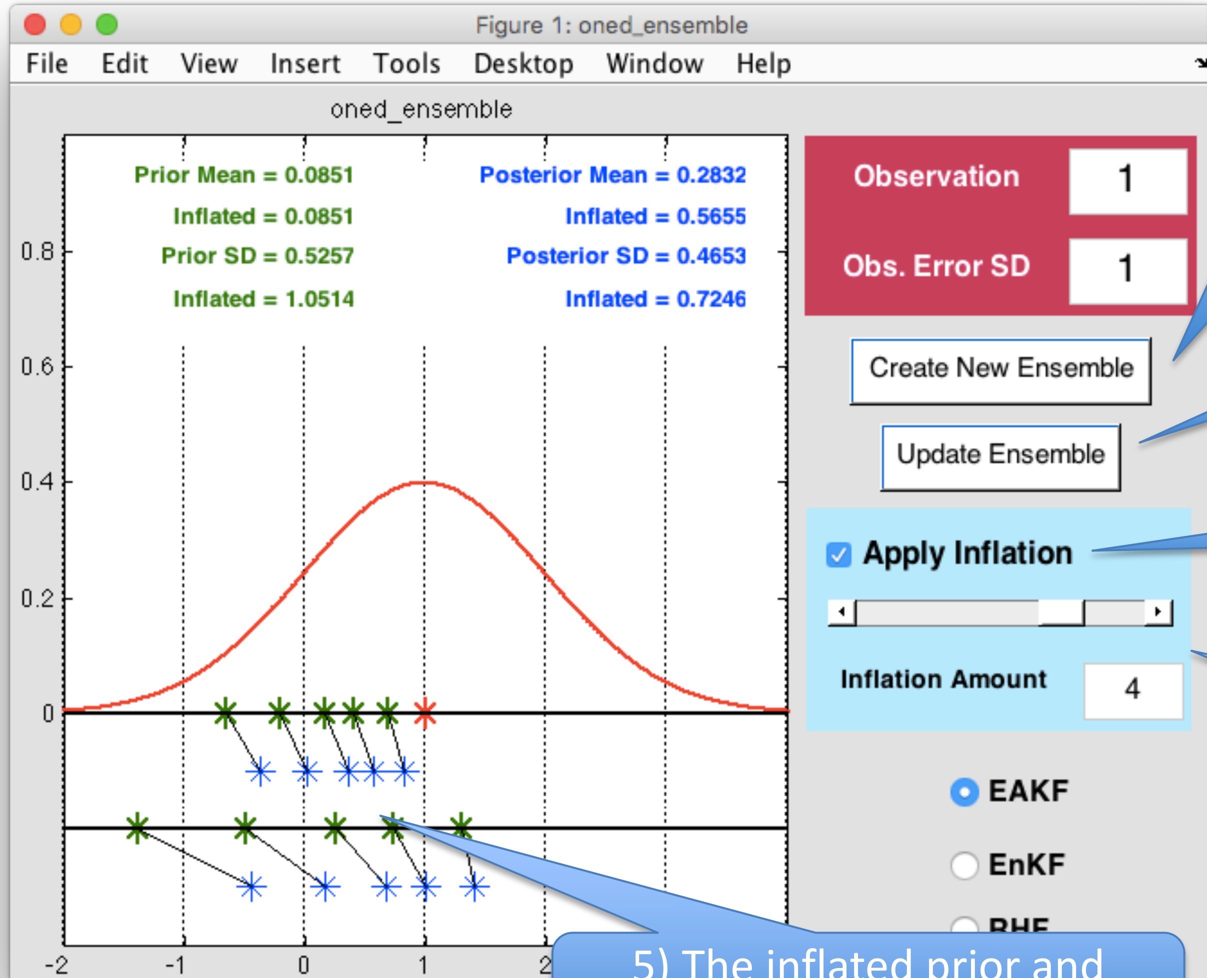
Dealing with systematic error: Variance Inflation

Naïve solution: increase the spread in the prior.

Give more weight to the observation, less to the prior.



Matlab Hands-On: oned_ensemble exploring prior inflation



1) Create a new ensemble.

4) Do assimilation.

2) Turn inflation on.

3) Set an inflation value.

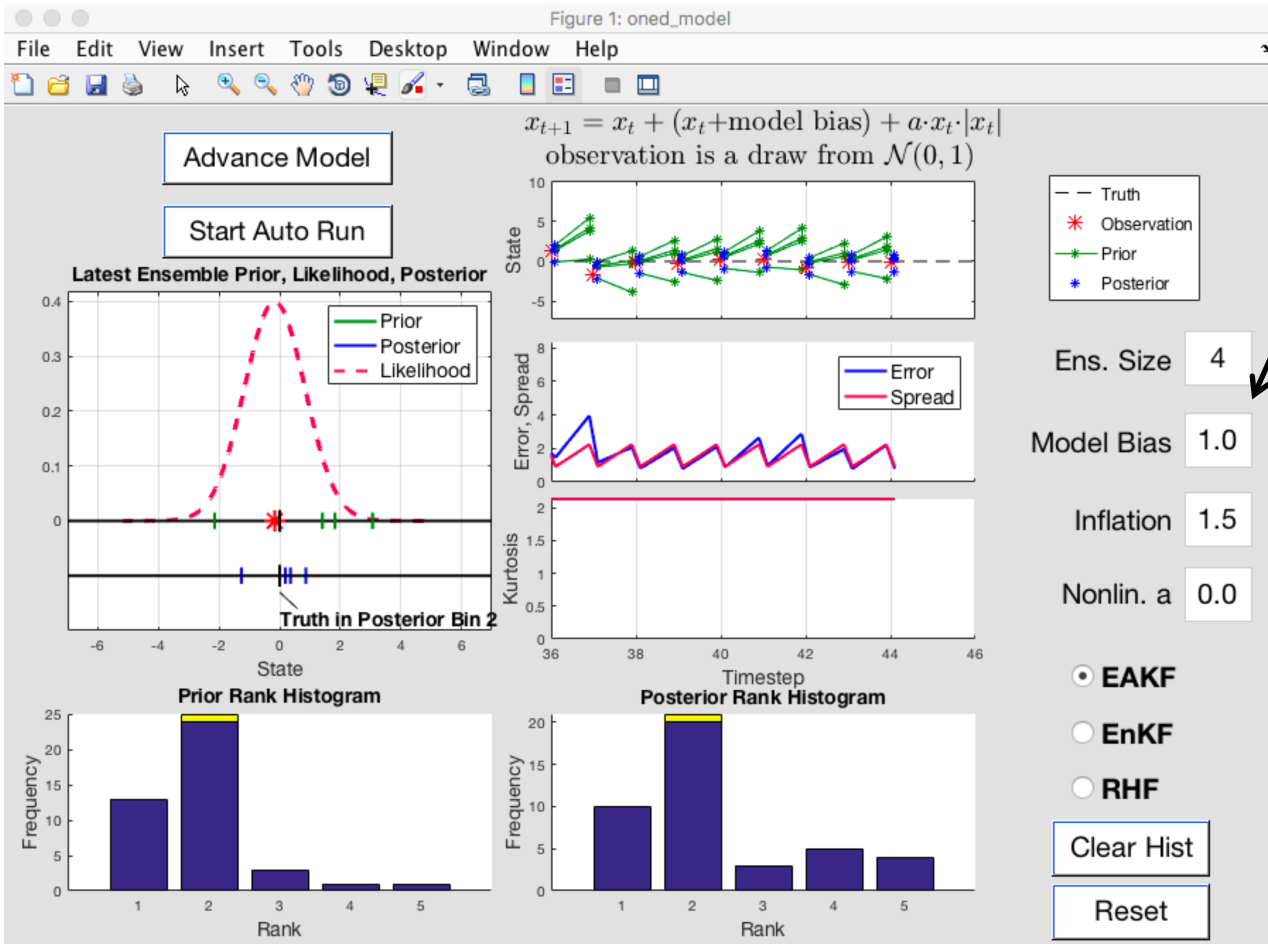
5) The inflated prior and posterior shows up here.

Matlab Hands-On: exploring prior inflation with oned_ensemble

Explorations:

- See how increasing inflation (> 1) changes the posterior mean and standard deviation.
- Look at priors that are not shifted but have small spread compared to the observation error distribution.
- Look at priors that are shifted from the observation.

Matlab Hands-On: oned_model using inflation to deal with systematic error



1. Add some model bias to simulate systematic error.
2. Run an assimilation and observe the error, spread, and rank histograms.
3. Add some inflation (try starting with 1.5) and observe how behavior changes.
4. What happens with too much inflation?

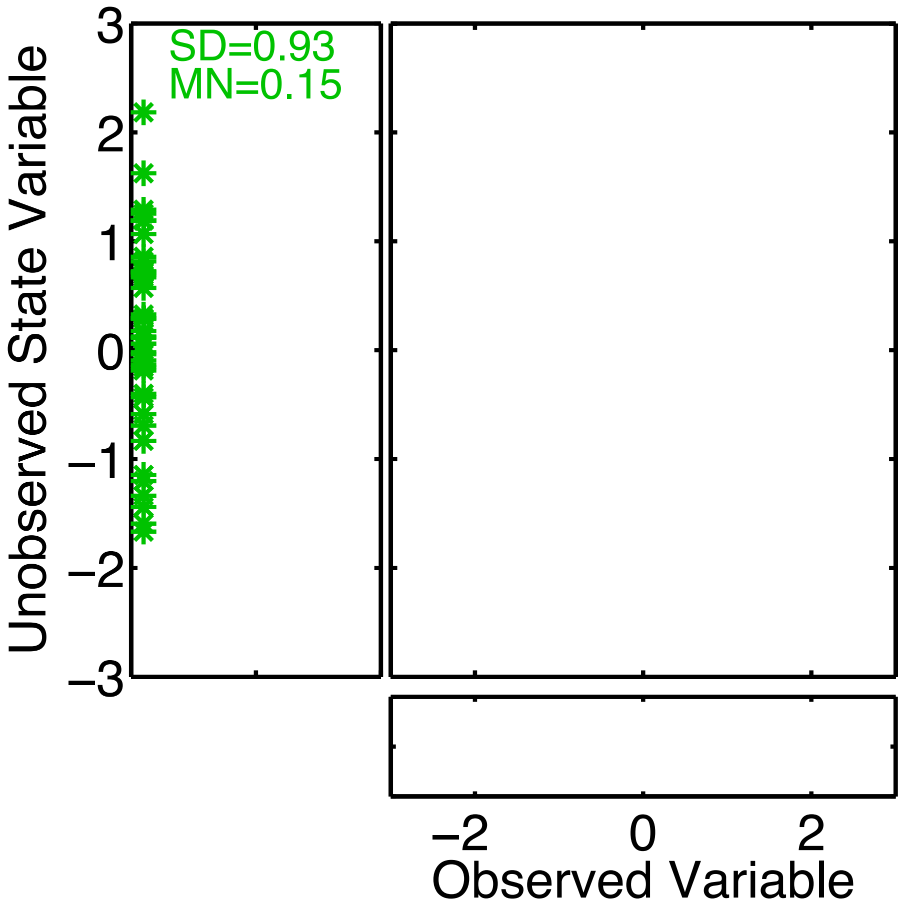
Note: The spread is increased by the square root of the variance inflation.

Matlab Hands-On: oned_model using inflation to deal with systematic error

Explorations:

- Try a variety of model bias and inflation settings.
- Try using inflation with a nonlinear model.

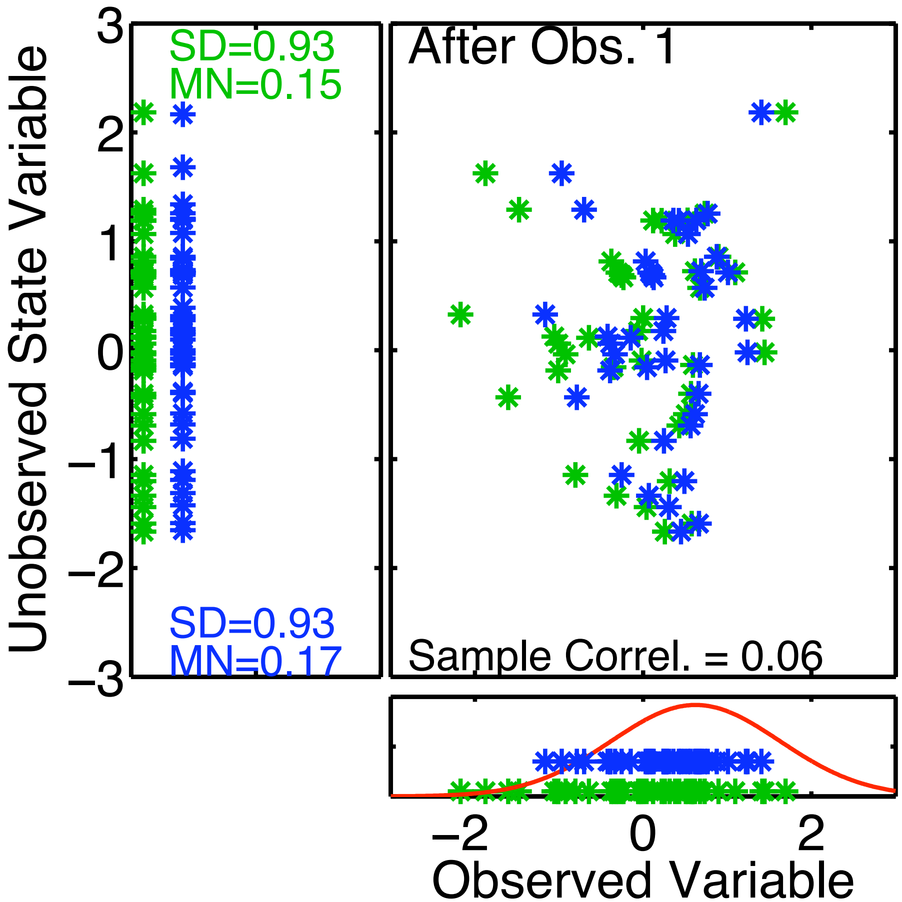
Regression Sampling Error



Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved variable should remain unchanged.

Regression Sampling Error

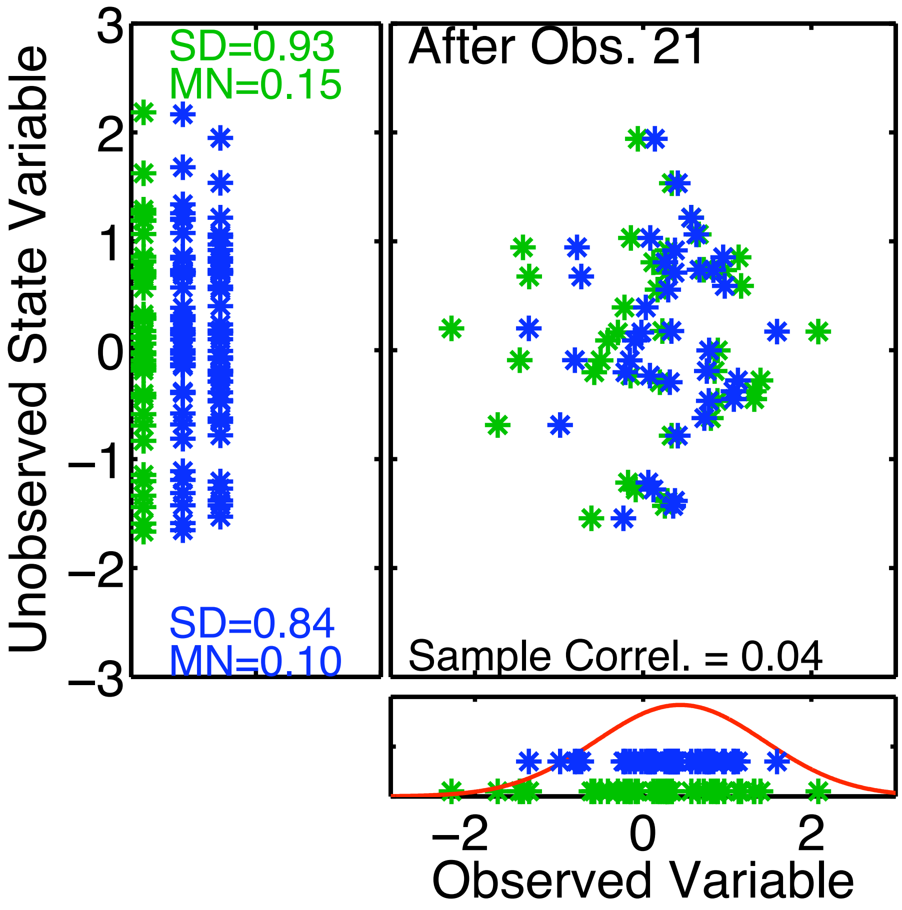


Suppose unobserved state variable is known to be unrelated to observed variables.

Finite samples from joint distribution have non-zero correlation, expected $|corr| \approx 0.19$ for 20 samples.

After one observation, unobserved variable mean and standard deviation change.

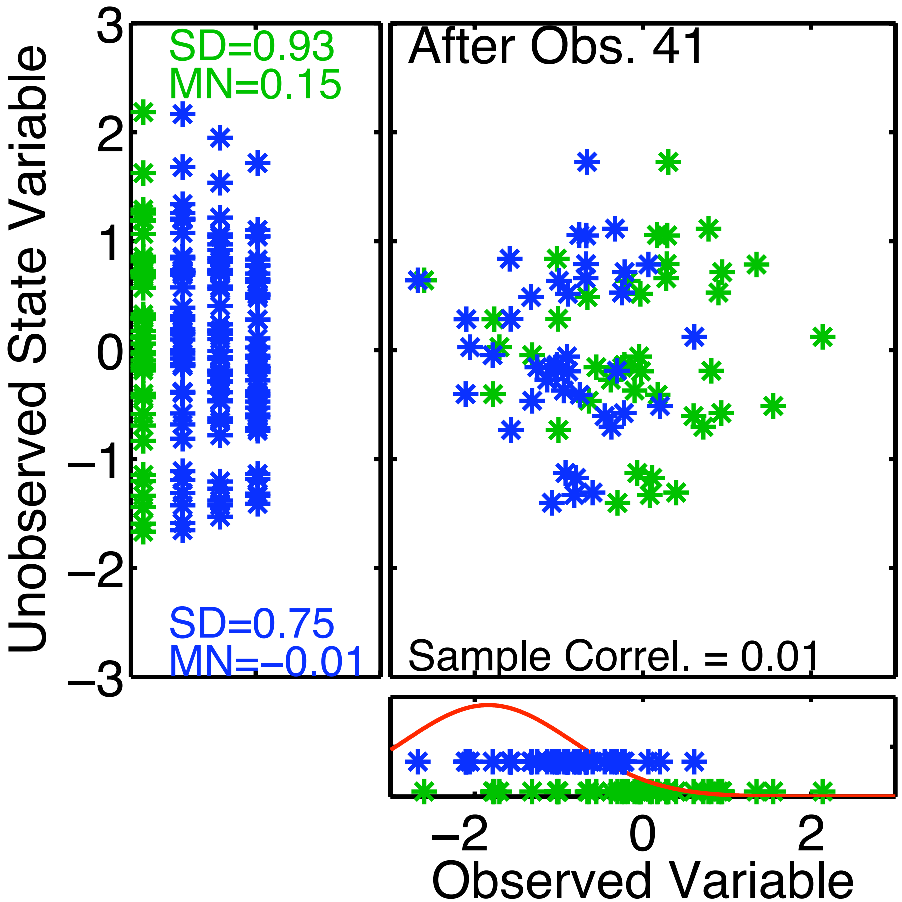
Regression Sampling Error



Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved mean follows a random walk as more observations are used.

Regression Sampling Error

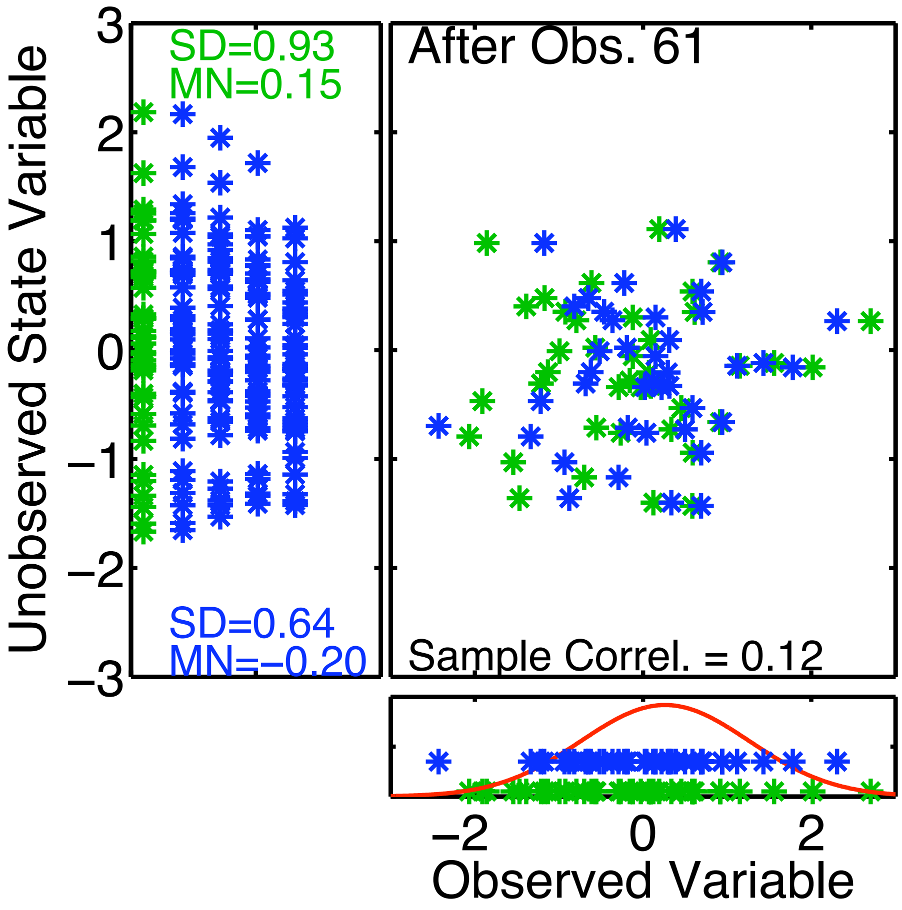


Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved mean follows a random walk as more observations are used.

Unobserved standard deviation consistently decreases.

Regression Sampling Error

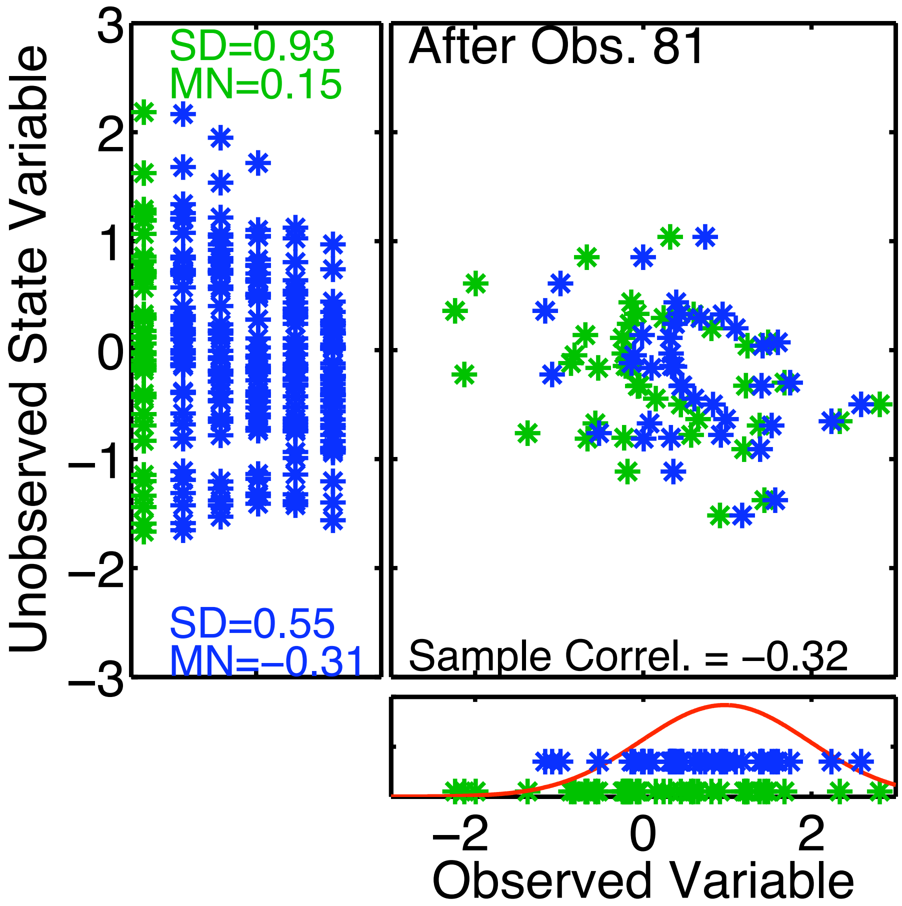


Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved mean follows a random walk as more observations are used.

Unobserved standard deviation consistently decreases.

Regression Sampling Error

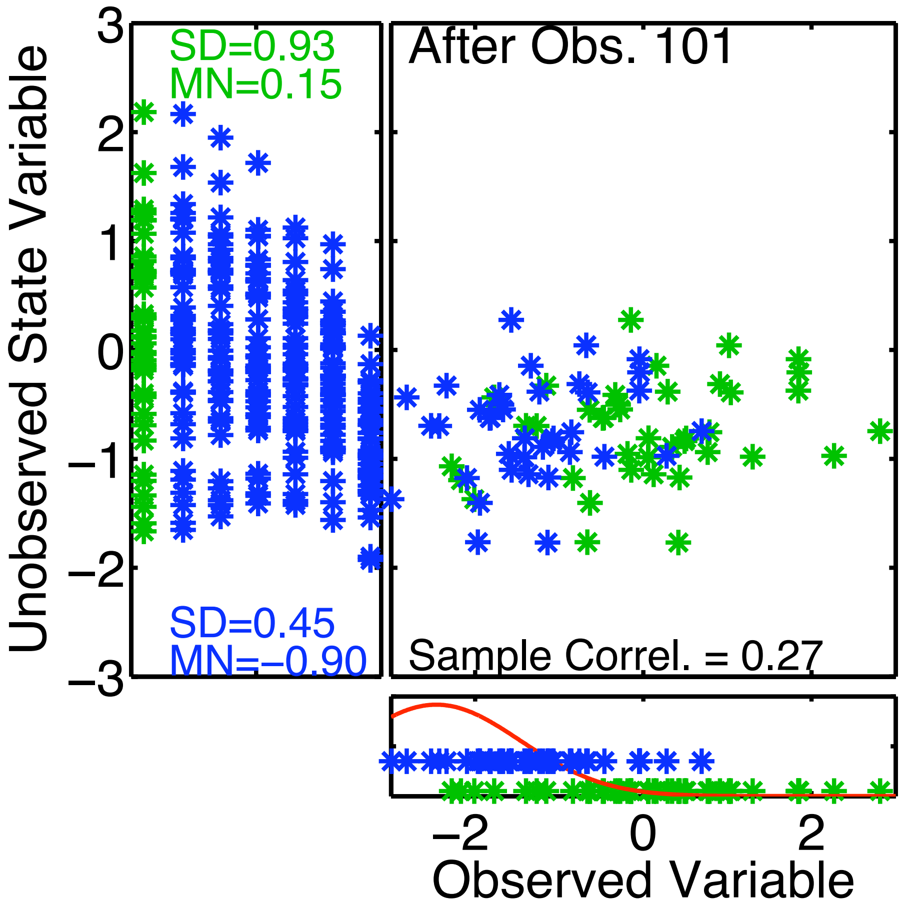


Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved mean follows a random walk as more observations are used.

Unobserved standard deviation consistently decreases.

Regression Sampling Error

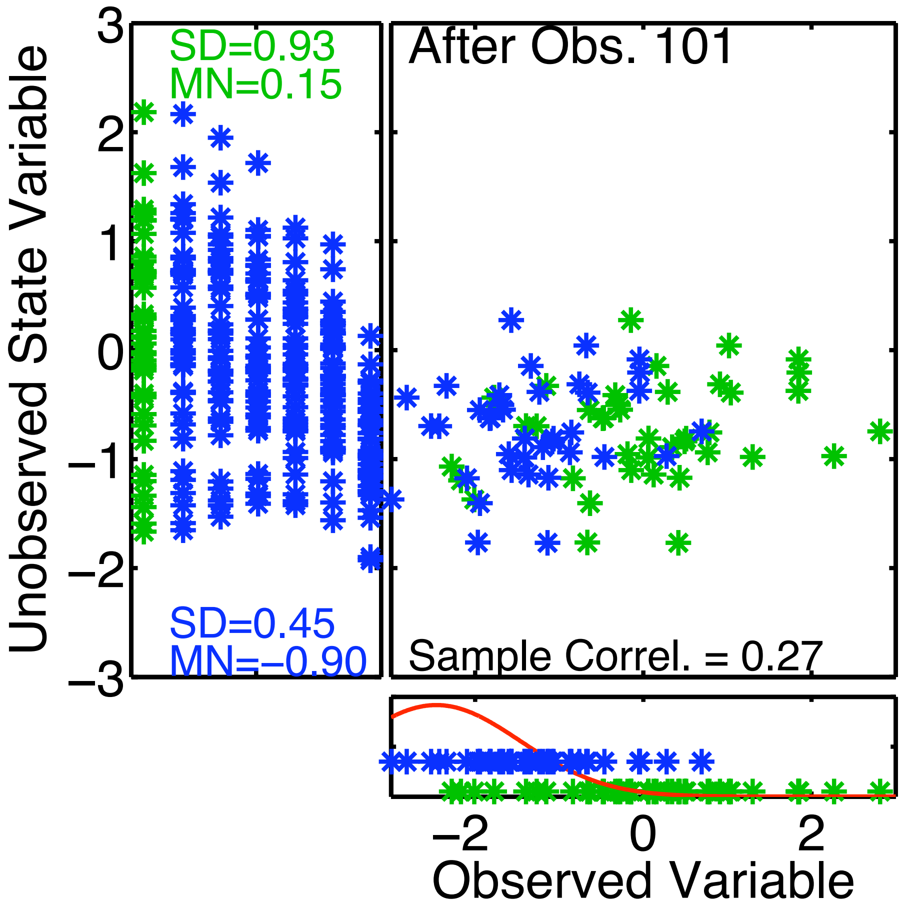


Suppose unobserved state variable is known to be unrelated to observed variables.

Unobserved mean follows a random walk as more observations are used.

Unobserved standard deviation consistently decreases.

Regression Sampling Error



Suppose unobserved state variable is known to be unrelated to observed variables.

- Estimates of unobserved are too confident.
- Give less weight to subsequent meaningful observations.
- Meaningful observations can end up being ignored.

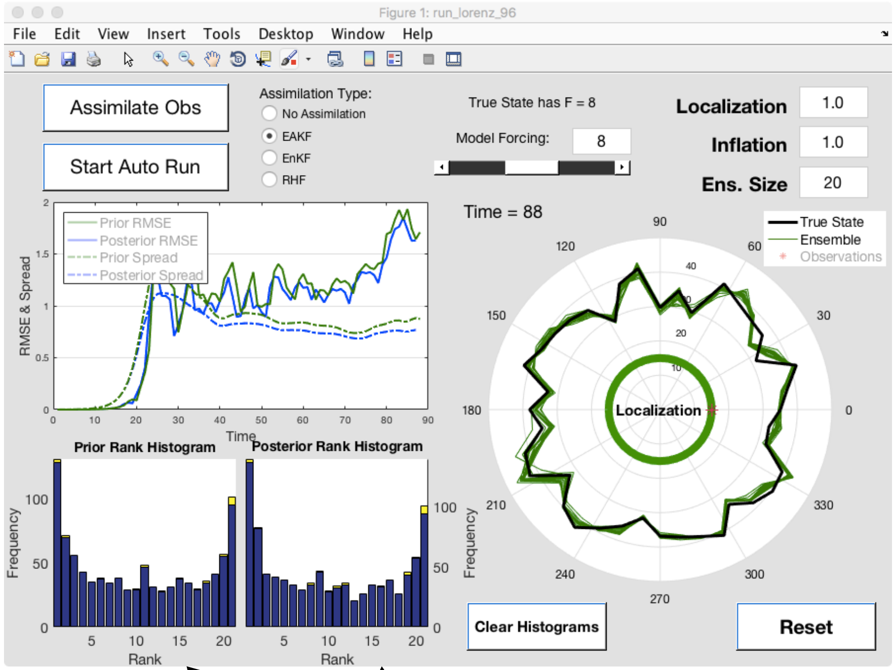
Regression Sampling Error

Ignoring meaningful observations due to insufficient spread is one form of filter divergence.

This could be seen in the Lorenz-96 assimilation example from the end of DART_LAB Section 2.

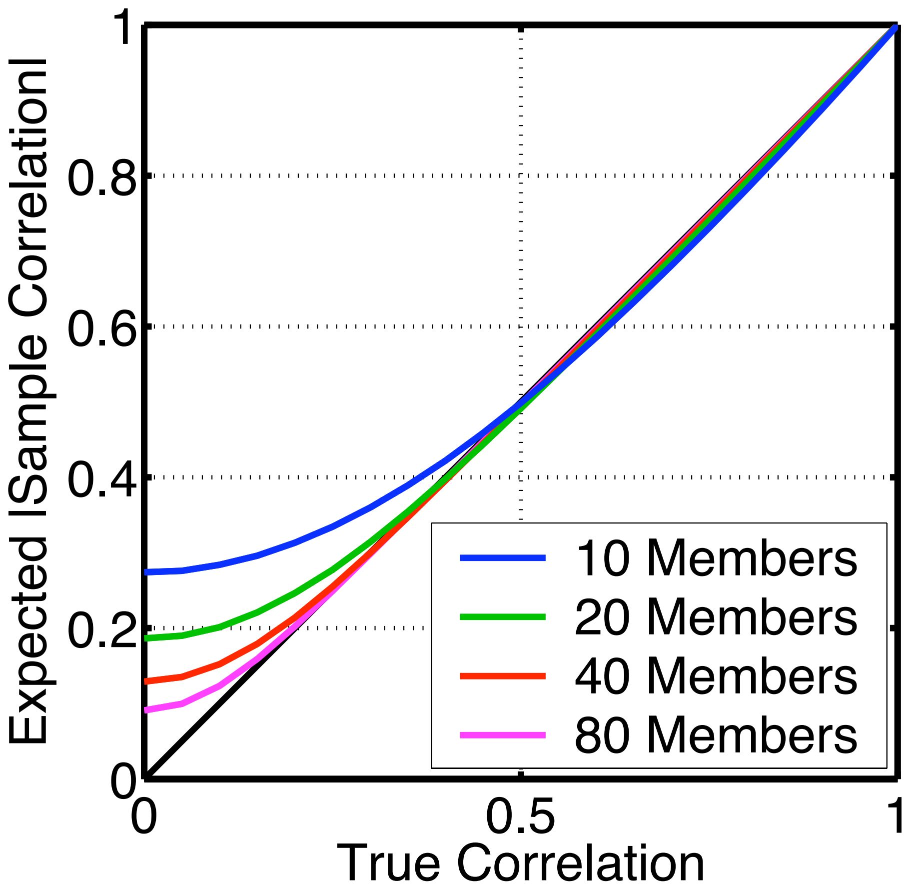
Spread became small, filter thinks it has good estimate.

Error was large because good observations were ignored.



Look how often the truth was outside the ensemble!

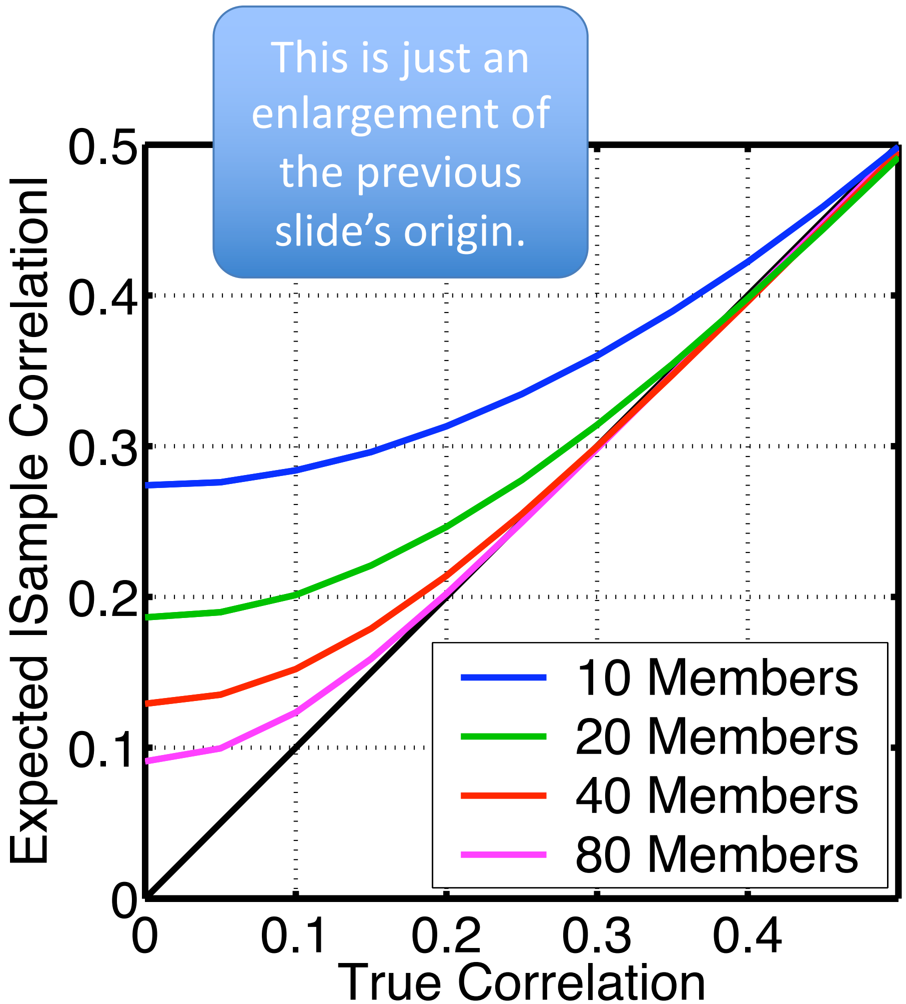
Regression Sampling Error



Absolute value of expected sample correlation vs. true correlation.

Errors decrease for large ensembles and for correlations with absolute value close to 1.

Regression Sampling Error



For small true correlations, sampling errors are undesirably large even for 80 members!

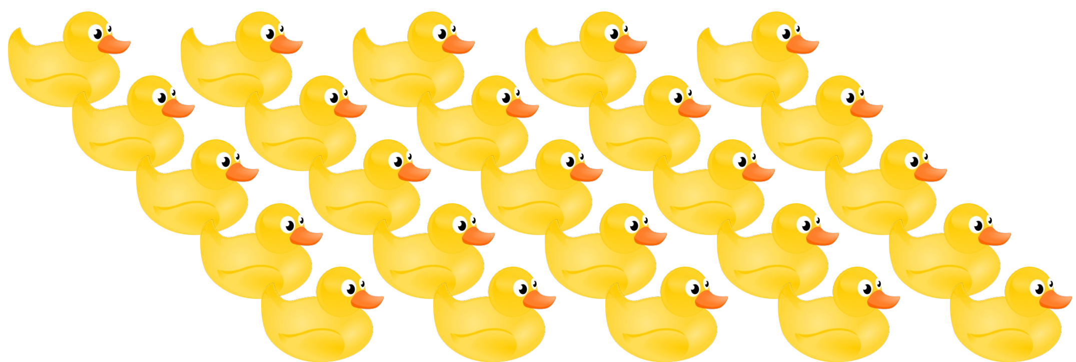
Ways to deal with Regression Sampling Error

1. Ignore it if number of observations that have small correlations with state variables is small and there is a way to maintain prior variance. Worked in Lorenz-63.

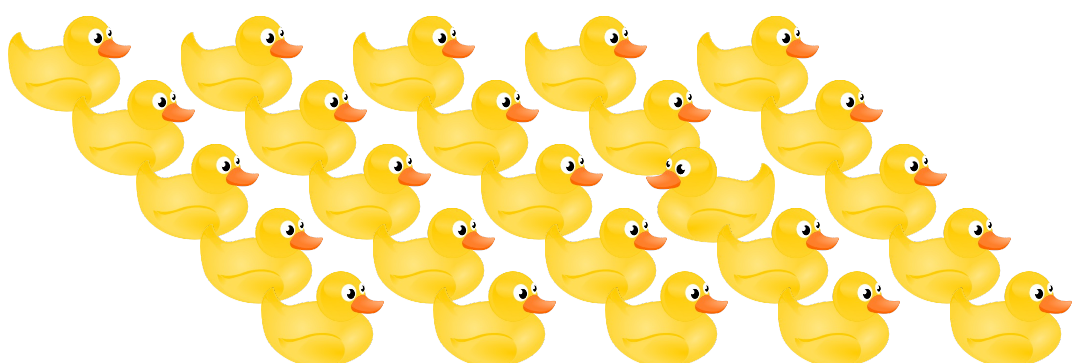
?

It may work.
It may not.

Ways to deal with Regression Sampling Error

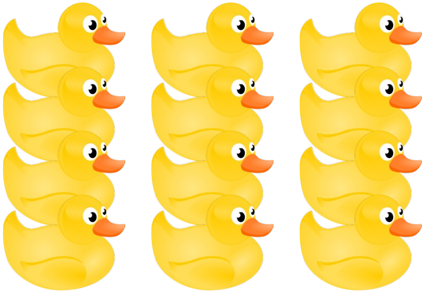


2. Use larger ensembles; expensive for large models.



Ways to deal with Regression Sampling Error

1. Ignore it if number of weakly correlated observations is small and there is a way to maintain prior variance. Worked in Lorenz-63.



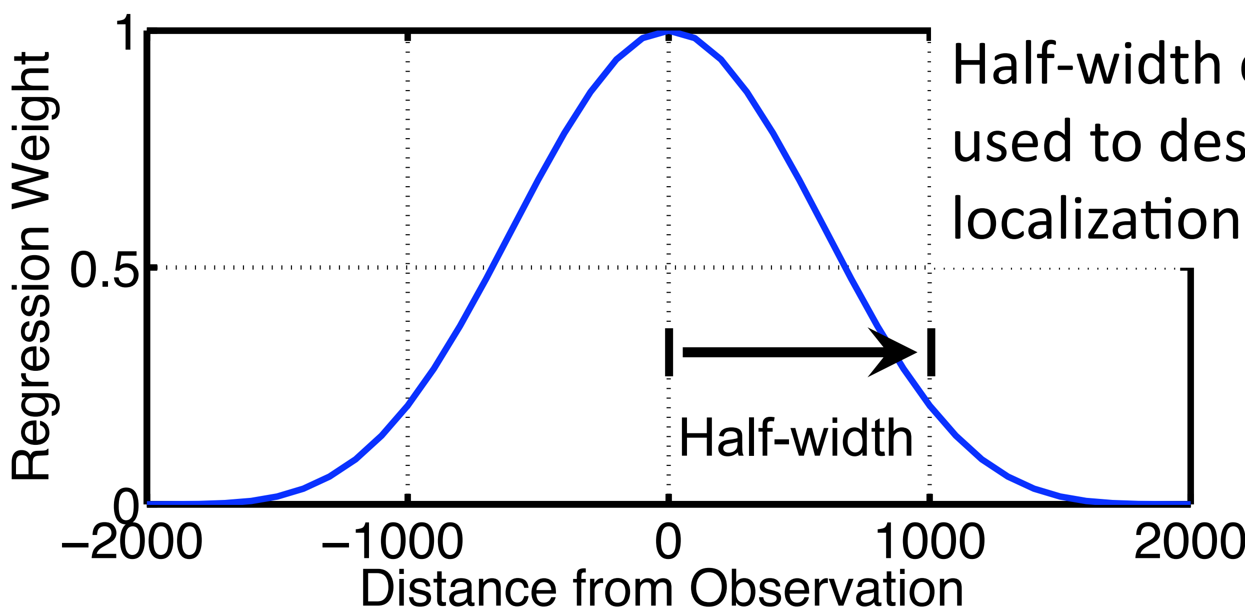
2. Use larger ensembles; expensive for large models.

3. Use additional a priori information about relation between observations and state variables.
   Don't let an observation impact a state variable if they are <u>known</u> to be unrelated.

Using additional a priori information: Localization

Try reducing regression factor as function of distance between observation and state variable.

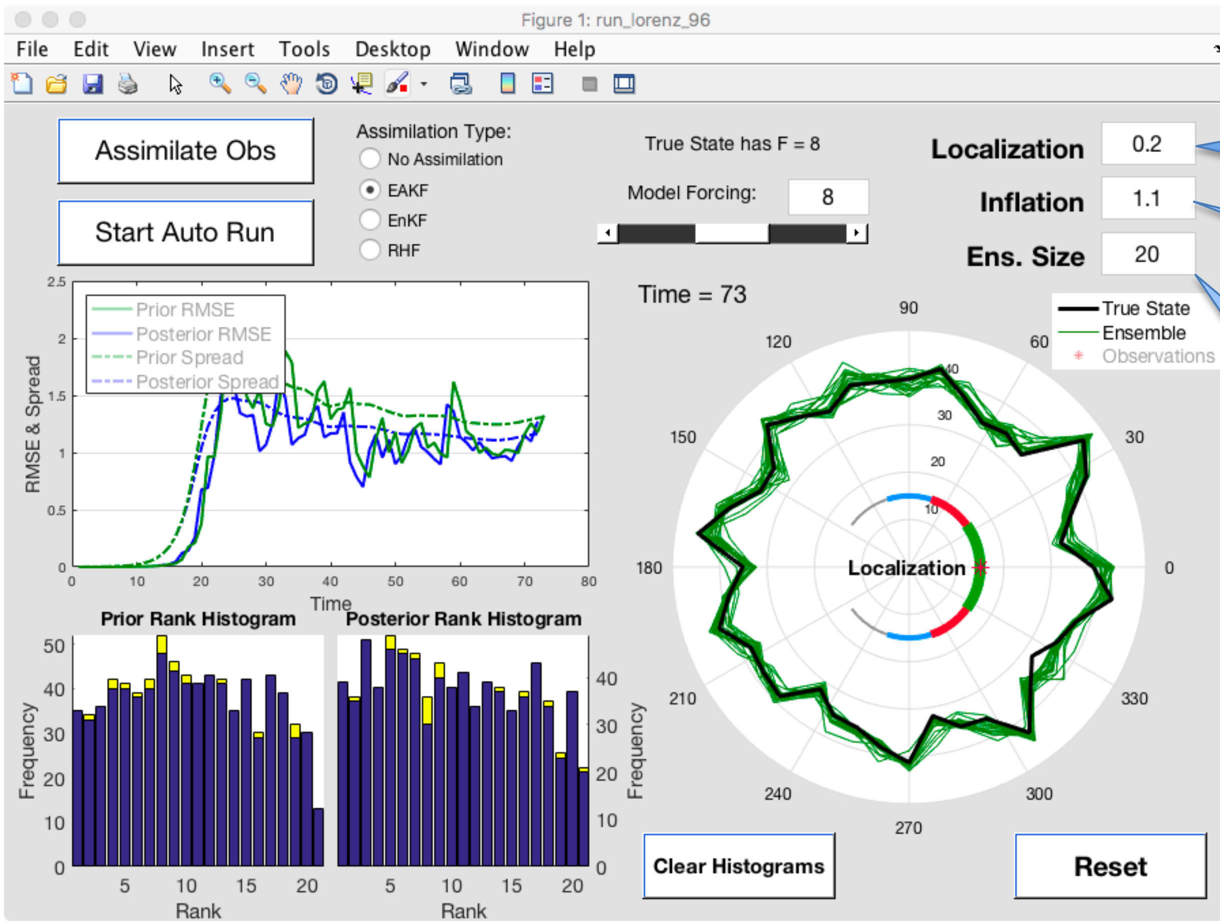
Compactly-supported 5th order polynomial (Gaspari-Cohn) is most commonly used for geophysics.



Half-width commonly used to describe localization function.

Matlab Hands-On: **run_lorenz_96** exercise 2

Purpose: Explore how localization, ensemble size, and inflation impact Lorenz-96 assimilation.



Choose a Localization. Units are fraction of domain.

Choose an Inflation.

Choose an ensemble size.

Matlab Hands-On: **run_lorenz_96** exercise 2

Explore!

- Do an extended free run to see error growth in the ensemble.
- Select EAKF and set localization to 0.2, try an assimilation. Note: the distance around the periodic domain is 1. A 0.2 half-width means an observation has no impact on state variables on the opposite side of the domain.
- Turn the localization off (set it to 1000000) and try a larger ensemble. This may be slow with Matlab; you really need to use the real Fortran DART!
- Try reducing the ensemble size to 10 and varying localization.
- Try adding in some inflation with localization.
- Try selecting model error by changing forcing for the assimilation.