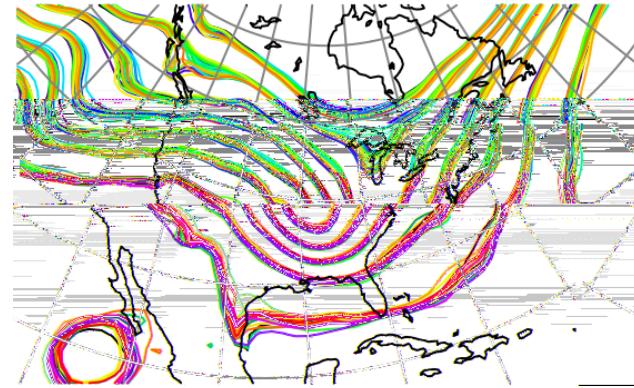


D  
a  
t  
a  
  
A  
s  
s  
i  
m  
i  
l  
o  
n  
  
R  
e  
s  
e  
a  
r  
c  
h  
  
T  
e  
s  
t  
b  
e  
d



# DART Tutorial Section 10: Regression and Non-linear Effects



©UCAR

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  
Atmospheric Research



# Updating additional prior state variables

Two primary error sources:

1. Linear approximation is invalid.

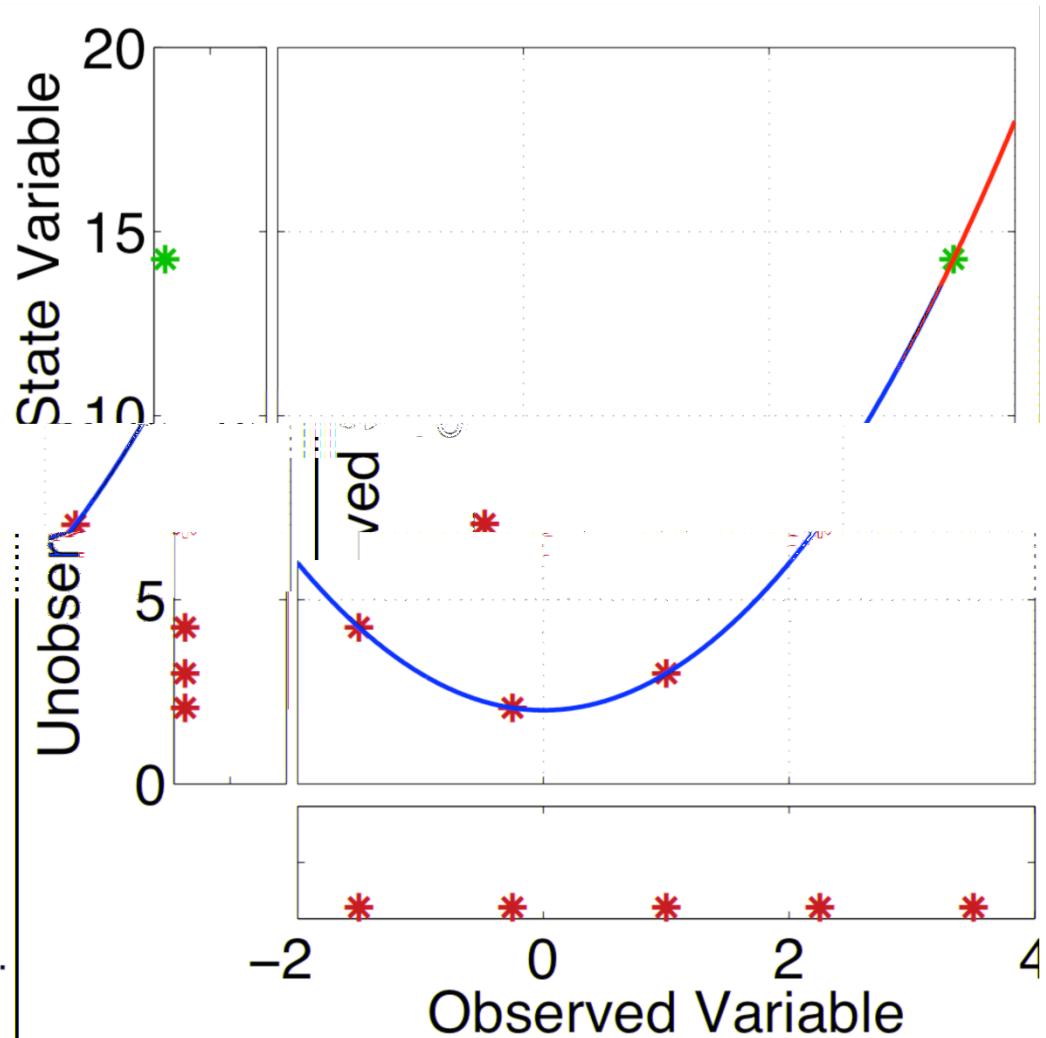
Substantial nonlinearity in ‘true’ relation over range of prior.

2. Sampling error due to noise (we’ve already looked at this).

Even if linear relation, sample regression coefficient imprecise.

May need to address both issues for good performance.

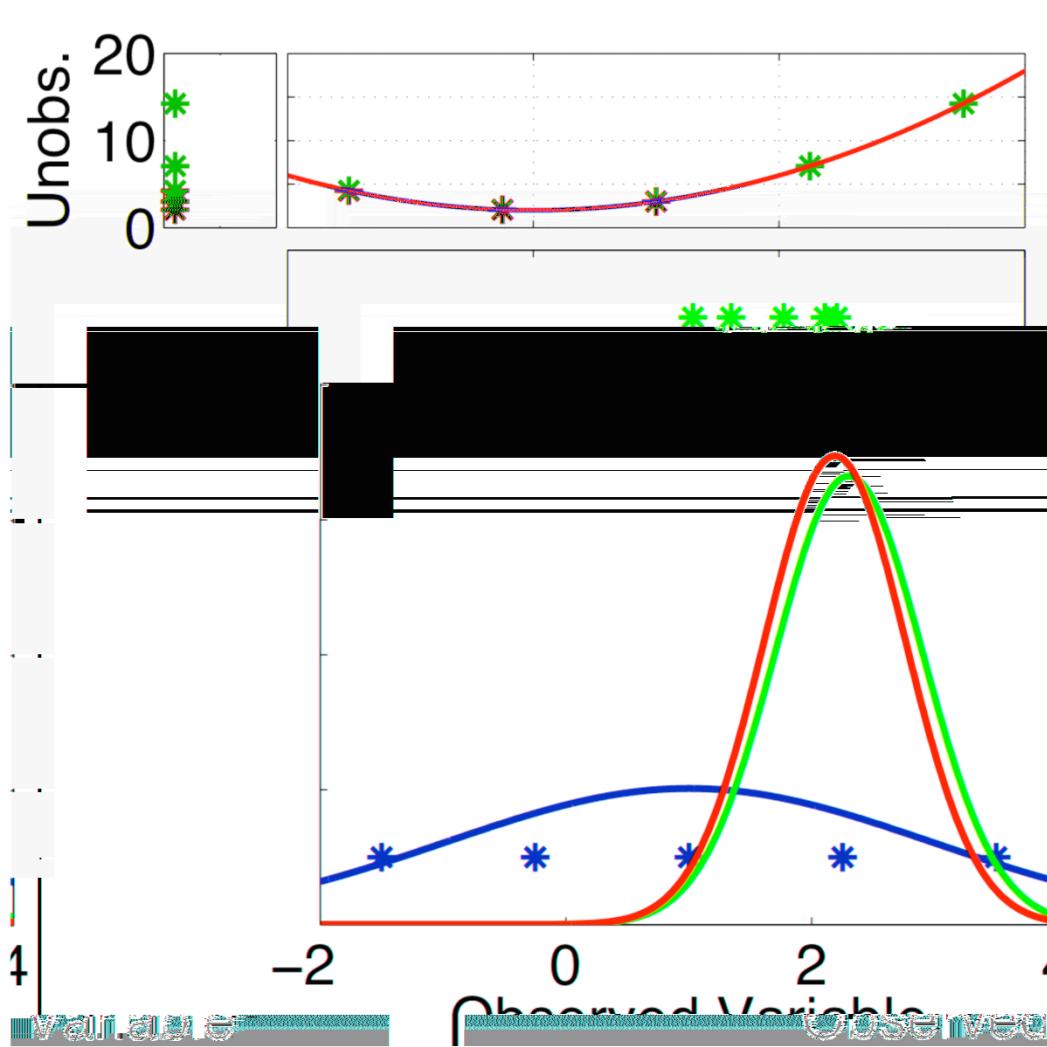
# Nonlinear relations between variables: Sorting increments



Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

# Nonlinear relations between variables: Sorting increments

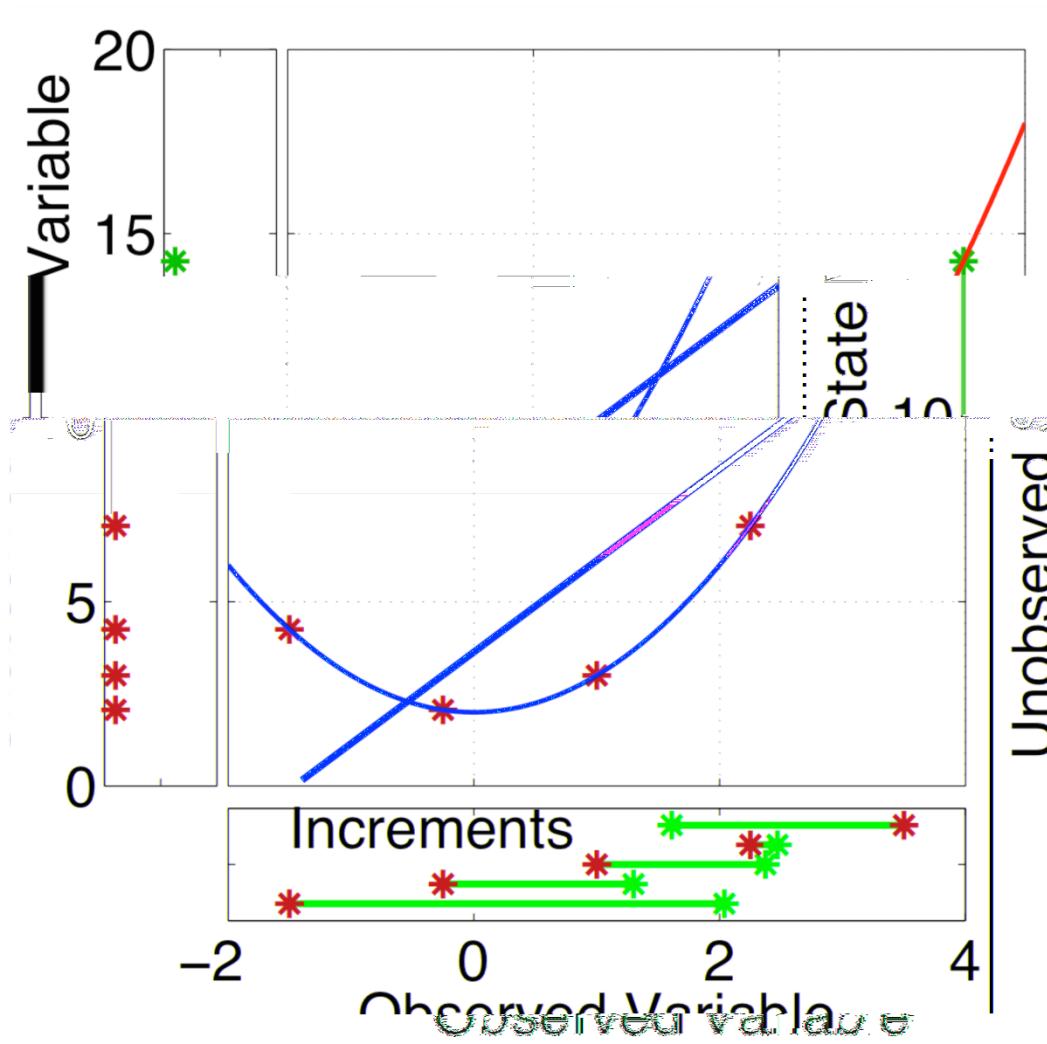


Suppose prior sample has  
NO noise.

But, relation between  
un/observed variables is  
non-linear.

Update observed sample  
and compute increments.

# Nonlinear relations between variables: Sorting increments

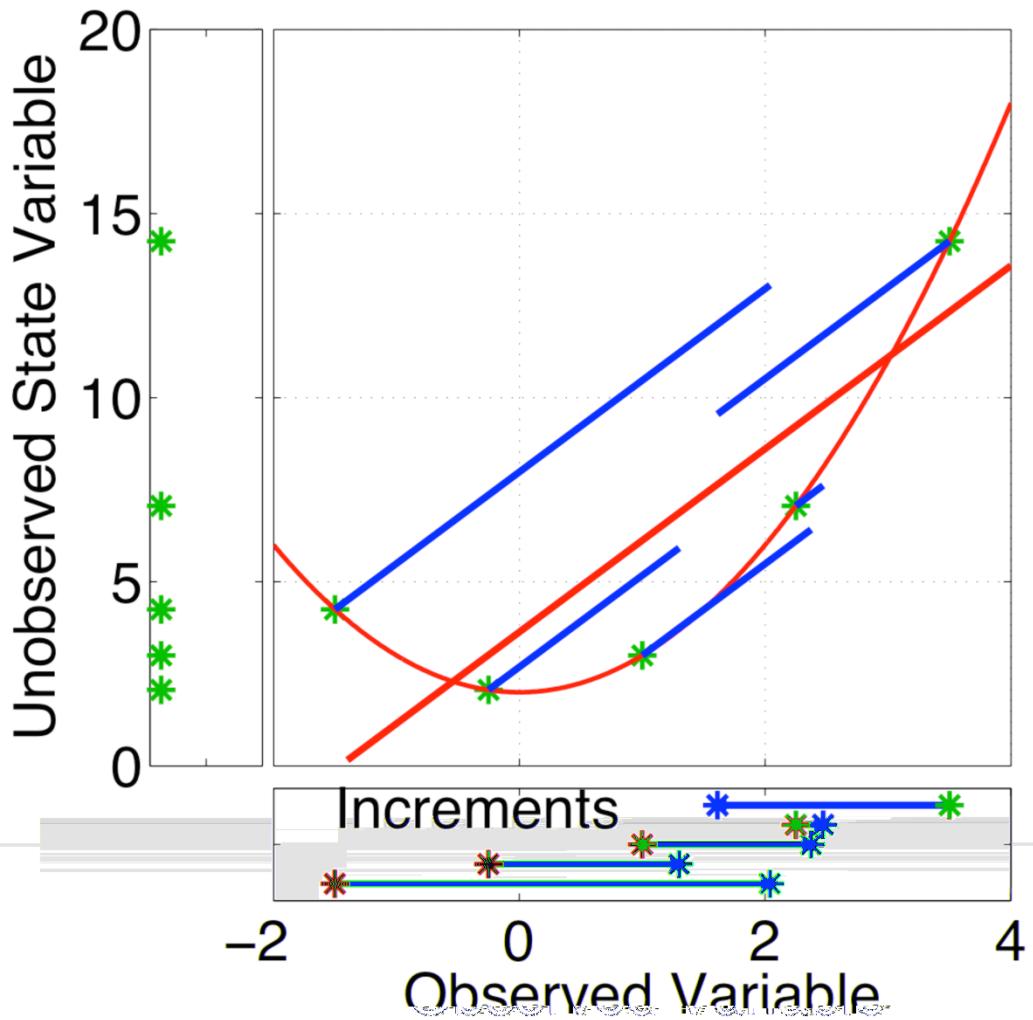


Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Regression error varies with value of observed variable.

# Nonlinear relations between variables: Sorting increments



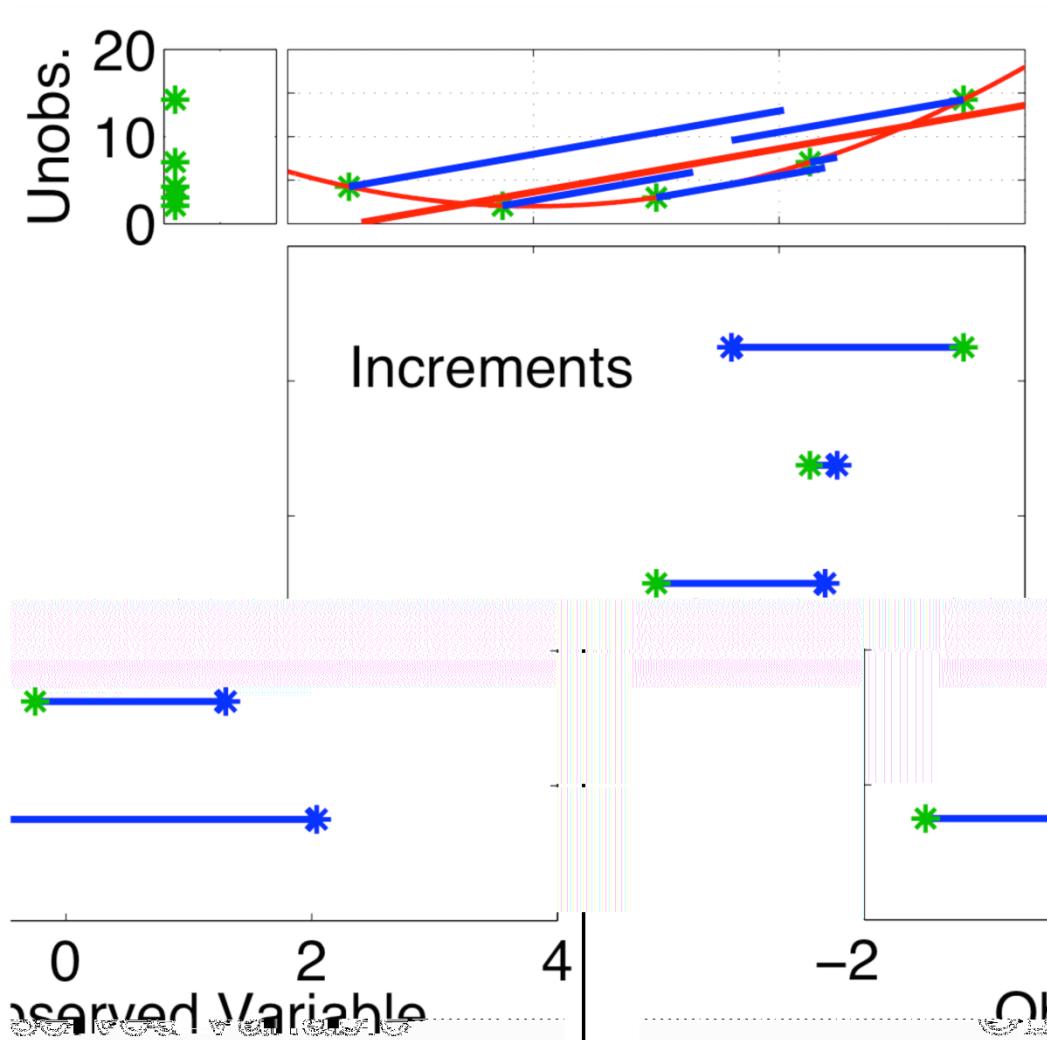
Suppose prior sample has  
NO noise.

But, relation between  
un/observed variables is  
non-linear.

Regression error varies  
with value of observed  
variable.

Smaller increments have  
smaller expected errors.

# Nonlinear relations between variables: Sorting increments



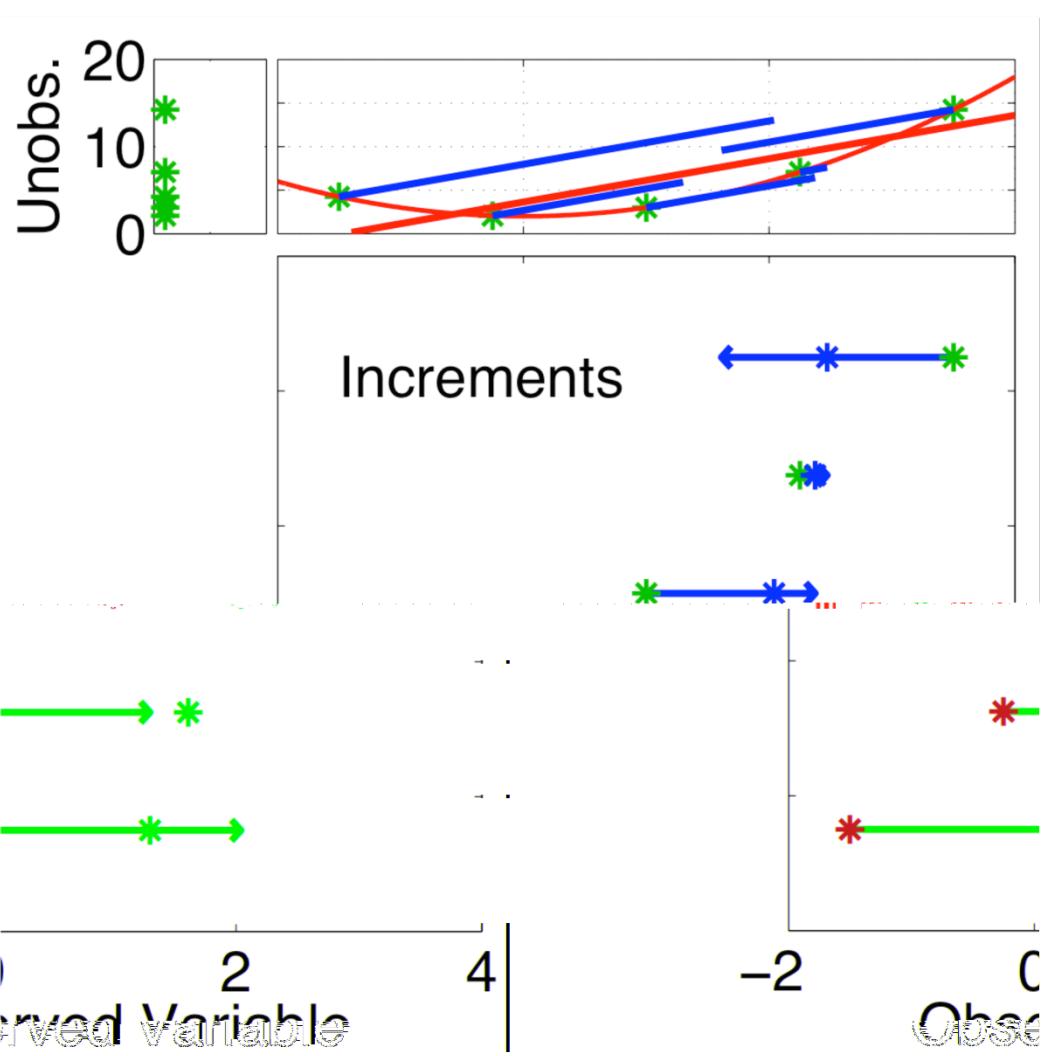
Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Pairing between prior and posterior sample of observed variable can be viewed as arbitrary.

Posterior is same sample however it's paired.

# Nonlinear relations between variables: Sorting increments



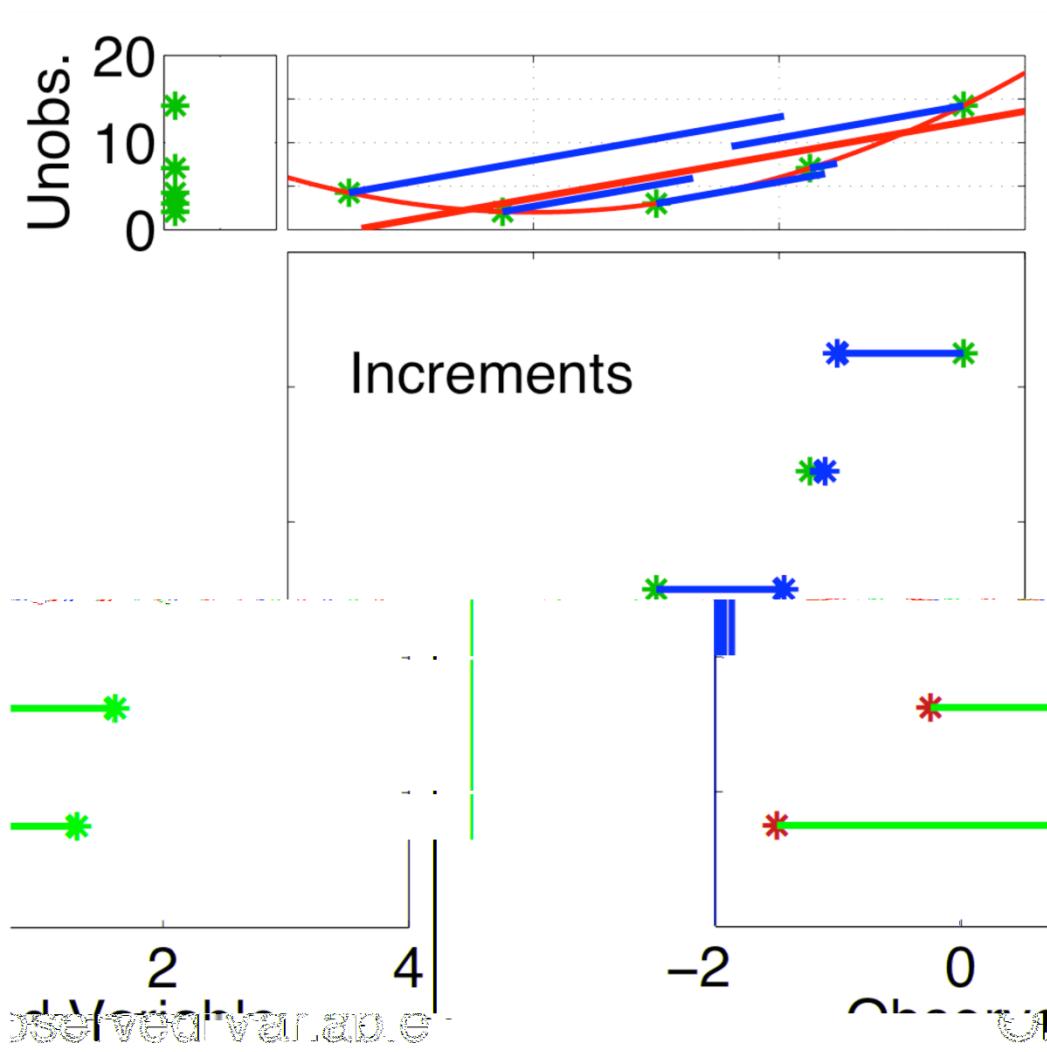
Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Can minimize increments by changing pairing.

Sorting prior and posterior and pairing samples minimizes one norm of increment size (could do other methods)

# Nonlinear relations between variables: Sorting increments



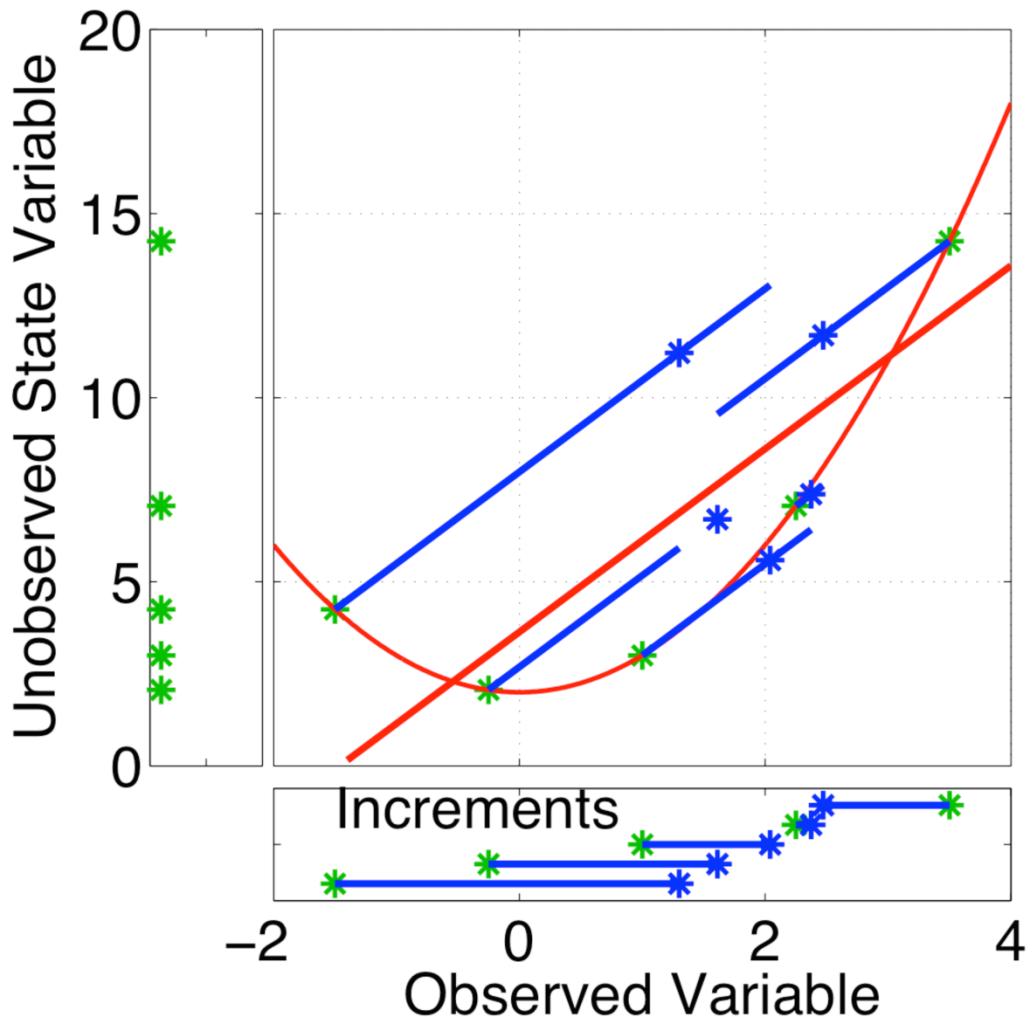
Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Can minimize increments by changing pairing.

Sorting prior and posterior and pairing samples minimizes one norm of increment size (could do other methods)

# Nonlinear relations between variables: Sorting increments



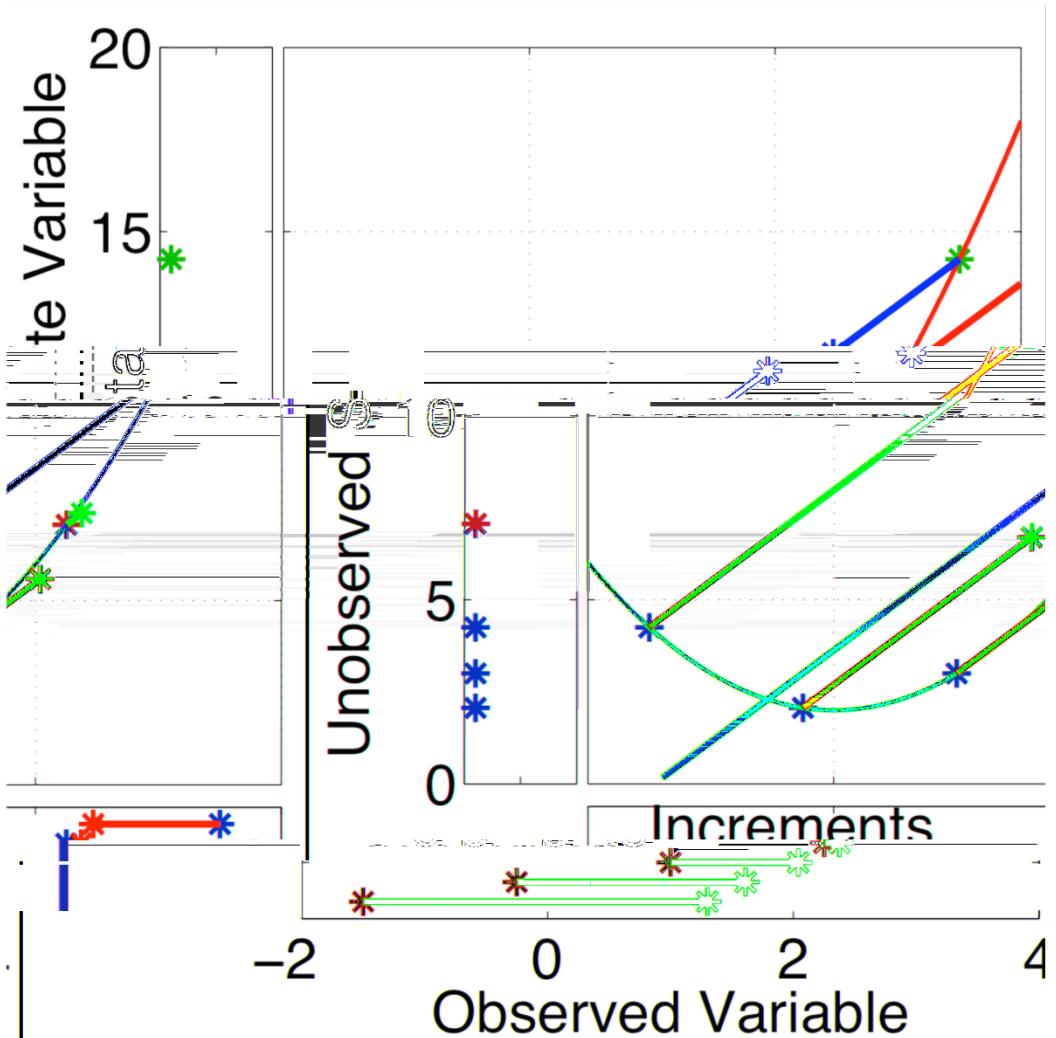
Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Resulting regression error is minimized.

Impact of sorting can be very large when posterior selected by 'random' algorithms.

# Nonlinear relations between variables: Sorting increments



Suppose prior sample has NO noise.

But, relation between un/observed variables is non-linear.

Resulting regression error is minimized.

Impact of sorting can be very large when posterior selected by 'random' algorithms.

# Nonlinear relations between variables: Sorting increments

Can see this impact nicely in 9var model.

models/9var/work/

Try *filter\_kind* = 2 in *assim\_tools\_nml* with:

*sort\_obs\_inc* = .true. (increments minimized) and

*sort\_obs\_inc* = .false.

```
&assim_tools_nml
    filter_kind          = 2
    sort_obs_inc        = .false.
    ...
    ...
```

change between .true. and .false.,

Examine the amount of noise in different time series.

Impact on RMS may not be what was expected.

There may be surprises in other low-order models when trying this.

# Nonlinear relations between variables: Sorting increments

Also can examine in Lorenz 96.

models/lorenz\_96/work/

Try *filter\_kind = 2* in *assim\_tools\_nml* with:

*sort\_obs\_inc = .true.* (increments minimized) and

*sort\_obs\_inc = .false.*

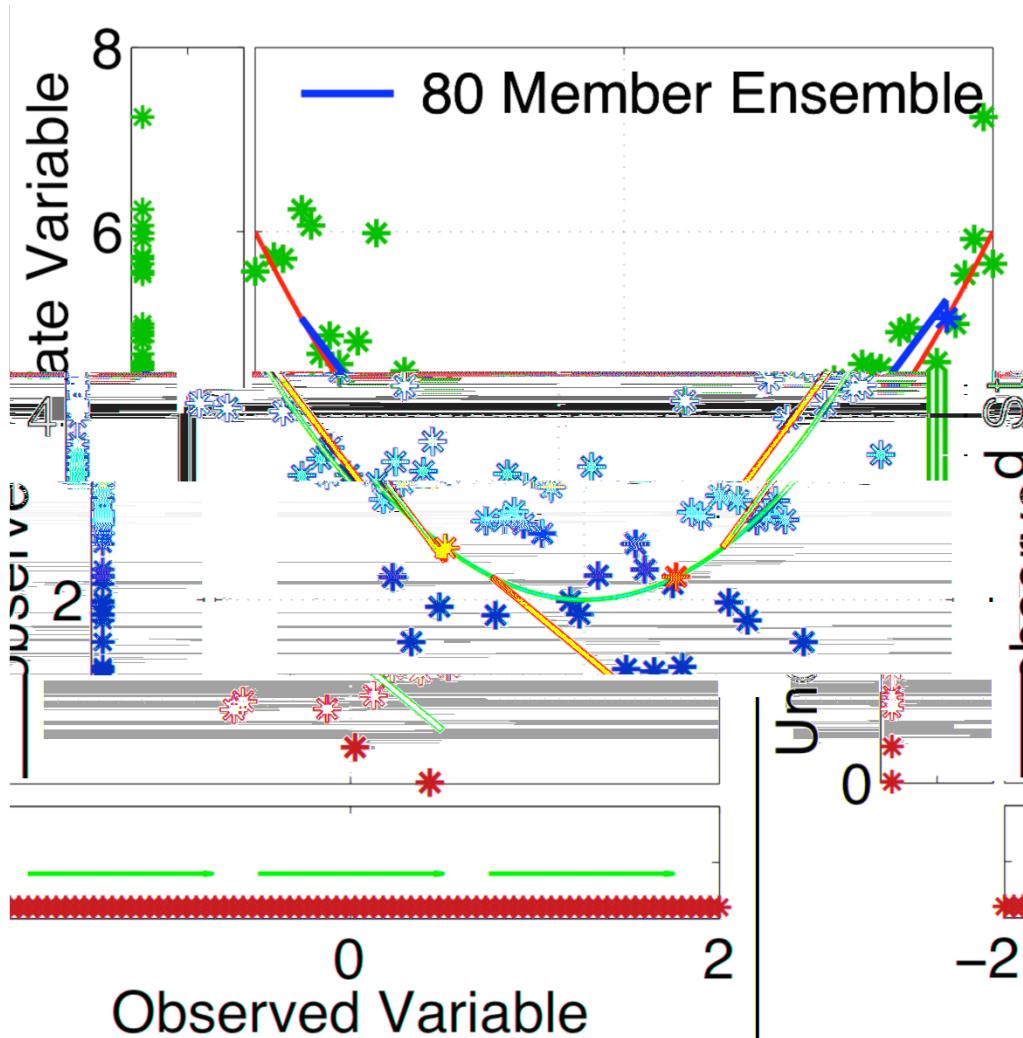
Try a case with no localization (large cutoff)

Vary inflation with and without the sorting.

```
&assim_tools_nml
    filter_kind          = 2
    sort_obs_inc        = .true.
    cutoff              = 1000000.0
    ...

```

# Nonlinear relations between variables: Local regression



Prior sample is noisy.

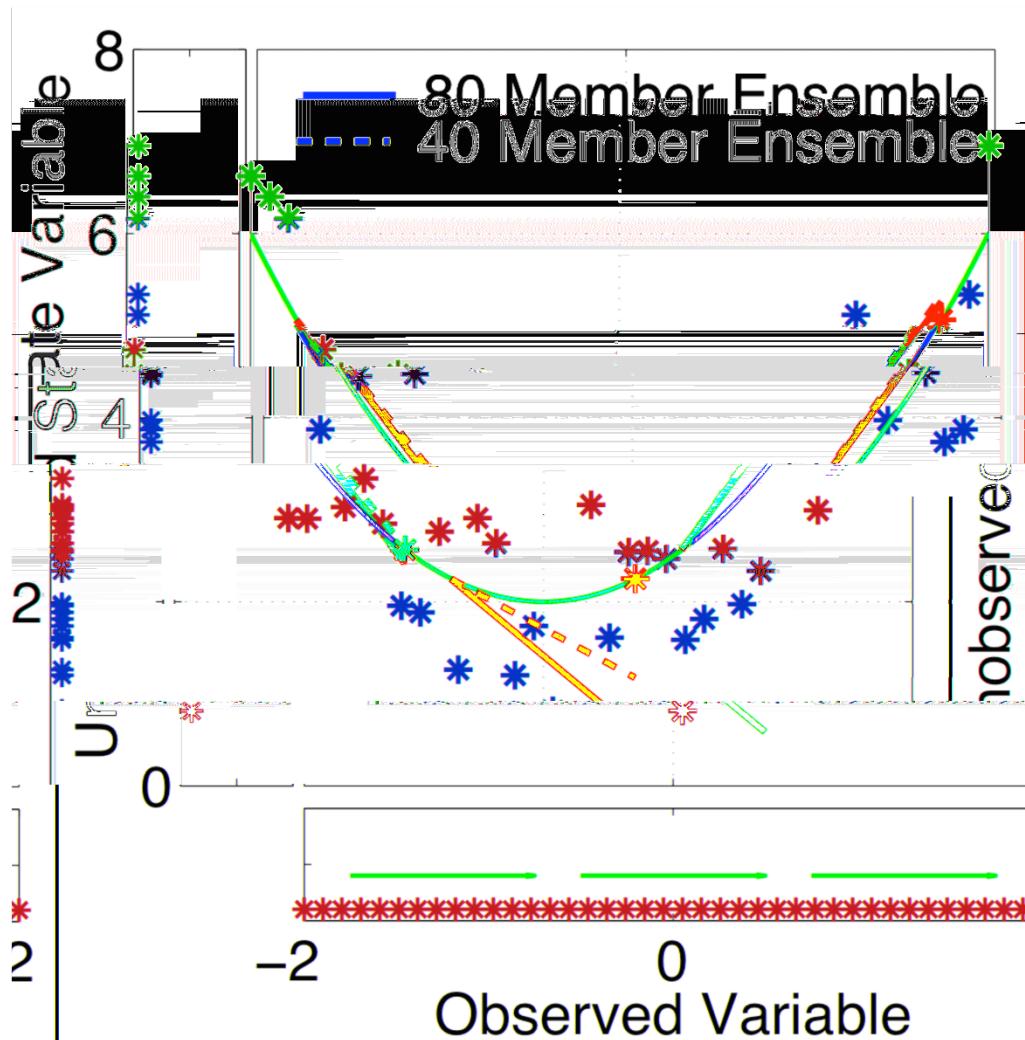
Un/observed relation is non-linear.

Doing global regression would be BAD here.

Can do regression only for points that lie in range of update increment.

Could also pick local sets in other ways.

# Nonlinear relations between variables: Local regression



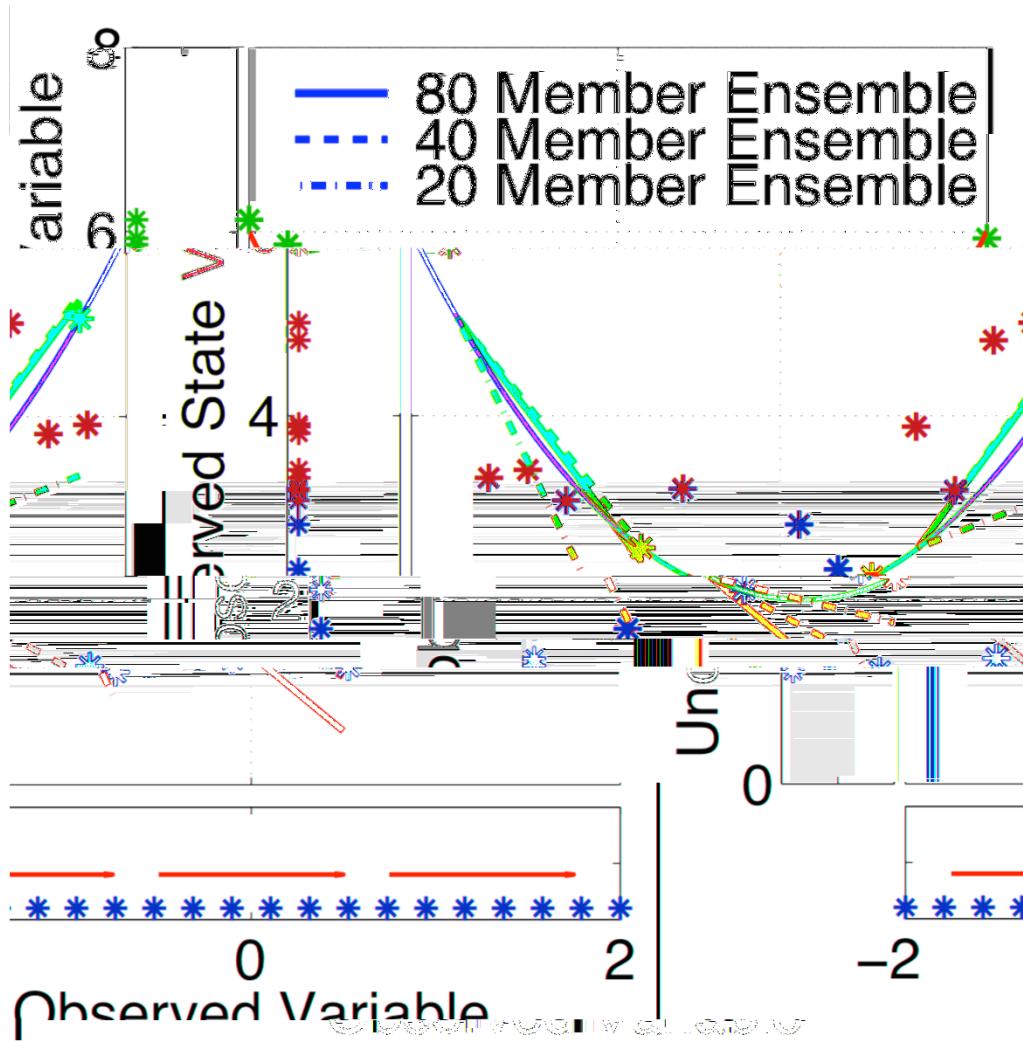
Prior sample is noisy.

Un/observed relation is non-linear.

For larger ensembles, local regressions can work well.

Error is largest where signal is weakest (near bottom of parabola here).

# Nonlinear relations between variables: Local regression



Prior sample is noisy.

Un/observed relation is non-linear.

As sample size decreases, error grows.

(Except where it was rotten to start).

Applications where local regression is useful are unknown to me.

## Nonlinear relations between variables: Local regression

# DART Tutorial Index to Sections

1. Filtering For a One Variable System
2. The DART Directory Tree
3. DART Runtime Control and Documentation
4. How should observations of a state variable impact an unobserved state variable?  
Multivariate assimilation.
5. Comprehensive Filtering Theory: Non-Identity Observations and the Joint Phase Space
6. Other Updates for An Observed Variable
7. Some Additional Low-Order Models
8. Dealing with Sampling Error
9. More on Dealing with Error; Inflation
10. Regression and Nonlinear Effects
11. Creating DART Executables
12. Adaptive Inflation
13. Hierarchical Group Filters and Localization
14. Quality Control
15. DART Experiments: Control and Design
16. Diagnostic Output
17. Creating Observation Sequences
18. Lost in Phase Space: The Challenge of Not Knowing the Truth
19. DART-Compliant Models and Making Models Compliant
20. Model Parameter Estimation
21. Observation Types and Observing System Design
22. Parallel Algorithm Implementation
23. Location module design (not available)
24. Fixed lag smoother (not available)
25. A simple 1D advection model: Tracer Data Assimilation