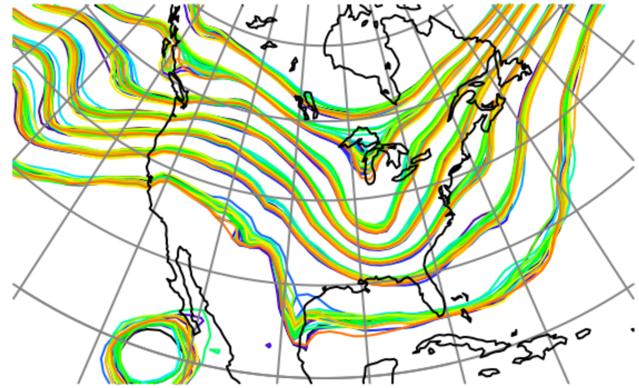


Exploiting Nonlinear Relations between Observations and State Variables in Ensemble Filters

Jeff Anderson, NCAR Data Assimilation Research Section

AMS, 9 January 2018

Schematic of a Sequential Ensemble Filter

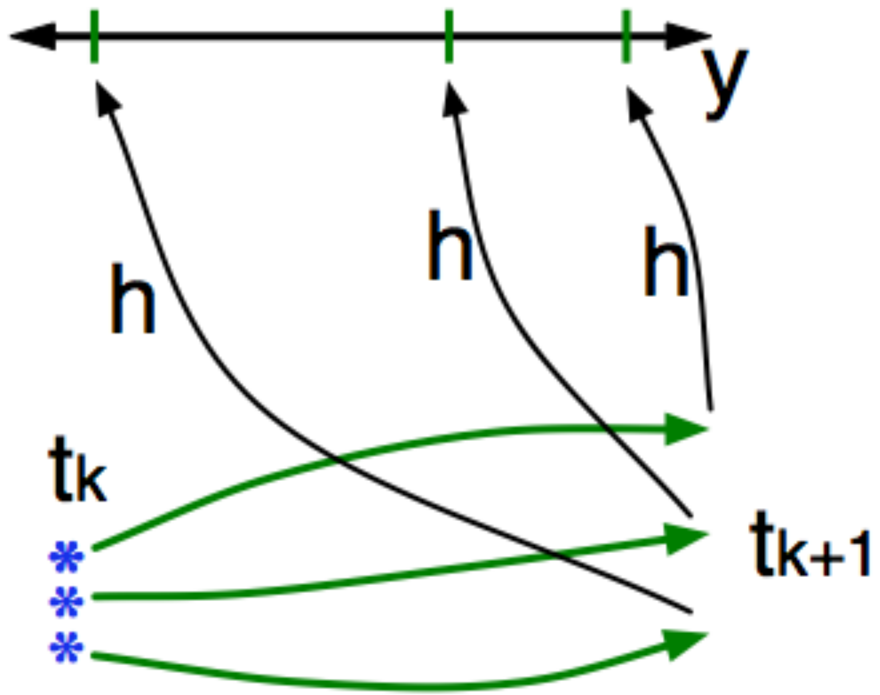
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.

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.

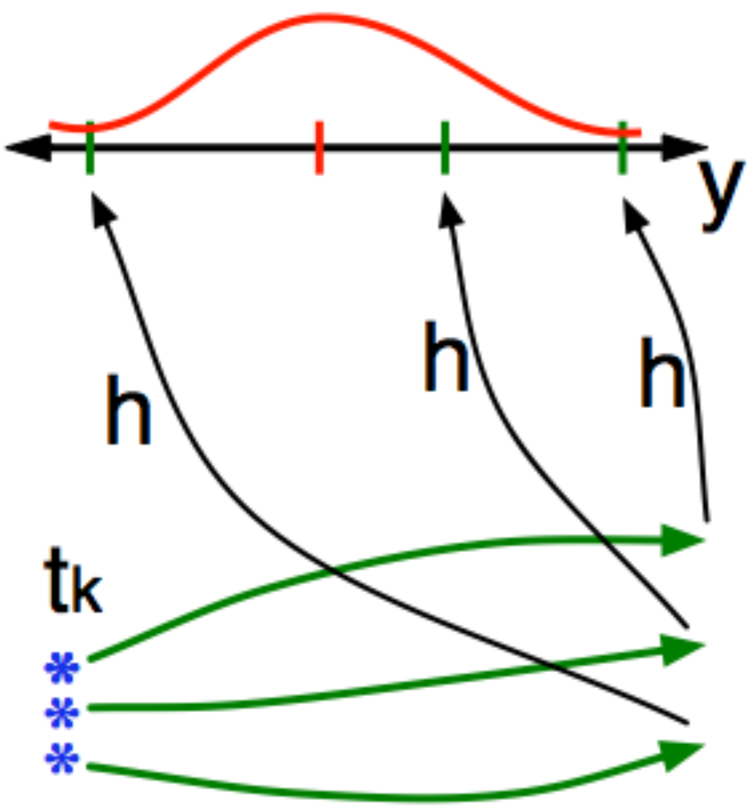


Theory: observations from instruments with uncorrelated errors can be done sequentially.

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

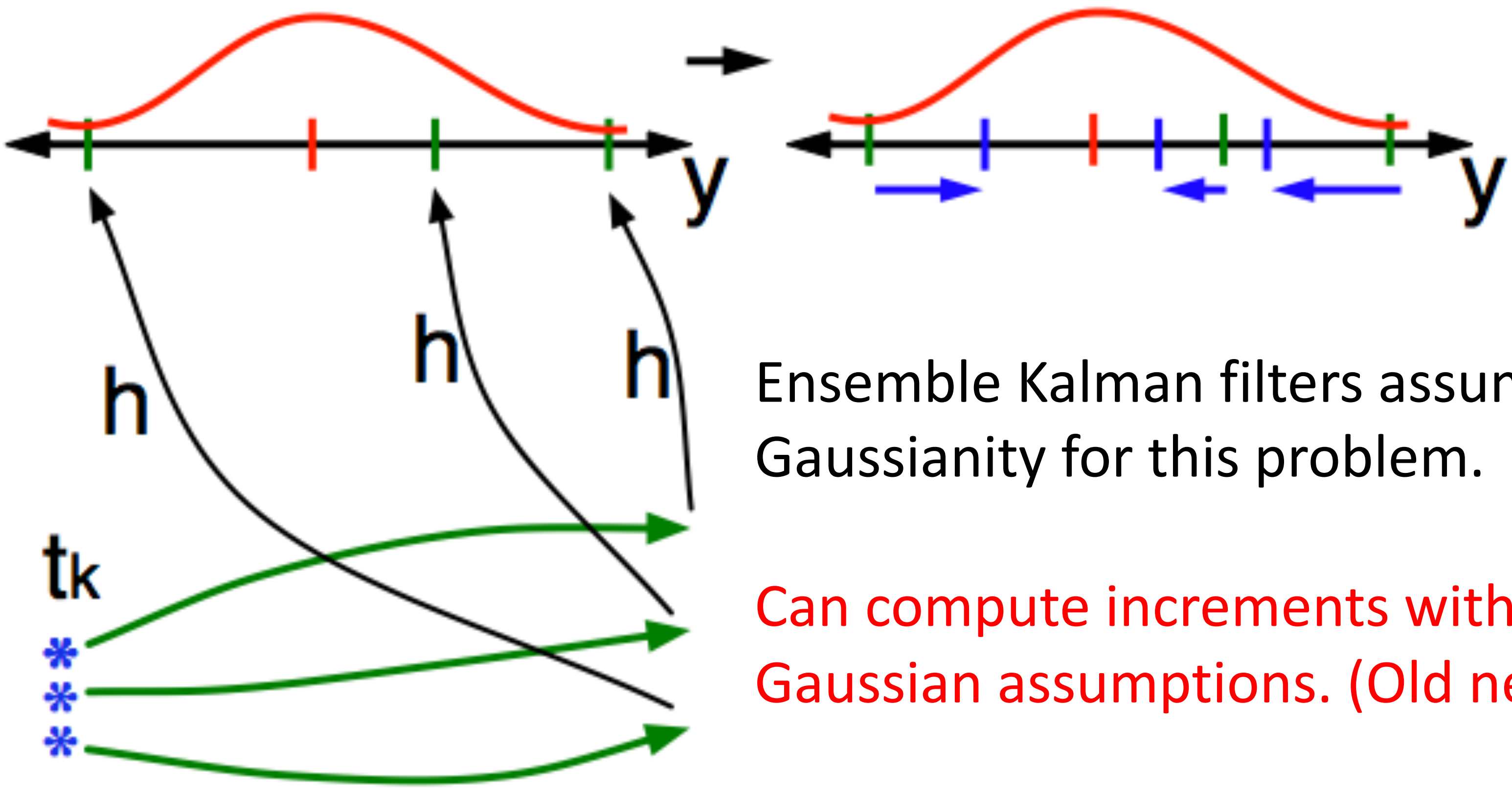
Schematic of a Sequential Ensemble Filter

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).

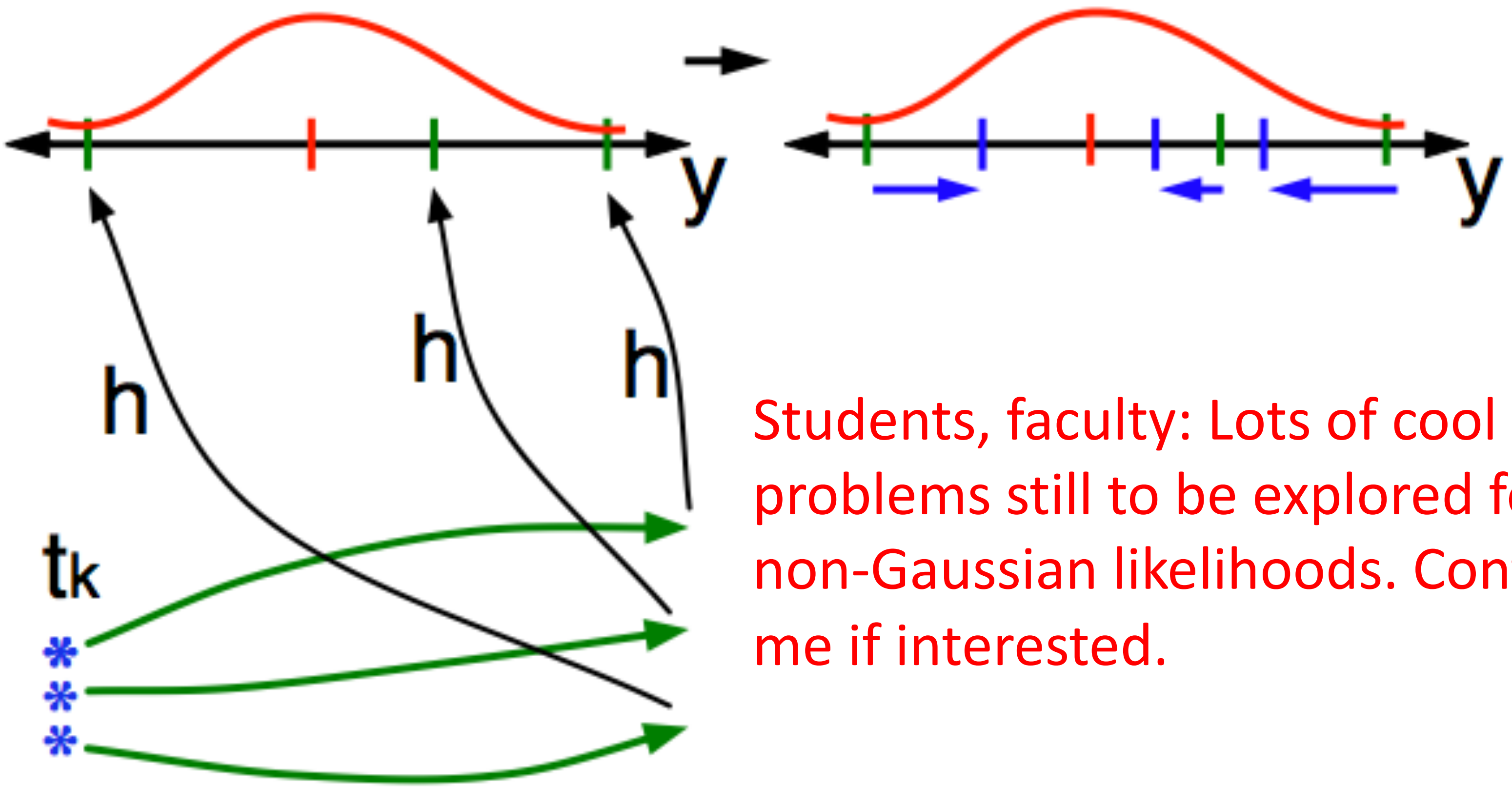


Ensemble Kalman filters assume Gaussianity for this problem.

Can compute increments without Gaussian assumptions. (Old news).

Schematic of a Sequential Ensemble Filter

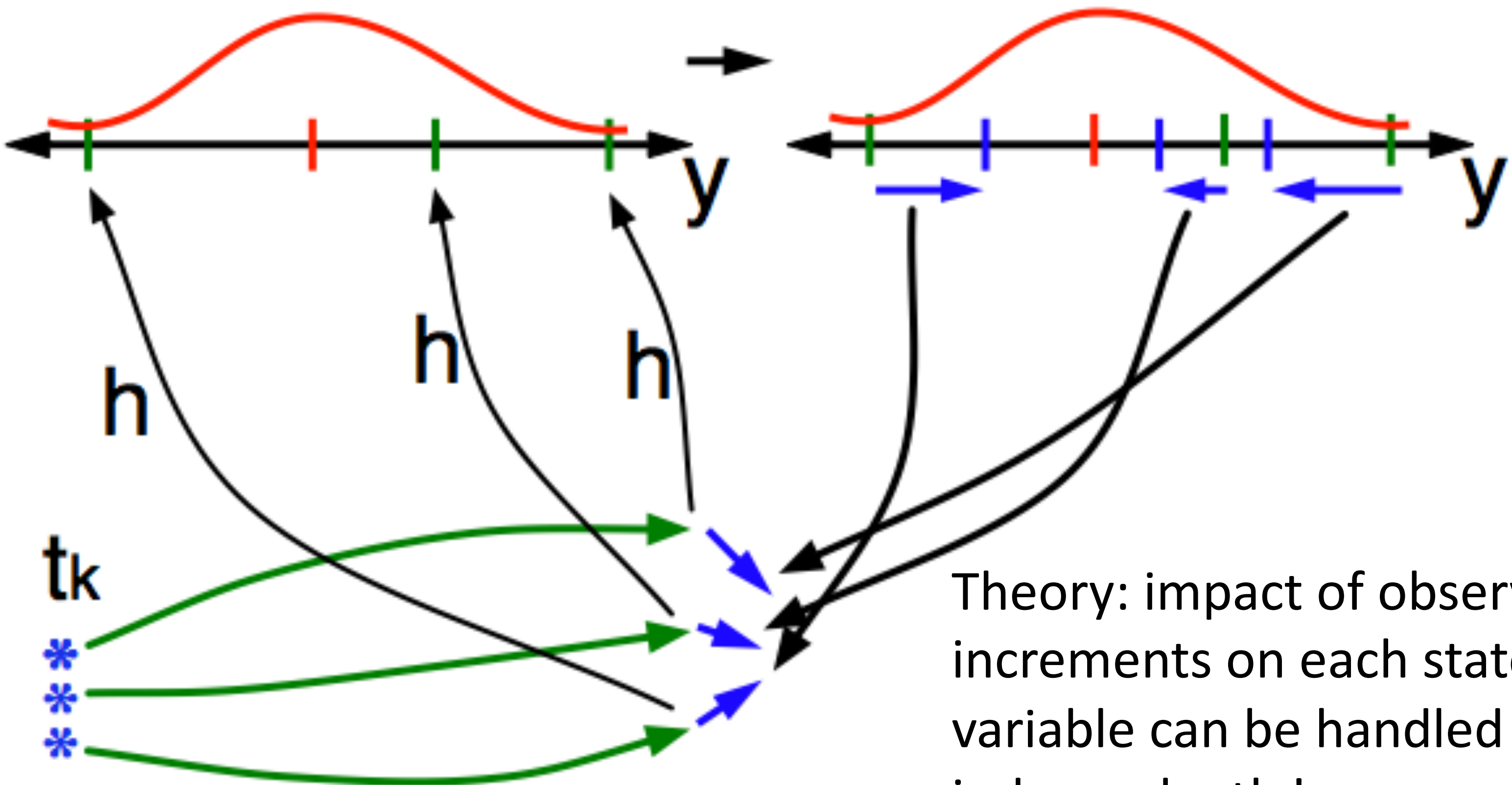
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).



Students, faculty: Lots of cool problems still to be explored for non-Gaussian likelihoods. Contact me if interested.

Schematic of a Sequential Ensemble Filter

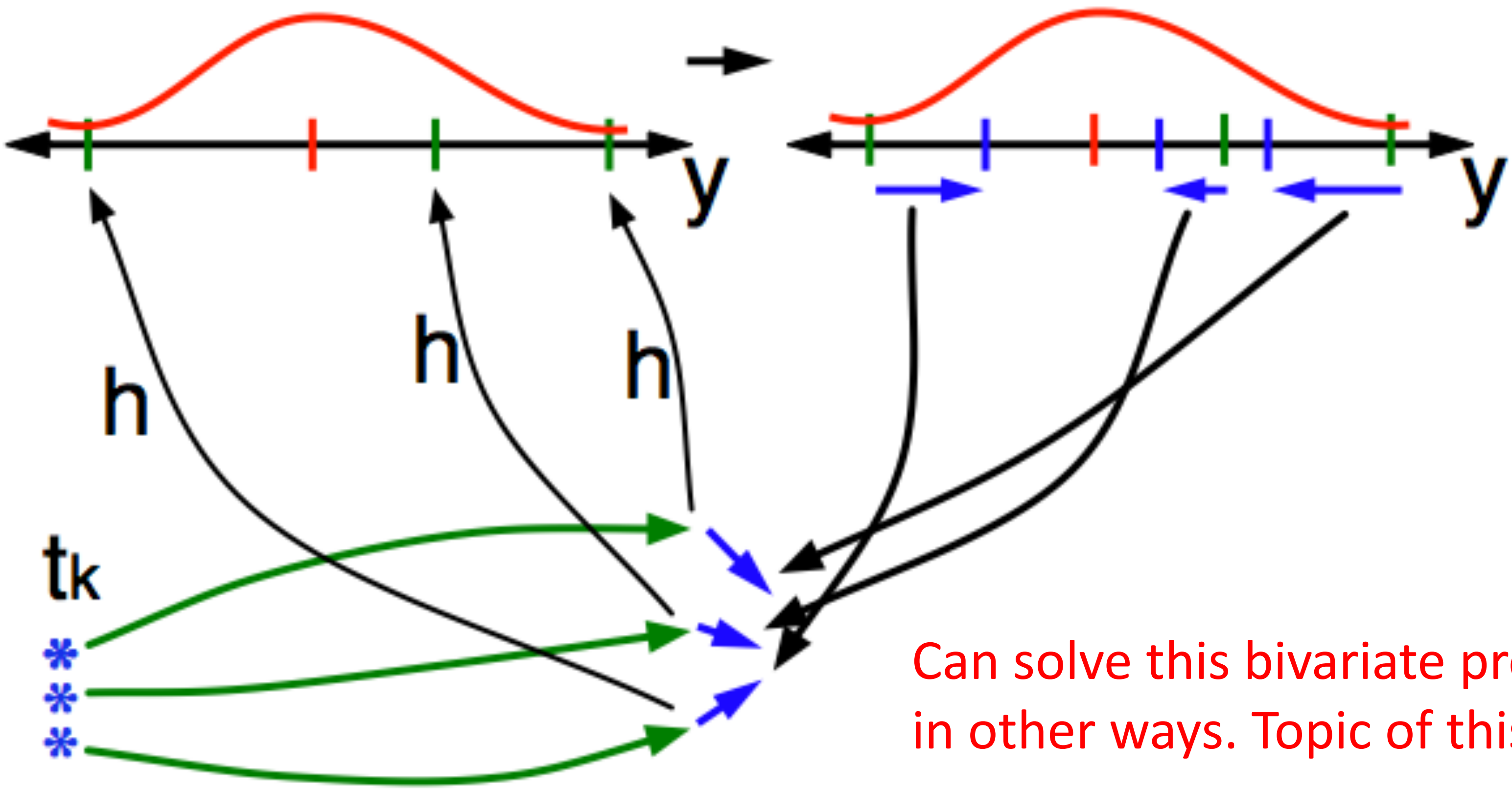
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



Theory: impact of observation increments on each state variable can be handled independently!

Schematic of a Sequential Ensemble Filter

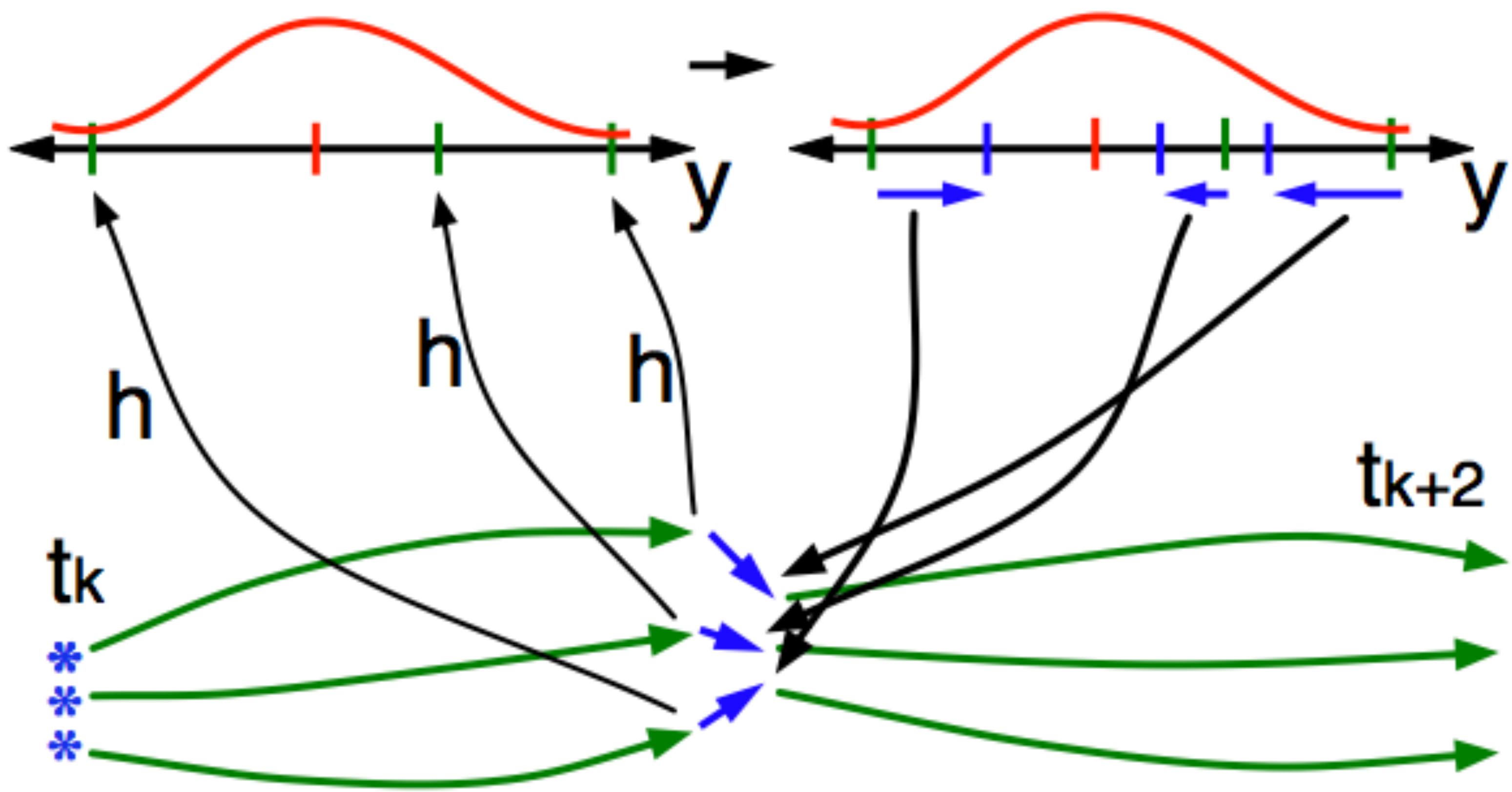
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments.



Can solve this bivariate problem in other ways. Topic of this talk.

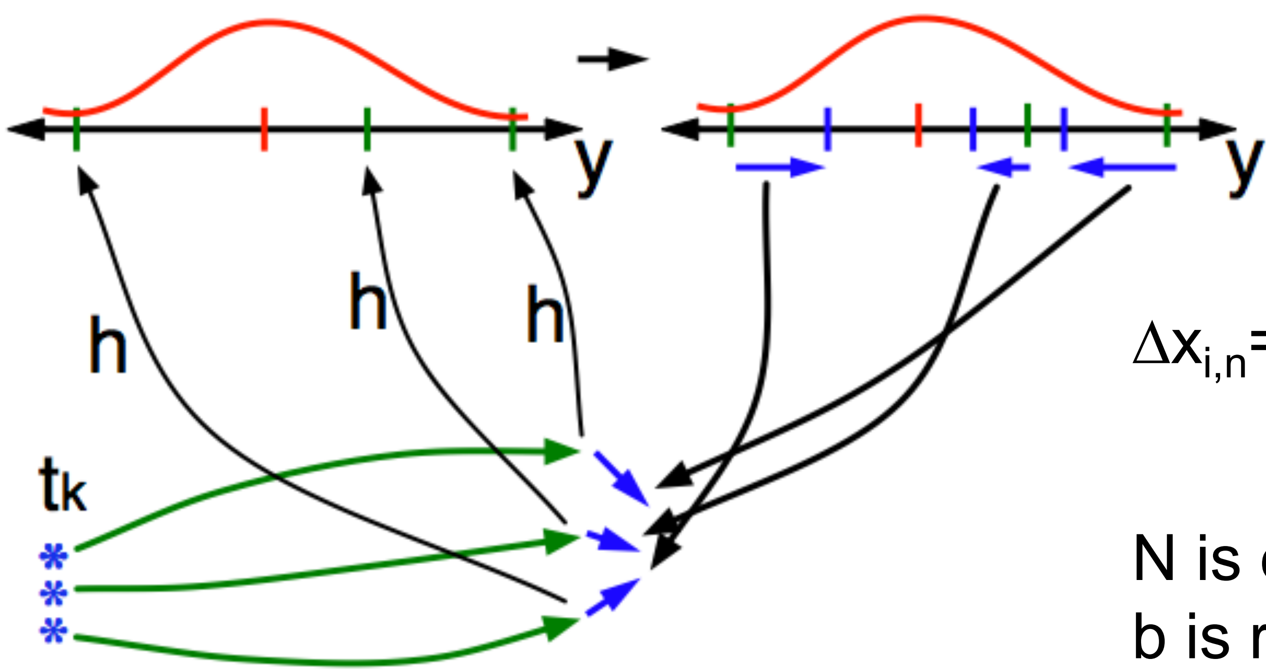
Schematic of a Sequential Ensemble Filter

6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation …



Focus on the Regression Step

Standard ensemble filters just use bivariate sample regression to compute state increments.

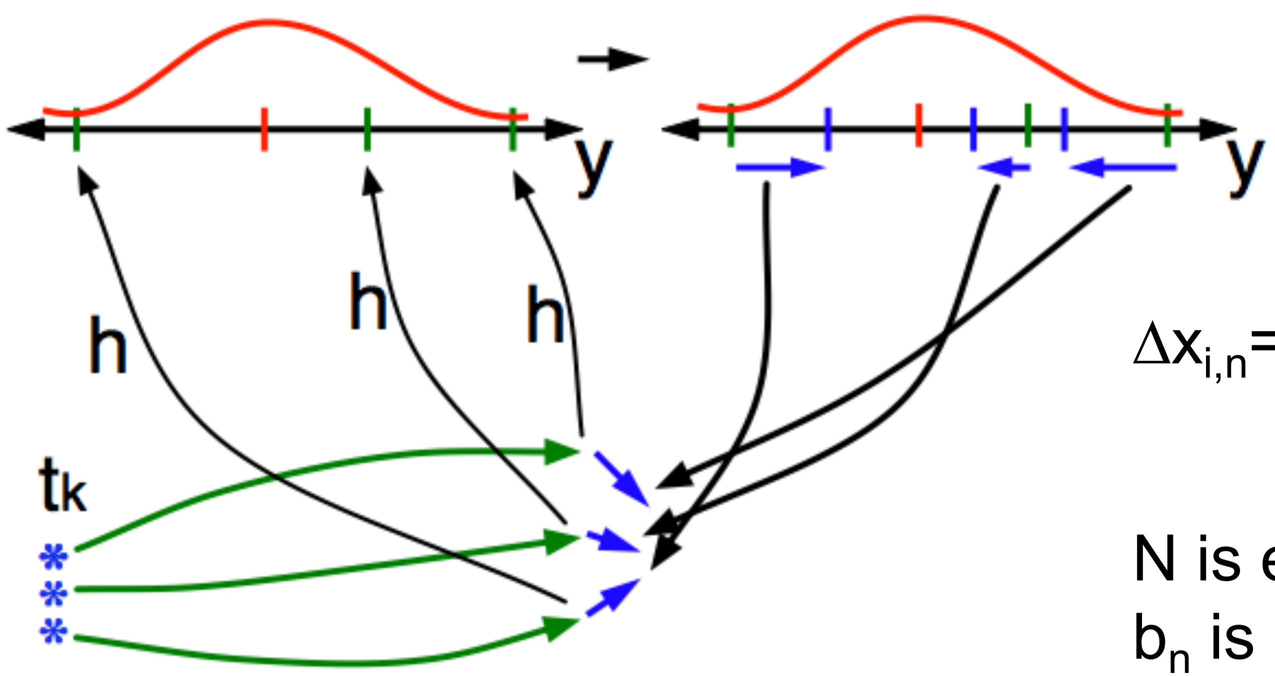


$\Delta x_{i,n} = b\Delta y_n$,
$n=1,\ldots N$.

N is ensemble size.
b is regression coefficient.

Focus on the Regression Step

Will explore using different 'regression' for each ensemble member to compute increments for x_i

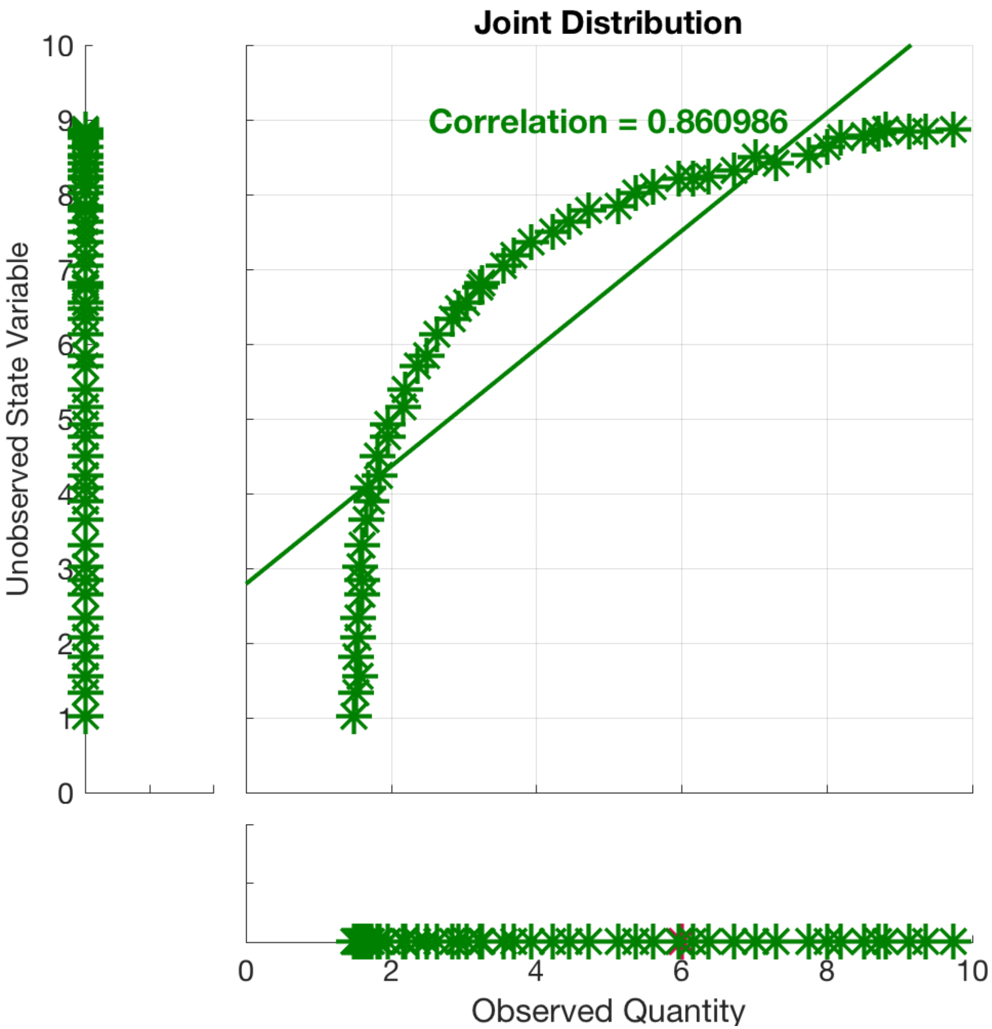


$\Delta x_{i,n} = b_n \Delta y_n$, $n=1,\ldots N$.

N is ensemble size.

b_n is 'local' regression coefficient.

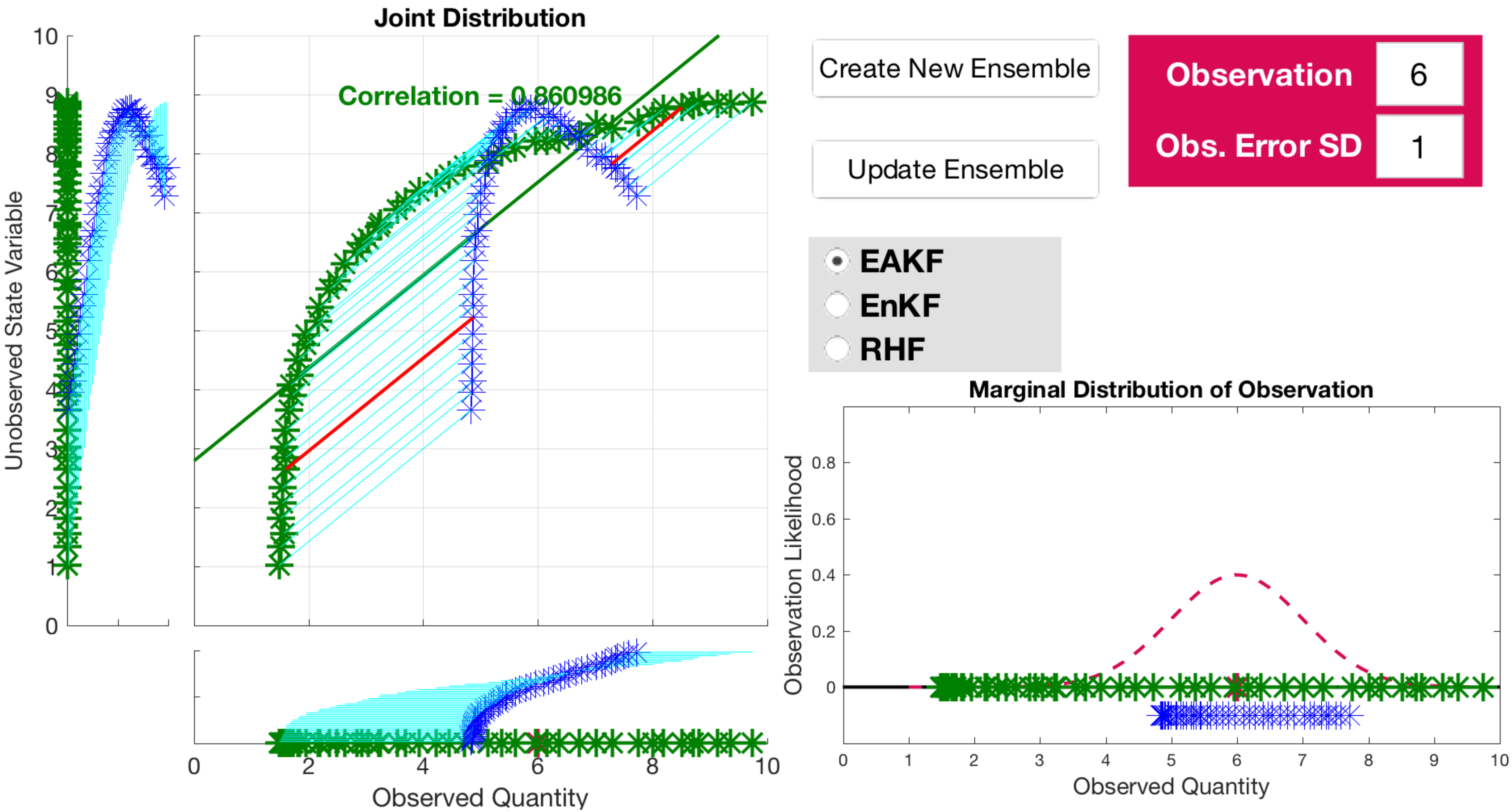
Nonlinear Regression Example



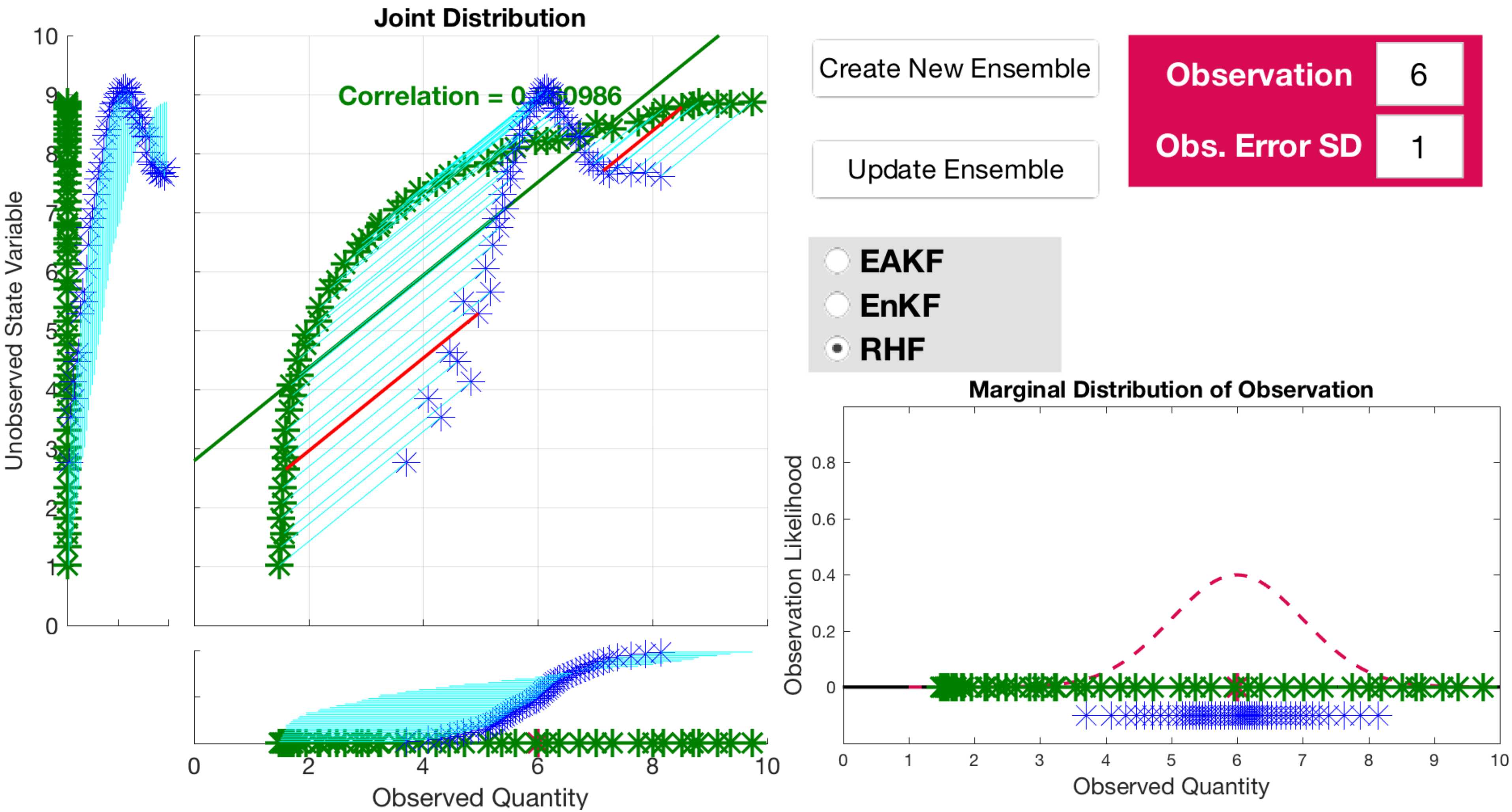
Try to exploit nonlinear prior relation between a state variable and an observation.

Example: Observation $y \sim x^n$, for instance $y = T^4$.

Standard Ensemble Kalman Filter (EAKF)



Nongaussian Filter (Rank Histogram Filter)



Local Linear Regression

Relation between observation and state is nonlinear.

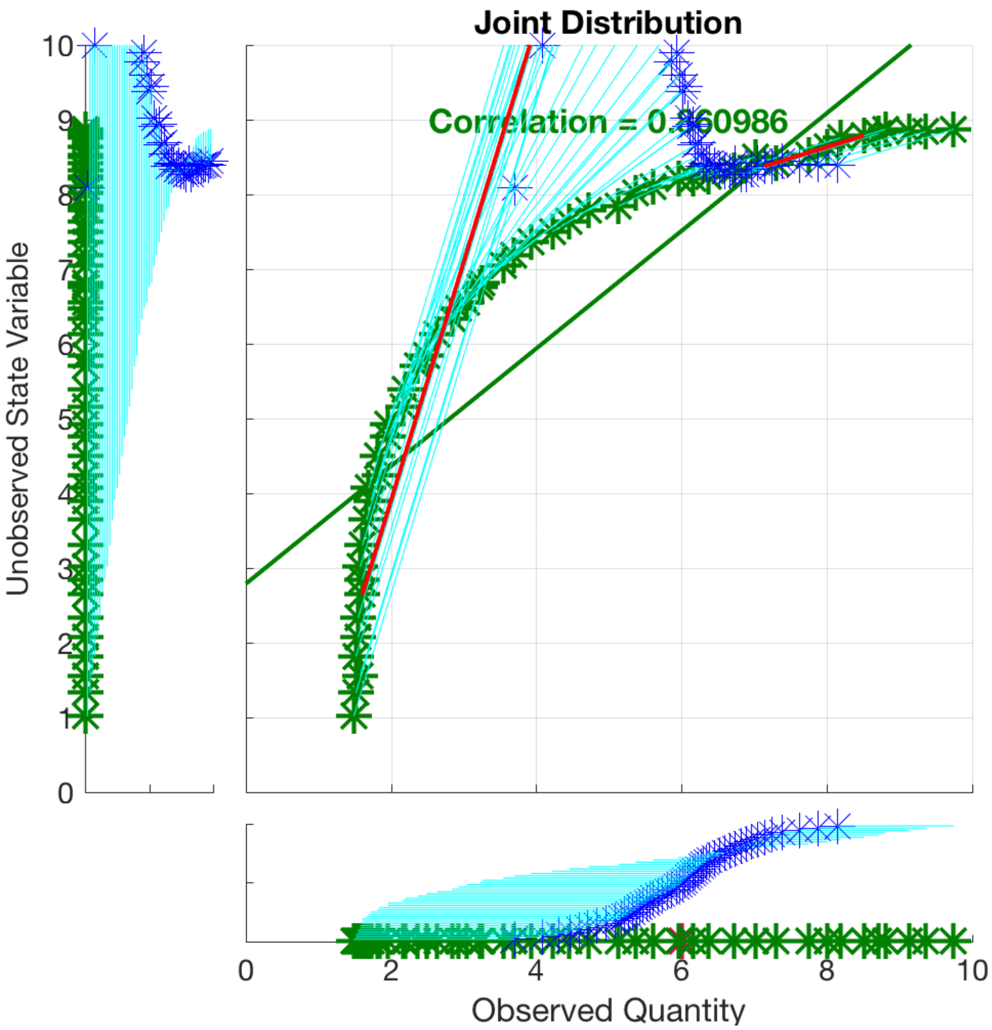
Try using 'local' subset of ensemble to compute regression.

What kind of subset?

Cluster that contains ensemble member being updated.

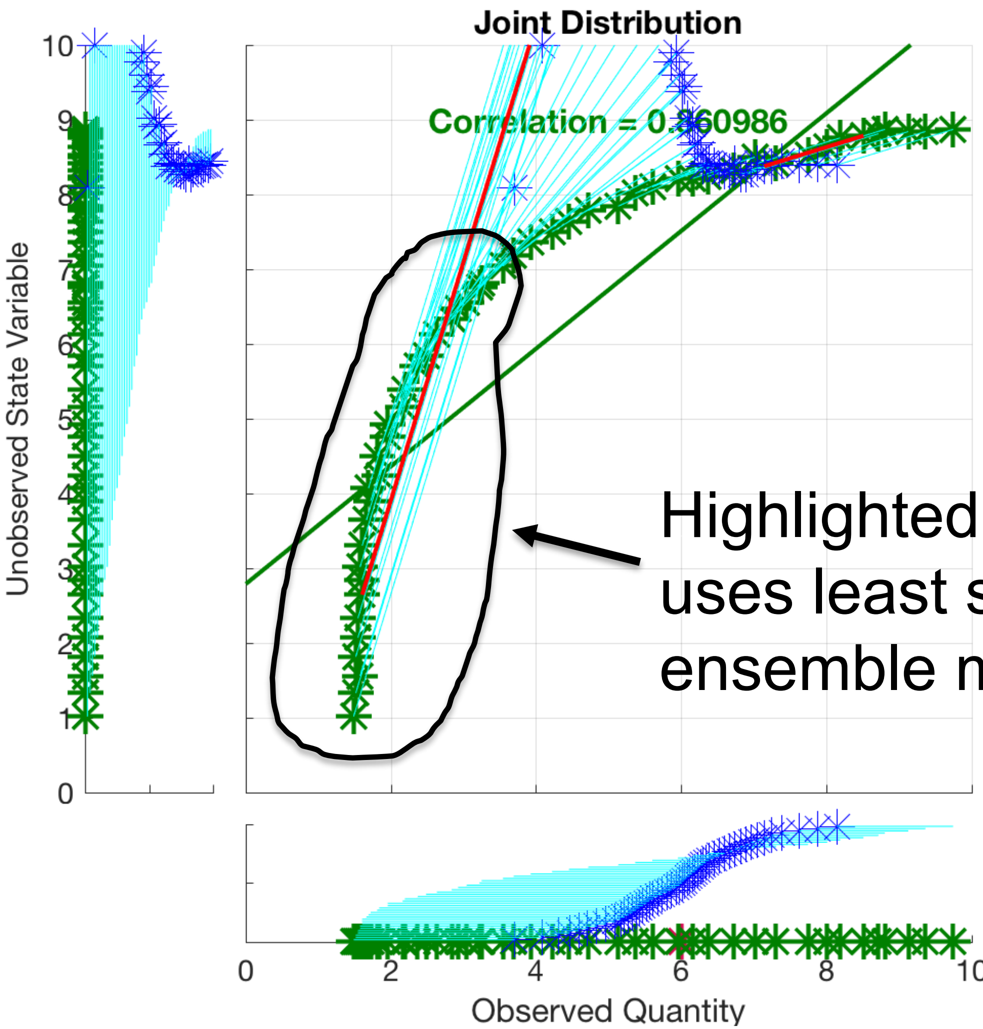
- Lots of ways to define clusters.
- Here, use naïve closest neighbors in (x,y) space.
- Vary number of nearest neighbors in subset.

Local Linear Regression



Local ensemble subset is nearest ½ . Regression approximates local slope of the relation.

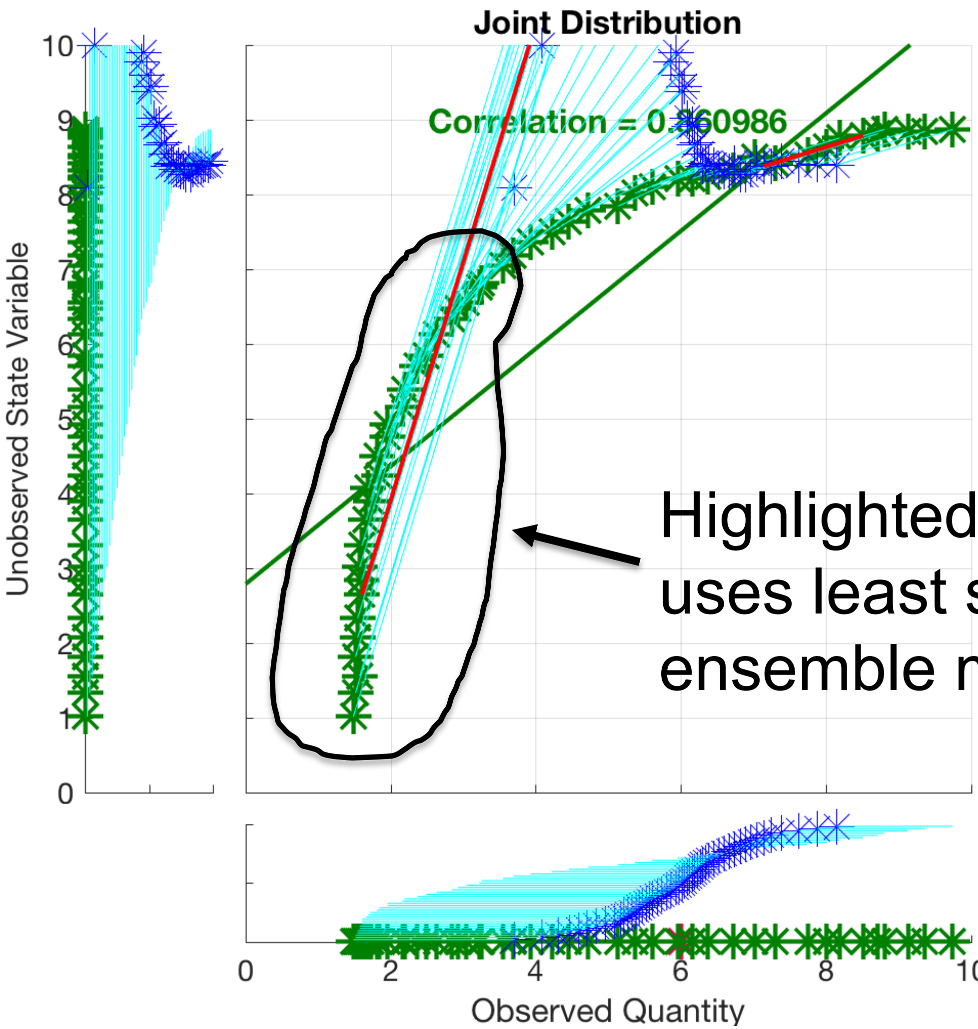
Local Linear Regression



Local ensemble subset is nearest ½ . Regression approximates local slope of the relation.

Highlighted red increment uses least squares fit to ensemble members in region.

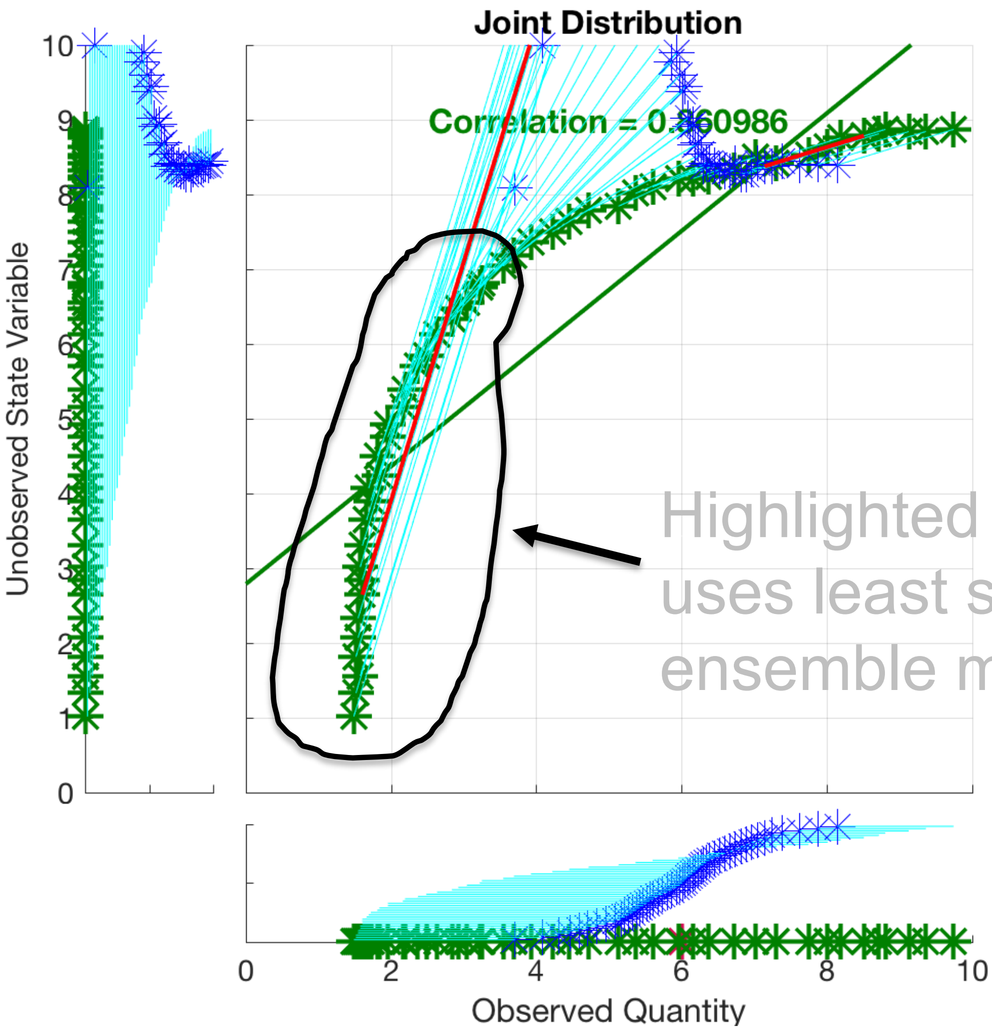
Local Linear Regression



Slope more accurate locally, but a disaster globally.

Highlighted red increment uses least squares fit to ensemble members in region.

Local Linear Regression



Note similarity to Houtekamer's method, except local ensemble members are used, rather than non-local.

Highlighted red increment uses least squares fit to ensemble members in region.

Local Linear Regression with Incremental Update

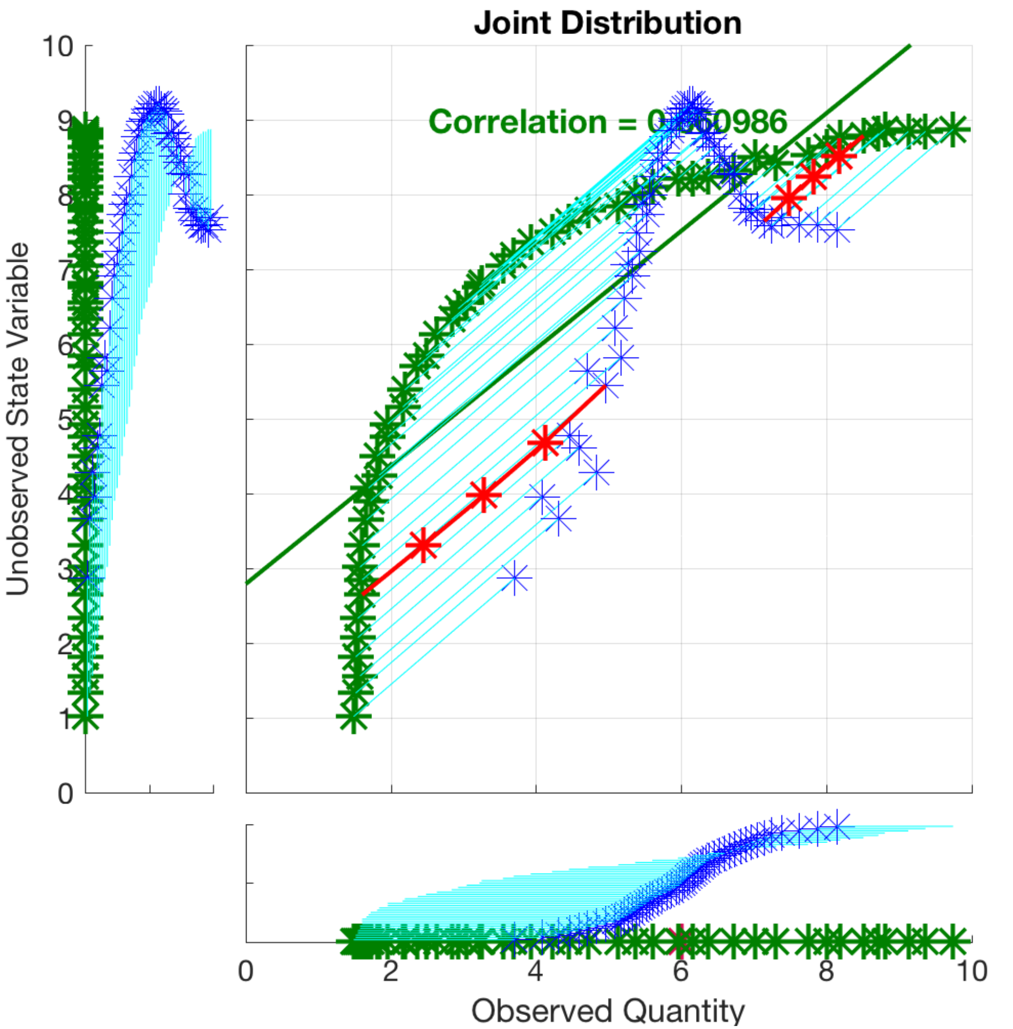
Local slope is just that, local.

Following it for a long way is a bad idea.

Will use a Bayesian consistent incremental update.

- Observation with error variance *s*.
- Assimilate *k* observations with this value.
- Each of these has error variance *s/k*.

Incremental Update

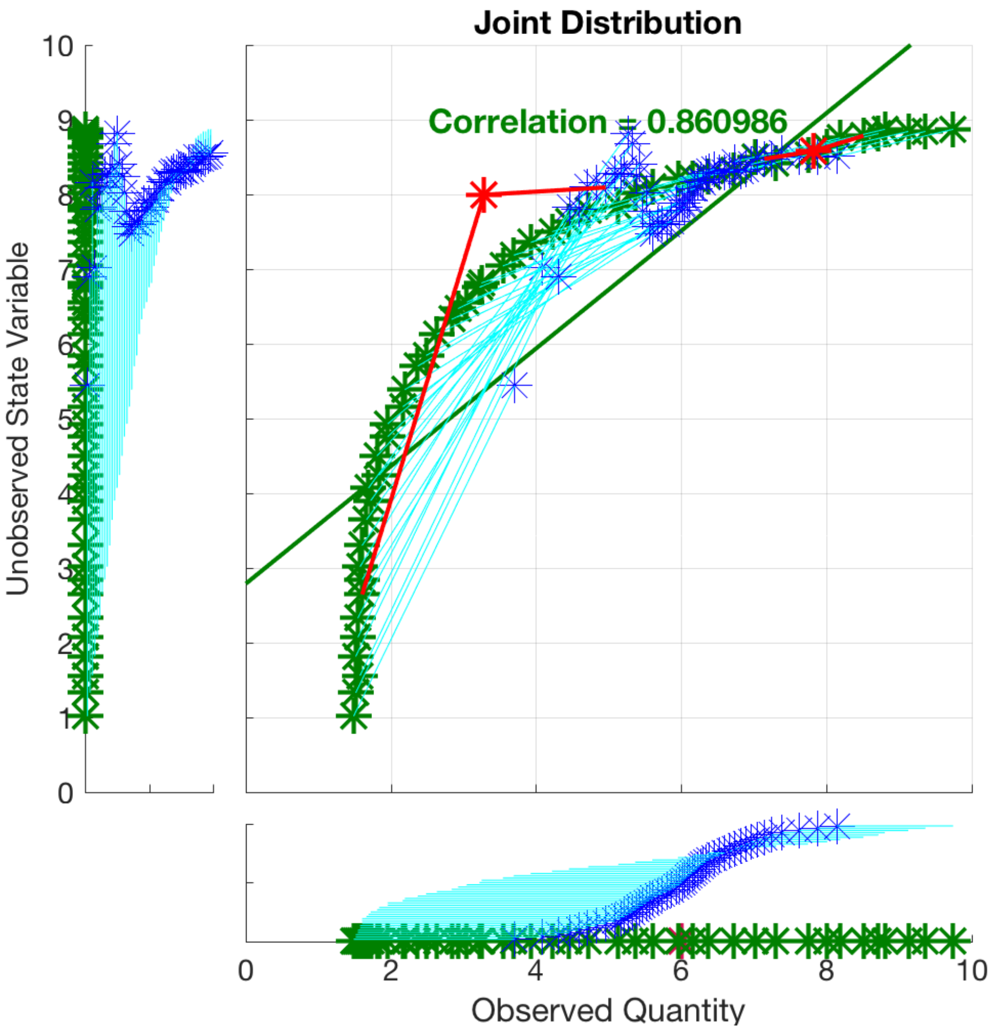


This is an RHF update with 4 increments. Individual increments highlighted for two ensemble members.

For an EAKF, posterior would be identical to machine precision.

Nearly identical for RHF.

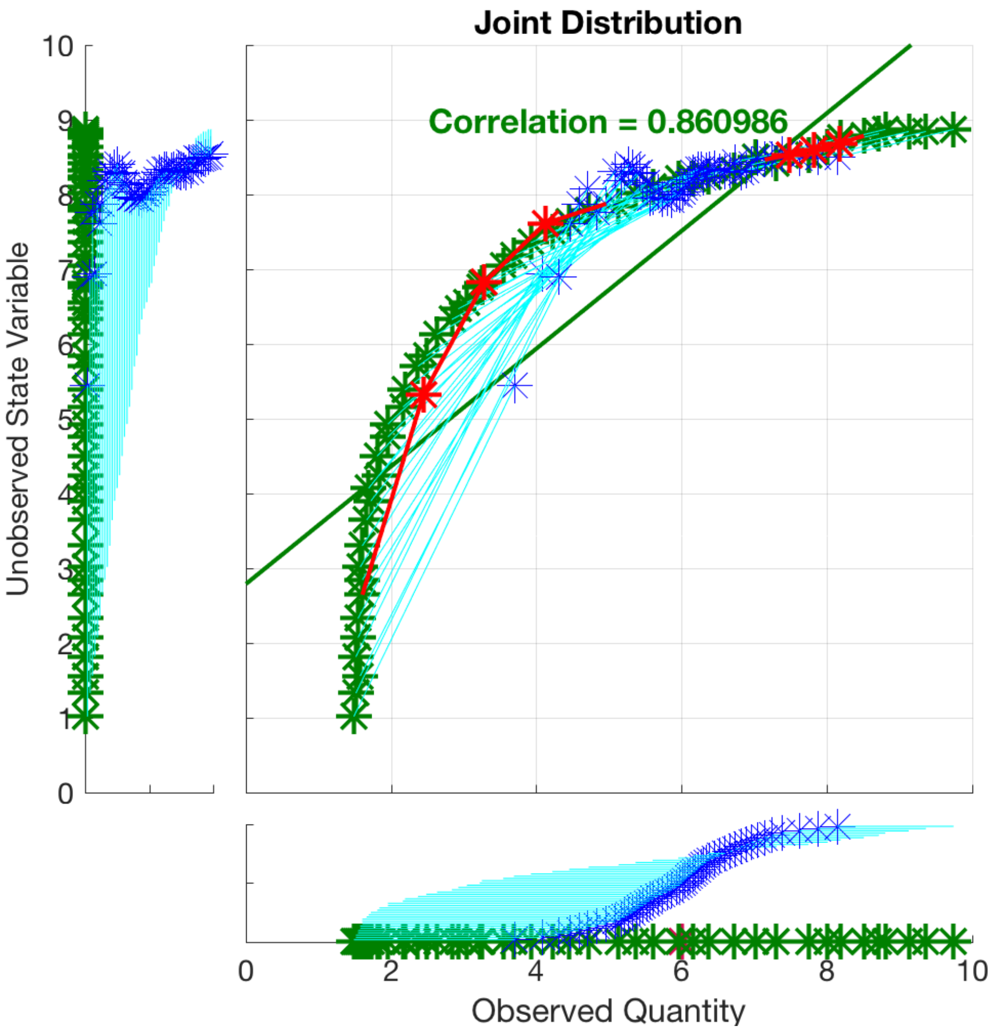
Local Linear Regression with Incremental Update



2 increments with subsets ½ ensemble.

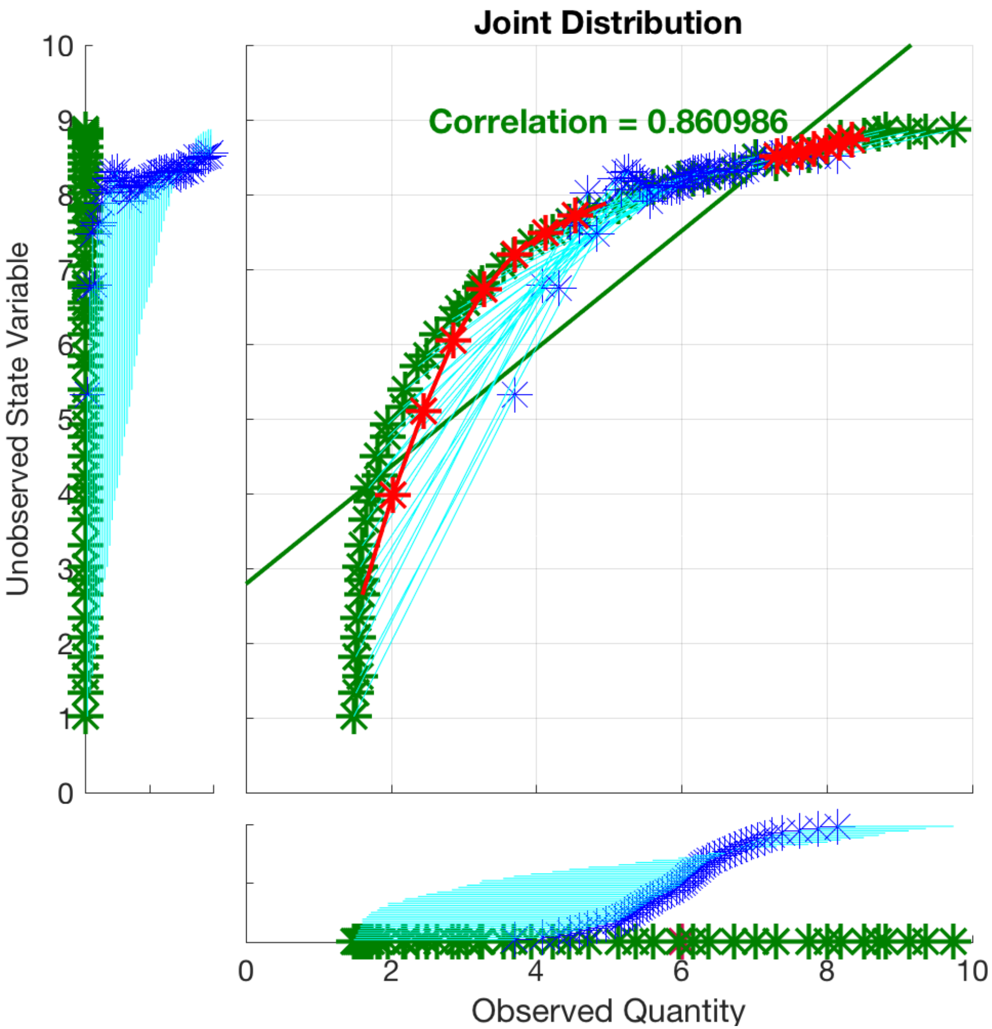
Posterior for state qualitatively improving.

Local Linear Regression with Incremental Update



4 increments with subsets ½ ensemble.

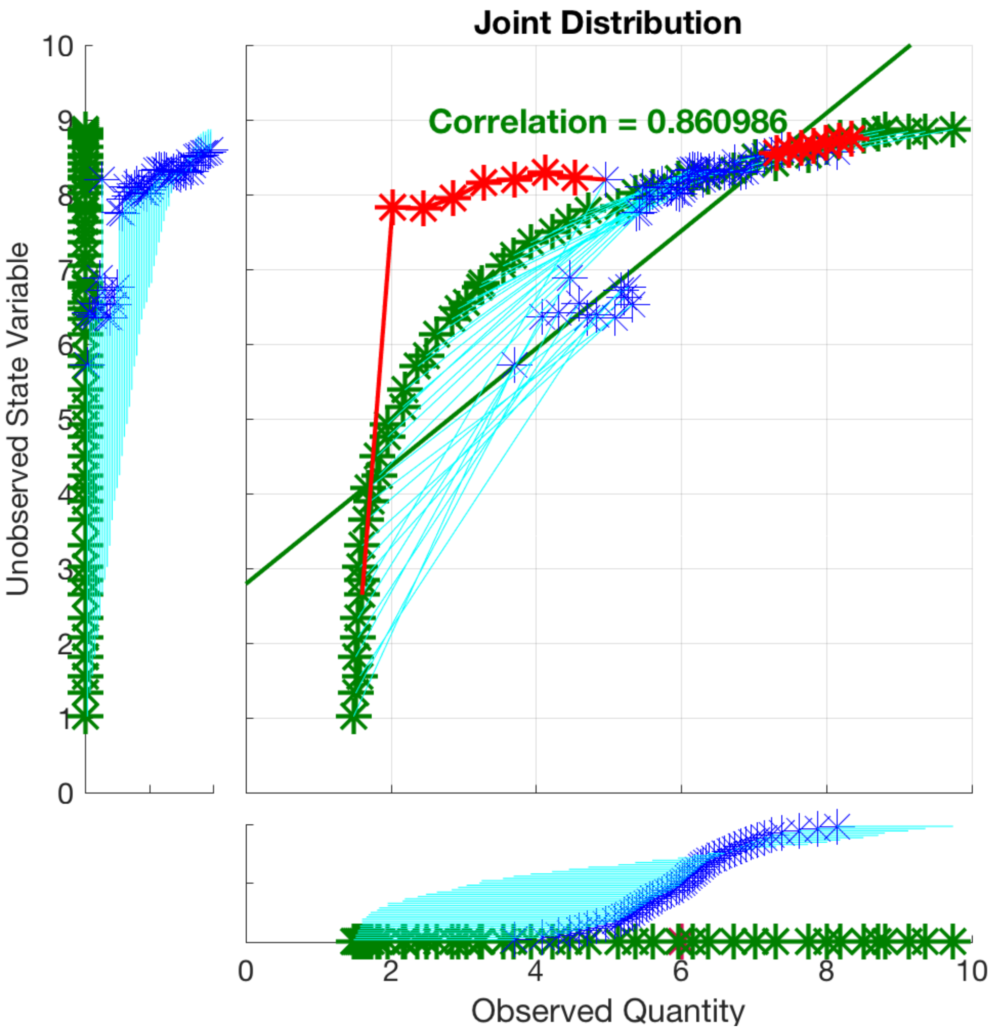
Posterior for state qualitatively improving.



8 increments with subsets ½ ensemble.

Posterior for state qualitatively improving.

Local Linear Regression with Incremental Update

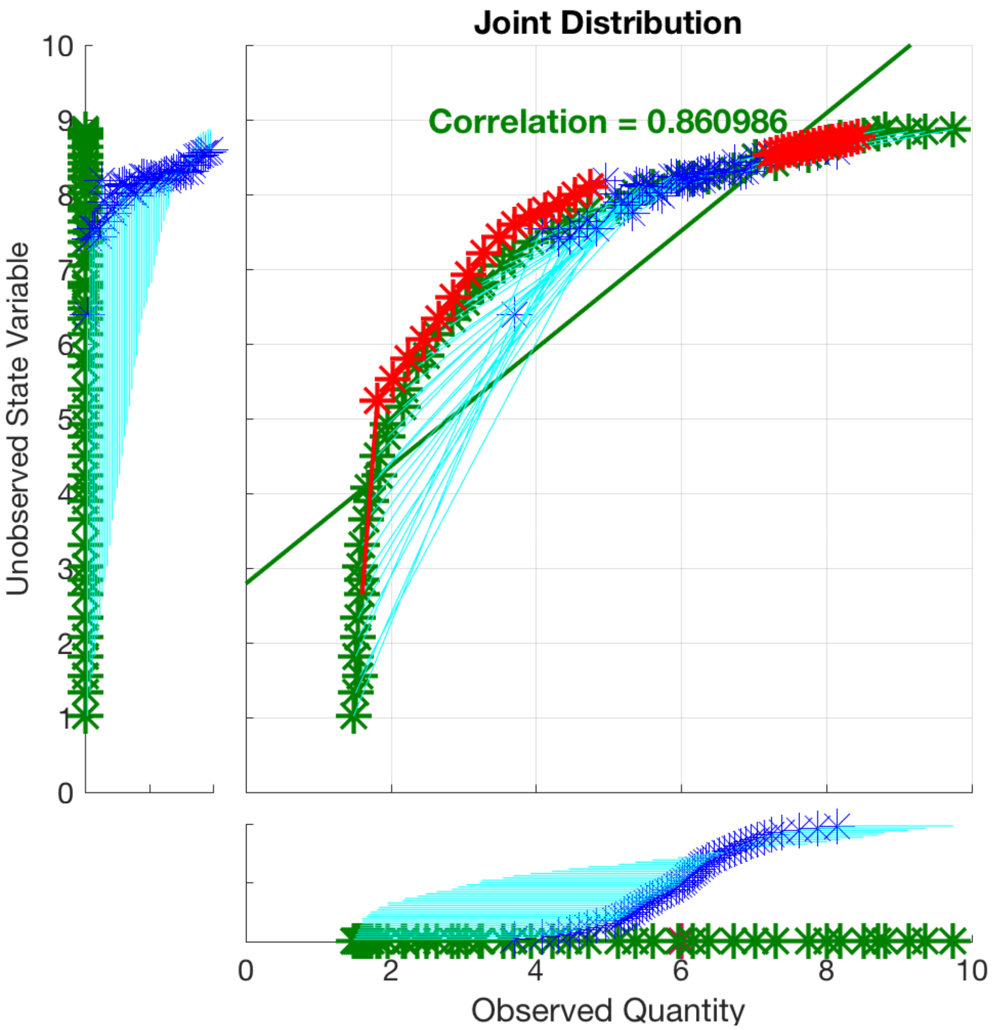


8 increments with subset 1/4 ensemble.

Posterior for state degraded.

Increment is moving outside of local linear validity.

Local Linear Regression with Incremental Update



16 increments with subset 1/4 ensemble.

Posterior for state improved.

Local Linear Regression with Incremental Update

If relation between observation and state is locally a continuous, smooth (first two derivatives continuous) function:
Then, in the limit of a large ensemble, fixed local subset size, and large number of increments:

The local linear regression with incremental update converges to the correct posterior distribution.

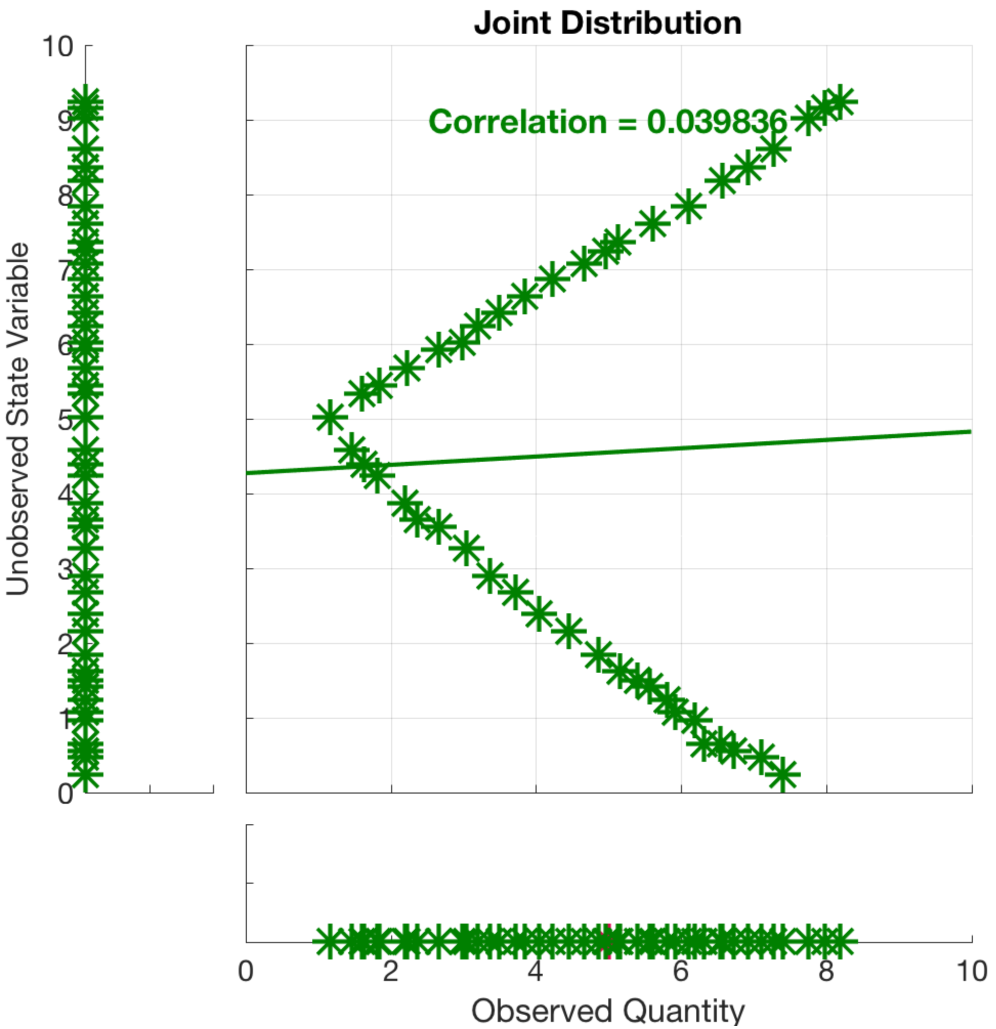
Local Linear Regression with Incremental Update

If relation between observation and state is locally a continuous, smooth (first two derivatives continuous) function:
Then, in the limit of a large ensemble, fixed local subset size, and large number of increments:

The local linear regression with incremental update converges to the correct posterior distribution.

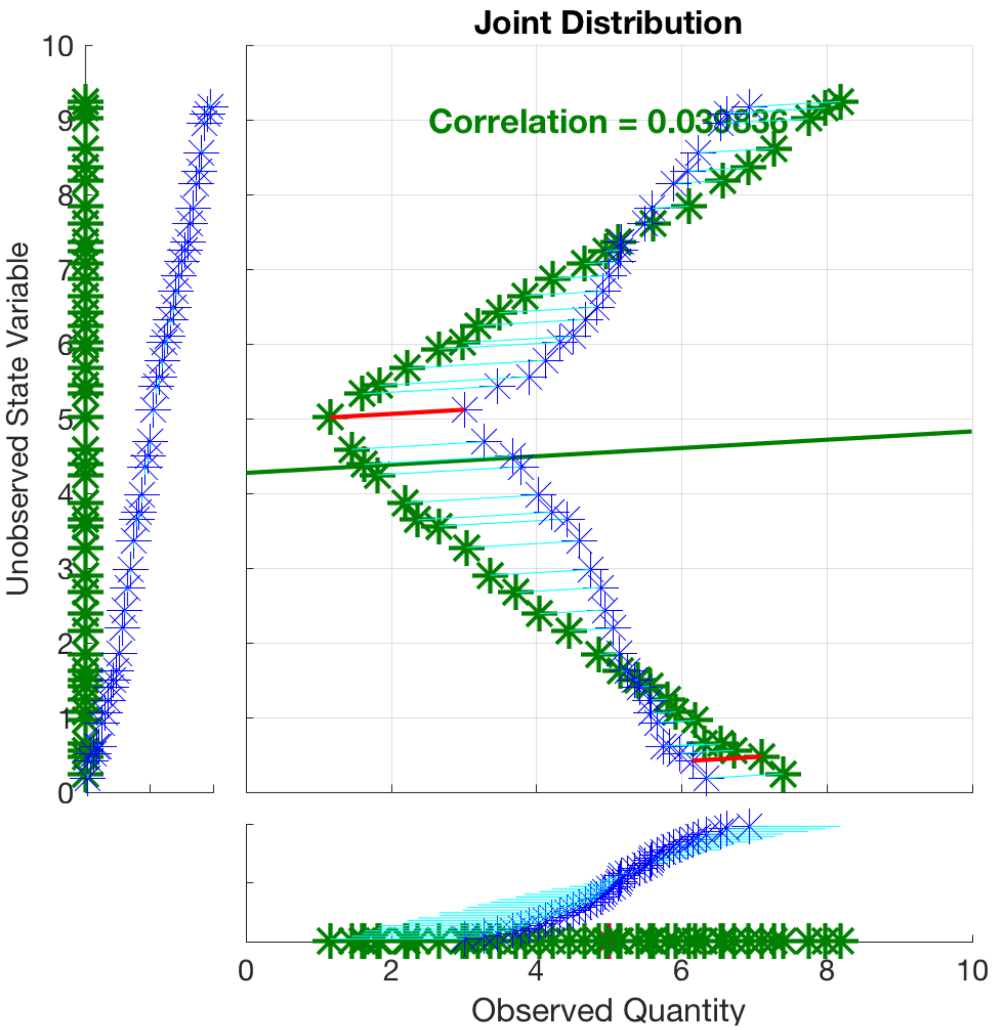
This could be very expensive,
No guarantees about what goes on in the presence of noise.

Multi-valued, not smooth example.



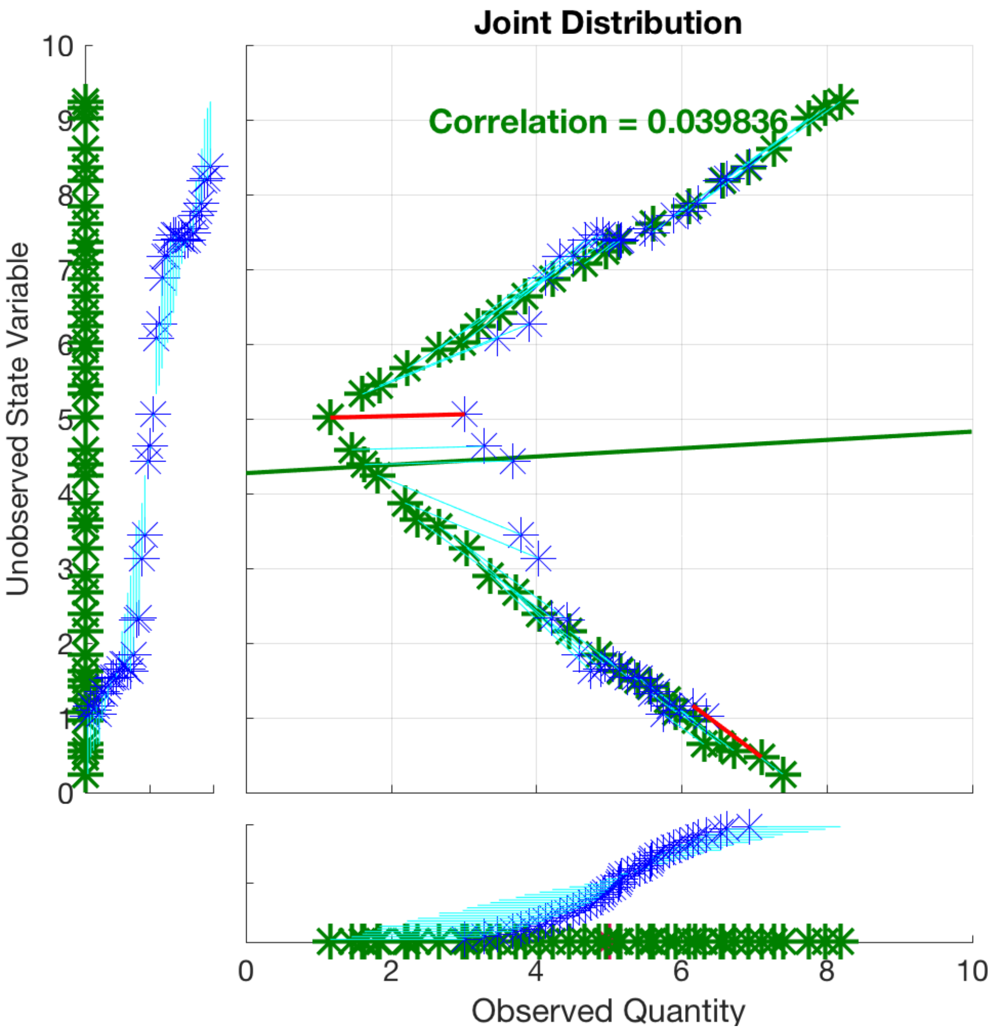
Similar in form to a wind speed observation with state velocity component.

Multi-valued, not smooth example.



Standard regression does not capture bimodality of state posterior.

Multi-valued, not smooth example.

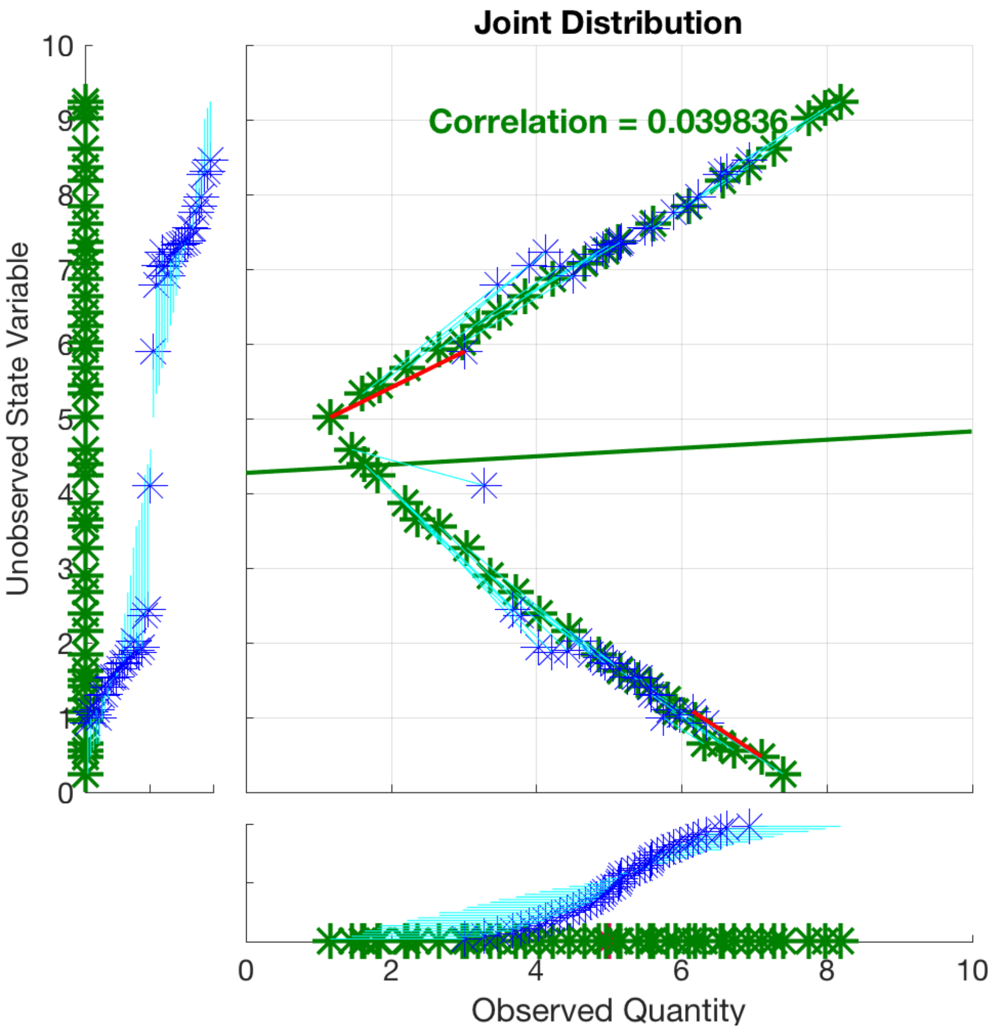


Local regression with ½ of the ensemble does much better.

Captures bimodal posterior.

Note problems where relation is not smooth.

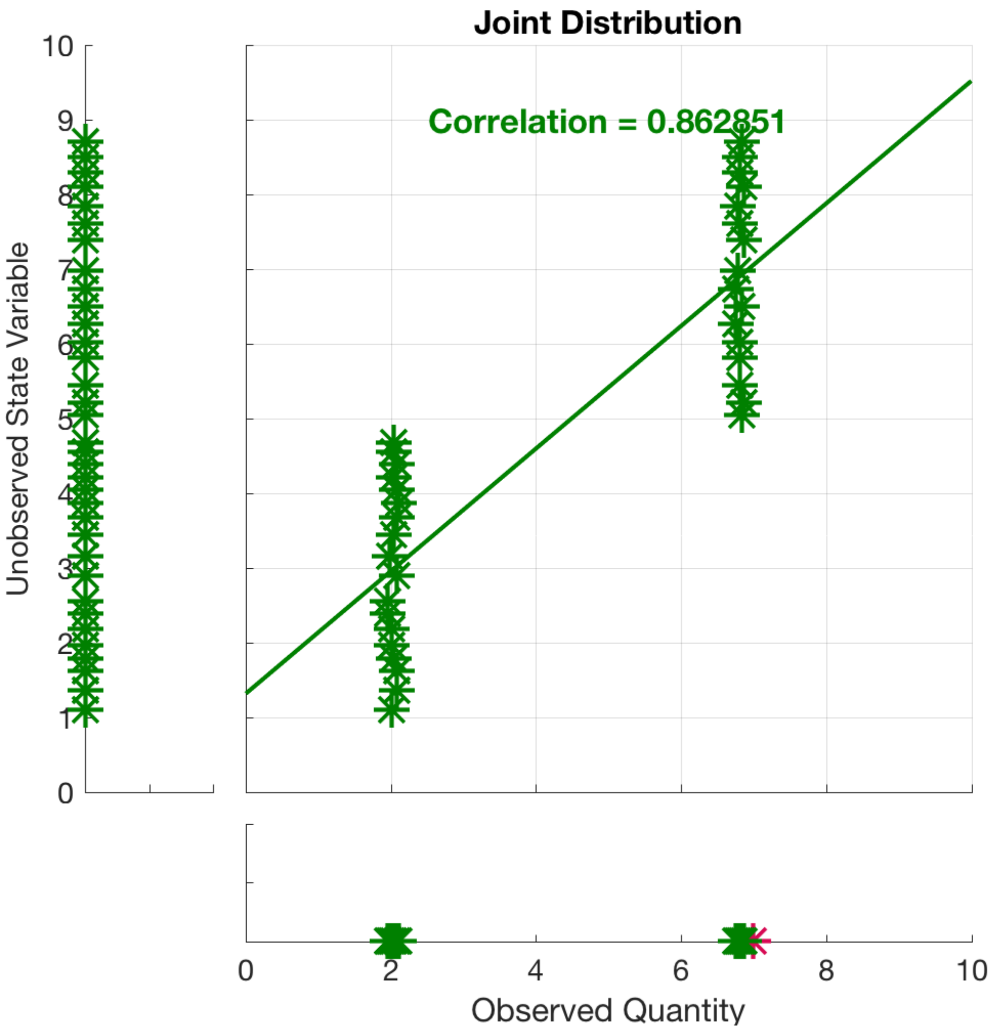
Multi-valued, not smooth example.



Local regression with1/4 of the ensemble does even better.

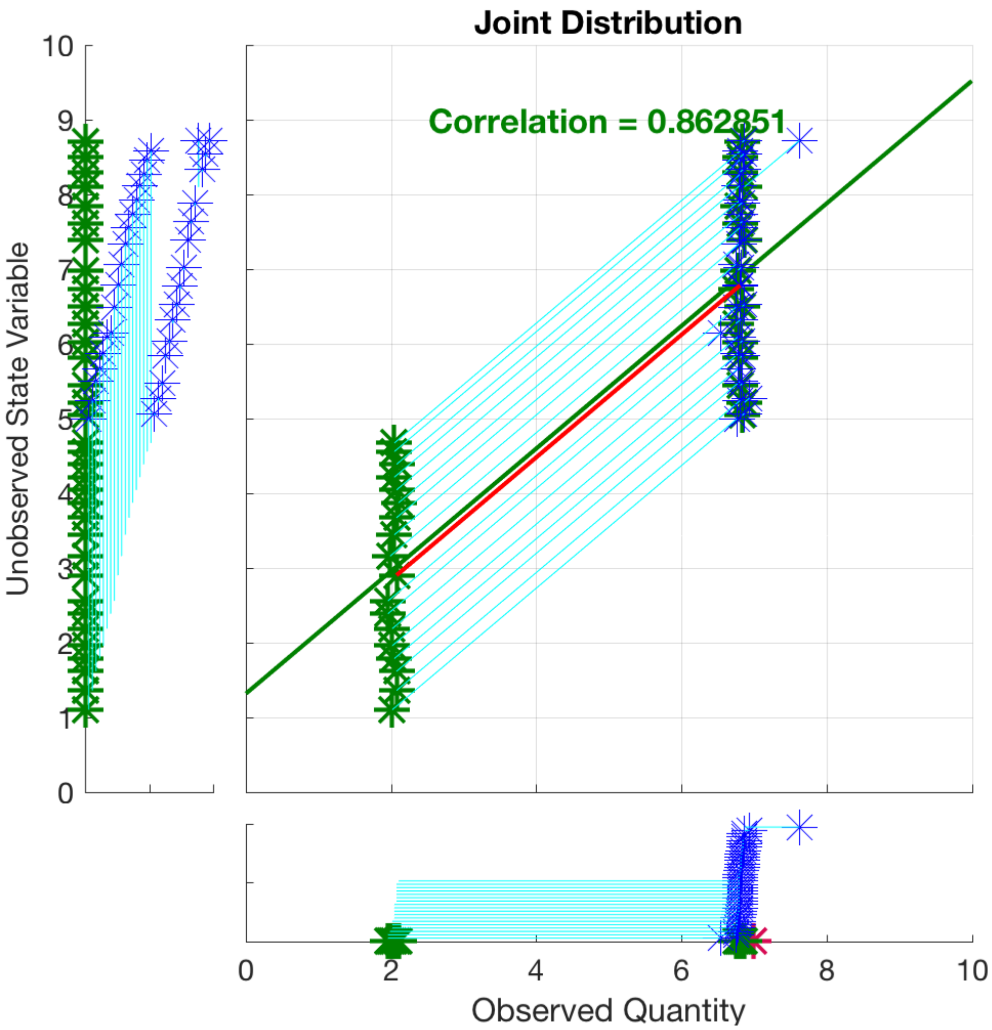
No need for incremental updates here.

Discontinuous example.



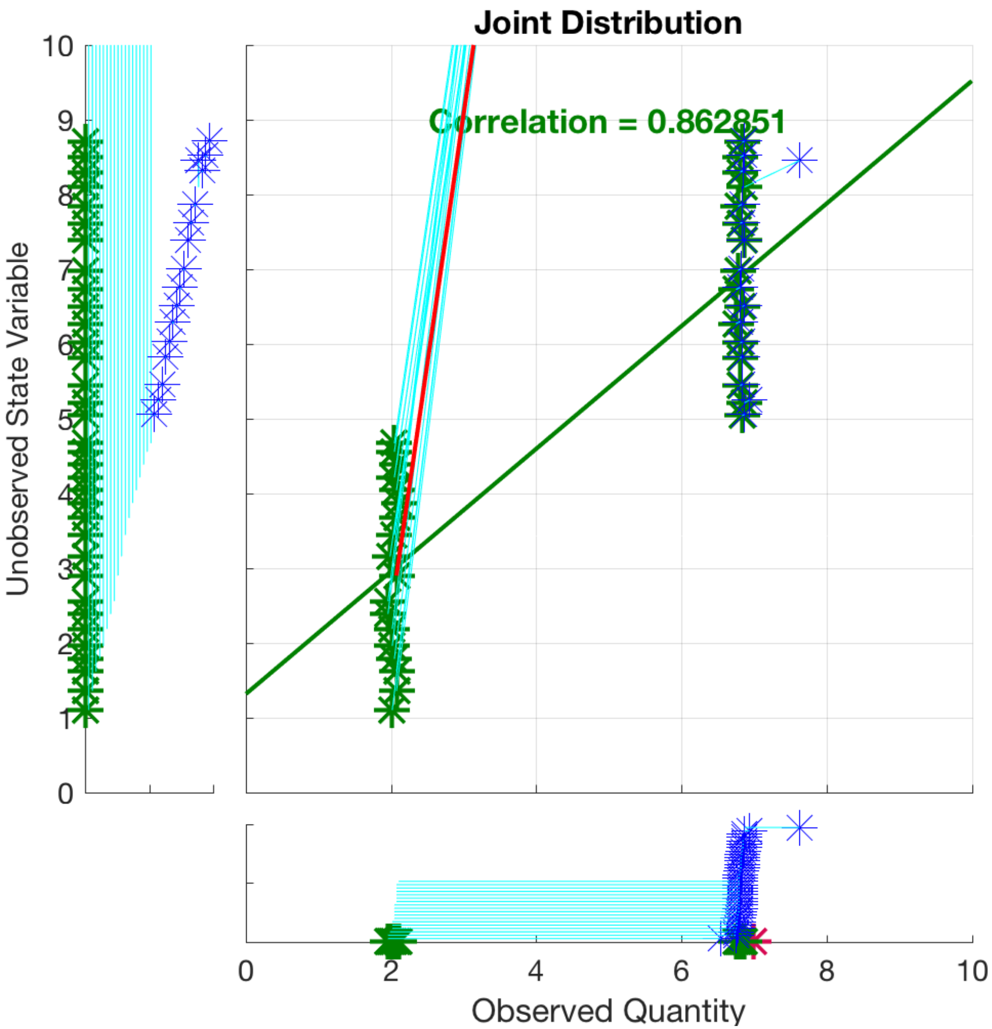
Example of threshold process. Like measurement of rainfall rate and state variable of low-level temperature.

Discontinuous example.



Basic RHF does well on this, but it's an 'accident'.

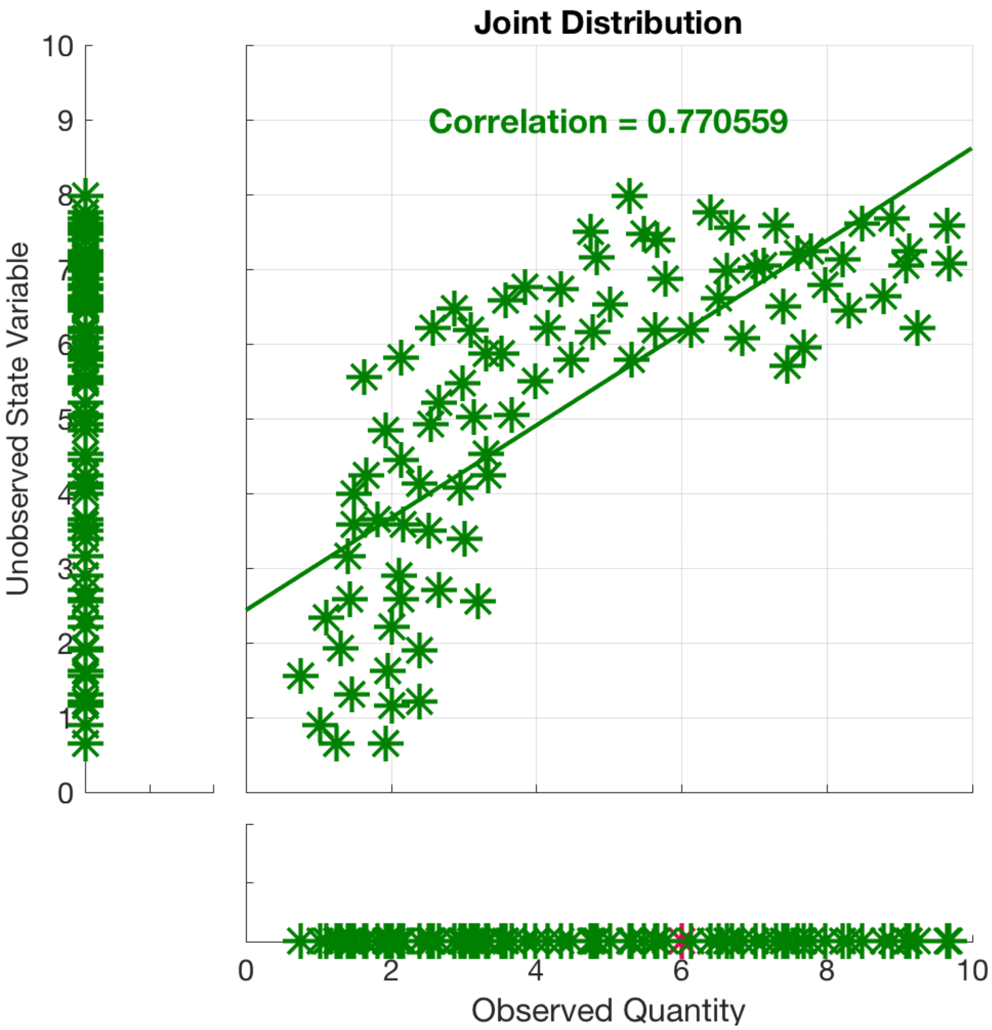
Discontinuous example.



Local regression with the two obvious clusters leads to ridiculous posterior.

Lack of continuity of bivariate prior makes any regression-like update problematic.

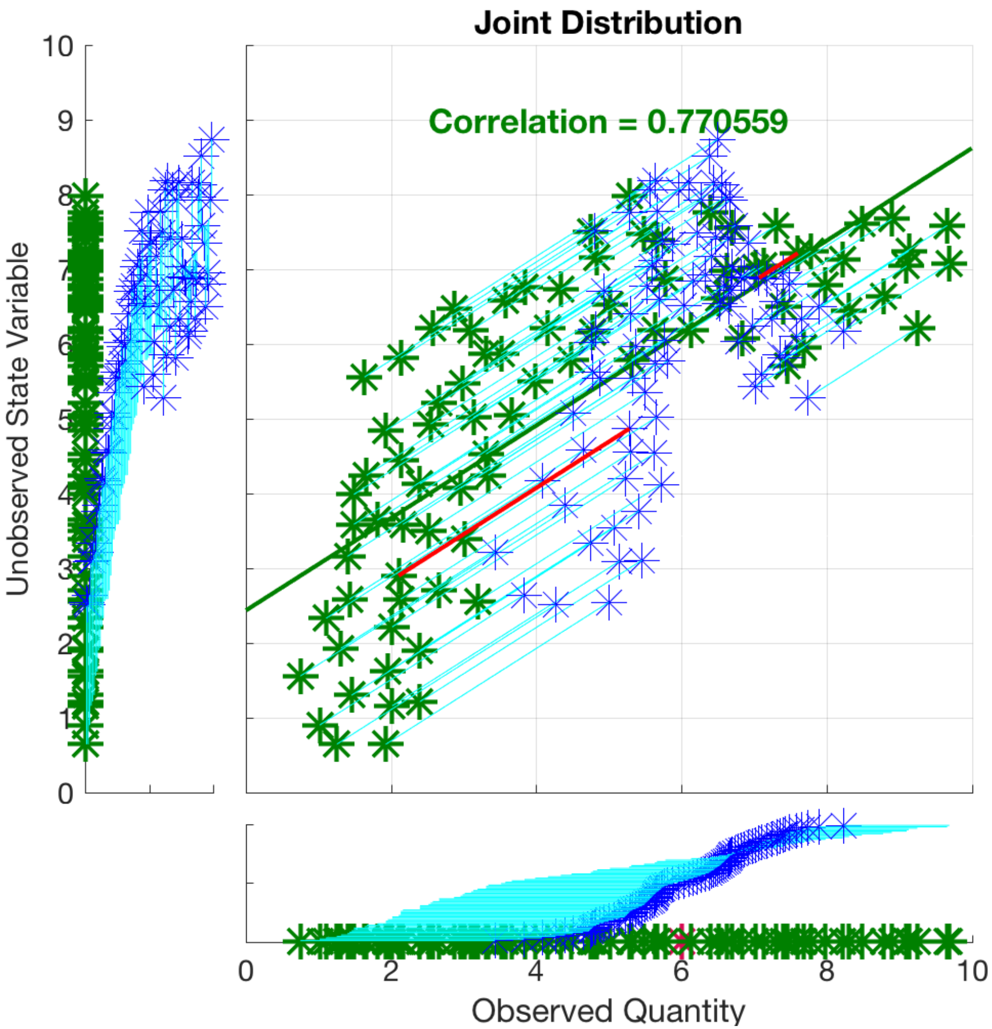
Local Regression with Noisy Priors



Most geophysical applications have noisy bivariate priors.

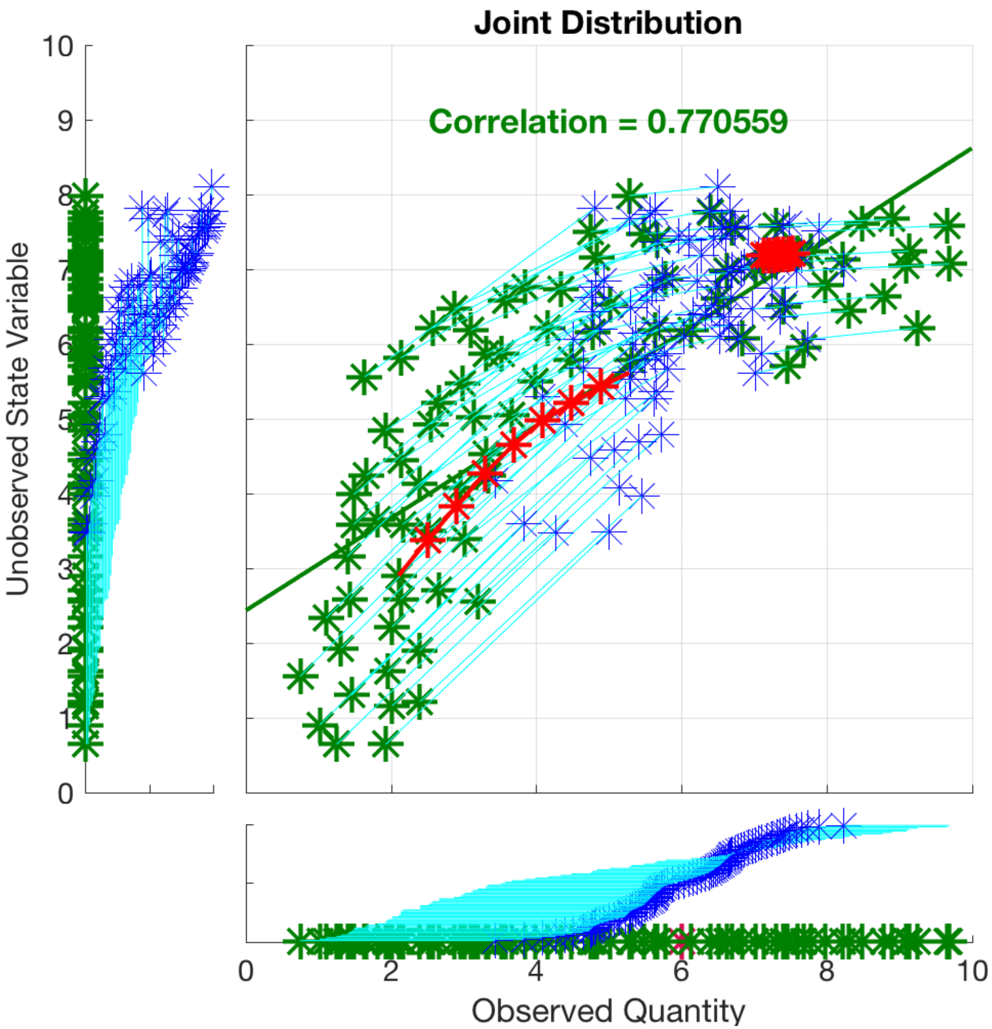
Usually hard to detect nonlinearity (even this example is still pretty extreme).

Local Regression with Noisy Priors



Basic RHF can still be clearly suboptimal.

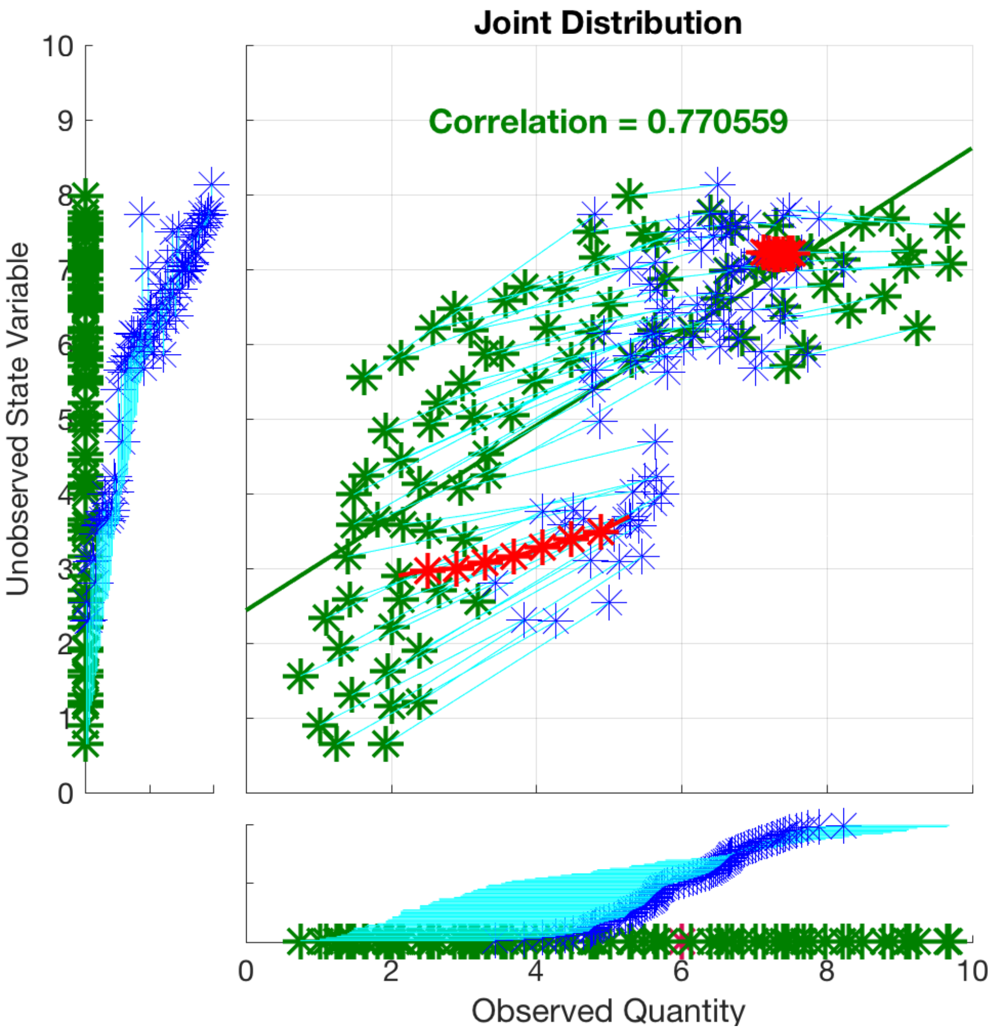
Local Regression with Noisy Priors



This result is for local ensemble with nearest ½ of ensemble and 8 increments.

Need bigger local ensembles to reduce sampling errors.

Local Regression with Noisy Priors



This result is for local ensemble with nearest 1/4 of ensemble and 8 increments.

The small ensemble subsets lead to large sampling error. Probably worse than standard RHF.

Results: Lorenz96

Standard model configuration, perfect model.
Observation at location of each of the state variables.
Two observation frequencies: every step, every 12 steps.
Three observation types: state, square of state, log of state.
Observation error variance tuned to give time mean RMSE of approximately 1.0 for basic RHF.
80 total ensemble members.

For local regression:

- Local subsets of 60 members (pretty large).
- 4 increments.

Tuned adaptive inflation standard deviation, localization for lowest RMSE in base RHF case.

Results: Lorenz96

Observations of state:

Results from base RHF and local regression statistically indistinguishable.

Observations of square of state:

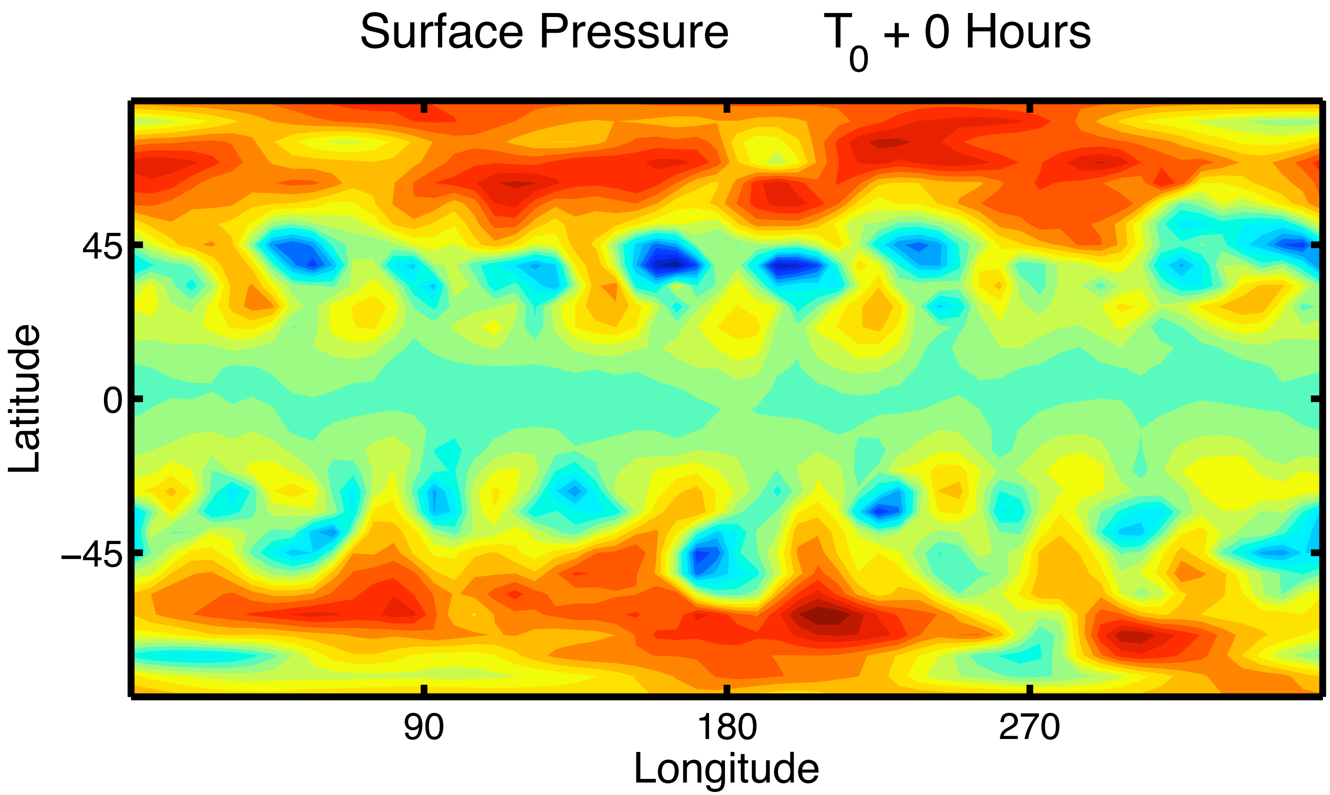
Results from local regression are significantly better, but filter suffers occasional catastrophic divergence.

Must be restarted in these cases.

Observations of log of state:

Results from local regression are significantly better.

Low-Order Dry Dynamical Core



Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Conclusions

Sequential ensemble filters can:

- Apply non-Gaussian methods in observation space,
- Nonlinear methods for bivariate regression.

Local regression with incremental update can be effective for locally smooth, continuous relations. Useful for:

- Nonlinear forward operators,
- Transformed state variables (log, anamorphosis, …).

Can be expensive for ’noisy’ bivariate priors:

- Requires large subsets (hence large ensembles),
- Subsets can be found efficiently,
- Incremental update is a multiplicative cost.

Conclusions

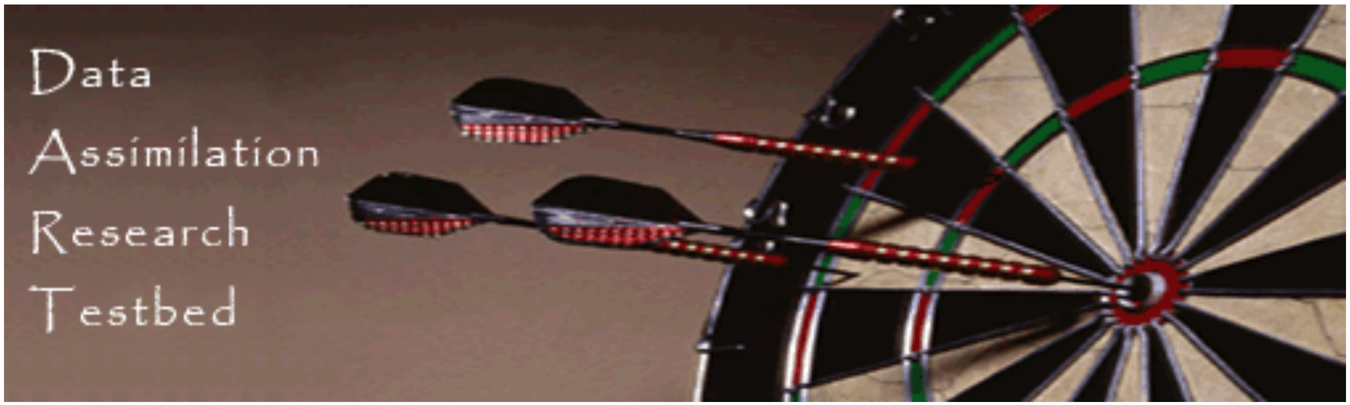
Hard to beat standard regression for many applications.

Being smart about when to apply nonlinear methods is key.

- Detect nonlinear bivariate priors automatically.
- Apply appropriate methods.

Multivariate relations may still be tricky.

All results here with DARTLAB tools freely available in DART.



www.image.ucar.edu/DAReS/DART

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Arellano, A., 2009: *The Data Assimilation Research Testbed: A community facility.* BAMS, **90**, 1283—1296, doi: 10.1175/2009BAMS2618.1