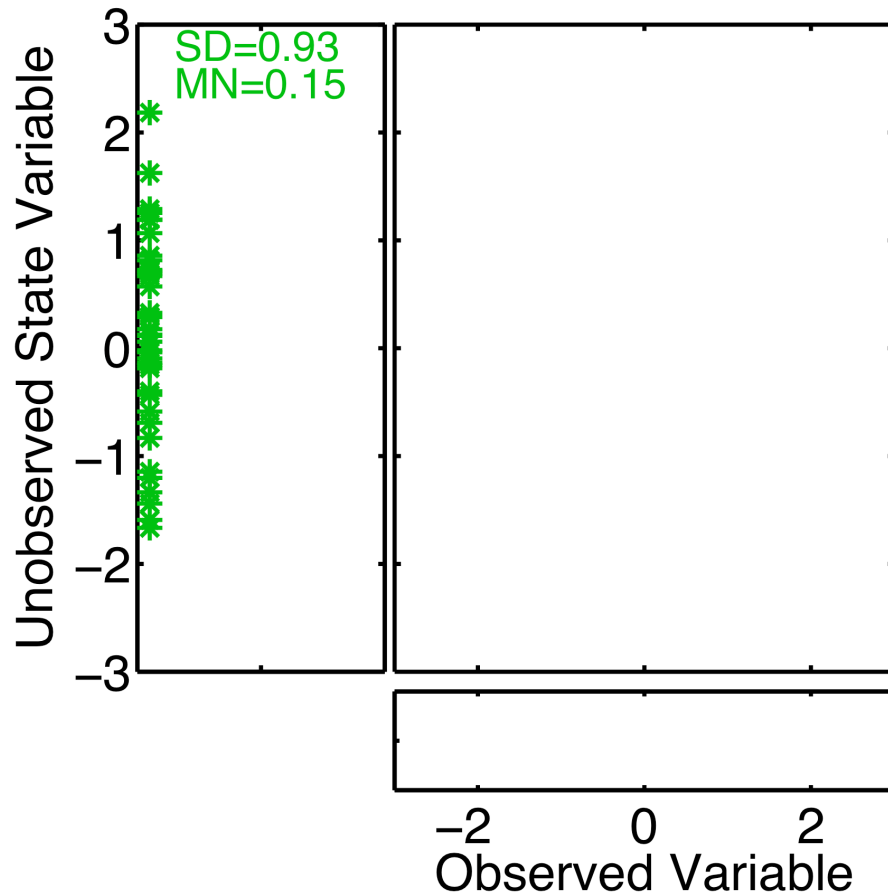


# Data Assimilation Research Testbed Tutorial



## Section 3: Sampling error and localization.

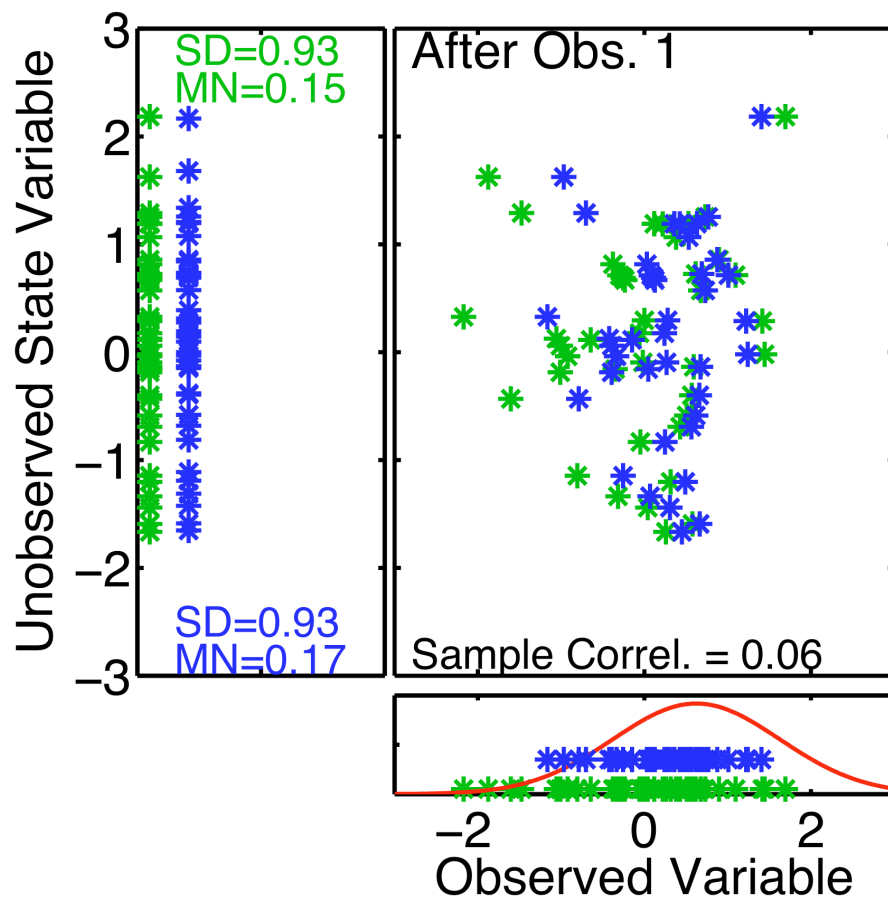
## Regression Sampling Error



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

Unobserved variable **should remain unchanged.**

## Regression Sampling Error

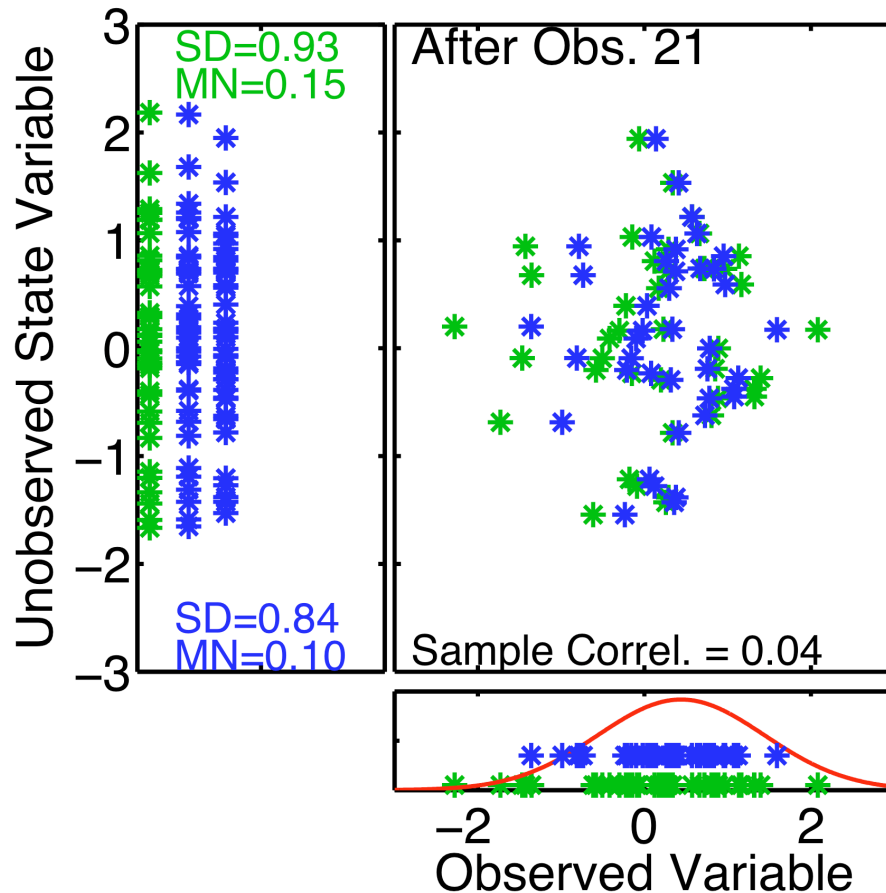


Suppose unobserved state variable is known to be unrelated to observed variables.

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

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

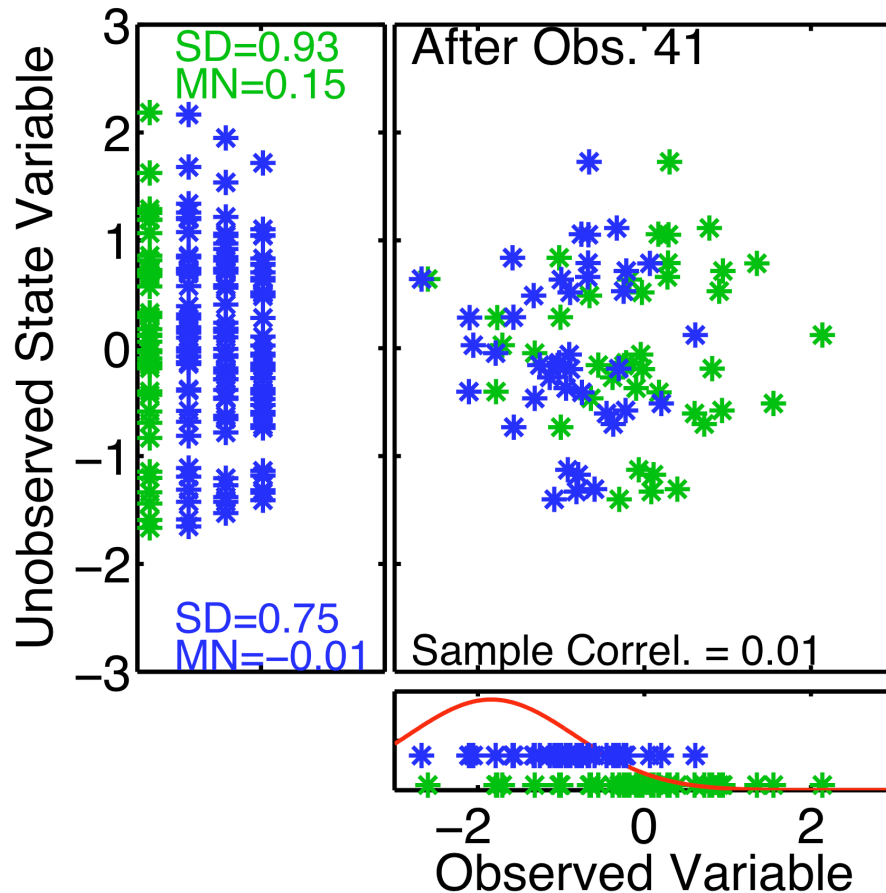
# Regression Sampling Error



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

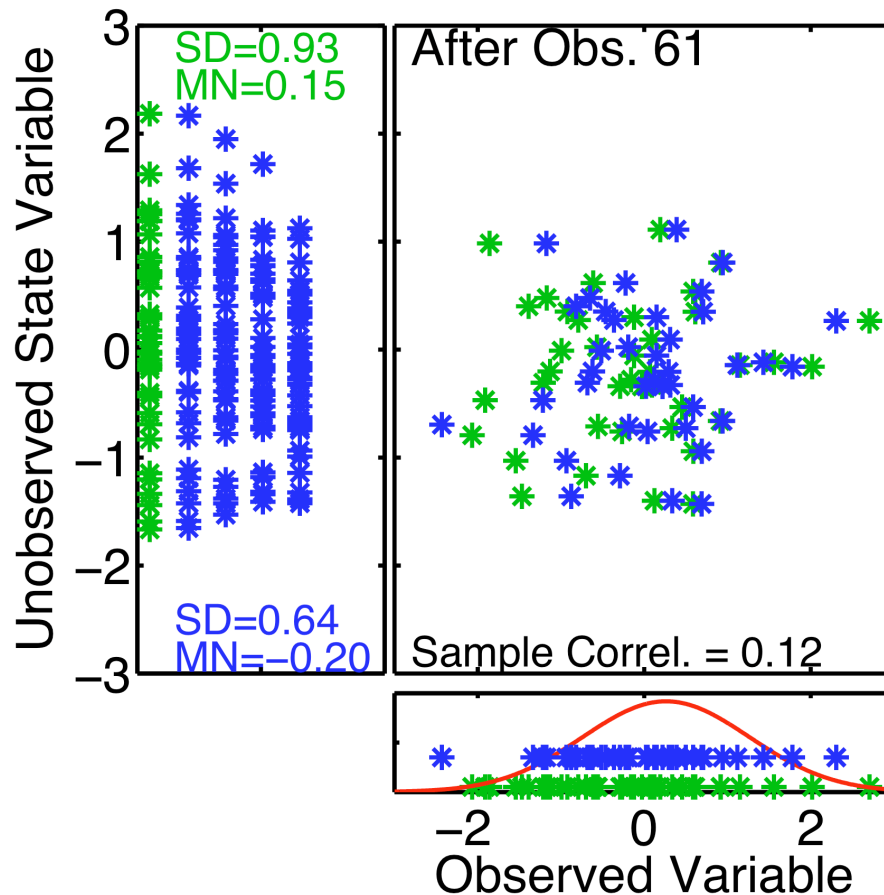


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

## Regression Sampling Error

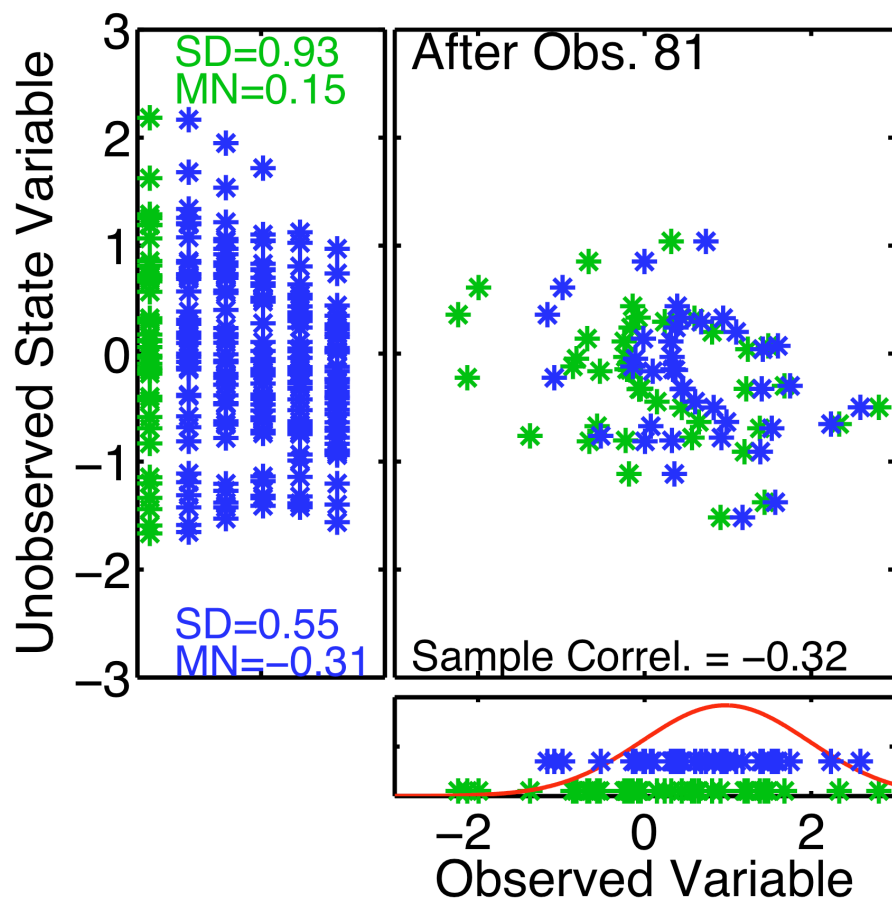


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

## Regression Sampling Error

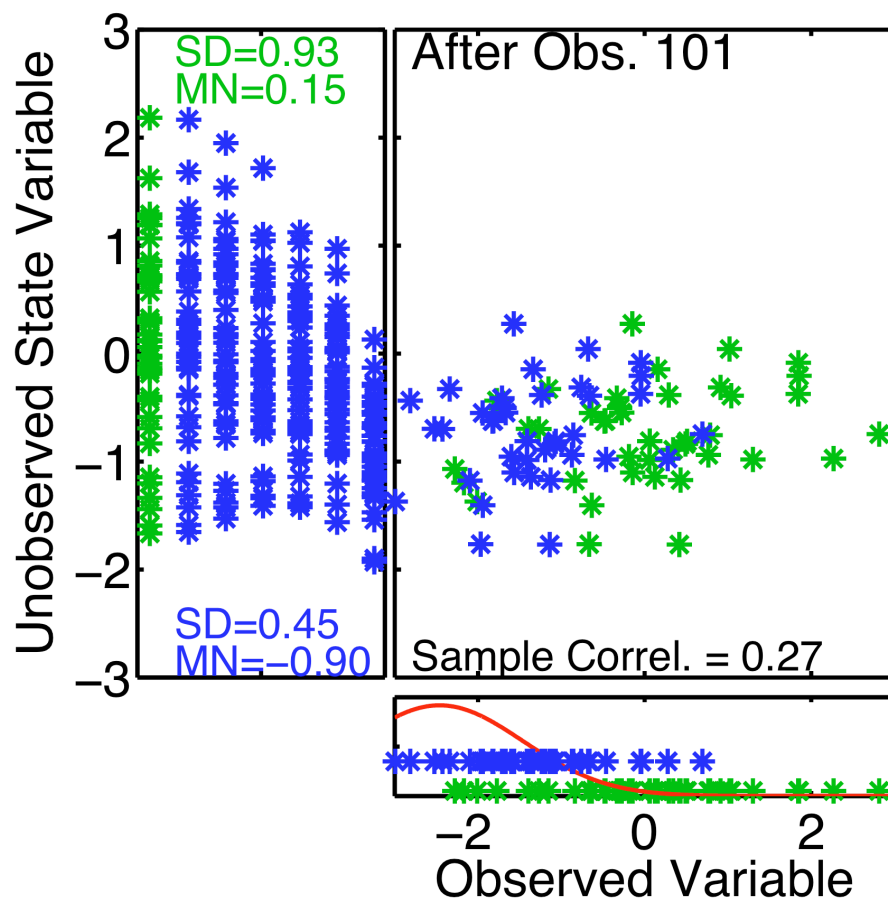


Suppose unobserved state variable is known to be unrelated to observed variables.

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

Unobserved standard deviation consistently decreases.

## Regression Sampling Error



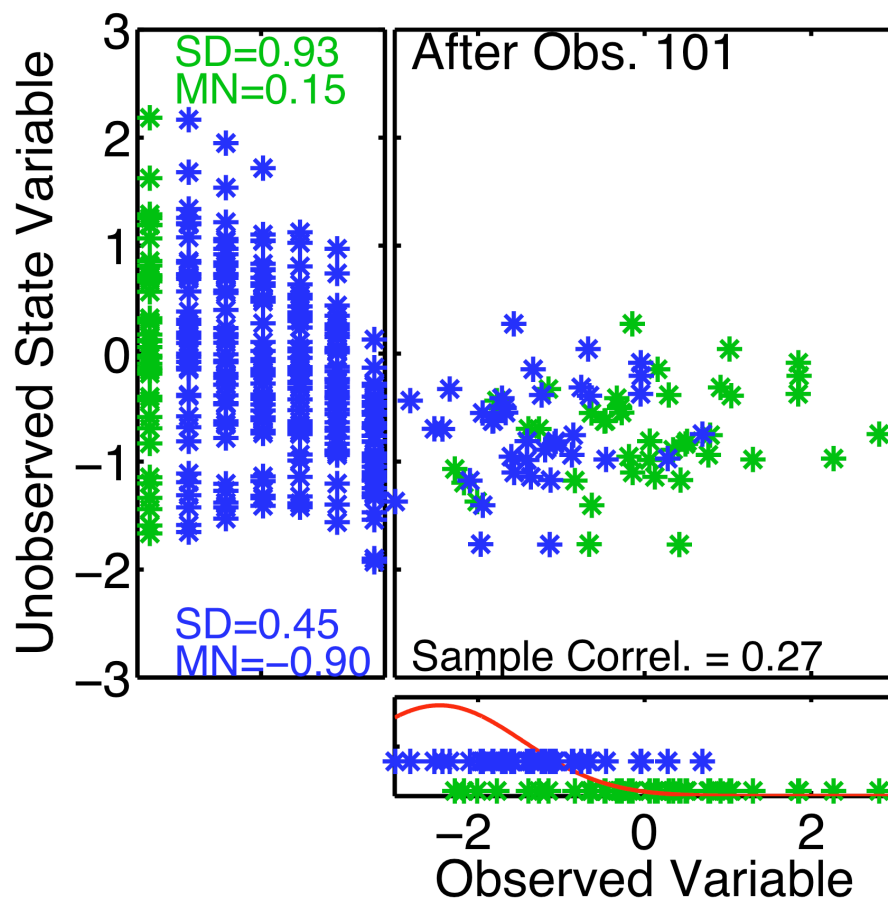
Suppose unobserved state variable is known to be unrelated to observed variables.

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 meaningful observations.

Meaningful observations can end up being ignored.

## Regression Sampling Error

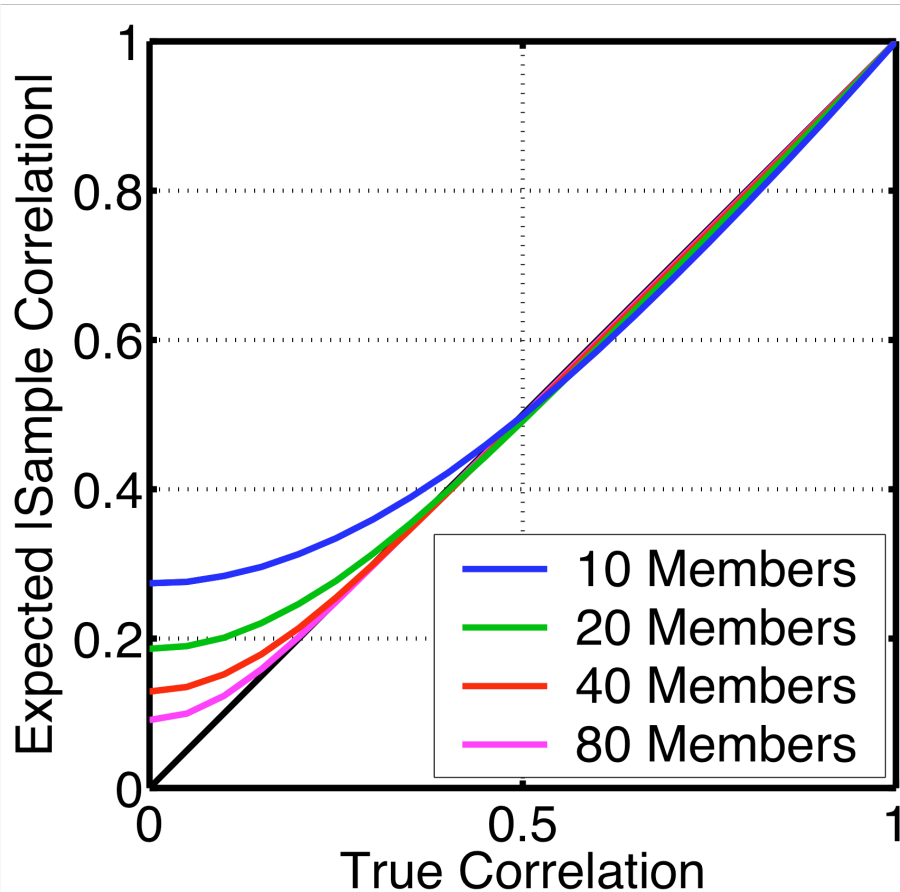
Ignoring meaningful observations due to insufficient spread is one form of filter divergence.

This could be seen in the Lorenz-96 assimilation.

Spread became small, filter thinks it has good estimate.

Error was large because good observations were ignored.

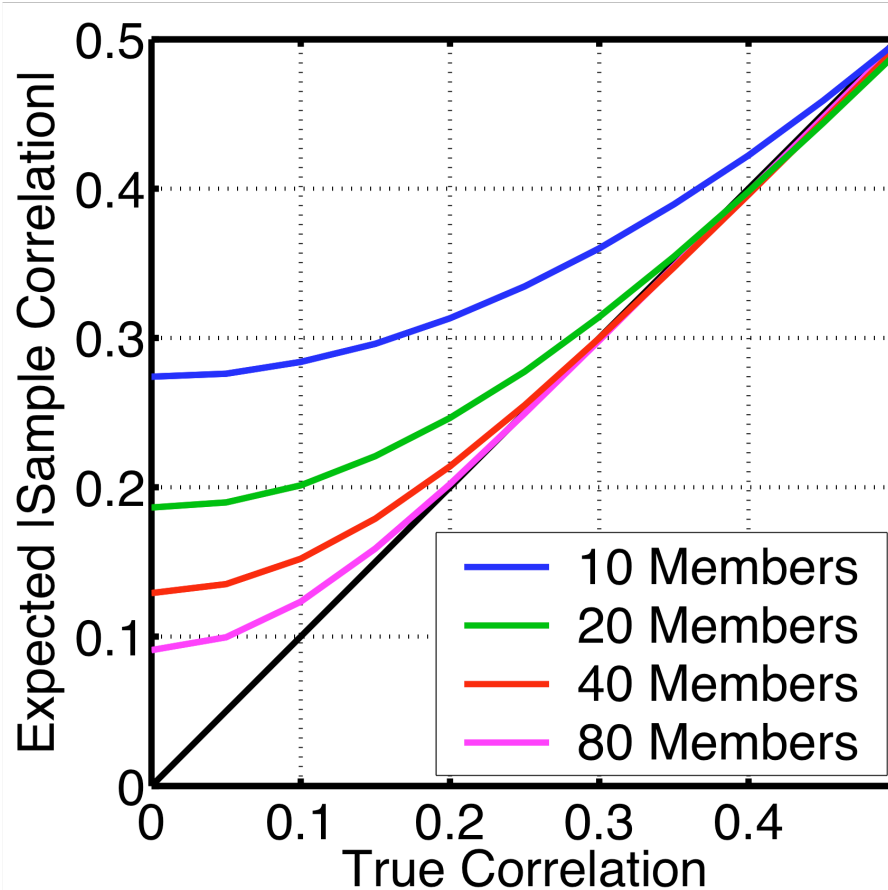
## Regression Sampling Error



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.

## 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.

## 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 known to be unrelated.

## Using additional a prior information: Localization.

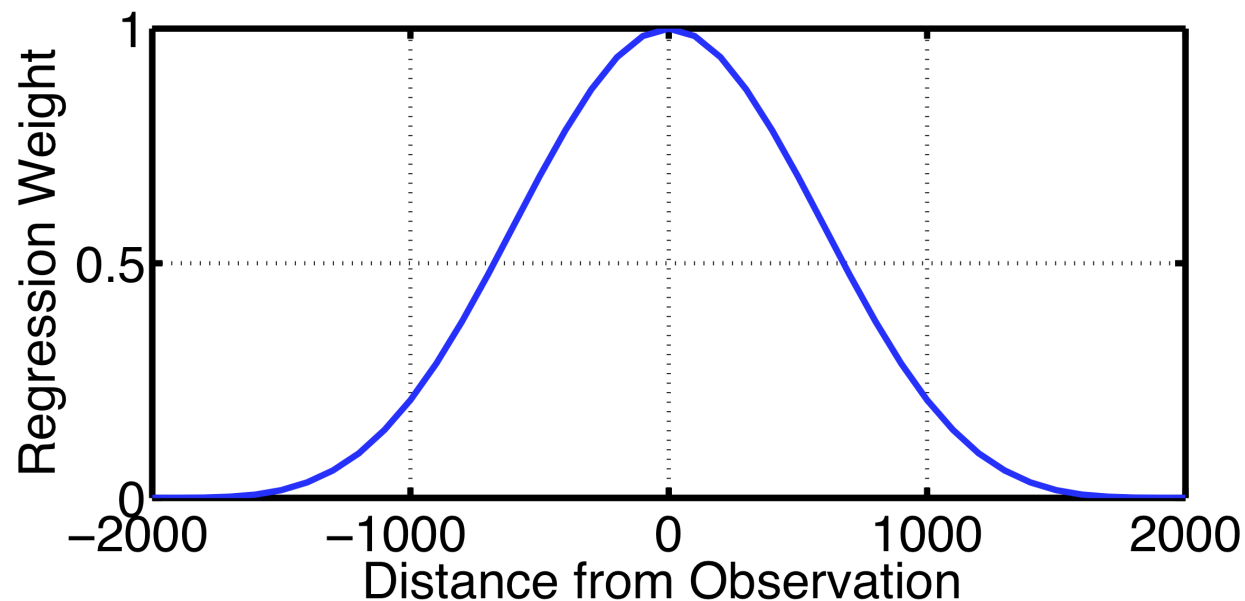
Try reducing regression factor as function of distance between observation and state variable.



## Using additional a prior information: Localization.

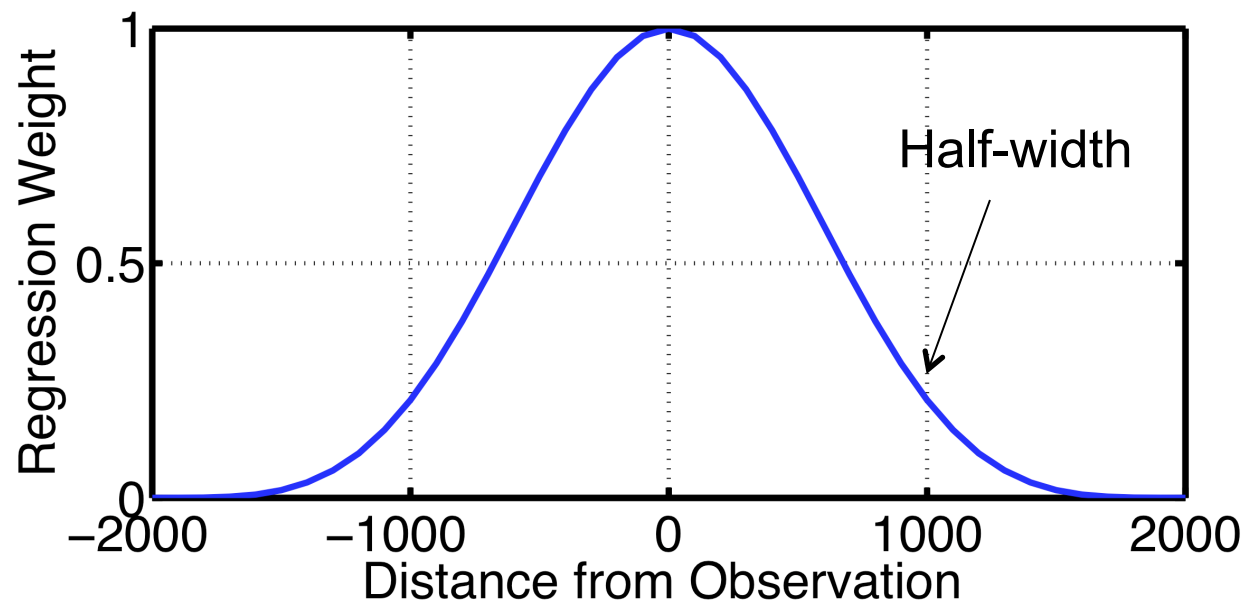
Try reducing regression factor as function of distance between observation and state variable.

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



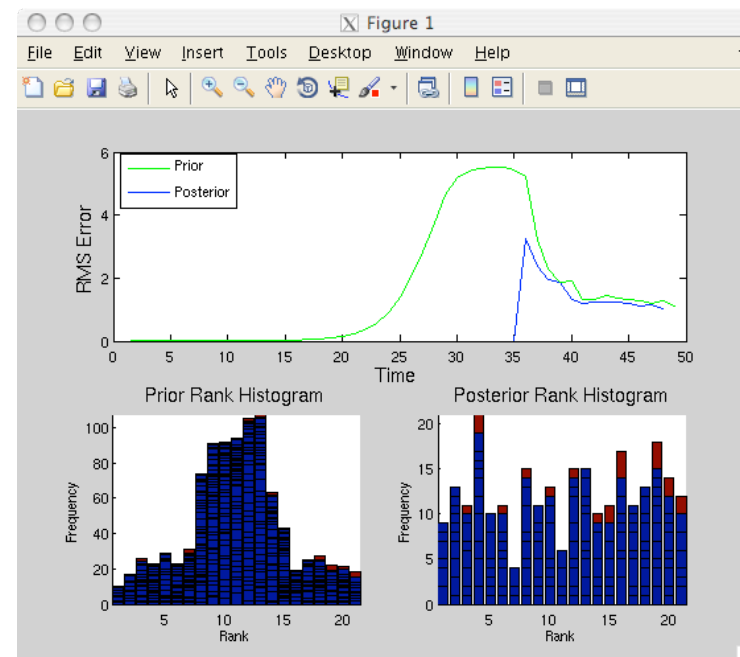
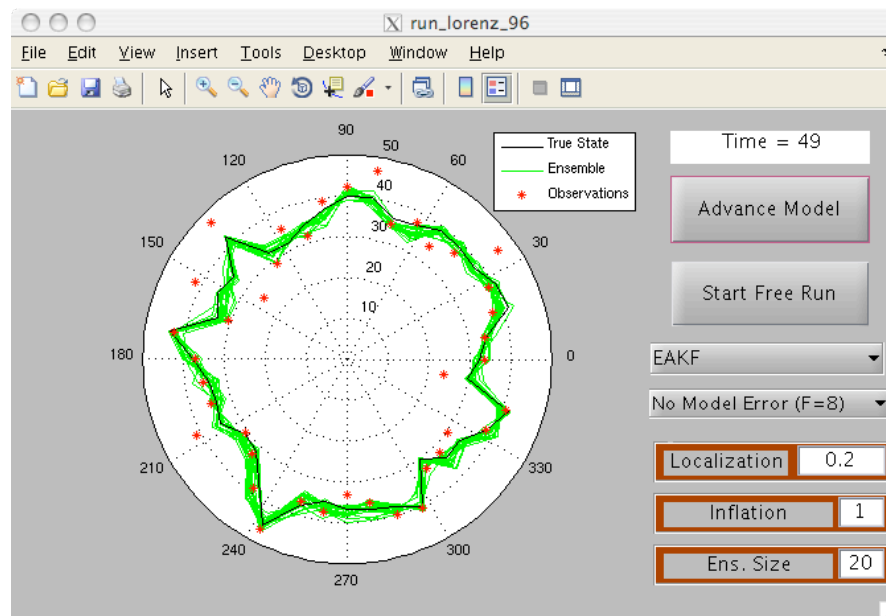
## Using additional a prior information: Localization.

Half-width commonly used to describe localization function.



## Matlab Hands-On: run\_lorenz\_96 (2)

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



## Matlab Hands-On: run\_lorenz\_96 (2)

### Procedure:

1. To see individual model advance and assimilation steps, select the top button on the menu window (it will alternate between **Advance Model** and **Assimilate Obs**).
2. Selecting **Start Free Run** starts a sequence of advance and assimilation steps.
3. Selecting **Stop Free Run** stops the sequence of steps.
4. Selecting **No Assimilation** runs truth and ensemble without assimilation. Selecting **EAKF** uses deterministic ensemble Kalman filter.
5. Inflation and ensemble size can be set with dialogue boxes.

## Matlab Hands-On: run\_lorenz\_96 (2)

What do I see?

The control panel shows the time evolution of the true state (black) and ensemble members (green).

At each observation time, the 40 observations are marked by red asterisks.

RMS error for the prior and posterior ensemble mean are in the top panel of figure 1.

Rank histograms for prior and posterior are in the lower panels.

Notes: All 40 state variables are observed. Error standard deviation is 4.

## Matlab Hands-On: run\_lorenz\_96 (2)

### Explorations:

1. Do an extended free run to see error growth in the ensemble.
2. 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.
3. 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!
4. Try reducing the ensemble size to 10 and varying localization.
5. Try adding in some inflation with localization.
6. Try selecting model error from the pull-down with other options.