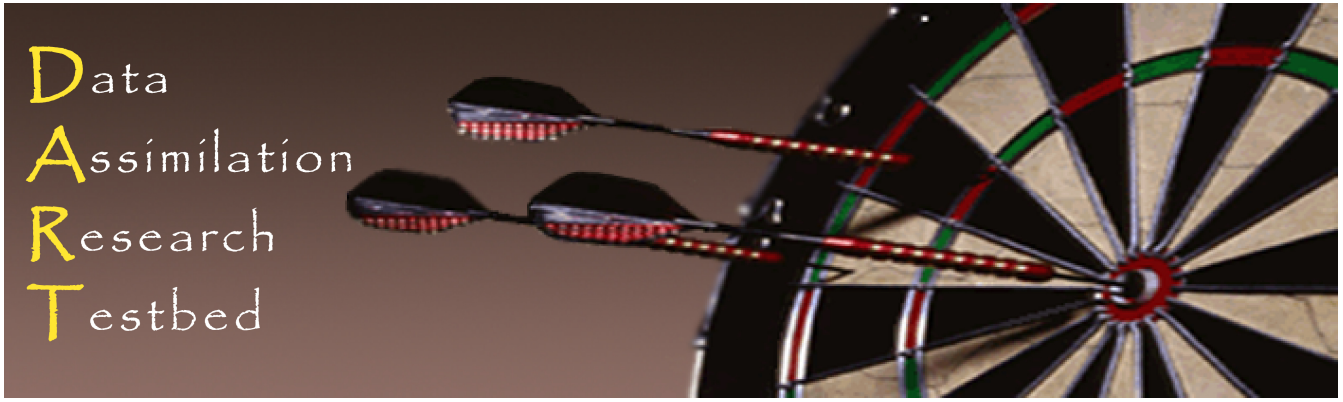
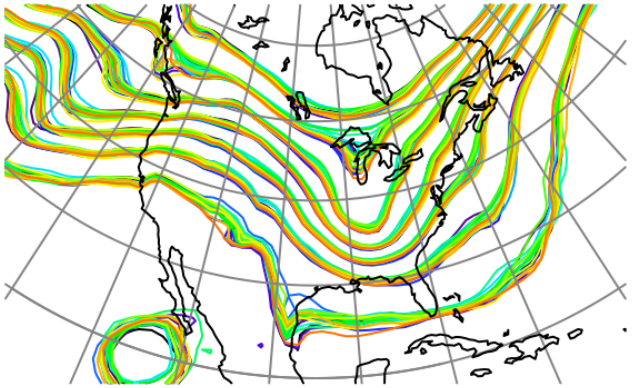


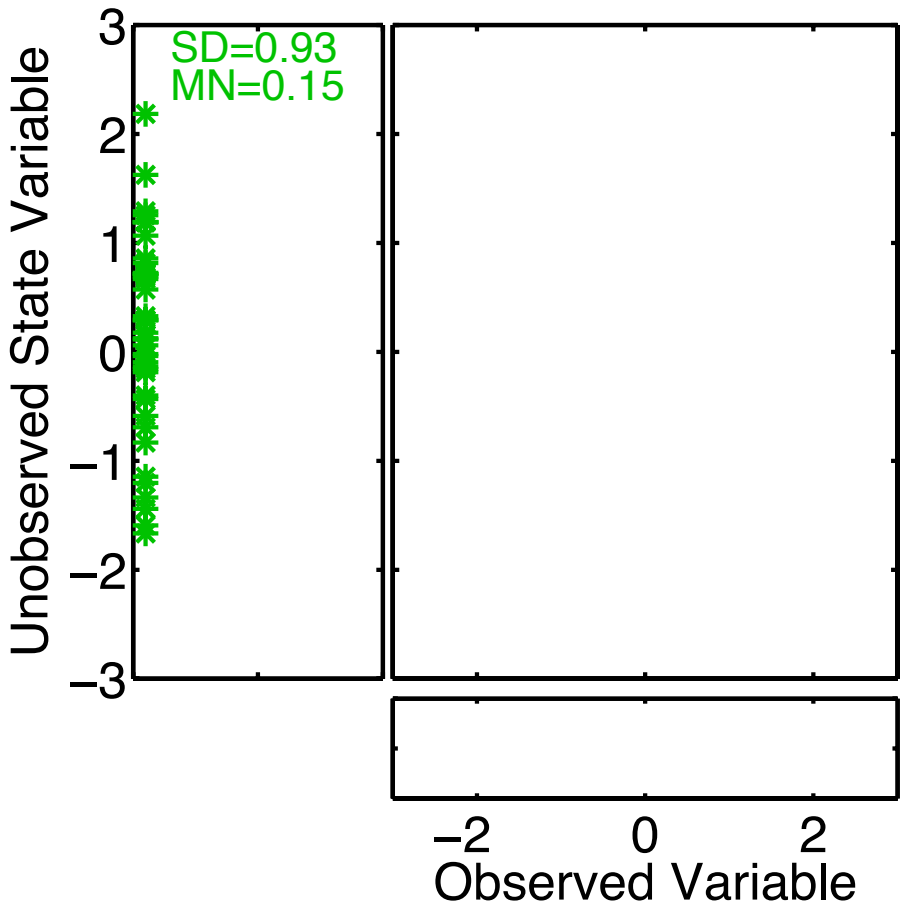
DART_LAB Tutorial Section 3: Sampling error and localization.



The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

NCAR | National Center for
UCAR | Atmospheric Research

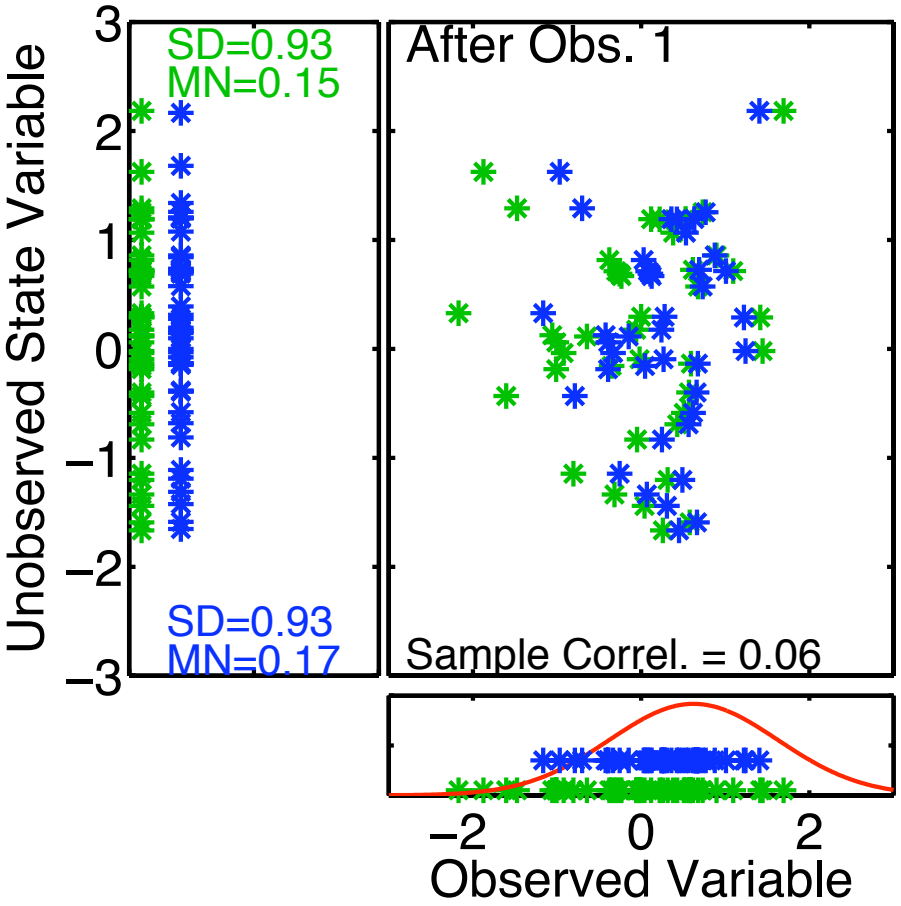
Regression Sampling Error



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

Unobserved variable should remain unchanged.

Regression Sampling Error

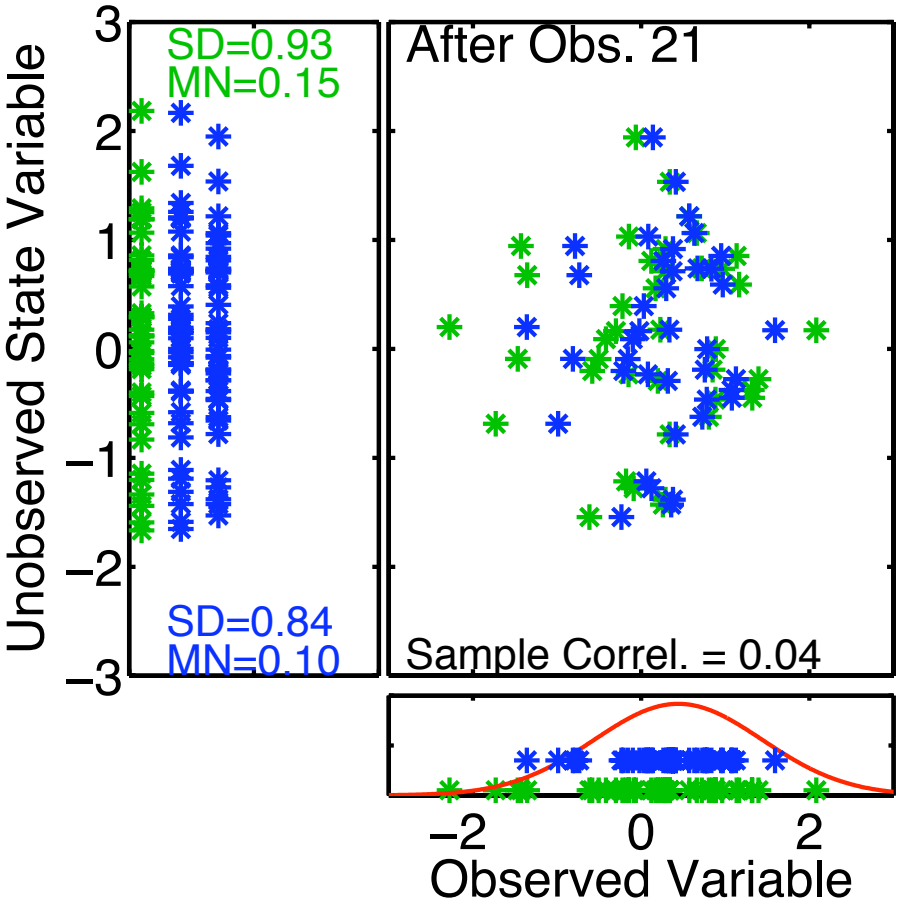


Suppose unobserved state variable is known to be unrelated to observed variables.

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

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

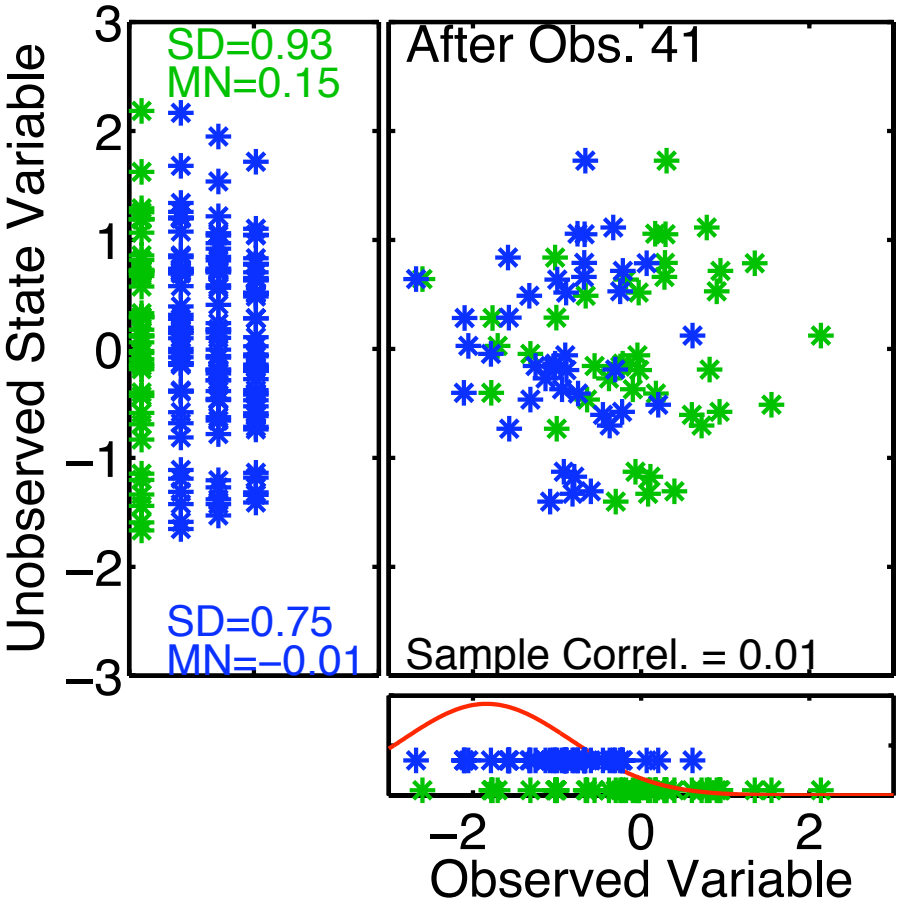
Regression Sampling Error



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

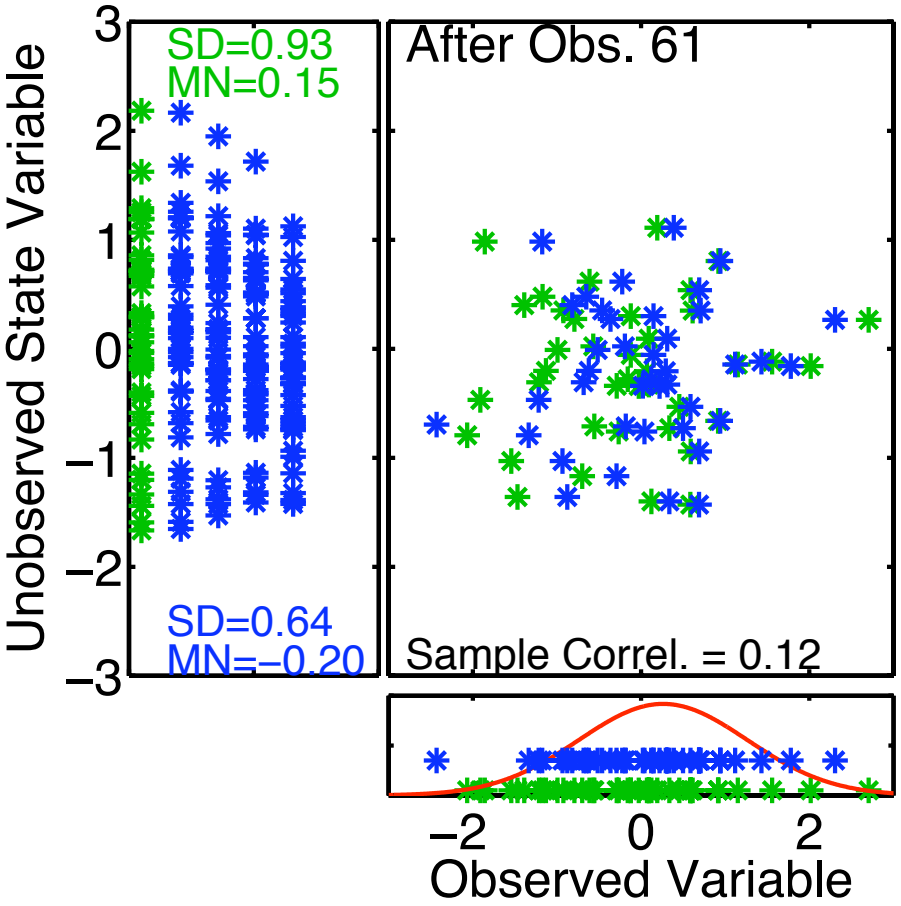


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

Regression Sampling Error

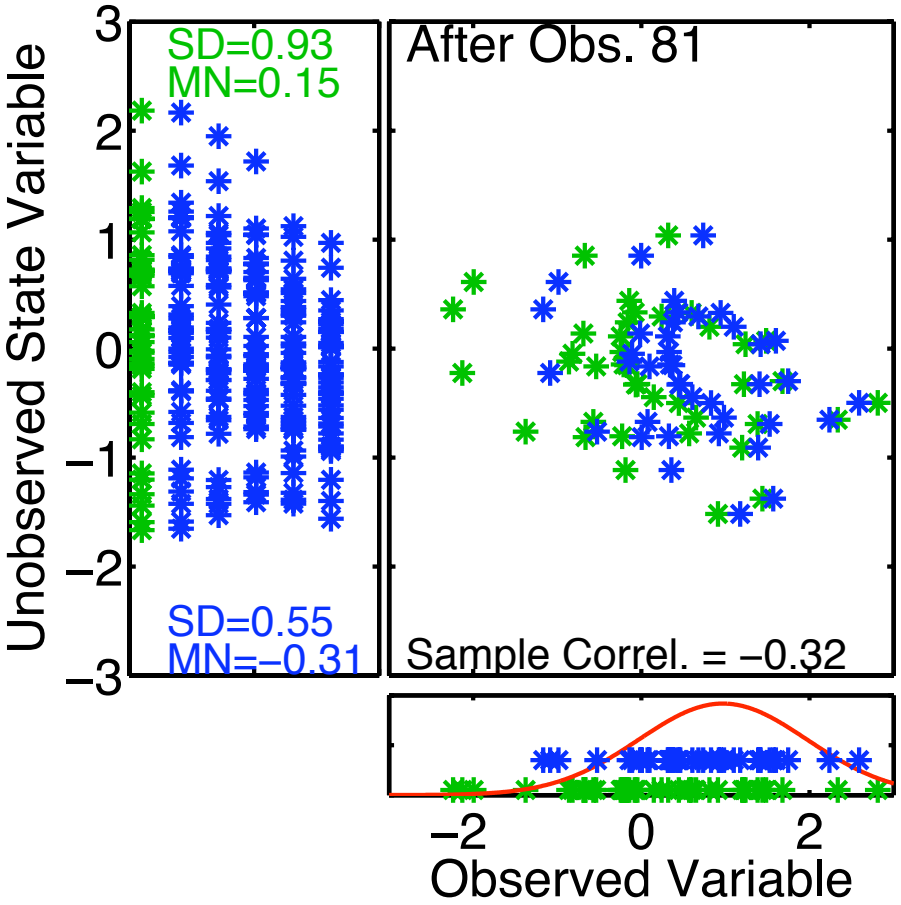


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

Regression Sampling Error

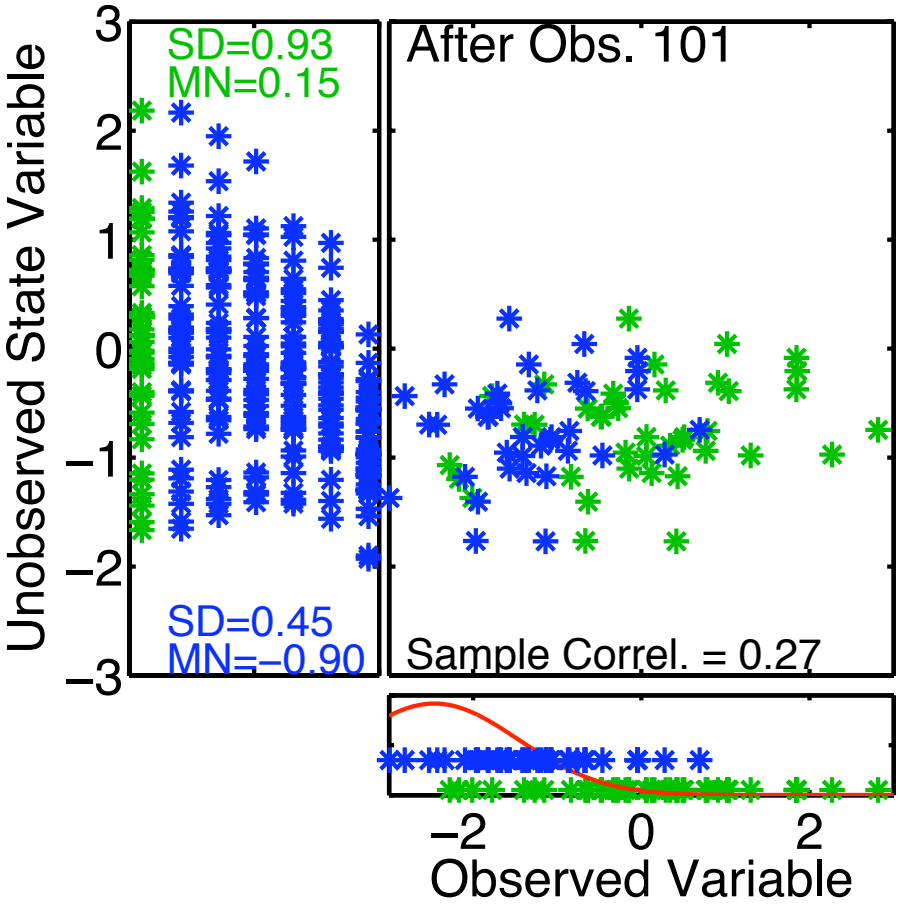


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

Regression Sampling Error

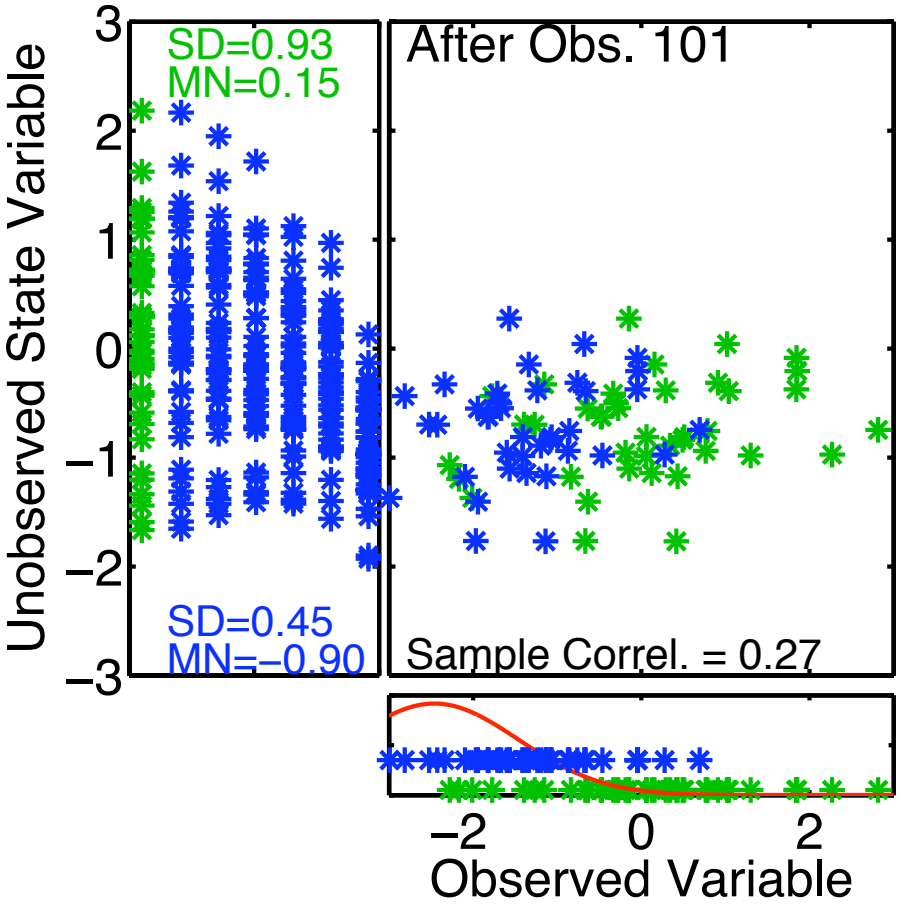


Suppose unobserved state variable is known to be unrelated to observed variables.

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.

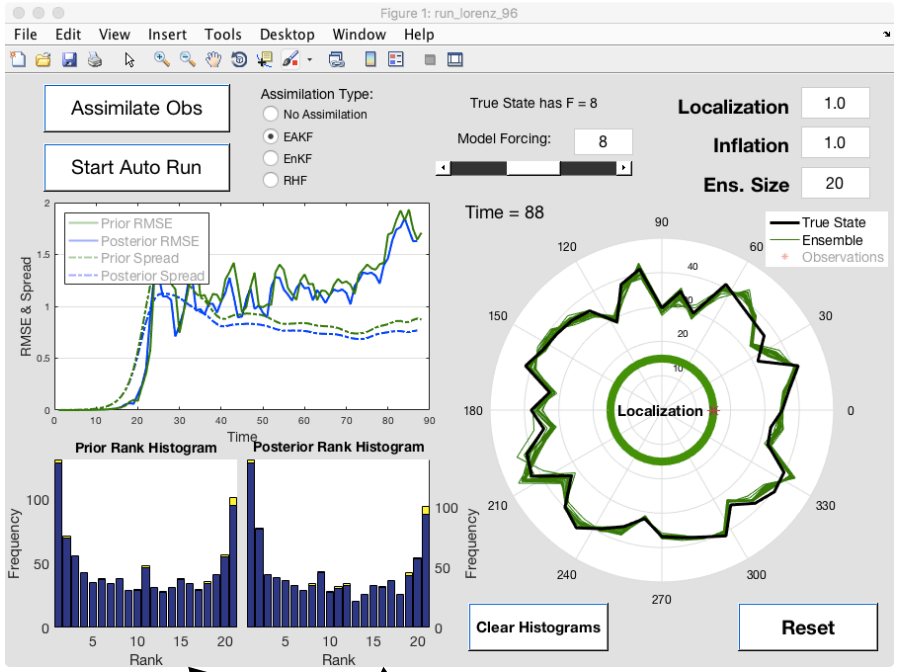
Regression Sampling Error

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.

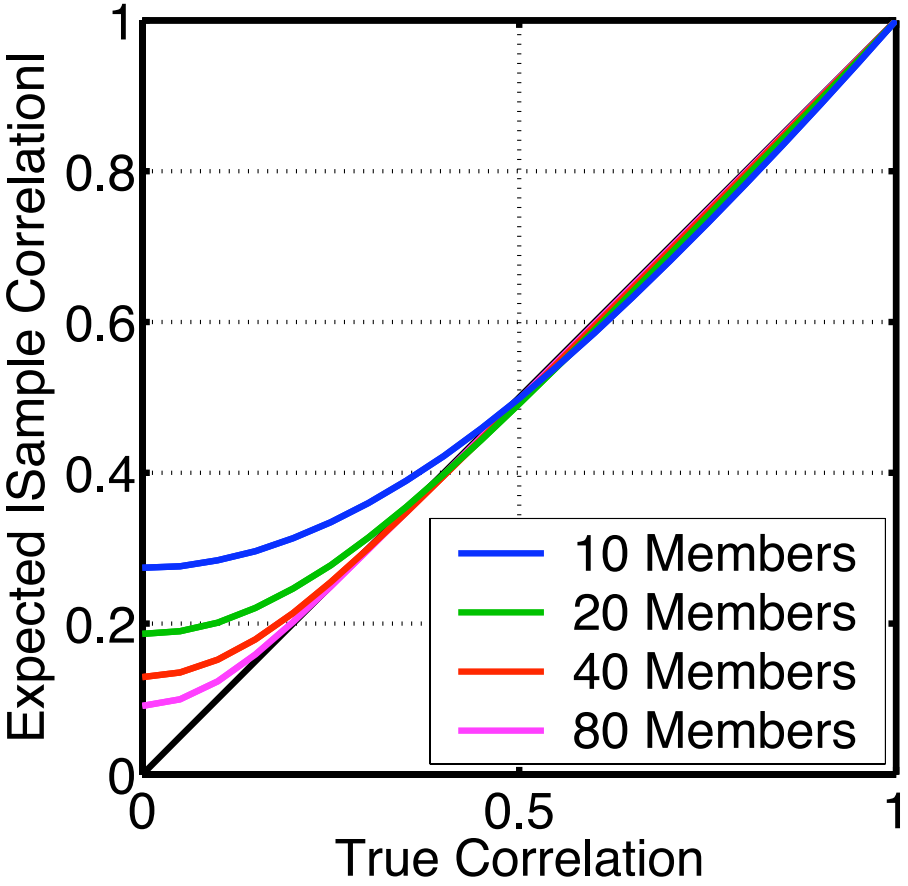
Spread became small, filter thinks it has good estimate.

Error was large because good observations were ignored.



Look how often the truth was outside the ensemble!

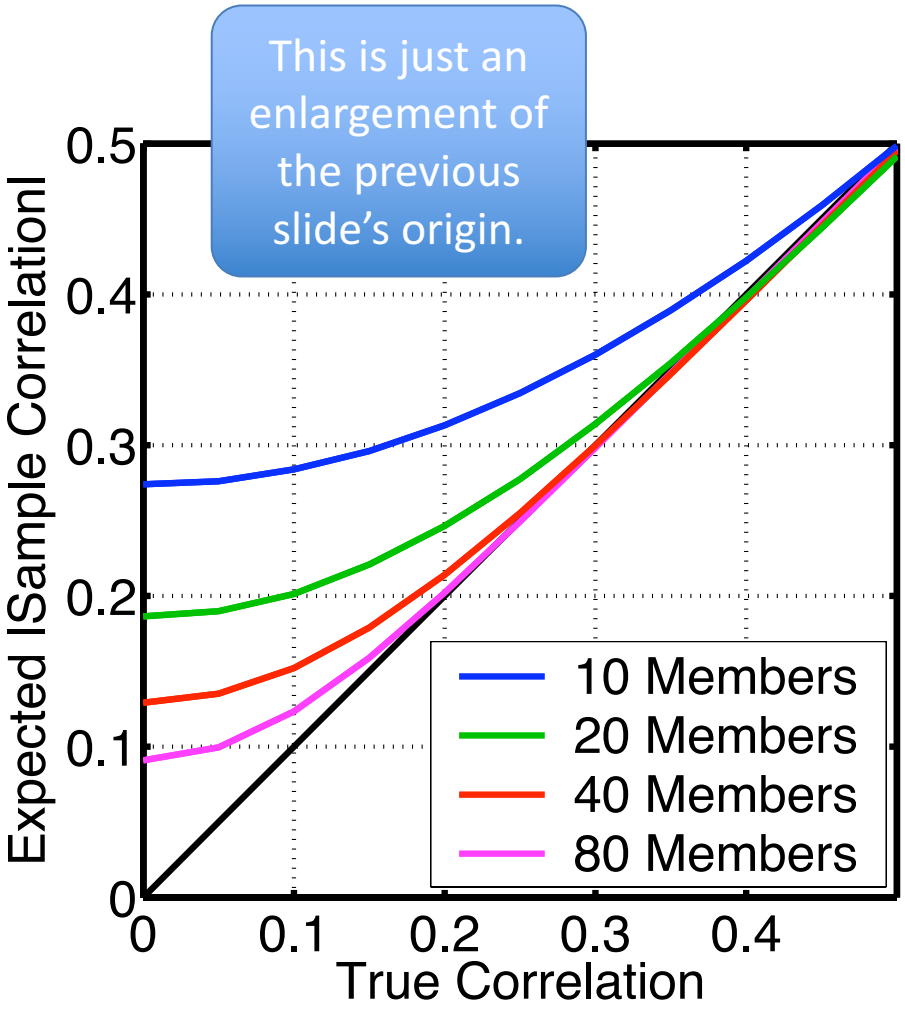
Regression Sampling Error



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!

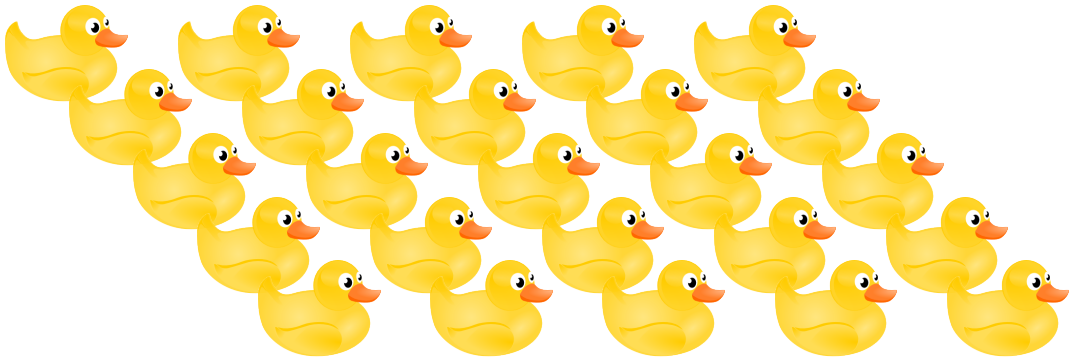
Ways to deal with Regression Sampling Error

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

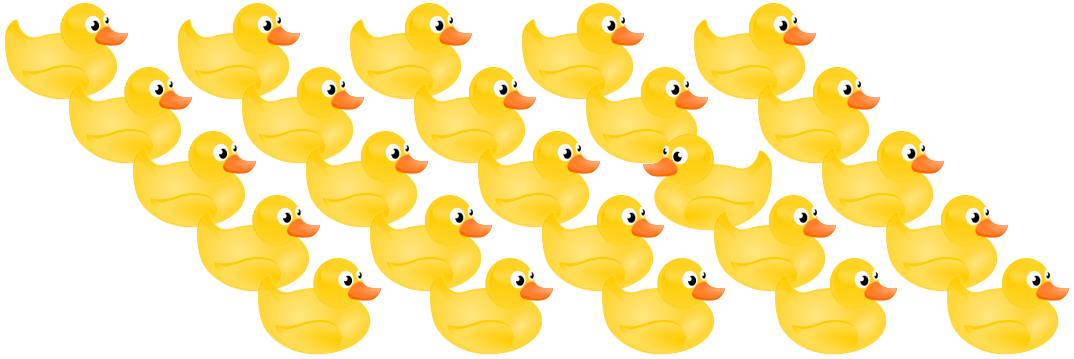
?

It may work.
It may not.

Ways to deal with Regression Sampling Error



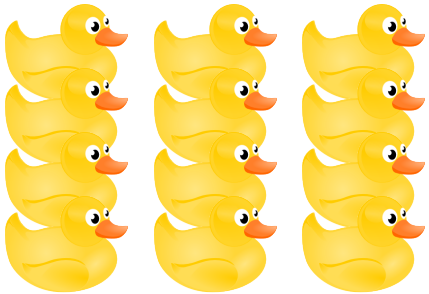
2. Use larger ensembles; expensive for large models.



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.

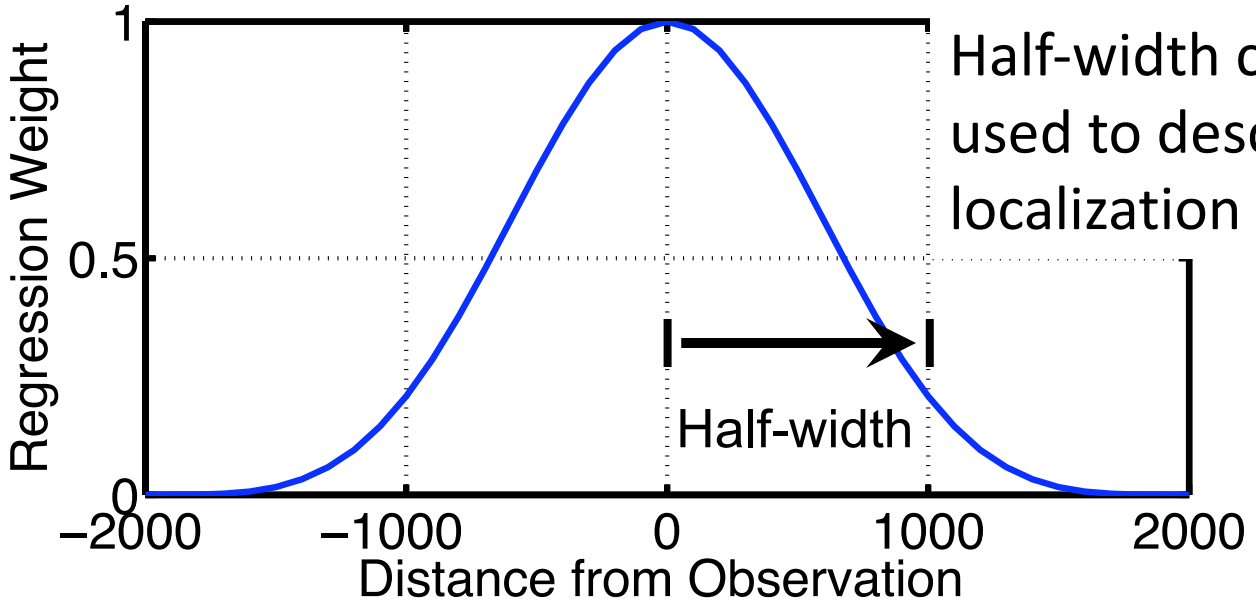


3. Use additional a priori information about relation between observations and state variables.
   Don't let an observation impact a state variable if they are <u>known</u> to be unrelated.

Using additional a priori information: Localization

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

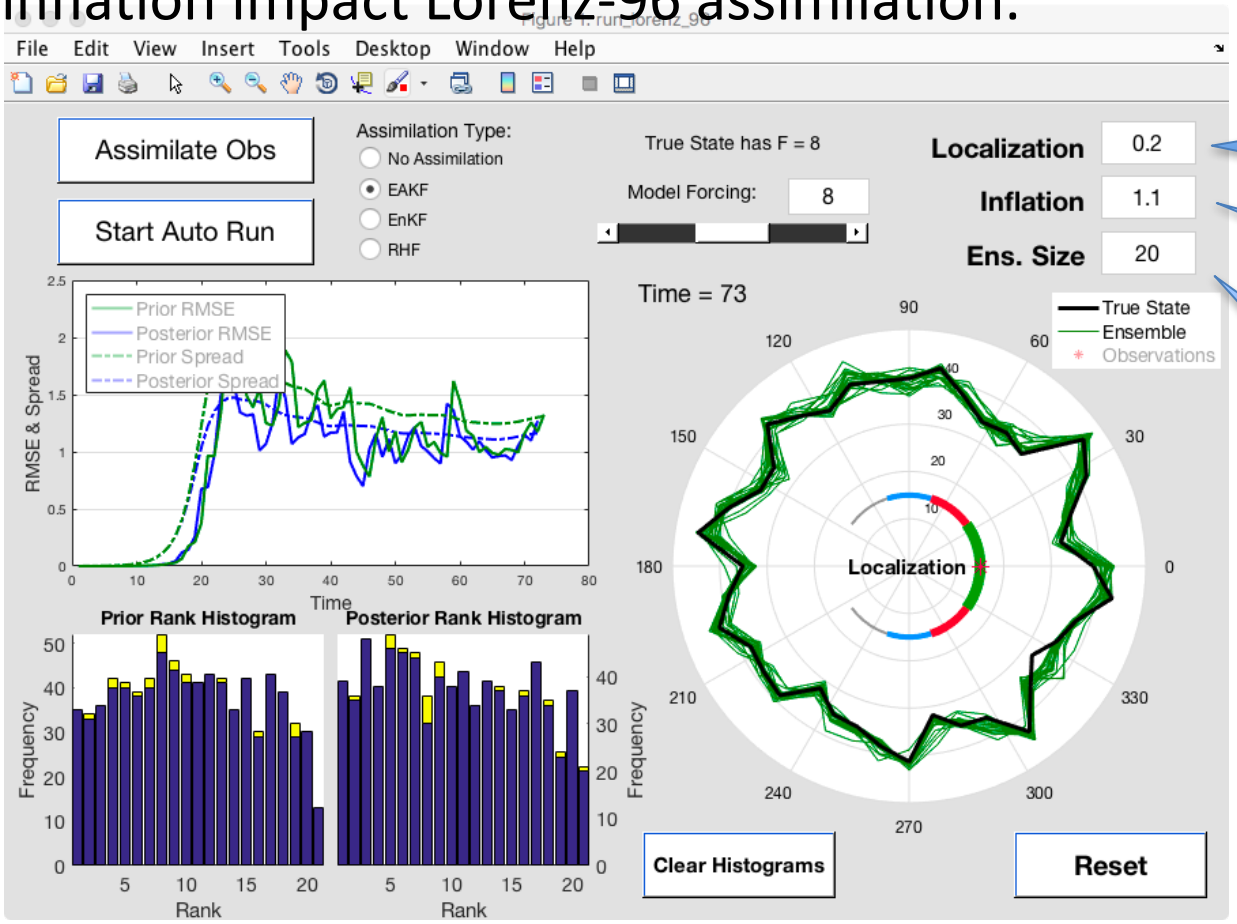
Compactly-supported 5th order polynomial (Gaspari-Cohn) is most commonly used for geophysics.



Half-width commonly used to describe localization function.

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

Purpose: Explore the 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 no weight is being given to state variables on the opposite side of the domain from an observation.
- 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.

DART_LAB Section 3