Hierarchical Bayesian Methods for Estimating Parameters of Ensemble Data Assimilation

Jeffrey Anderson
NCAR Data Assimilation Research Section (DAReS)

IMAGe TOY Workshop: 28 February, 2006

What I mean by Hierarchical Bayesian.

Two interdependent filters using the same observations.

Case 1: Adaptive error correction via inflation:

Ensemble filter for the state,

and

Continuous filter of parameter that corrects model error.

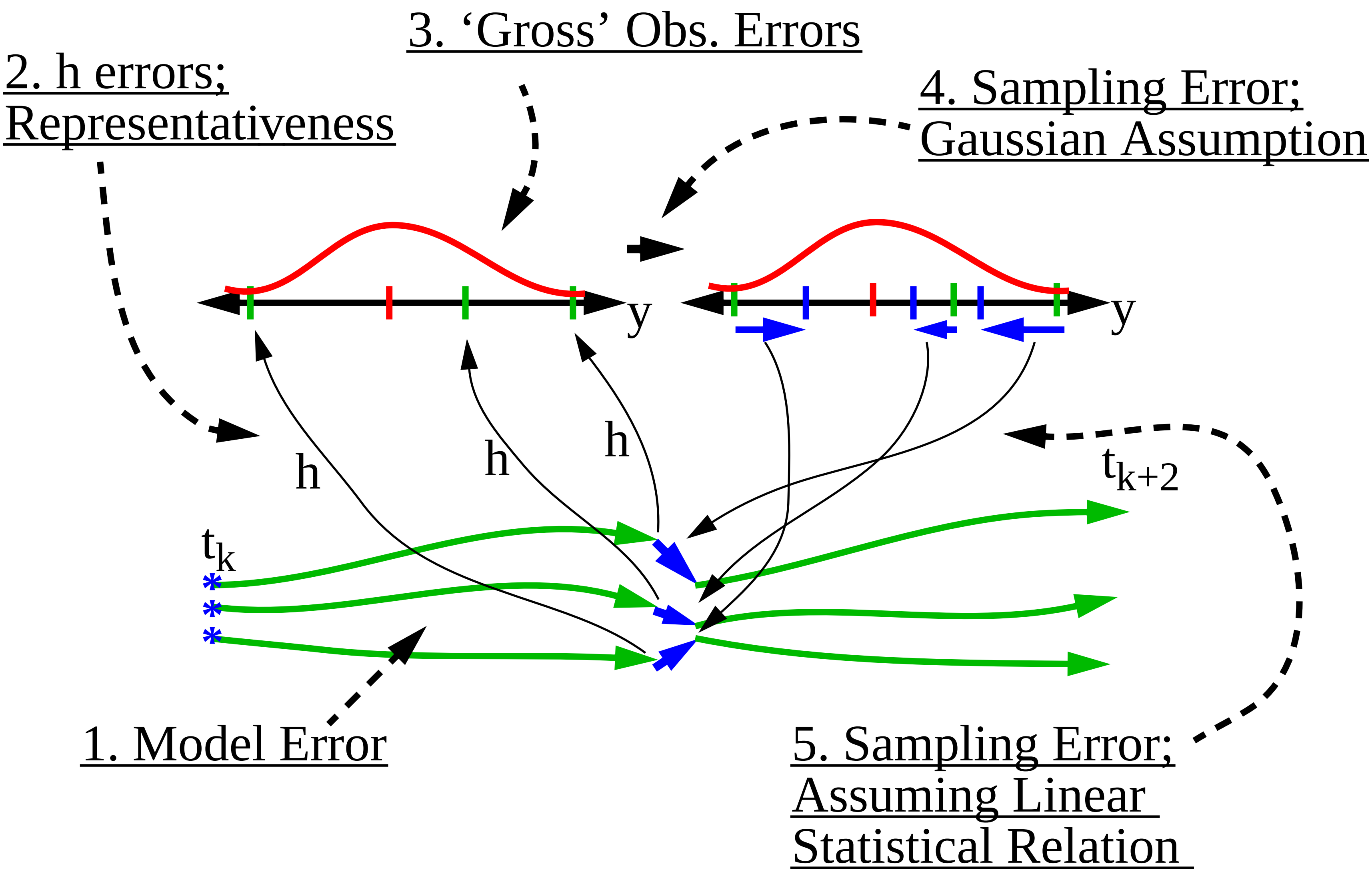
Case 2: Estimating sample regression error in ensemble filters

Ensemble filter for state,

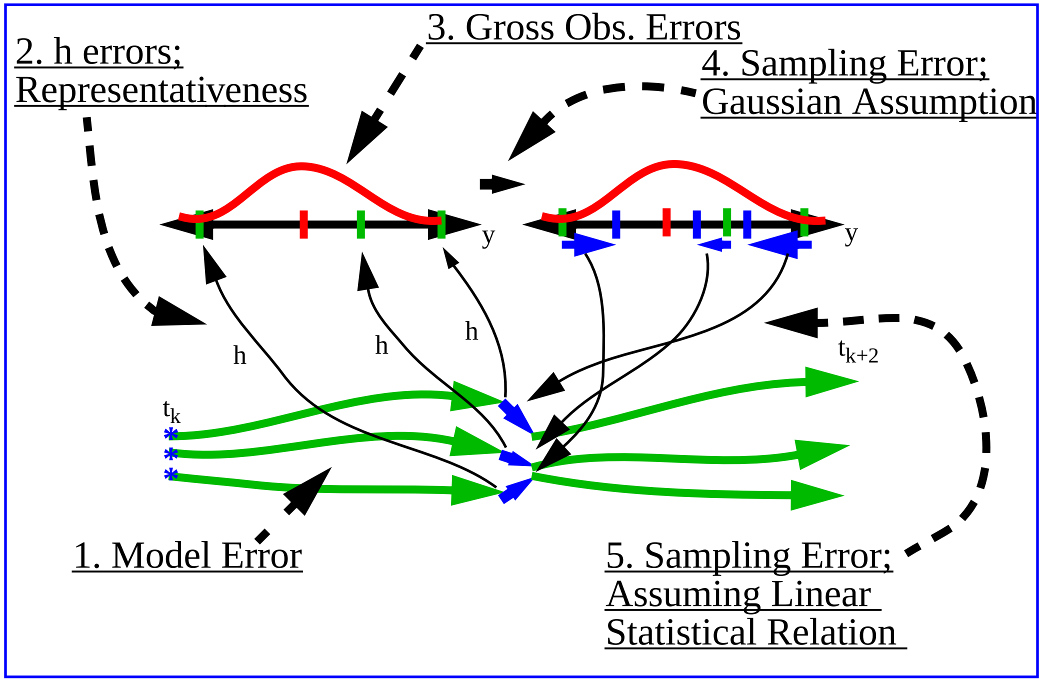
and

An ensemble of these for sampling error in regression.

Some Error Sources in Ensemble Filters



Dealing With Ensemble Filter Errors



Fix 1, 2, 3 independently HARD but ongoing.

Often, ensemble filters...

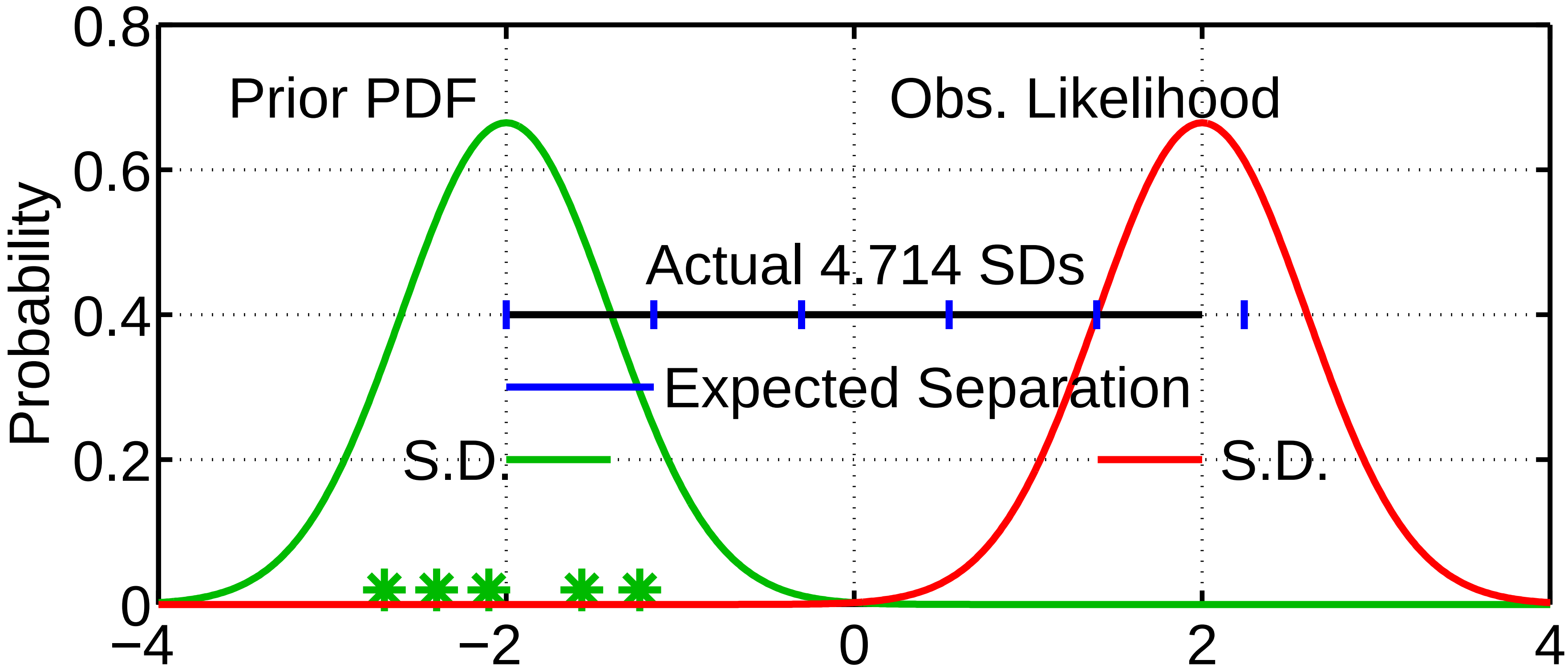
1-4: Covariance inflation, Increase prior uncertainty to give obs more impact.

5. ‘Localization’: only let obs. impact a set of ‘nearby’ state variables.

Often smoothly decrease impact to 0 as function of distance.

<u>Variance inflation for Observations: An Adaptive Error Tolerant Filter</u>

1. For observed variable, have estimate of prior-observed inconsistency



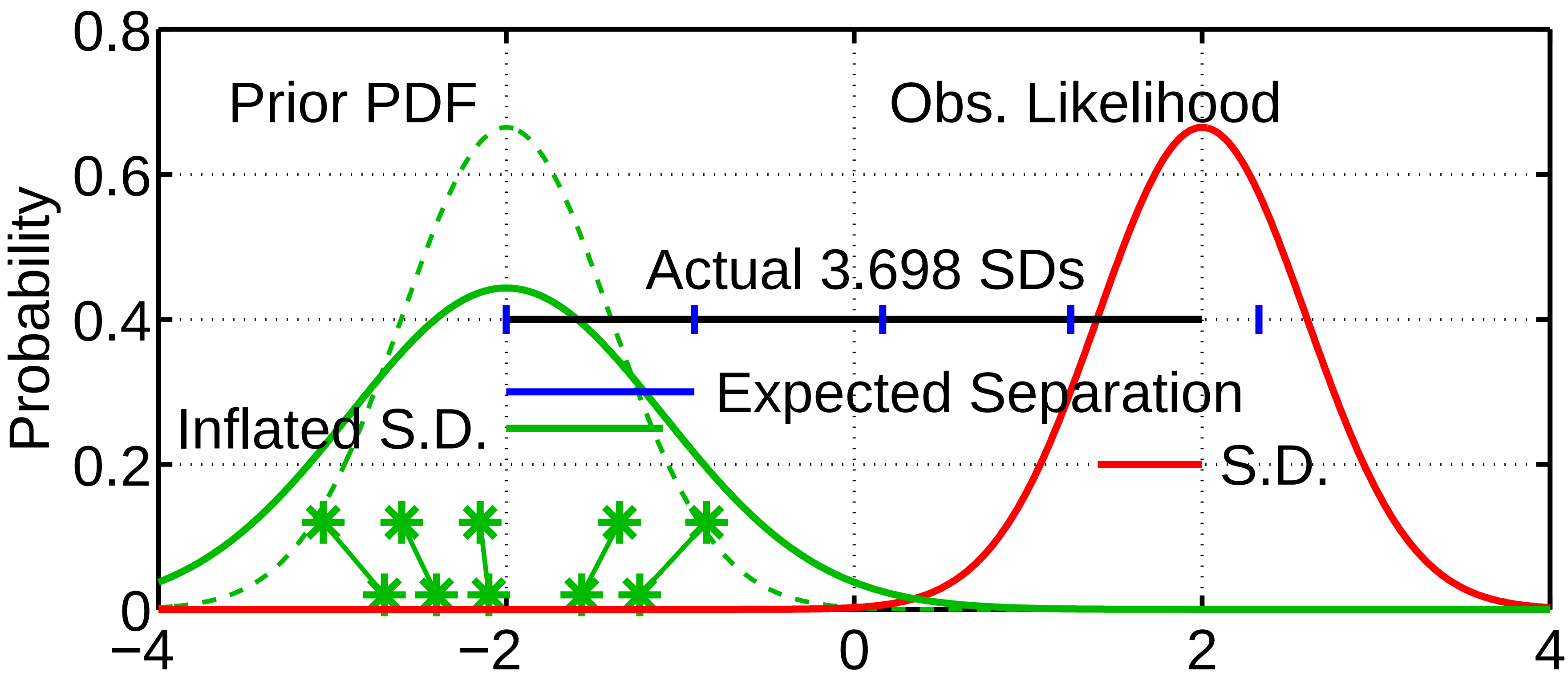
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

Assumes that prior and observation are supposed to be unbiased.

Is it model error or random chance?

Variance inflation for Observations: An Adaptive Error Tolerant Filter

1. For observed variable, have estimate of prior-observed inconsistency



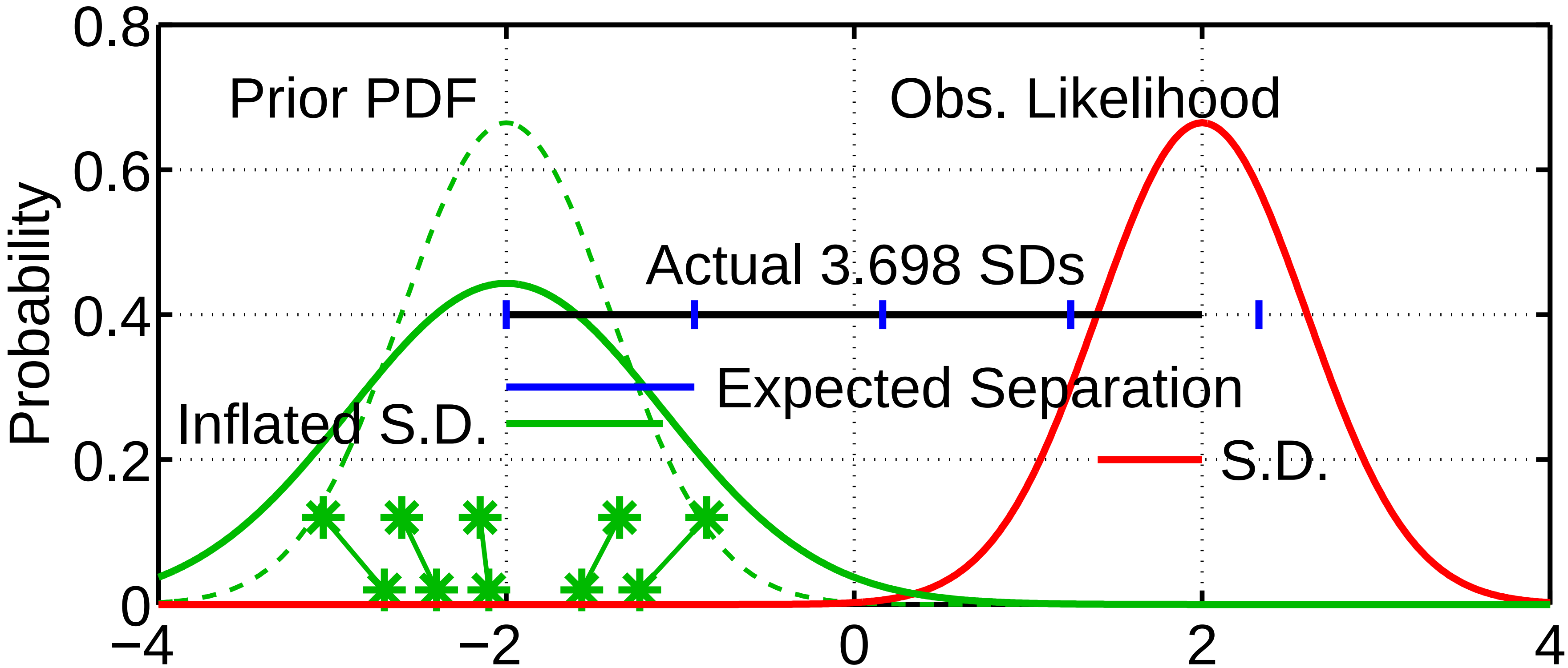
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

3. Inflating increases expected separation.

Increases 'apparent' consistency between prior and observation.

Variance inflation for Observations: An Adaptive Error Tolerant Filter

1. For observed variable, have estimate of prior-observed inconsistency

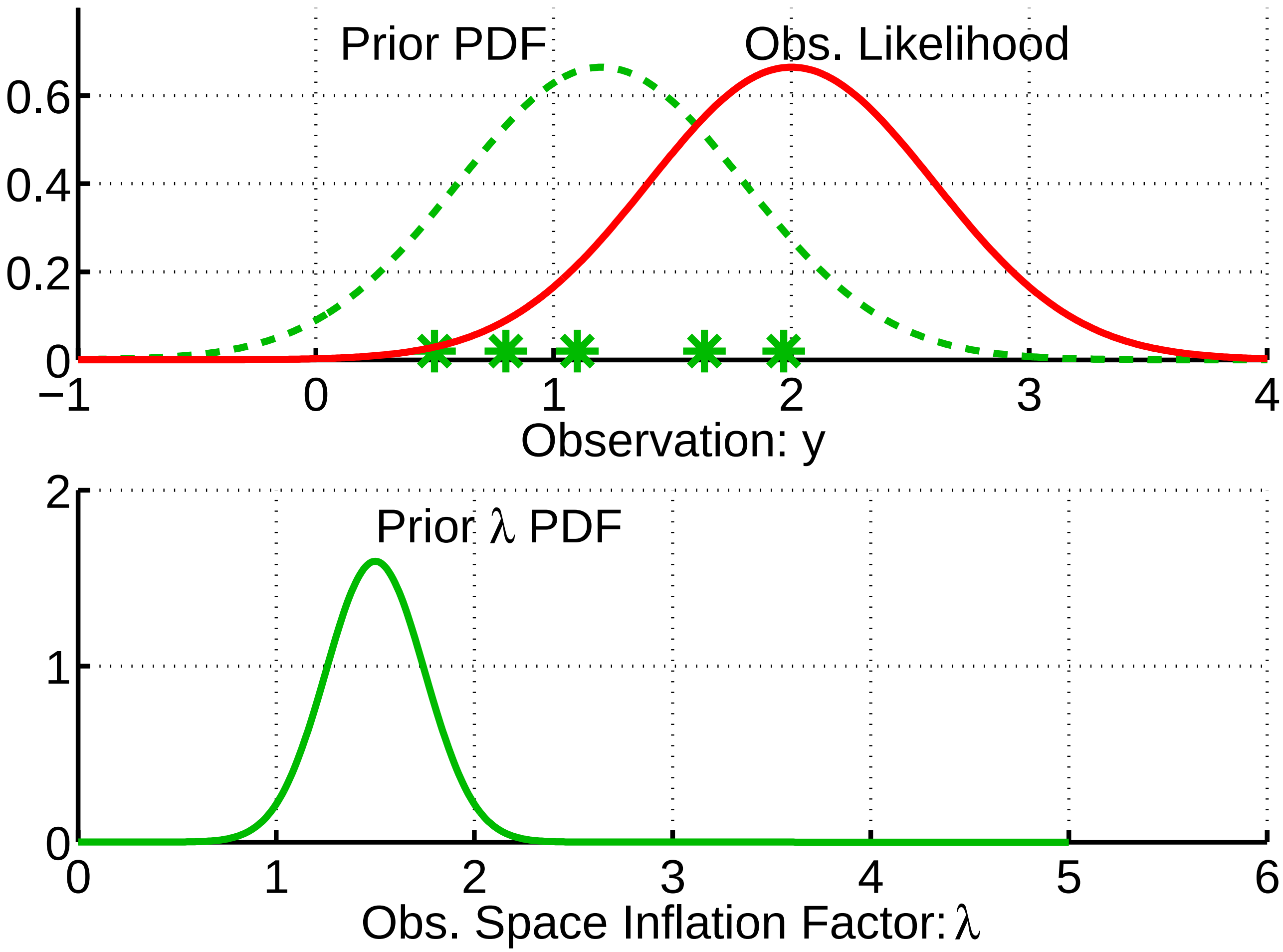


Distance, D, from prior mean y to obs. is $N\left(0, \sqrt{\lambda\sigma_{prior}^2 + \sigma_{obs}^2}\right) = N(0, \theta)$

Prob. y_o is observed given λ: $p(y_o|\lambda) = (2\Pi\theta^2)^{-1/2}\exp(-D^2/2\theta^2)$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

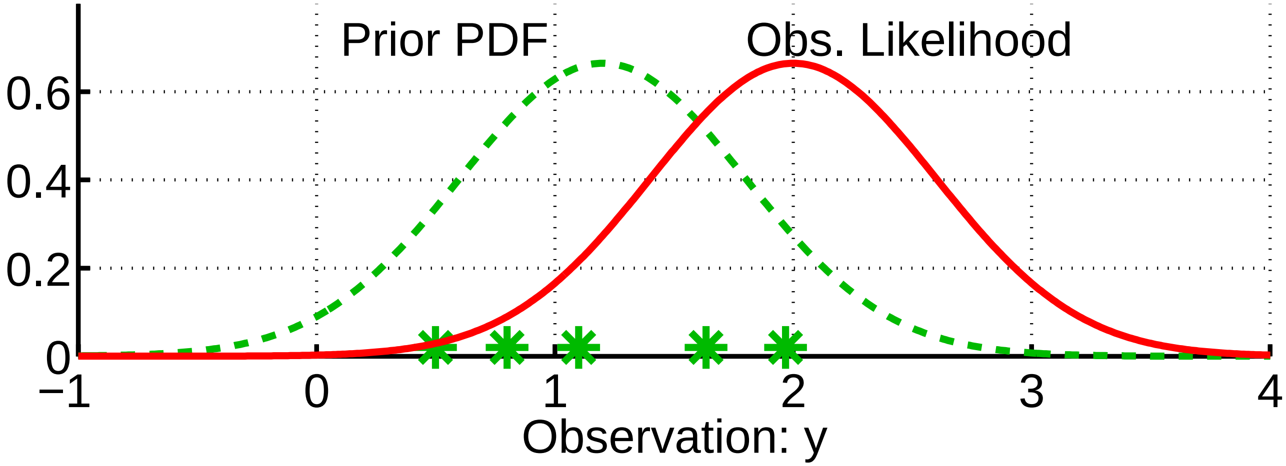
Use Bayesian statistics to get estimate of inflation factor, λ.



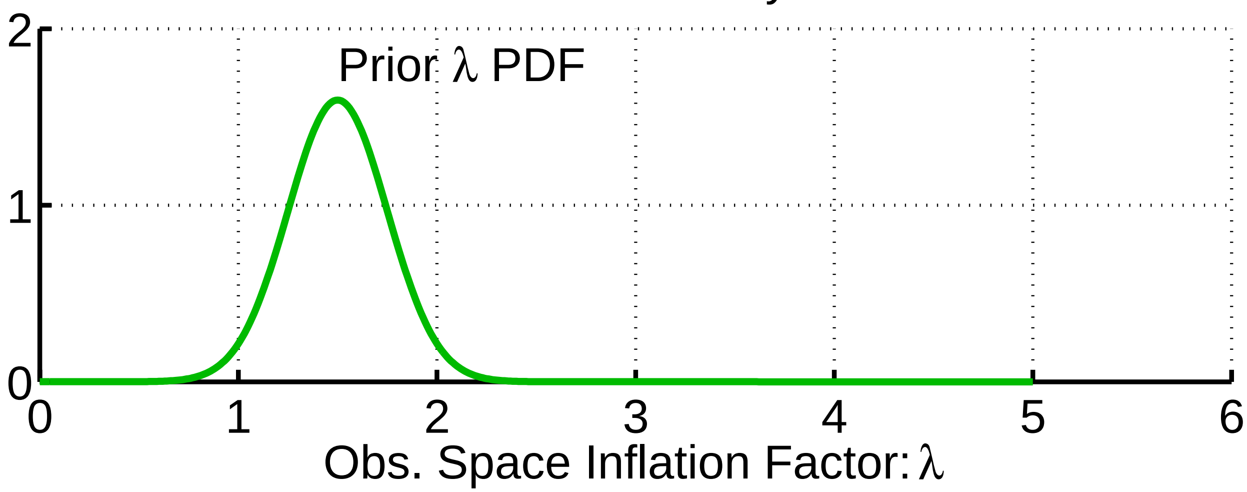
Assume prior is gaussian; $p(\lambda, t_k | Y_{t_{k-1}}) = N(\bar{\lambda}_p, \sigma^2_{\lambda, p})$.

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



We've assumed a gaussian for prior $p(\lambda, t_k | Y_{t_{k-1}})$.

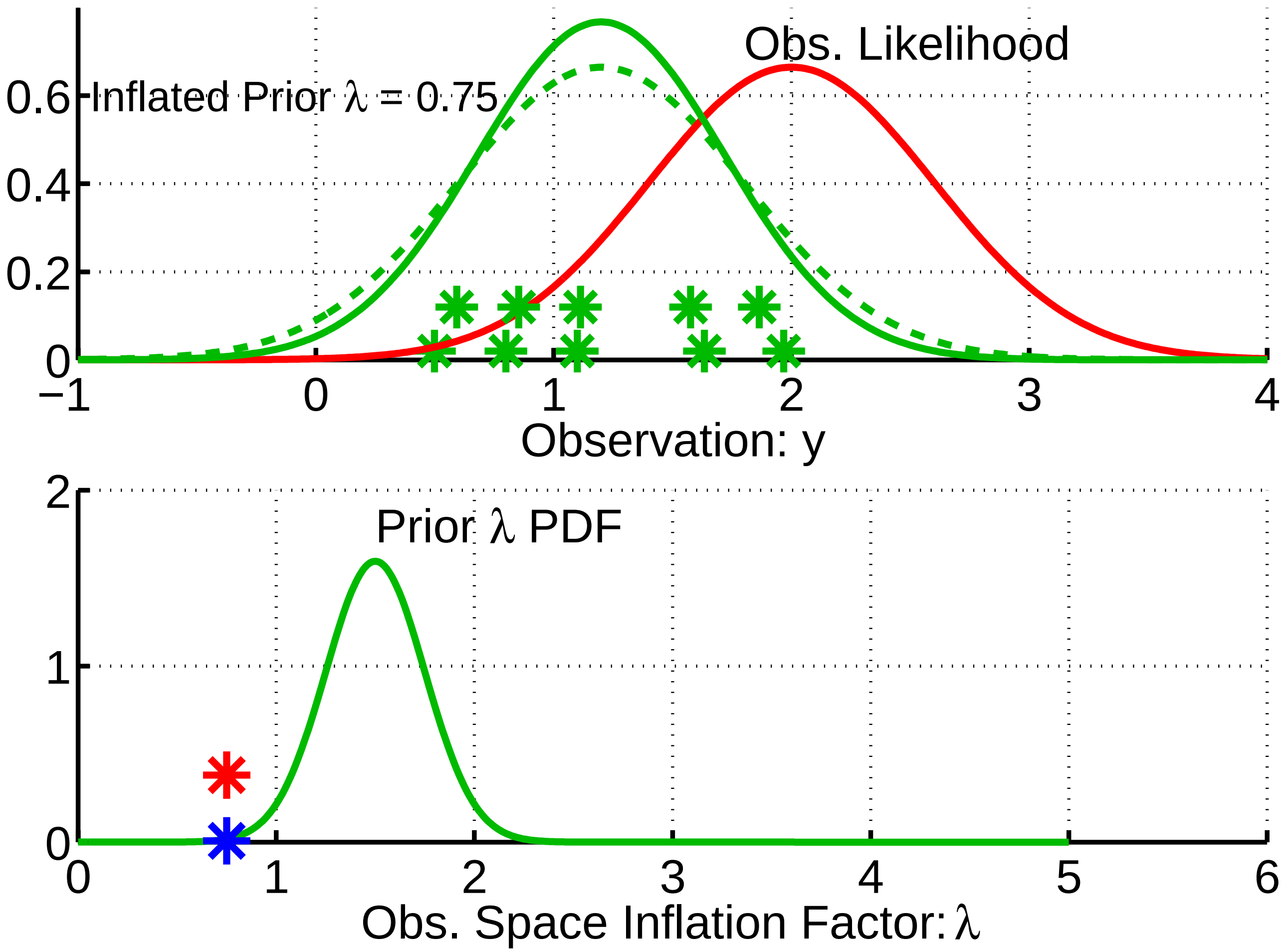


Recall that $p(y_k | \lambda)$ can be evaluated from normal PDF.

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



Get $p(y_k | \lambda = 0.75)$

from normal PDF.

Multiply by

$p(\lambda = 0.75, t_k | Y_{t_{k-1}})$

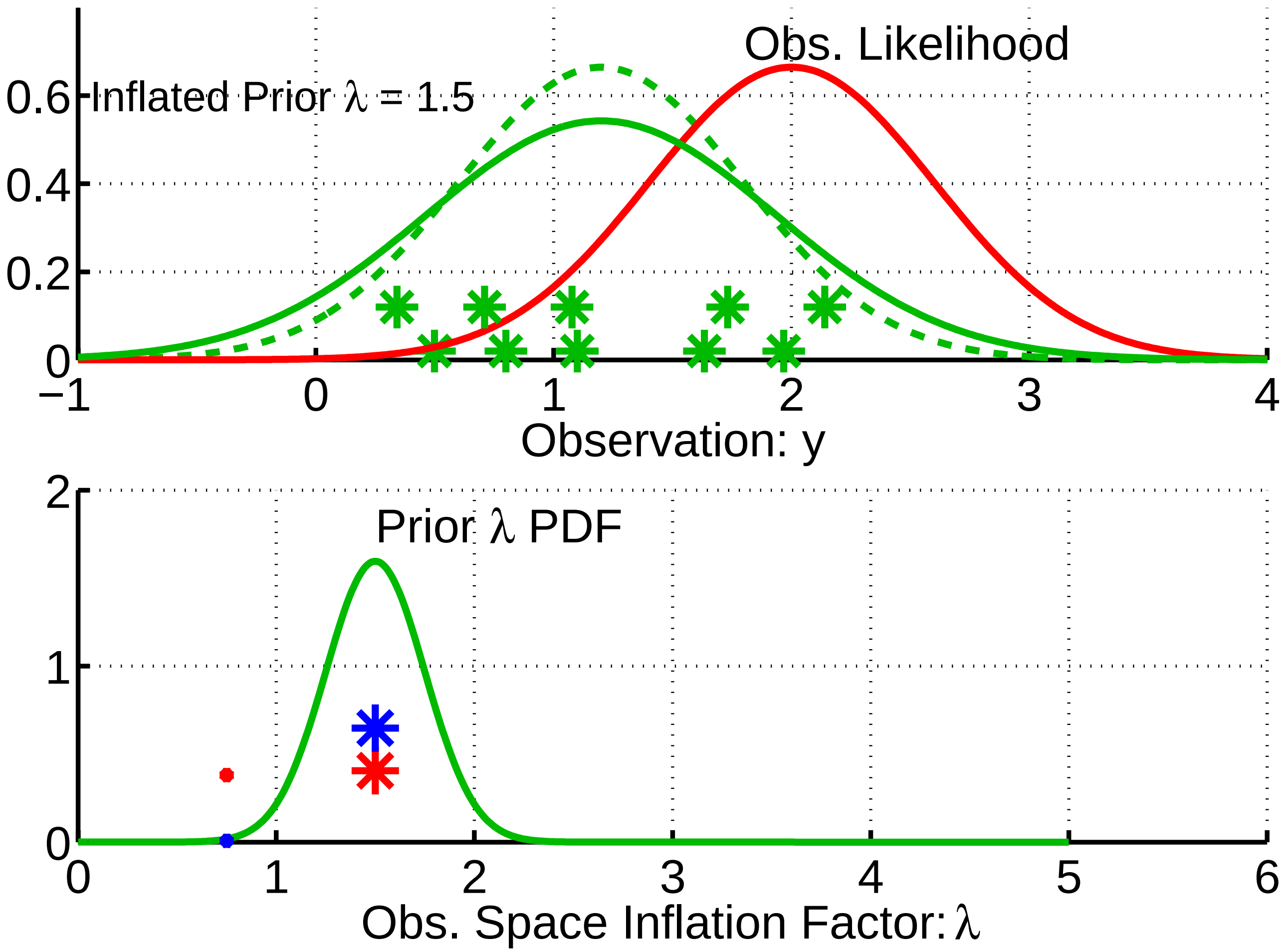
to get

$p(\lambda = 0.75, t_k | Y_{t_k})$

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



Get $p(y_k|\lambda = 1.50)$

from normal PDF.

Multiply by

$p(\lambda = 1.50, t_k|Y_{t_{k-1}})$

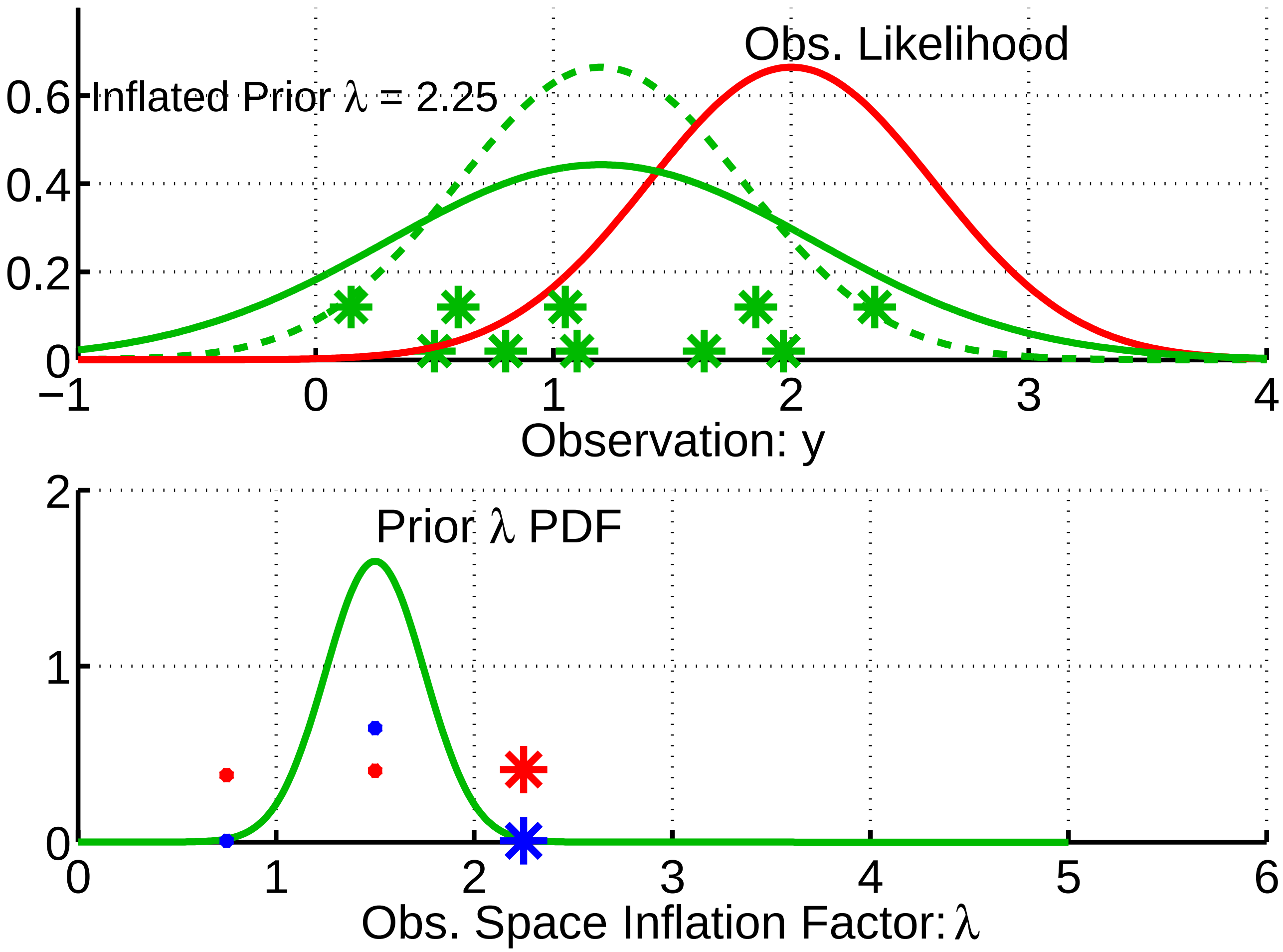
to get

$p(\lambda = 1.50, t_k|Y_{t_k})$

$$p(\lambda, t_k|Y_{t_k}) = p(y_k|\lambda)p(\lambda, t_k|Y_{t_{k-1}})/normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



Get $p(y_k|\lambda = 2.25)$

from normal PDF.

Multiply by

$p(\lambda = 2.25, t_k|Y_{t_{k-1}})$

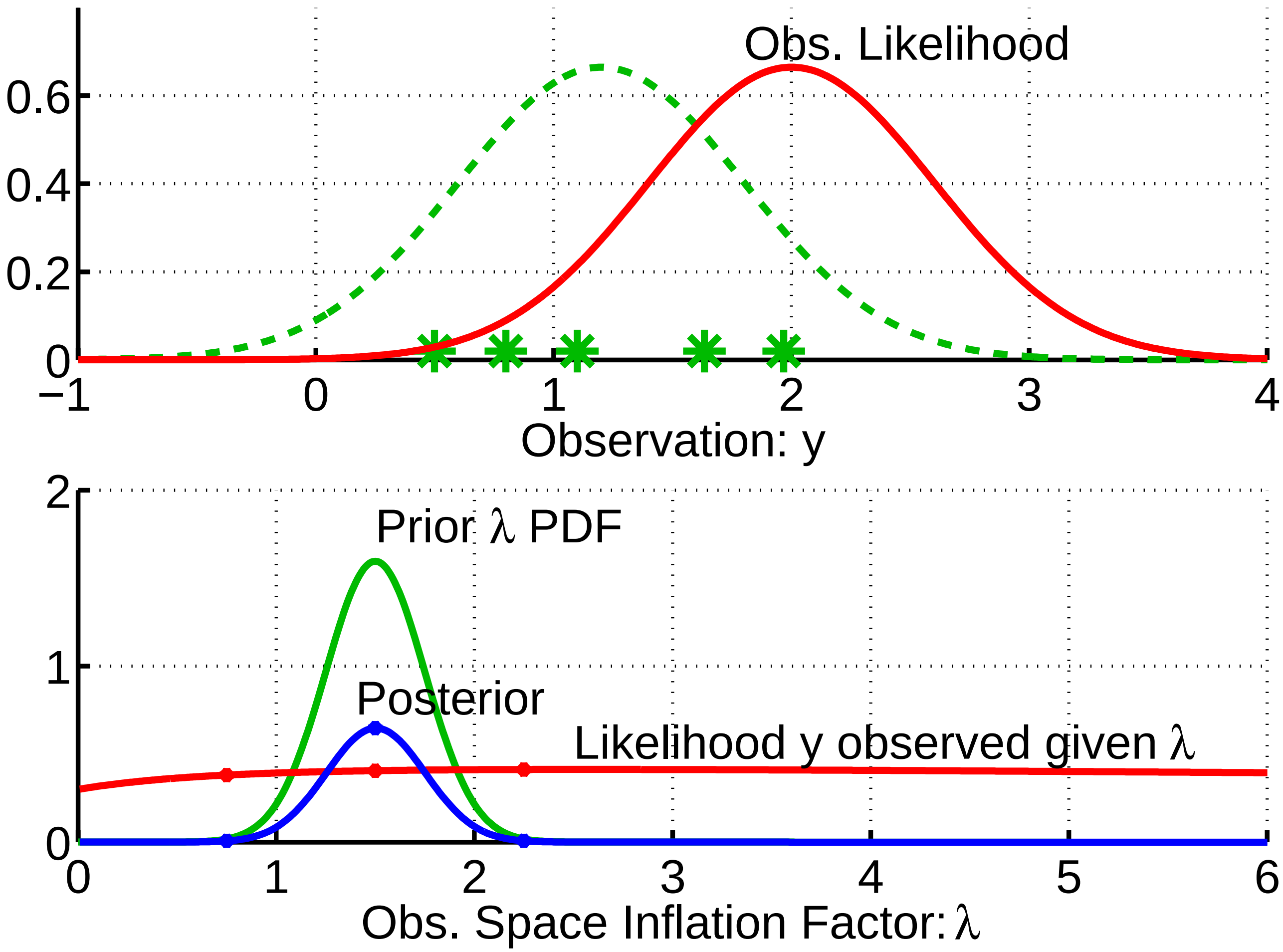
to get

$p(\lambda = 2.25, t_k|Y_{t_k})$

$$p(\lambda, t_k|Y_{t_k}) = p(y_k|\lambda)p(\lambda, t_k|Y_{t_{k-1}})/normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



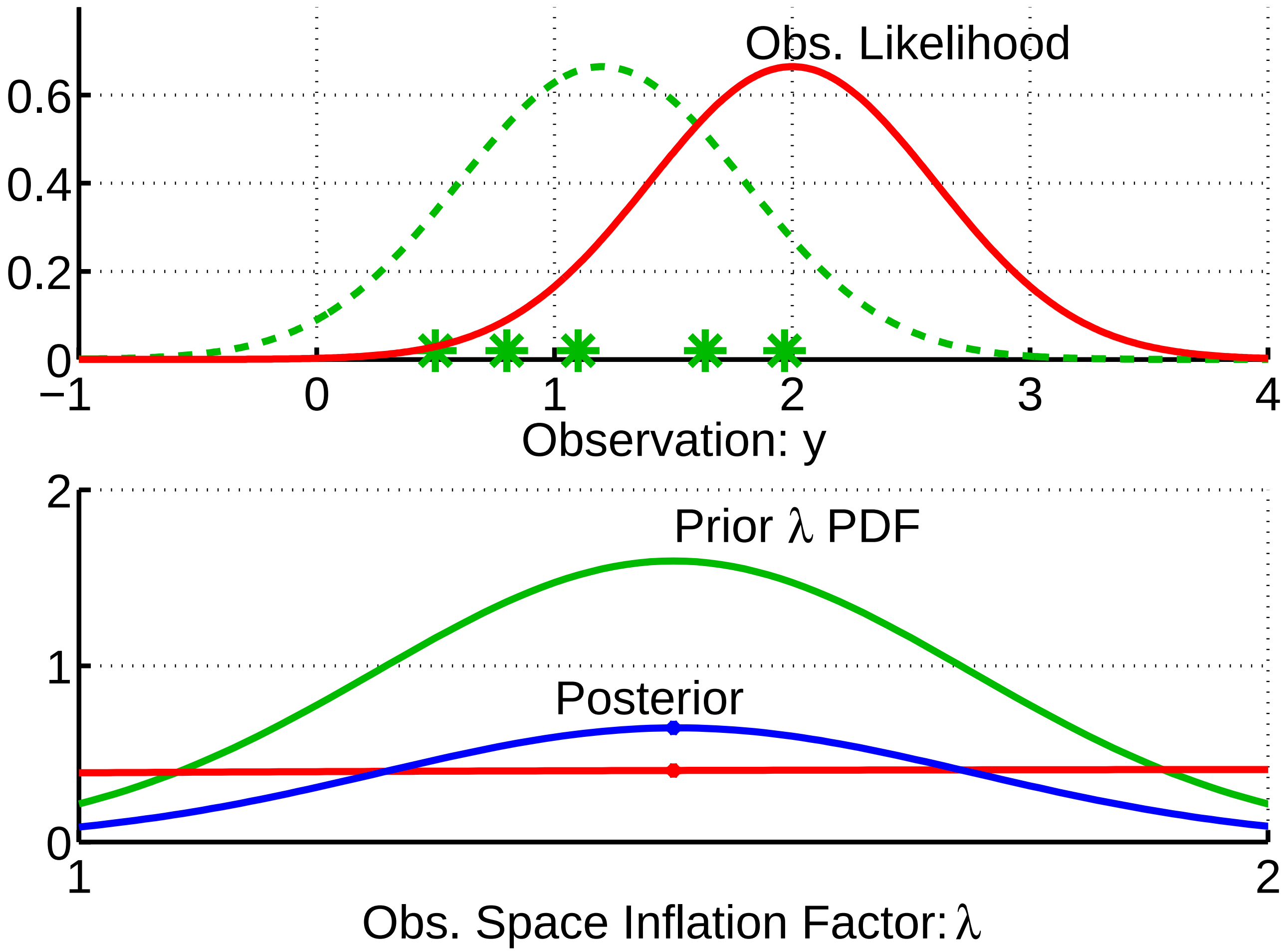
Repeat for a range of values of λ.

Now must get posterior in same form as prior (gaussian).

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



Very little information about λ in a single observation.

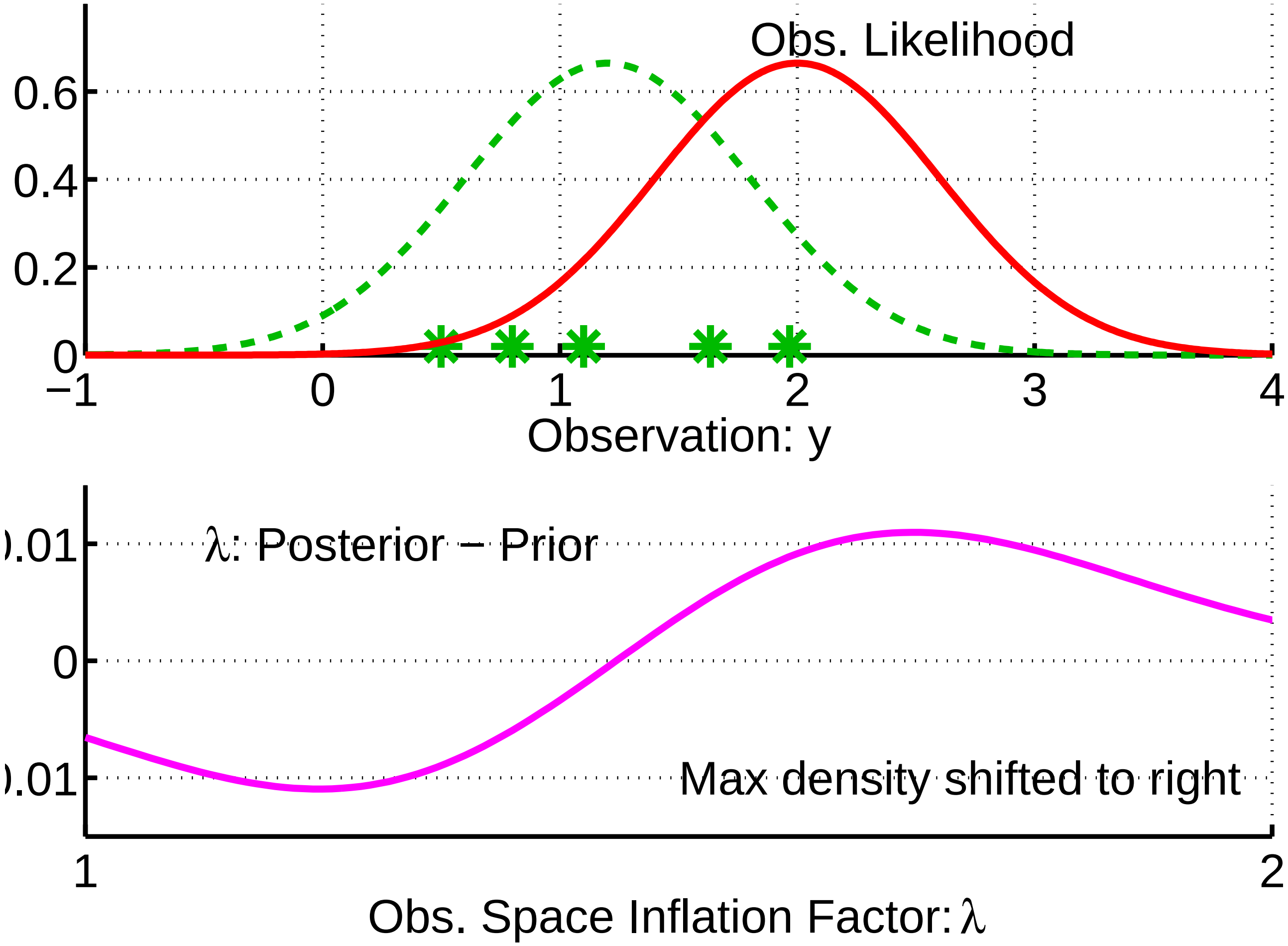
Posterior and prior are very similar.

Normalized posterior indistinguishable from prior.

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



Very little information about λ in a single observation.

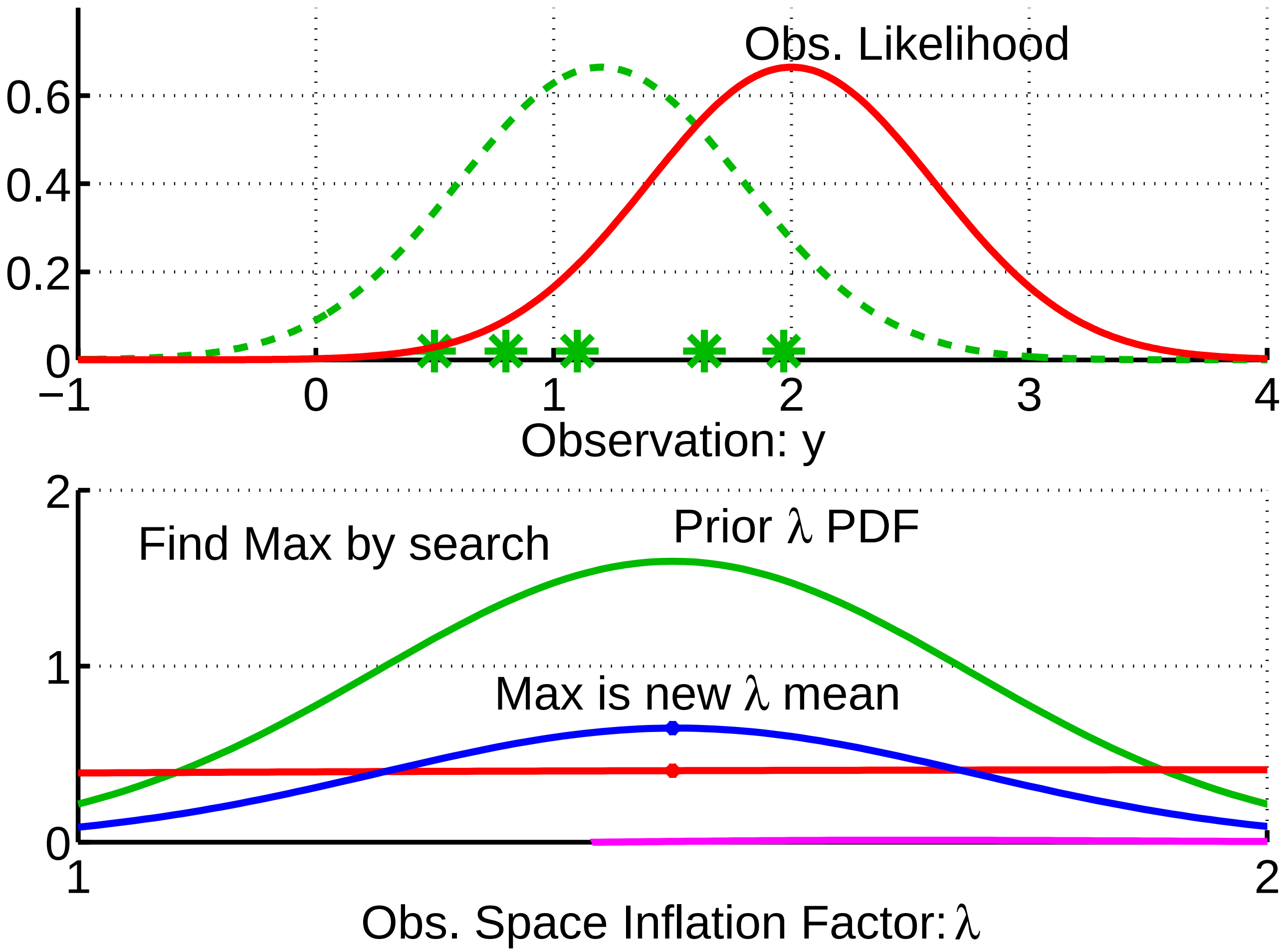
Posterior and prior are very similar.

Difference shows slight shift to larger values of λ.

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

Use Bayesian statistics to get estimate of inflation factor, λ.



One option is to use Gaussian prior for λ.

Select max (mode) of posterior as mean of updated Gaussian.

Do a fit for updated standard deviation.

$$p(\lambda, t_k | Y_{t_k}) = p(y_k | \lambda) p(\lambda, t_k | Y_{t_{k-1}}) / normalization.$$

Variance inflation for Observations: An Adaptive Error Tolerant Filter

A. Computing updated inflation mean, $\bar{\lambda}_u$.

Mode of $p(y_k|\lambda)p(\lambda, t_k|Y_{t_{k-1}})$ can be found analytically!

Solving $\partial\left[p(y_k|\lambda)p(\lambda, t_k|Y_{t_{k-1}})\right]/\partial\lambda = 0$ leads to 6th order poly in θ

This can be reduced to a cubic equation and solved to give mode.

New $\bar{\lambda}_u$ is set to the mode.

This is relatively cheap compared to computing regressions.

Variance inflation for Observations: An Adaptive Error Tolerant Filter

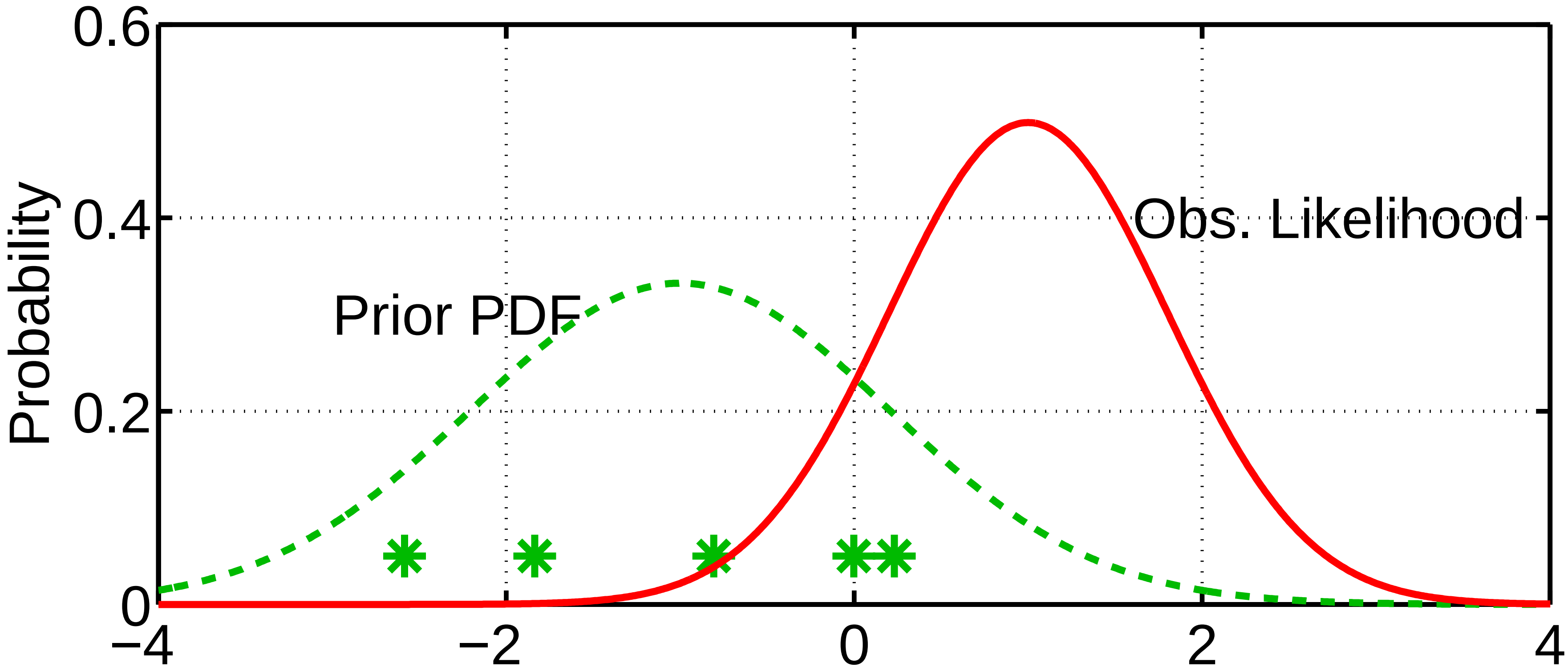
A. Computing updated inflation variance, $\sigma^2_{\lambda,\, u}$

1. Evaluate numerator at mean $\bar{\lambda}_u$ and second point, e.g. $\bar{\lambda}_u + \sigma_{\lambda,\, p}$.

2. Find $\sigma^2_{\lambda,\, u}$ so $N(\bar{\lambda}_u, \sigma^2_{\lambda,\, u})$ goes through $p(\bar{\lambda}_u)$ and $p(\bar{\lambda}_u + \sigma_{\lambda,\, p})$

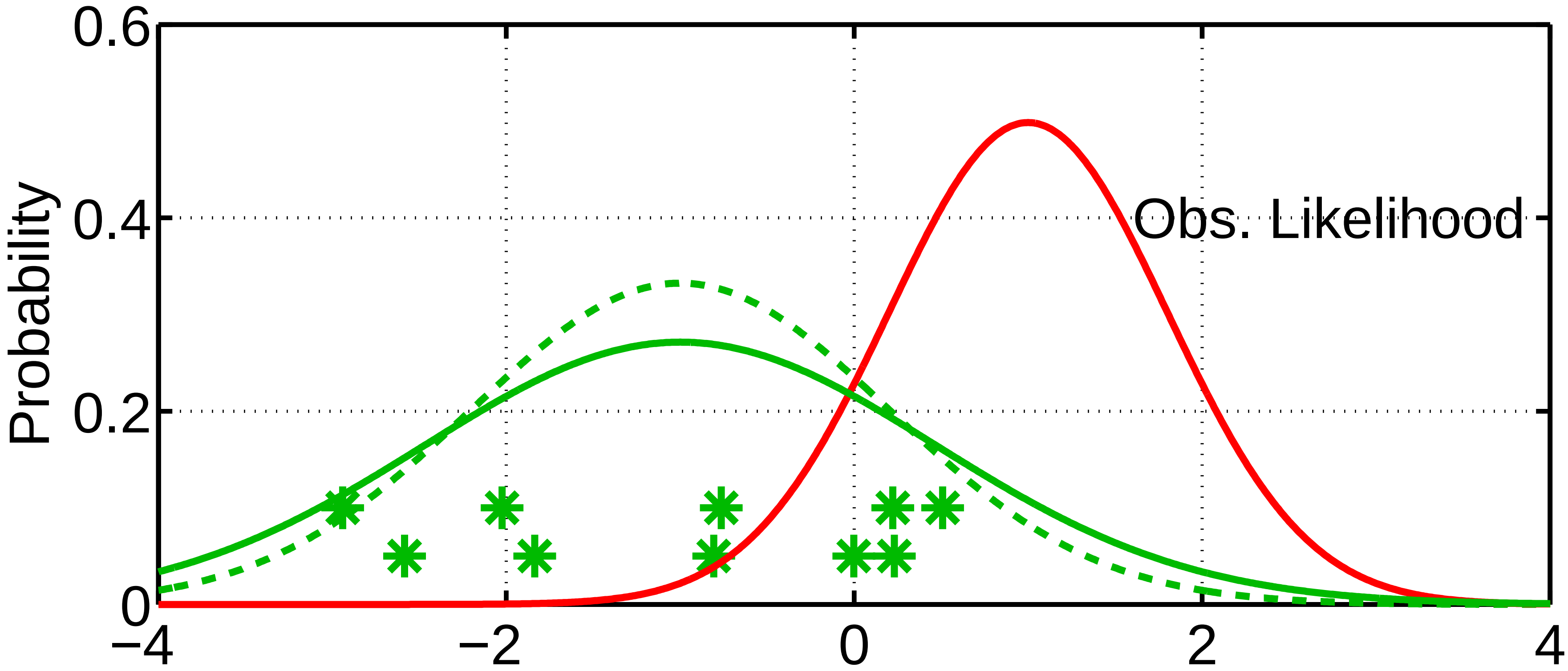
3. Compute as $\sigma^2_{\lambda,\, u} = -\sigma^2_{\lambda,\, p} / 2 \ln r$ where $r = p(\bar{\lambda}_u + \sigma_{\lambda,\, p}) / p(\bar{\lambda}_u)$

Observation Space Computations with Adaptive Error Correction



1. Compute updated inflation distribution, $p(\lambda, t_k | Y_{t_k})$.

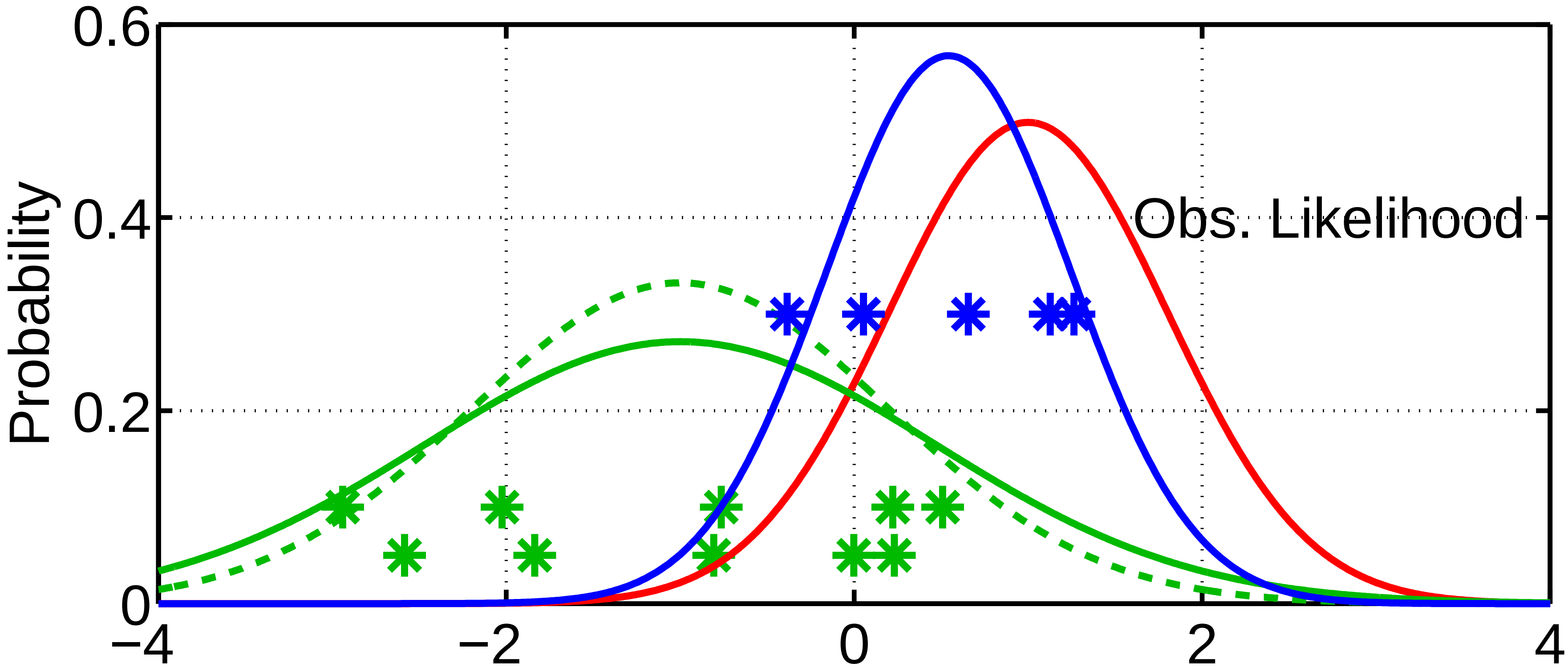
Observation Space Computations with Adaptive Error Correction



1. Compute updated inflation distribution, $p(\lambda, t_k | Y_{t_k})$.

2. Inflate ensemble using mean of updated λ distribution.

Observation Space Computations with Adaptive Error Correction

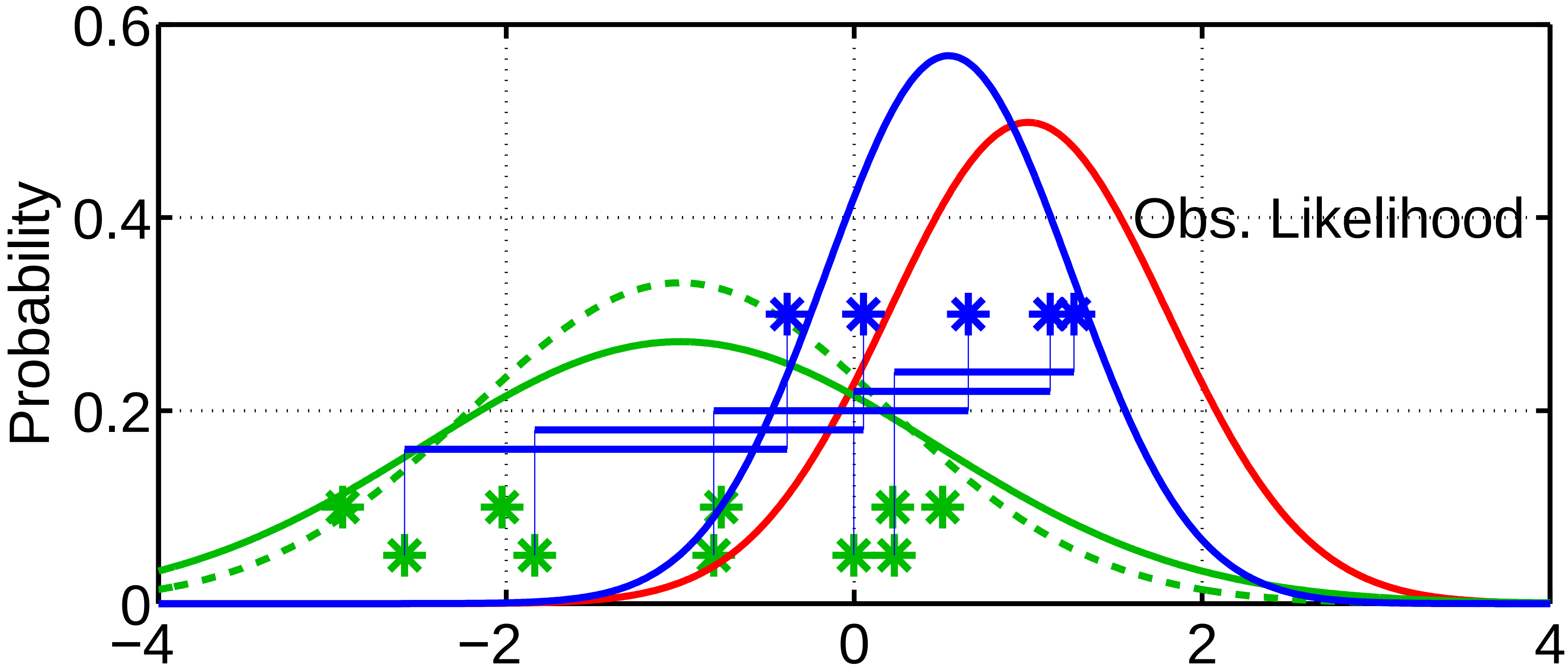


1. Compute updated inflation distribution, $p(\lambda, t_k | Y_{t_k})$.

2. Inflate ensemble using mean of updated λ distribution.

3. Compute posterior for y using inflated prior.

Observation Space Computations with Adaptive Error Correction



1. Compute updated inflation distribution, $p(\lambda, t_k | Y_{t_k})$.
2. Inflate ensemble using mean of updated λ distribution.
3. Compute posterior for y using inflated prior.
4. Compute increments from ORIGINAL prior ensemble.

Adaptive Observation Space Filter: Potential problems

1. Very heuristic.

2. Error model filter divergence (pretty hard to think about).

3. Equilibration problems, oscillations in λ with time.

4. Not clear that single distribution for all observations is right.

5. Amplifying unwanted model resonances (gravity waves)

Simulating Model Error in 40-Variable Lorenz-96 Model

Inflation can deal with all sorts of errors, including model error.

Can simulate model error in lorenz-96 by changing forcing.

Synthetic observations are from model with forcing = 8.0.

Use forcing to introduce model error.

Try forcing values of 7, 6, 5, 3 with and without adaptive inflation.

The F = 3 model is periodic, looks very little like F = 8.

Experimental design: Lorenz-96 Model Error Simulation

Truth and observations comes from long run with F=8

200 randomly located (fixed in time) ‘observing locations’

Independent 1.0 observation error variance

Observations every hour

σ_λ is 0.05, mean of λ adjusts but variance is fixed

4 groups of 20 members each (80 ensemble members total)

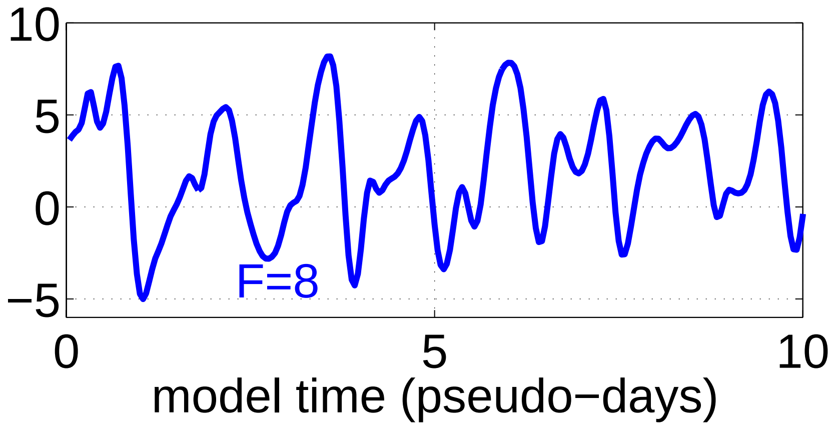
Results from 10 days after 40 day spin-up

Vary assimilating model forcing: F=8, 6, 3, 0

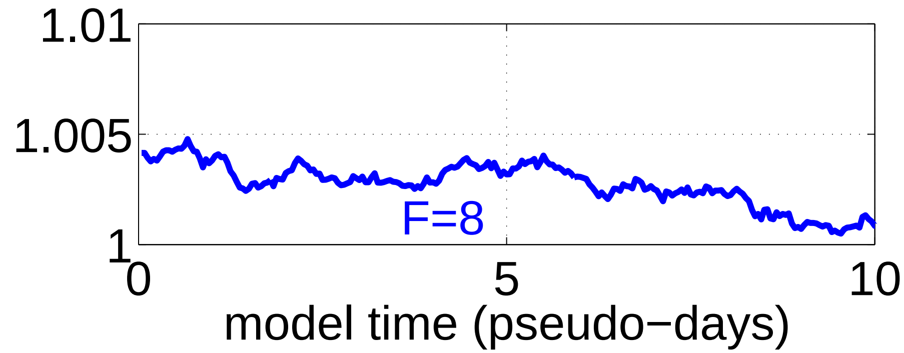
Simulates increasing model error

Assimilating F=8 Truth with F=8 Ensemble

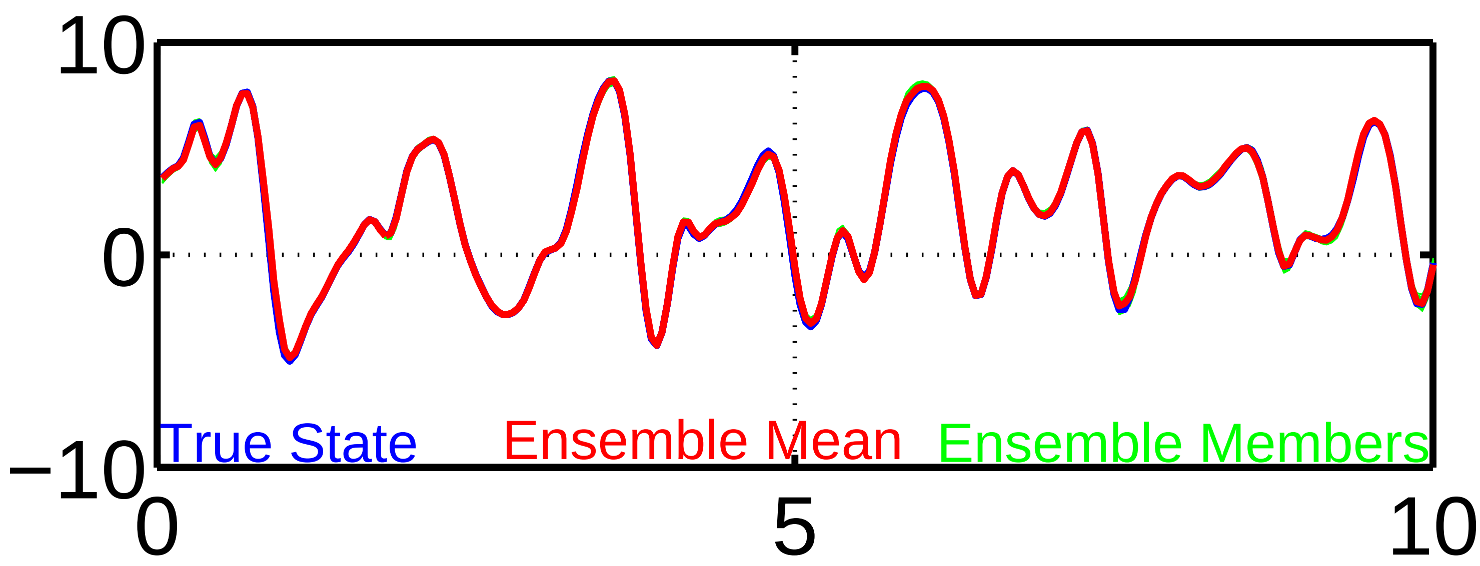
Model time series



Mean value of λ

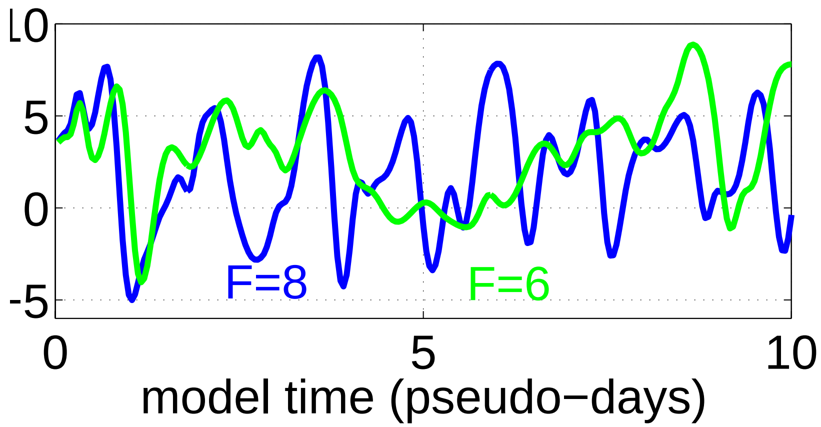


Assimilation Results

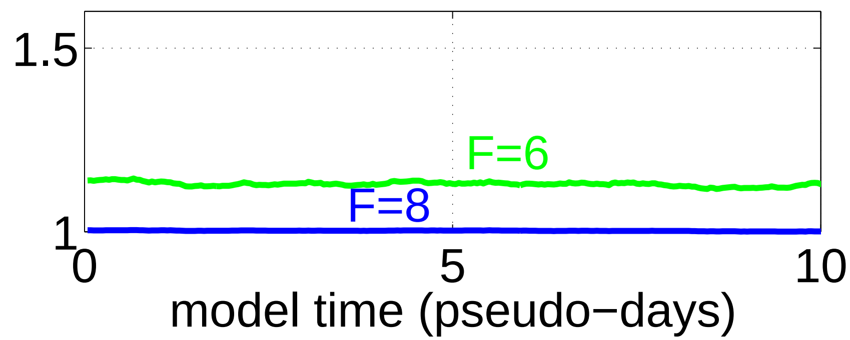


Assimilating F=8 Truth with F=6 Ensemble

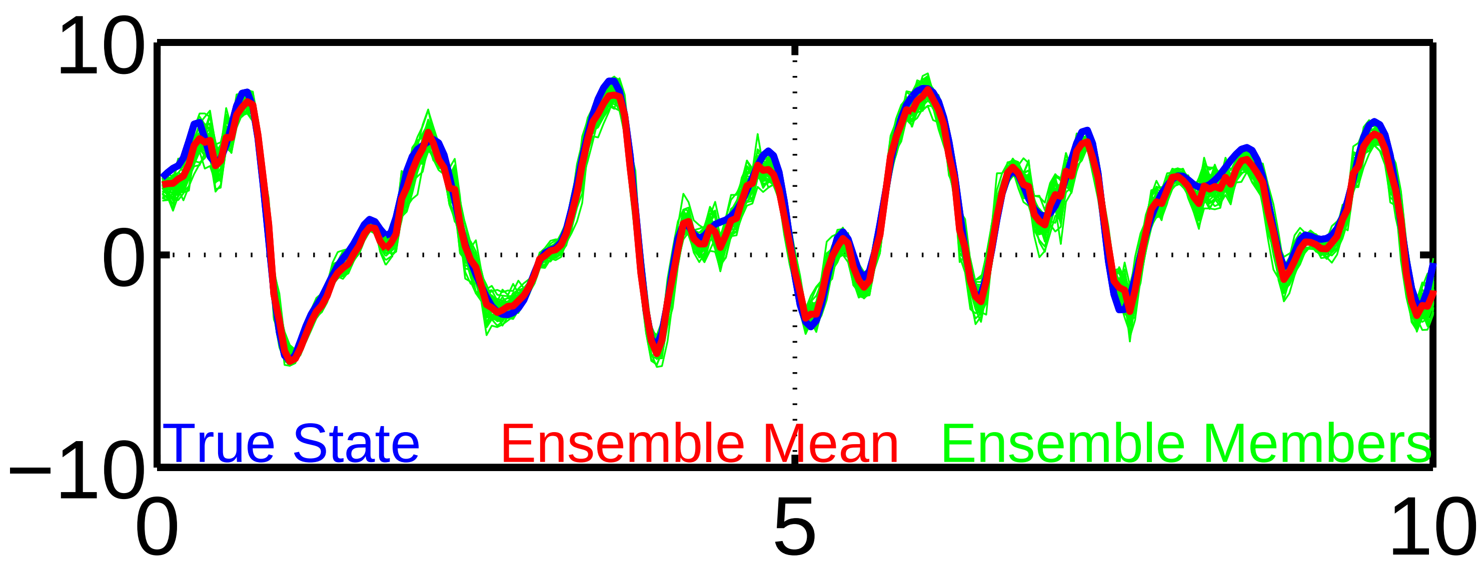
Model time series



Mean value of λ

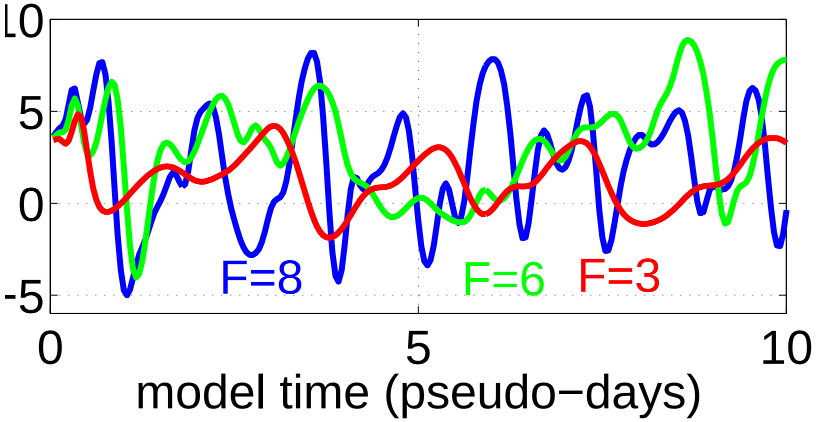


Assimilation Results

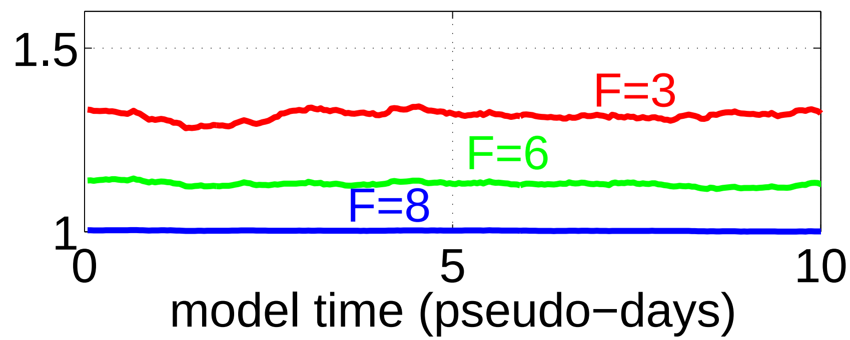


Assimilating F=8 Truth with F=3 Ensemble

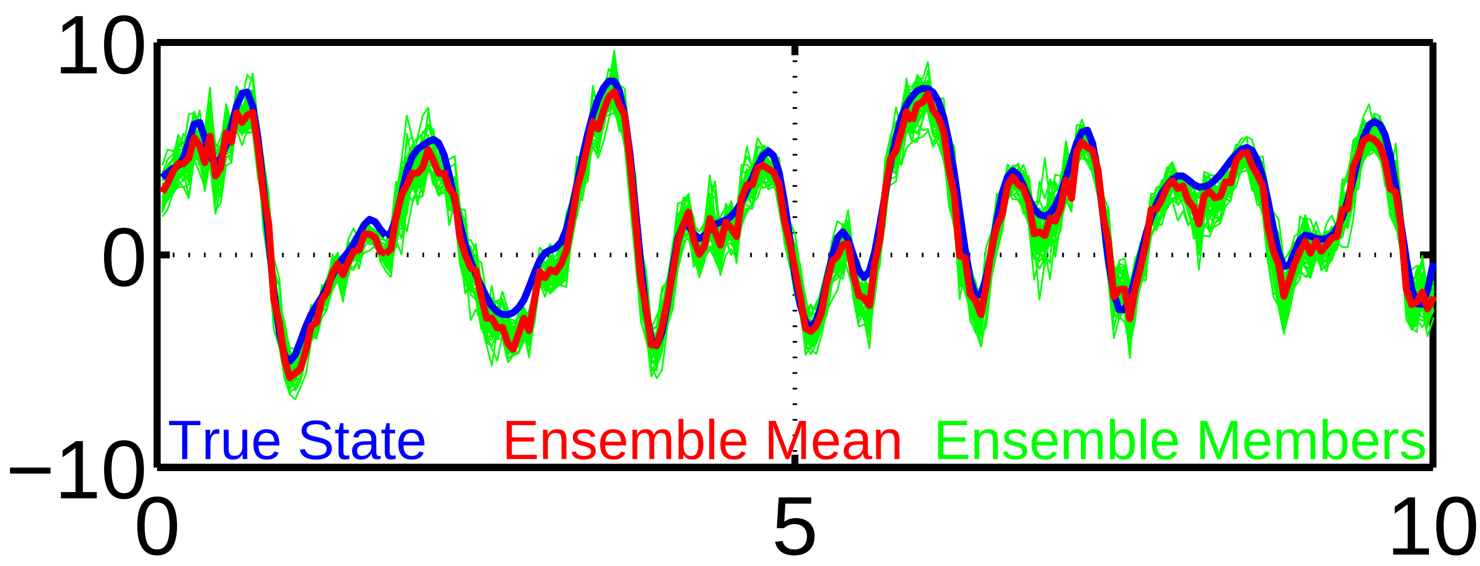
Model time series



Mean value of λ

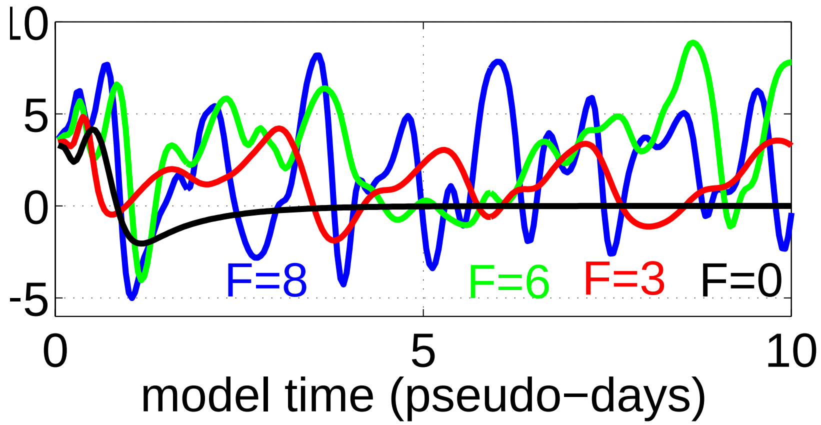


Assimilation Results

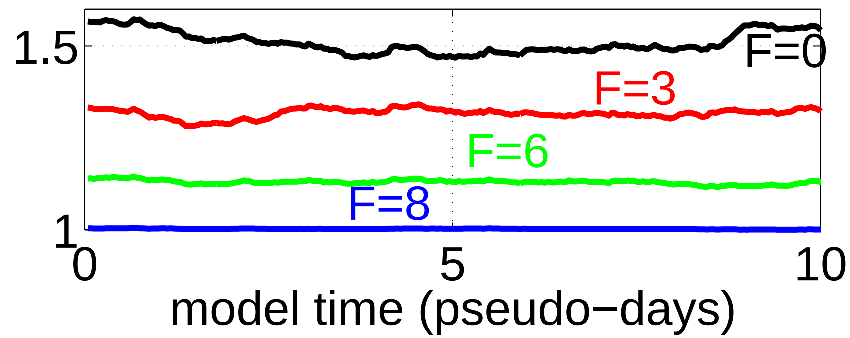


Assimilating F=8 Truth with F=0 Ensemble

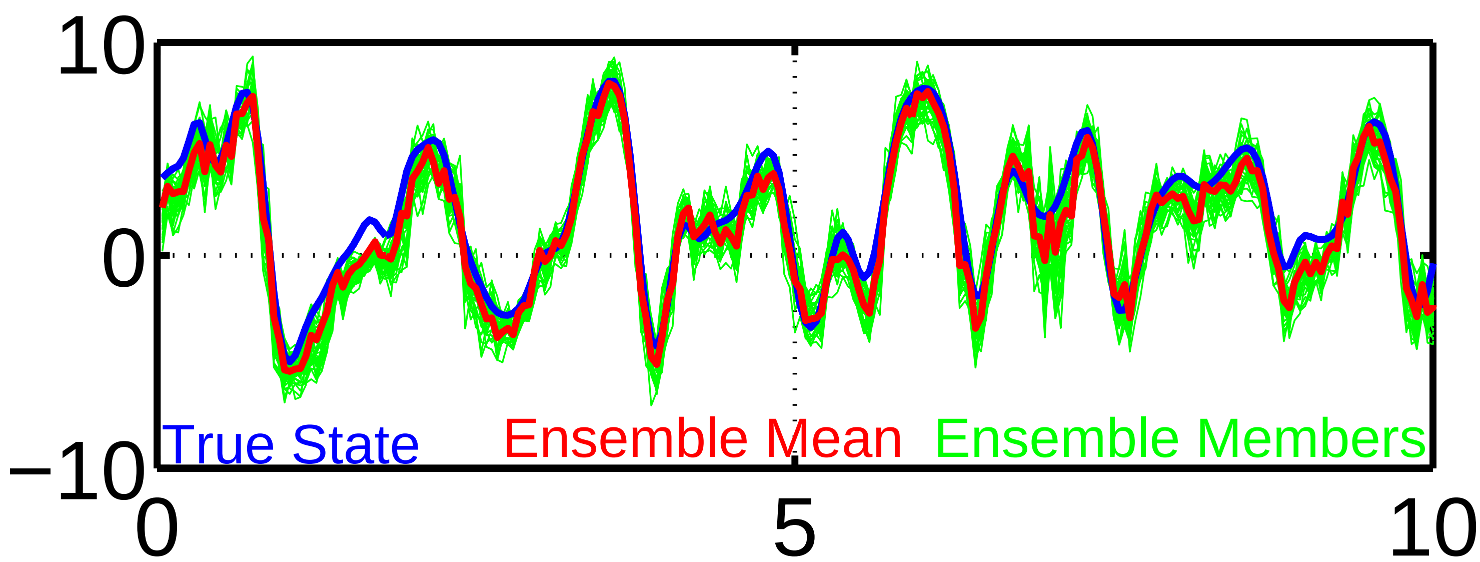
Model time series



Mean value of λ

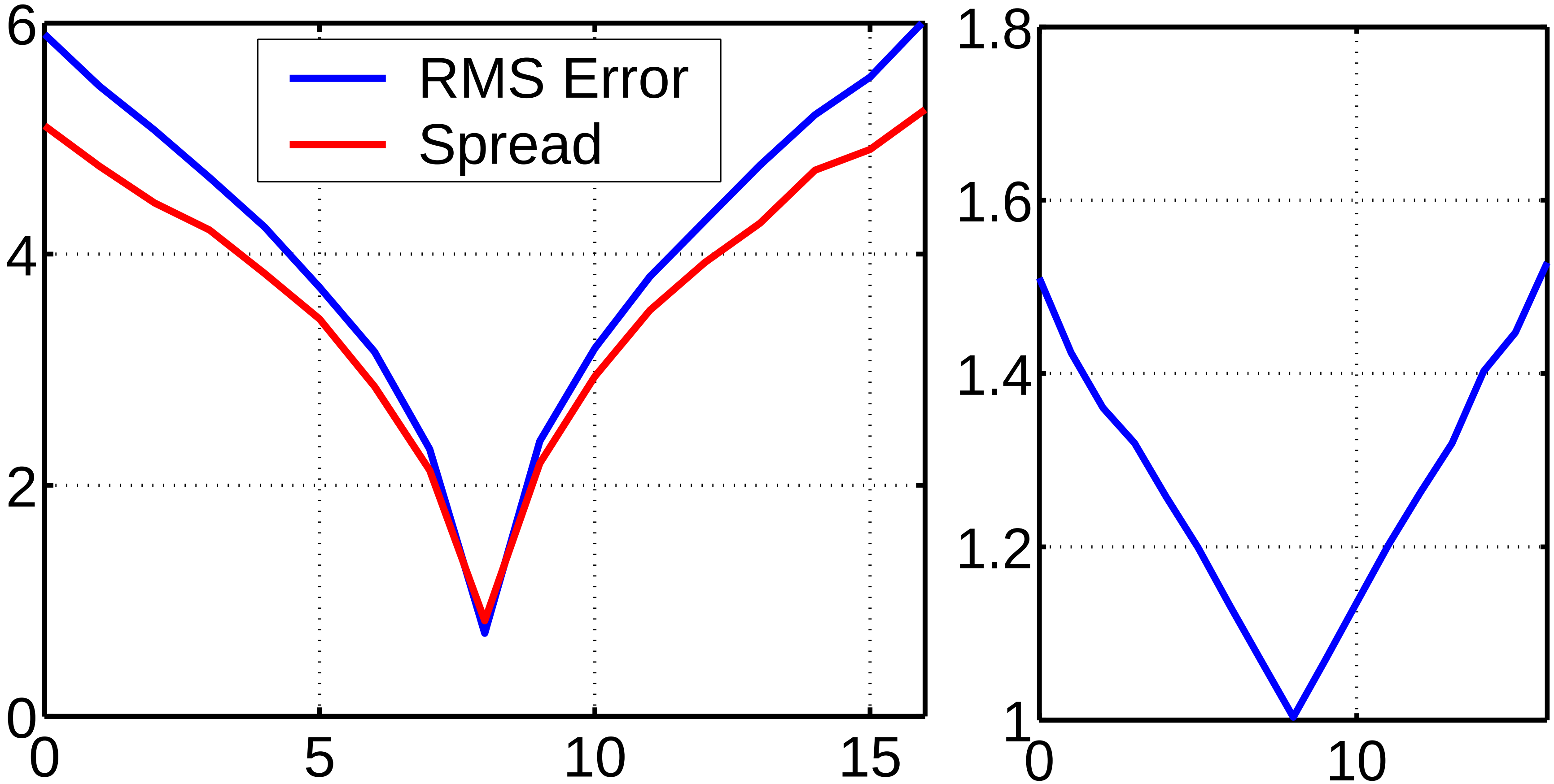


Assimilation Results



Prior RMS Error, Spread, and λ Grow as Model Error Grows

Base case: 200 randomly located observations per time



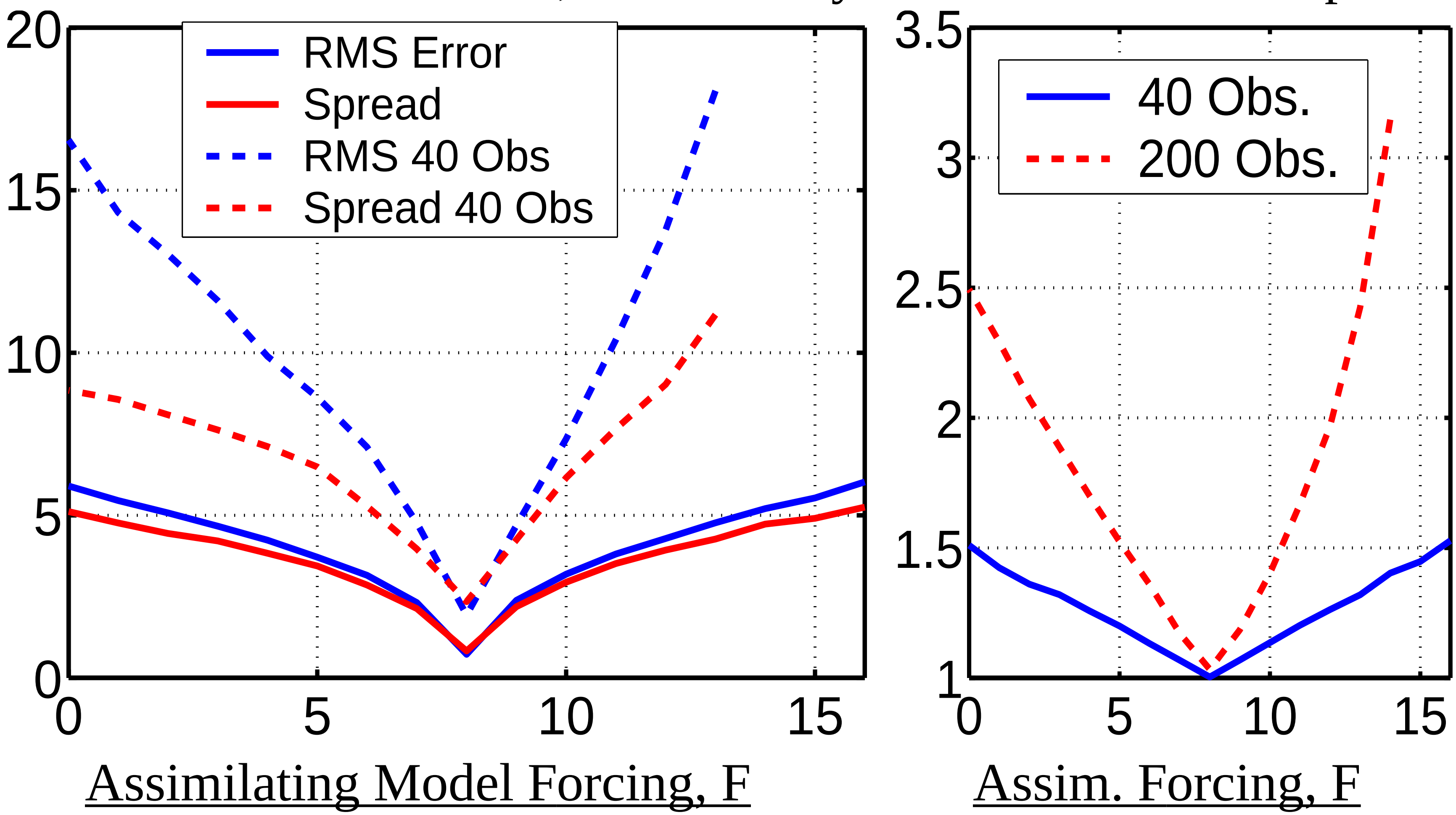
Assimilating Model Forcing, F

Assim. Forcing, F

(Error saturation is approximately 30.0)

Prior RMS Error, Spread, and λ Grow as Model Error Grows

Less well observed case, 40 randomly located observations per time



Adaptive State Space Inflation Algorithm

Suppose we want a global state space inflation, λ_s, instead.

Make same least squares assumption that is used in ensemble filter.

Inflation of λ_s for state variables inflates obs. priors by same amount.

Get same likelihood as before: $p(y_o|\lambda) = (2\Pi\theta^2)^{-1/2}\exp(-D^2/2\theta^2)$

$$\theta = \sqrt{\lambda_s\sigma_{prior}^2 + \sigma_{obs}^2}$$

Compute updated distribution for λ_s exactly as for observation space.

Assumes that inflating all state variables leads to corresponding inflation of all observation variables.

Implementation of Adaptive State Space Inflation Algorithm

1. Apply inflation to state variables with mean of λ_s distribution.

2. Do following for observations at given time sequentially:
   a. Compute forward operator to get prior ensemble.
   b. Compute updated estimate for λ_s mean and variance.
   c. Compute increments for prior ensemble.
   d. Regress increments onto state variables.

All the algorithmic variants could still be applied.
What are relative characteristics of these algorithms?

Spatially varying adaptive inflation algorithm:

Have a distribution for λ at for each state variable, $\lambda_{s,i}$

Use prior correlation from ensemble to determine impact of $\lambda_{s,i}$ on prior variance for given observation.

If γ is correlation between state variable i and observation then assume

$$\theta = \sqrt{[1 + \gamma(\sqrt{\lambda_{s,i}} - 1)]^2 \sigma_{prior}^2 + \sigma_{obs}^2}$$

Equation for finding mode of posterior is now full 12th order:
Can do Taylor expansion of θ around $\lambda_{s,i}$.
Retaining linear term is normally quite accurate.
There is an analytic solution to find mode of product in this case!

Ensemble and inflation filters now tightly interwoven!

Results from CAM 'Operational' Assimilation

Run like an operational Numerical Weather Prediction Model
Assimilate reanalyis obserations every 6 hours.

1. Radiosonde u,v,t,q
2. ACARS (airplane) u, v, t
3. Satellite drift u, v

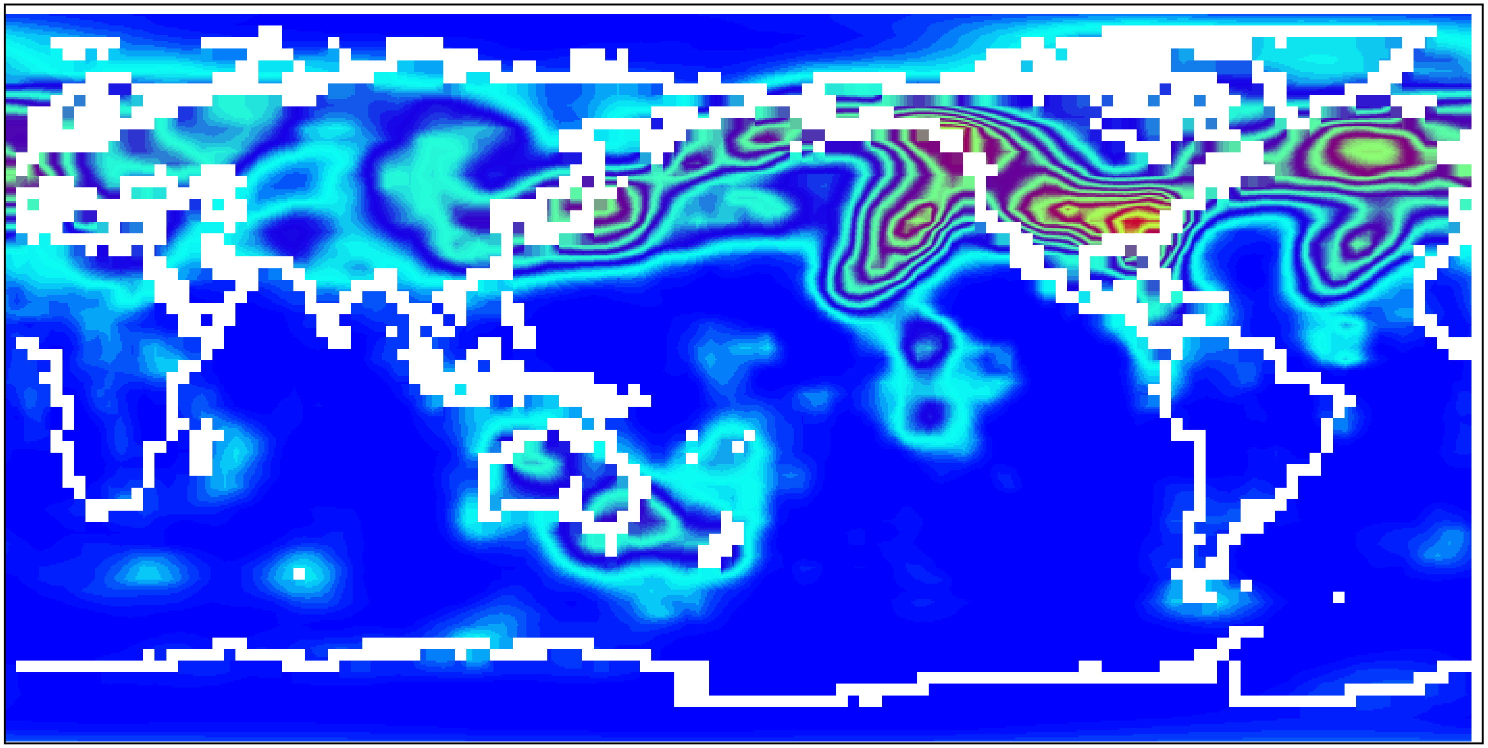
Sampling error leads to variance loss where observations are dense.

T42 CAM global model

Significant model error.

Results from CAM 'Operational' Assimilation

Mean inflation (range 1 to 3) for 500mb Temperature



Overall assimilation quality is improved.
Filter divergence is avoided.

Combined model and observational error variance adaptive algorithm

Is this really possible. Yes, in certain situations...
Is there enough information available?

Spatially-vary inflation for state

Inflation factor for different sets of observations (all radiosonde T's)

$$\theta = \sqrt{[1 + \gamma(\sqrt{\lambda_{s,i}} - 1)]^2 \sigma_{prior}^2 + \lambda_o \sigma_{obs}^2}$$

Different λ's see different observations

Initial tests in L96 with model error AND incorrect obs. error variance can correct for both!!!

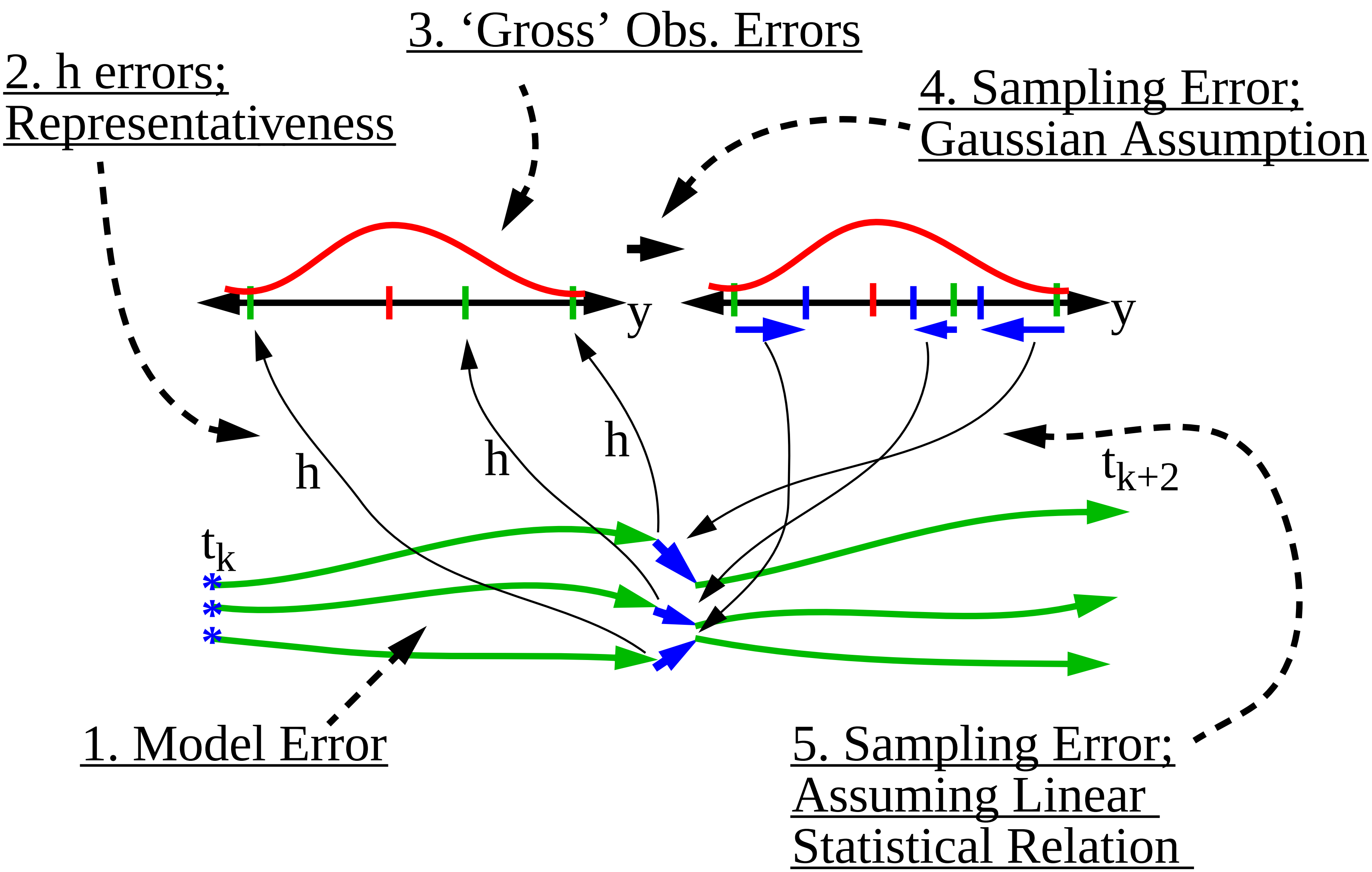
Case 2: Estimating sample regression error in ensemble filters

Ensemble filter for state,

and

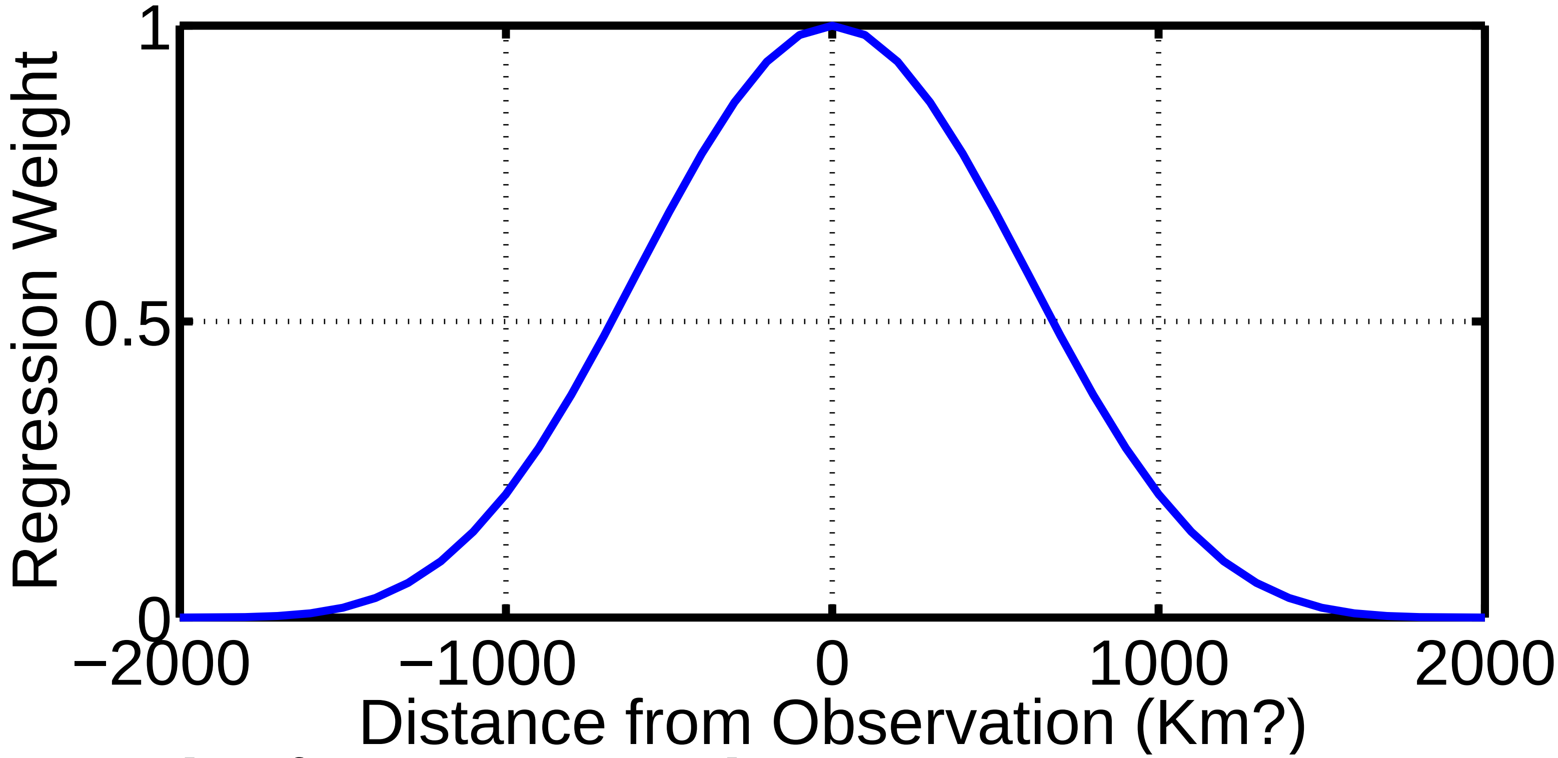
An ensemble of these for sampling error in regression.

Some Error Sources in Ensemble Filters



Ways to deal with regression sampling error:

3. Use additional a priori information about relation between observations and state variables.



Can use other functions to weight regression.

Unclear what *distance* means for some obs./state variable pairs.

Referred to as LOCALIZATION.

<u>Localization is function of expected correlation between obs and state.</u>

Often, don't know much about this.

Horizontal distance between same type of variable may be okay.

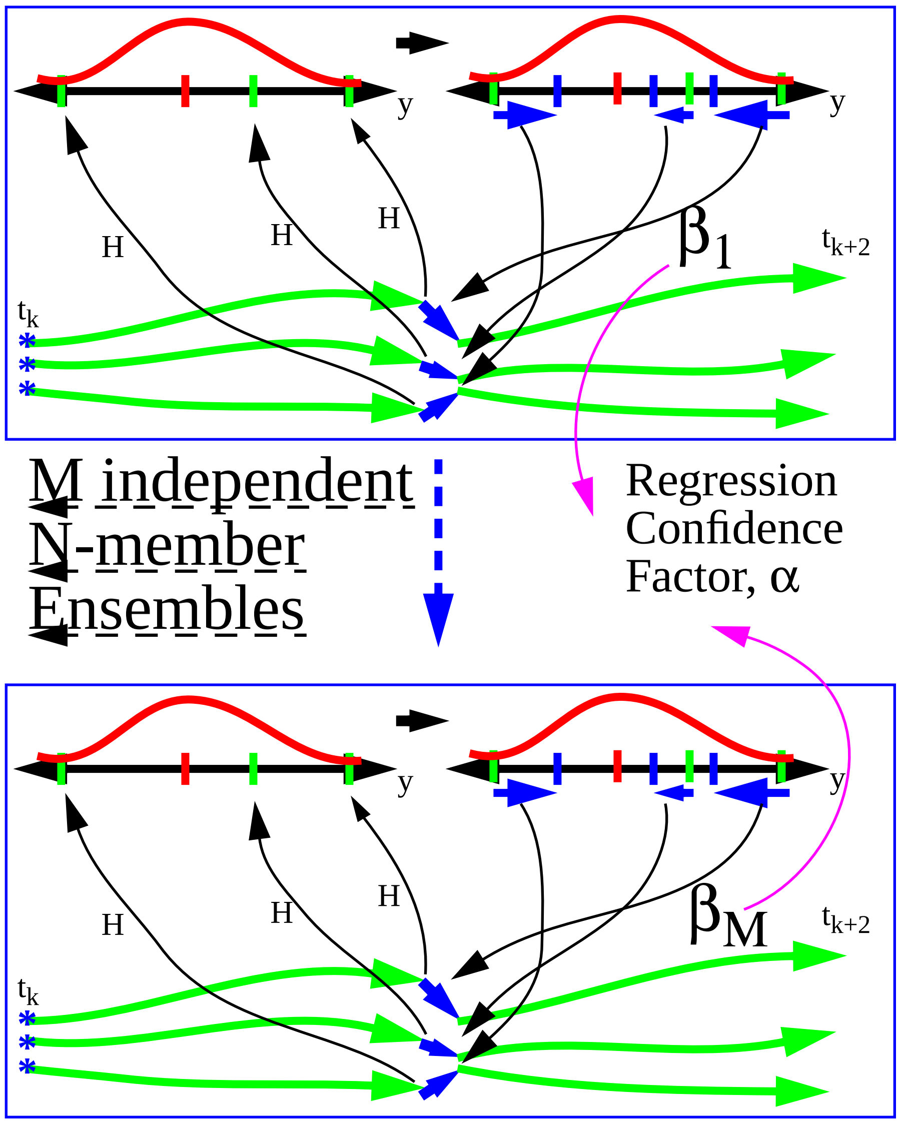
What is expected correlation for co-located temperature and pressure?

What about vertical localization? Looks pretty complex.

What about complicated forward operators:
Expected correlation of satellite radiance and wind component?

Ways to deal with regression sampling error:

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



M groups of N-member ensembles.

Compute obs. increments for each group.

For given obs. / state pair:

1. Have M samples of regression coefficient, β.
2. Uncertainty in β implies state variable increments should be reduced.
3. Compute regression confidence factor, α.

4C. Use hierarchical Monte Carlo: ensemble of ensembles.

Split ensemble into M independent groups.

For instance, 80 ensemble members becomes 4 groups of 20.

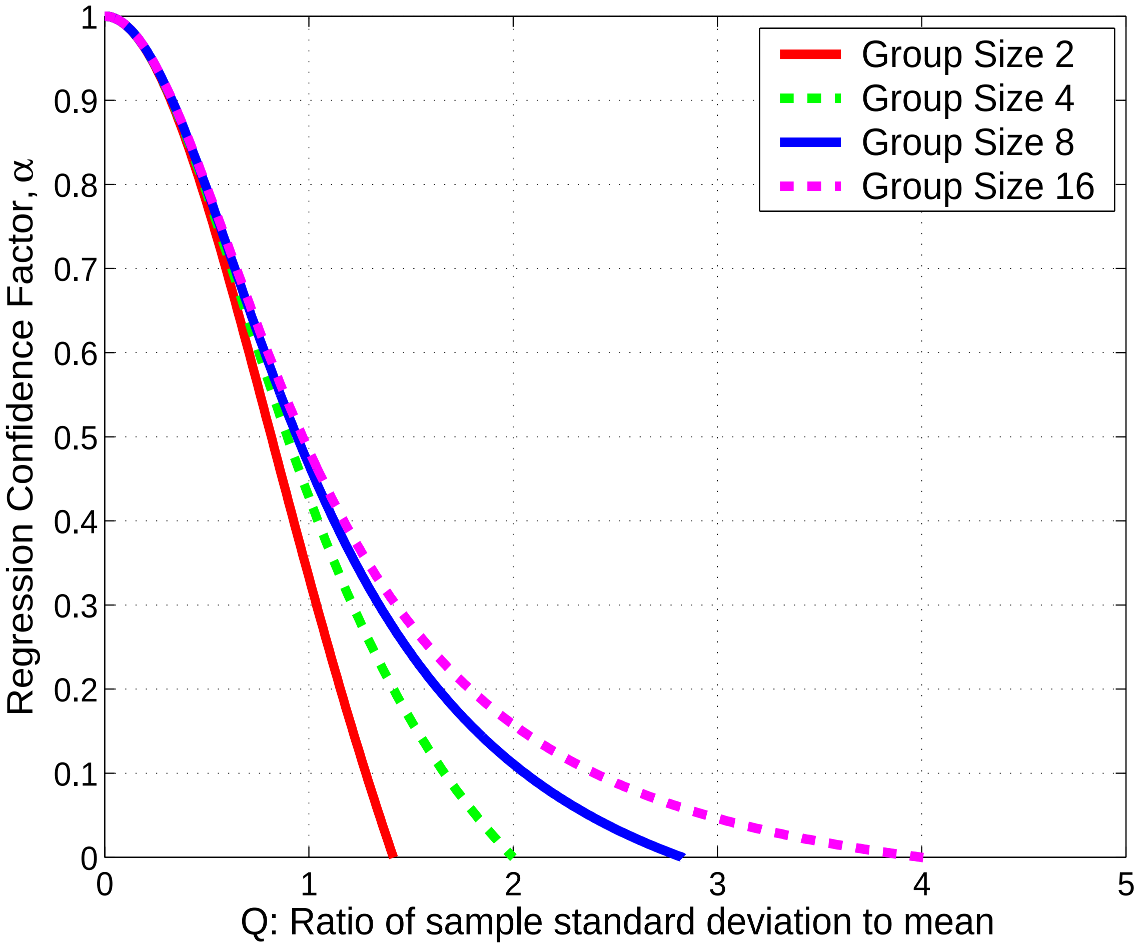
With M groups get M estimates of regression coefficient, β_i.

Find regression confidence factor α (weight) that minimizes:

$$\sqrt{\sum_{j=1}^{M} \sum_{i=1, i \neq j}^{M} [\alpha\beta_i - \beta_j]^2}$$

Minimizes RMS error in the regression (and state increments).

4C. Use hierarchical Monte Carlo: ensemble of ensembles.



Weight regression by α.

If one has repeated observations, can generate sample mean or median statistics for α.

Mean α can be used in subsequent assimilations as a localization.

α is function of M and $Q = \Sigma_{\beta} / \bar{\beta}$ (sample SD / sample mean regression)

Lorenz 96 Experimental Design

Initial ensemble members random draws from 'climatology'

Observations every time step

4000 step assimilations, results shown from second 2000 steps

Covariance inflation tuned for minimum RMS

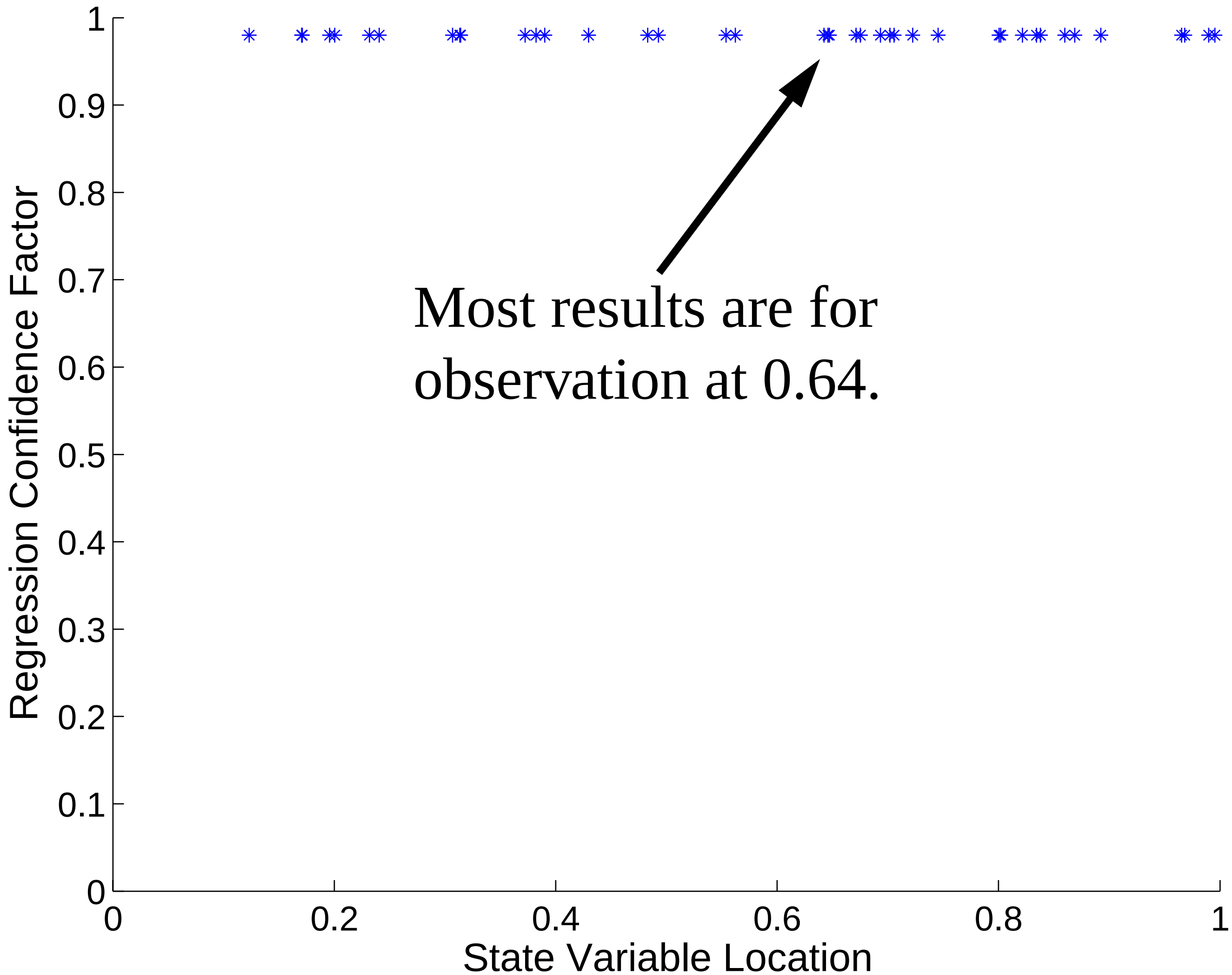
4 groups of ensembles used unless otherwise noted

40 Randomly located observations

Observation error variance 10^{-5}, 10^{-3}, 0.1, 1.0, 10.0, 10^{7}

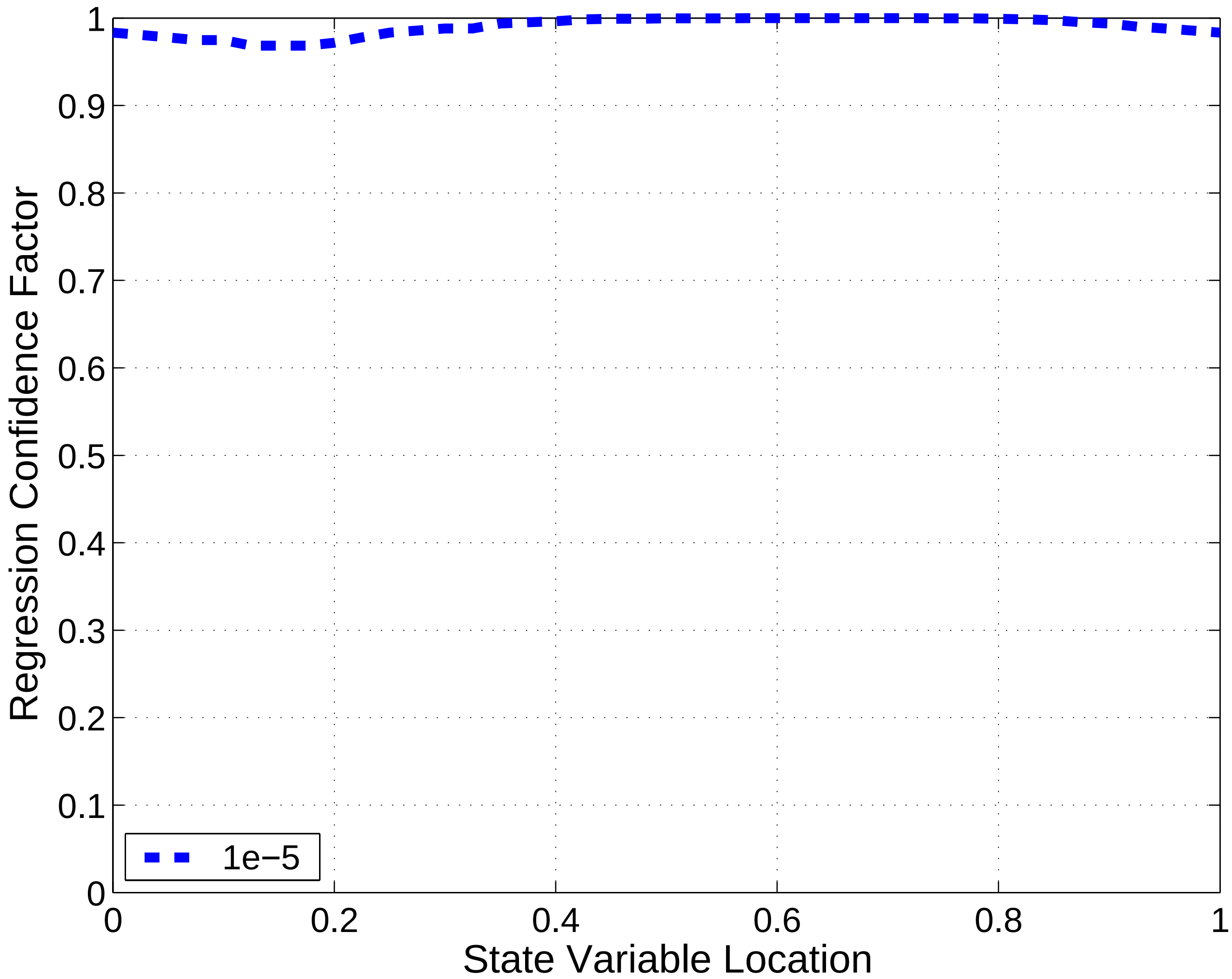
14 member ensembles; not degenerate

Time Mean Regression Confidence Envelopes



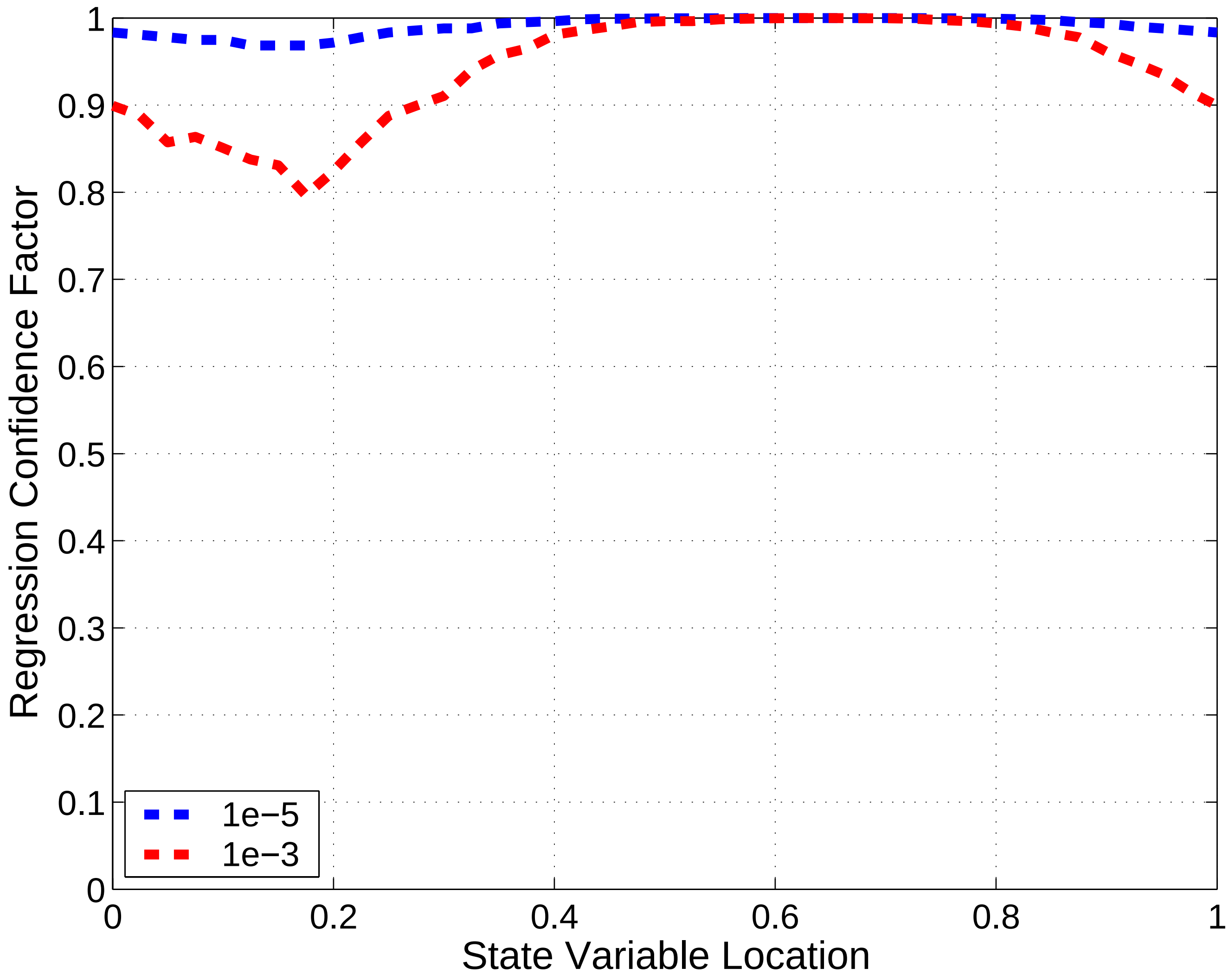
Location of 40 randomly located observations

Time Median Envelopes: Varying Obs. Error Variance

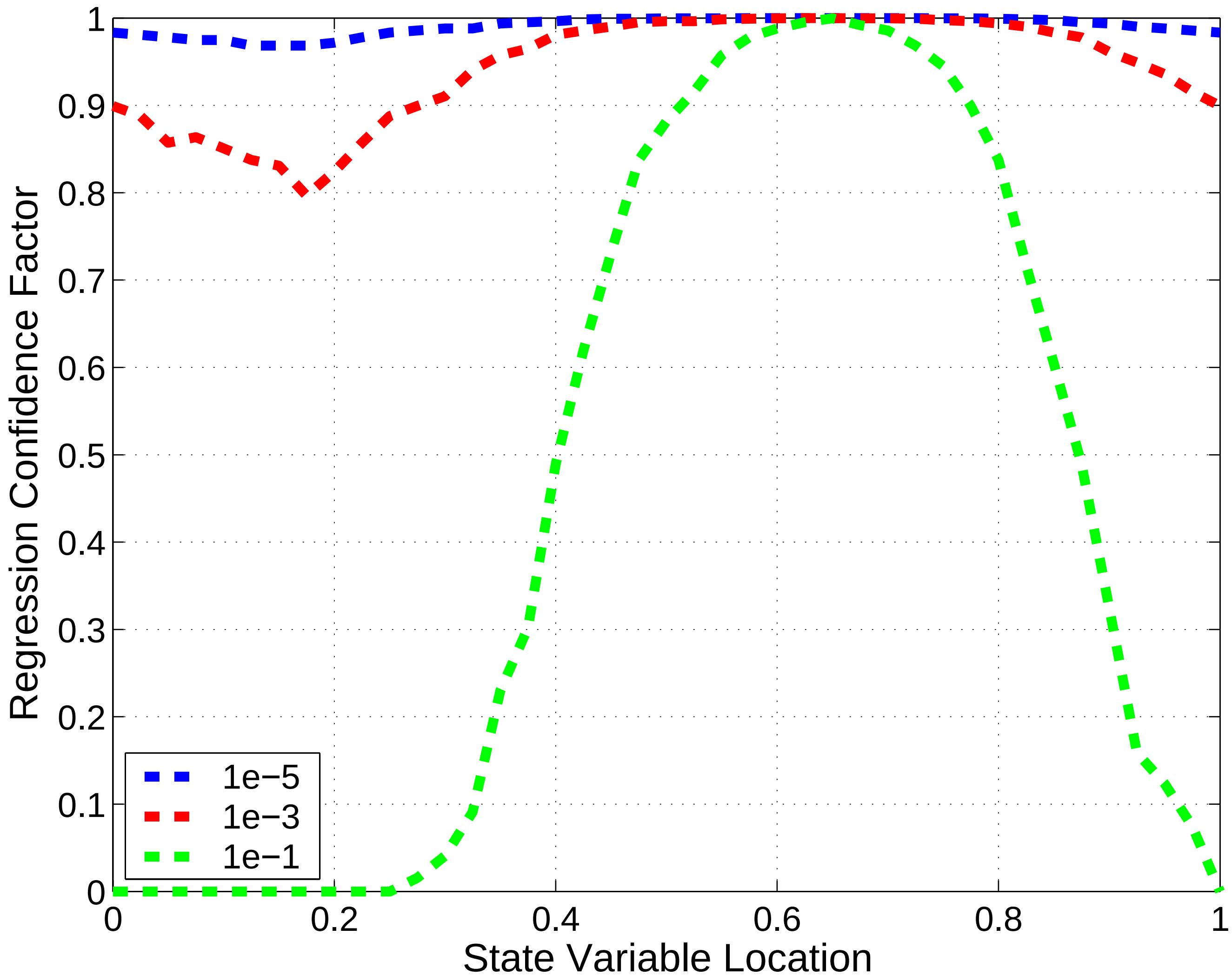


Small error implies no need for localization

Time Median Envelopes: Varying Obs. Error Variance

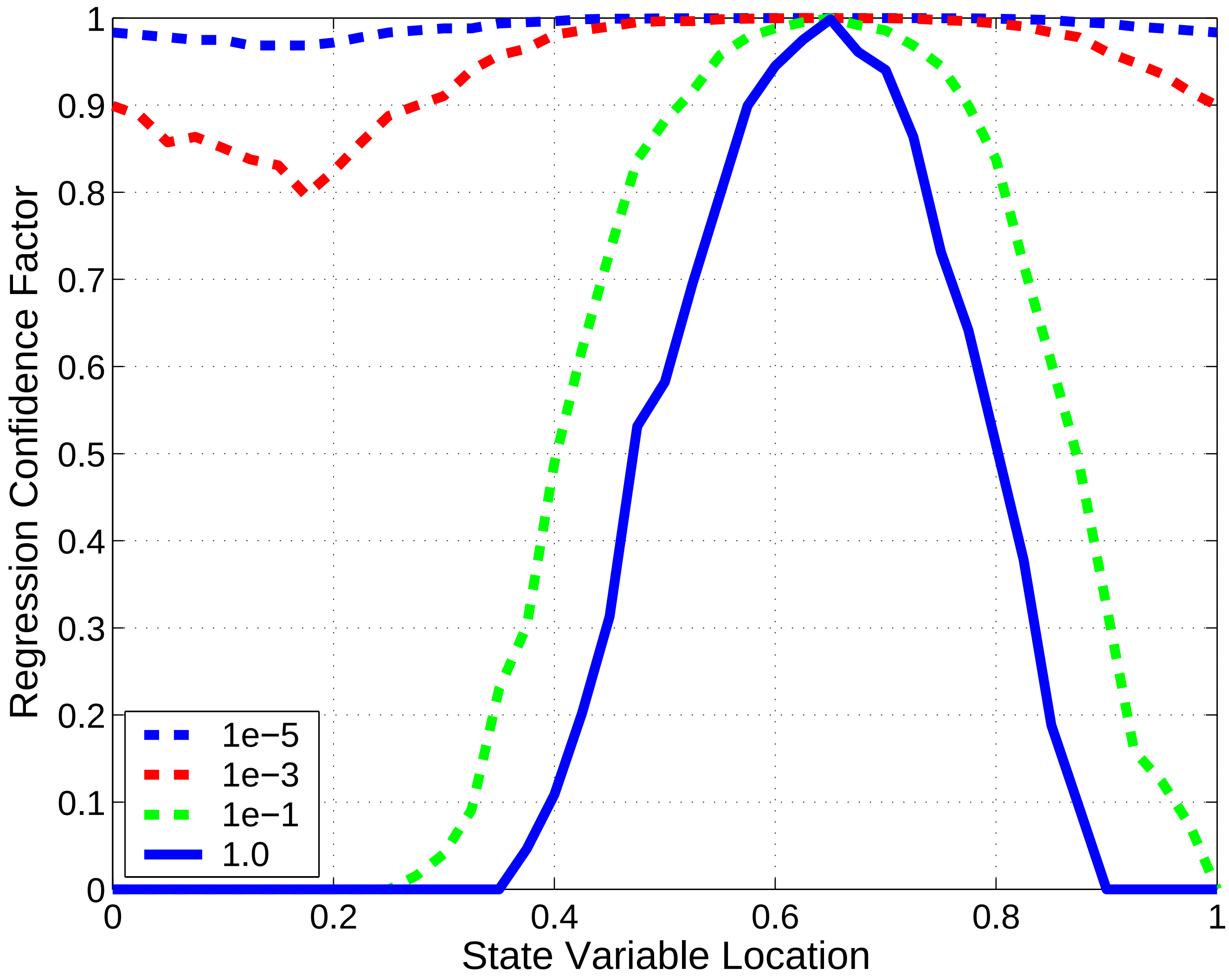


Time Median Envelopes: Varying Obs. Error Variance

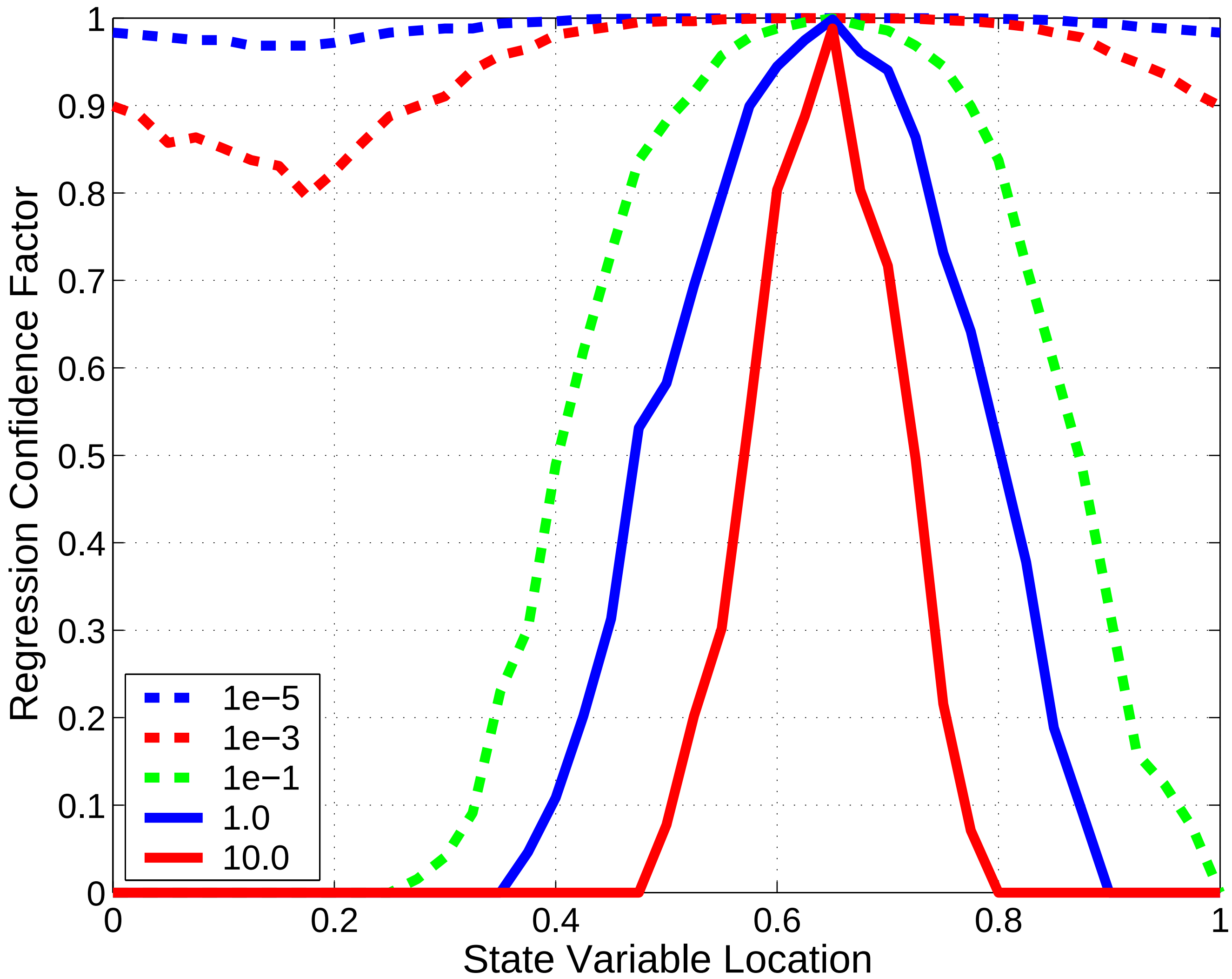


Increasing error implies increasing localization

Time Median Envelopes: Varying Obs. Error Variance

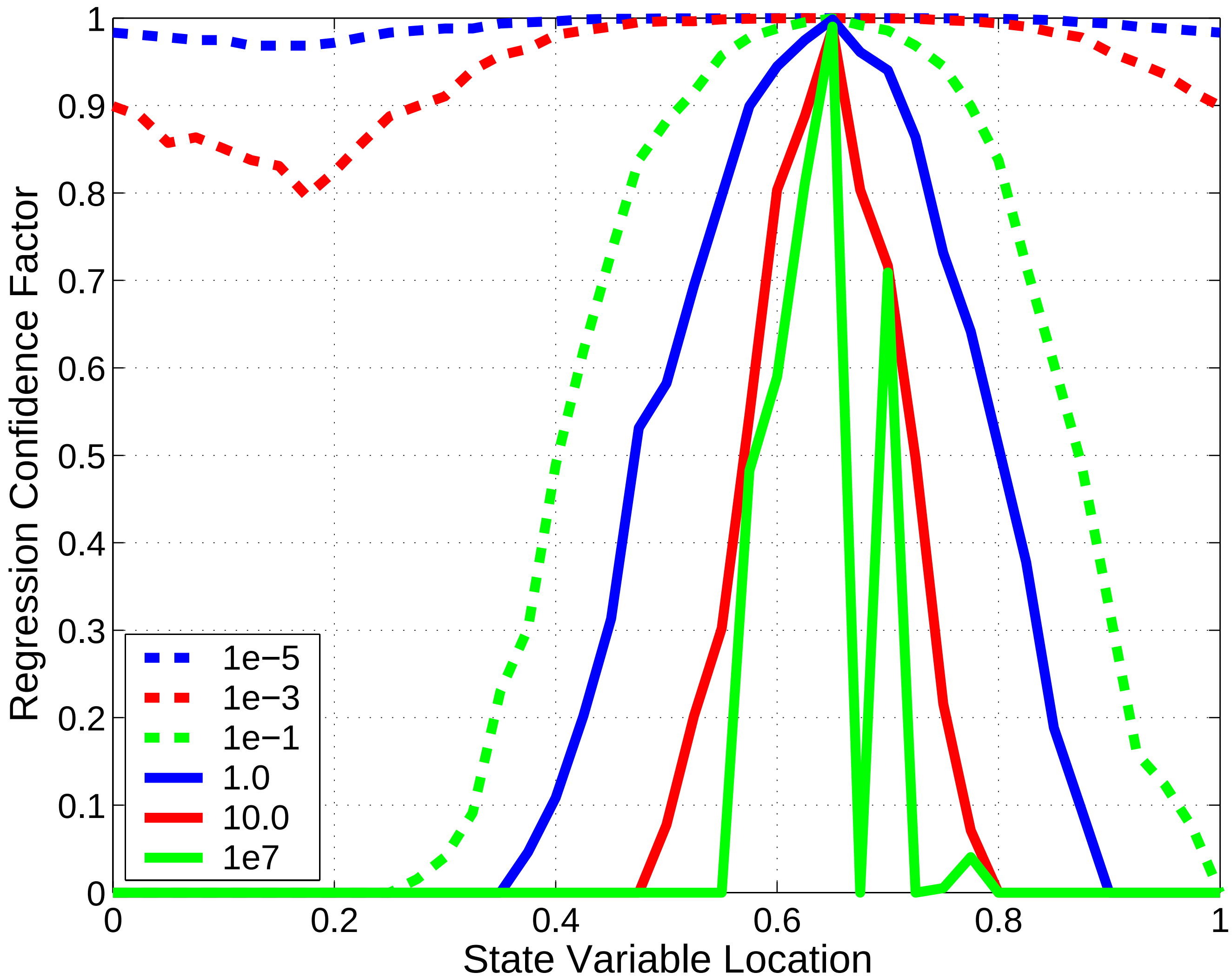


Time Median Envelopes: Varying Obs. Error Variance



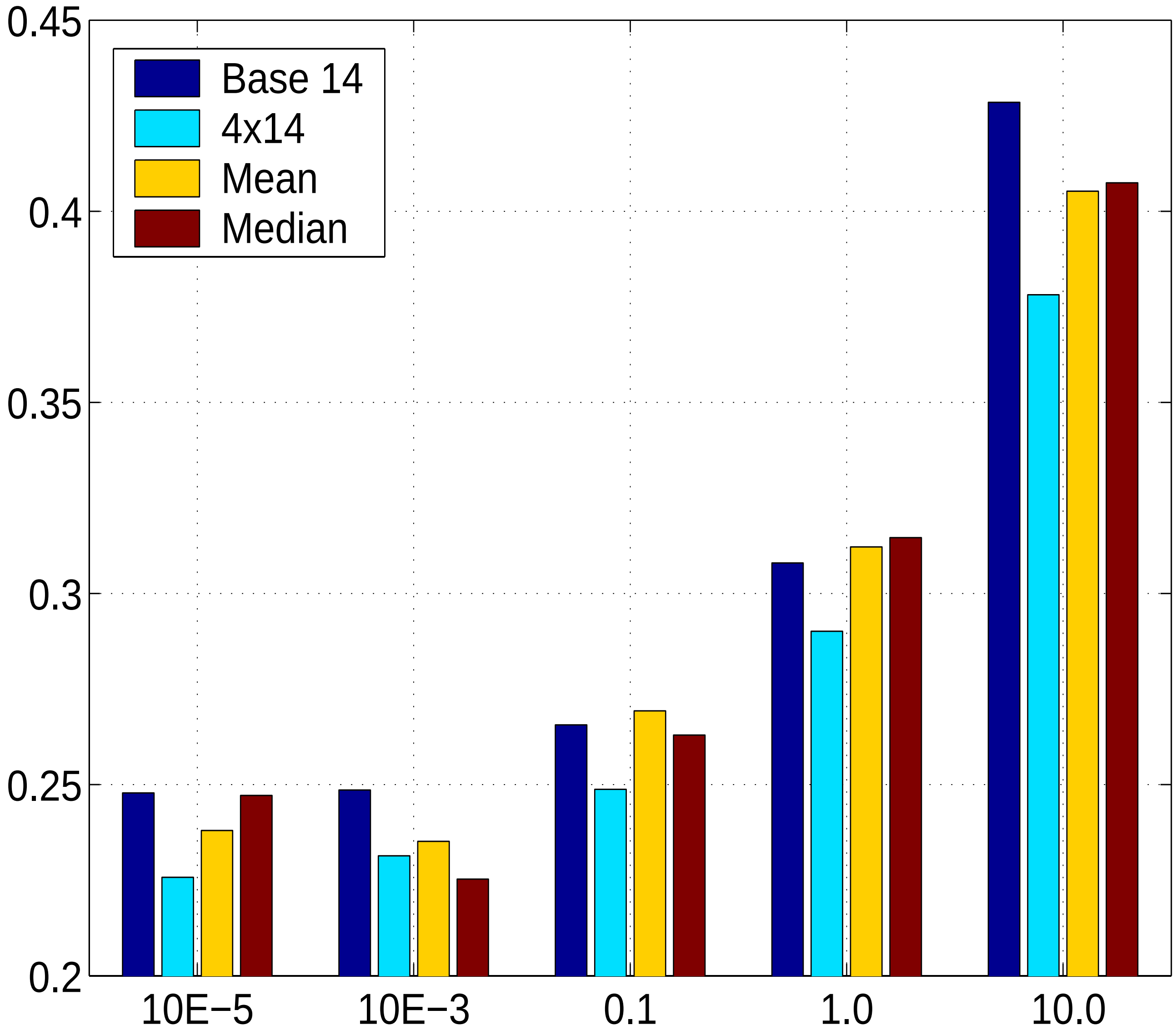
Single Gaspari Cohn half-width can't deal with this range of errors

Time Median Envelopes: Varying Obs. Error Variance



Climatological case is unique: Looks like time mean coherence

Time Median Envelopes: Varying Obs. Error Variance



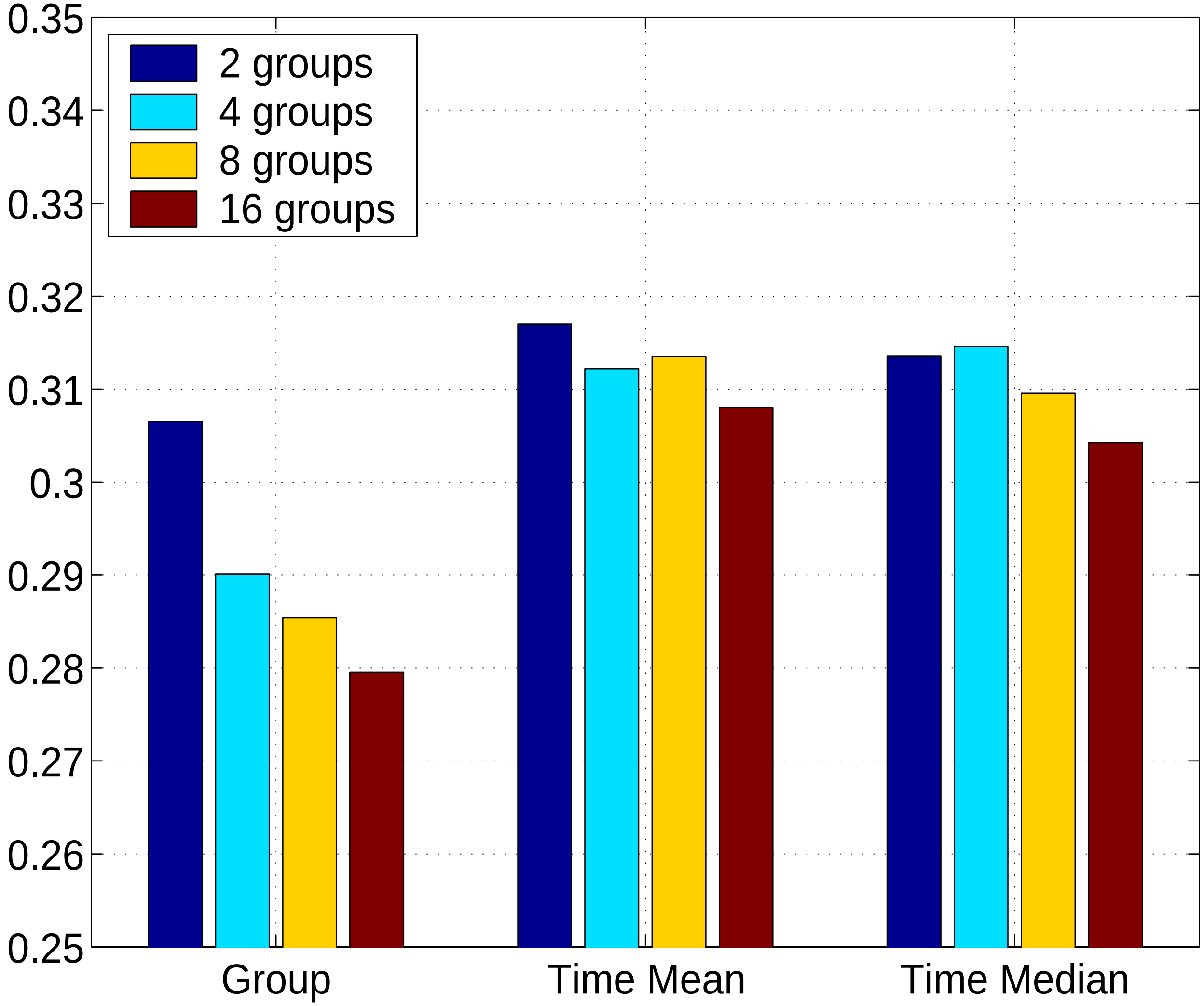
Error scaled by obs. error standard deviation

As error grows, filter becomes less ‘efficient’

Things are more ‘nonlinear’ for large errors

Group, mean and median compare favorably with tuned base for all cases!

Sensitivity of results to group size



Gradual improvement for increased group size

Time mean/median improve more slowly

Important that small groups give good results (for cost)

Group better than time mean implies some time varying information (time means should reduce noise)

Obs. error variance 1.0, 14 member ensemble case

Assimilating observations at times different from state estimate

Ensemble smoothers: use future observations

Targeted observations: examine impact of obs. in past

Real-time assimilation: use of late arriving observations in forecast

Expect correlations to diminish as time separation increases

Need a 'localization' in time, too

Group filter can provide this

Time ‘localization’: Experimental design

4 group, 14 ensemble member filter

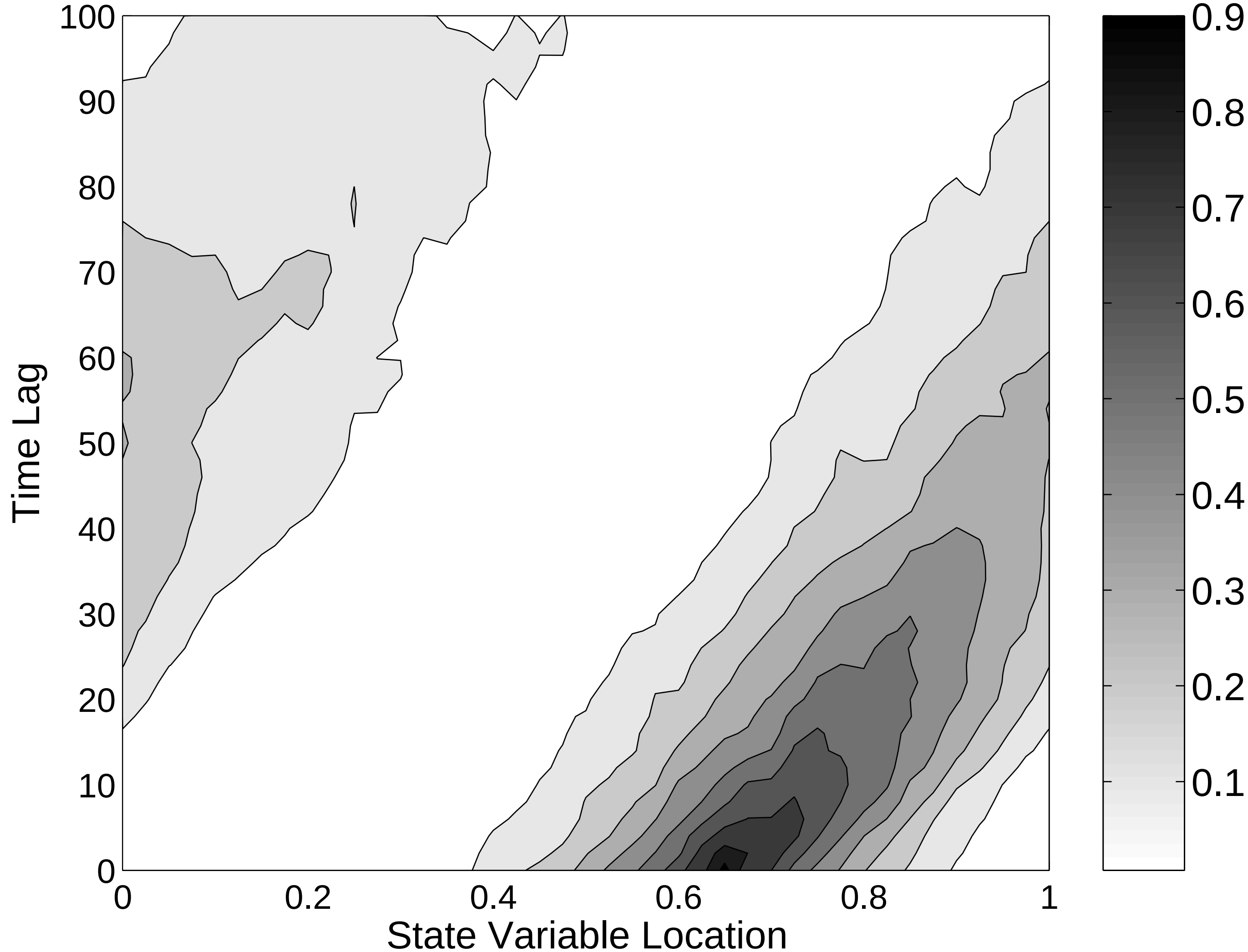
40 random obs. with 1.0 error variance

1 additional observation at location 0.642

The additional observation is from a prior time step

Time mean regression confidence envelope as function of time lag

Regression confidence factor as function of obs. lag time

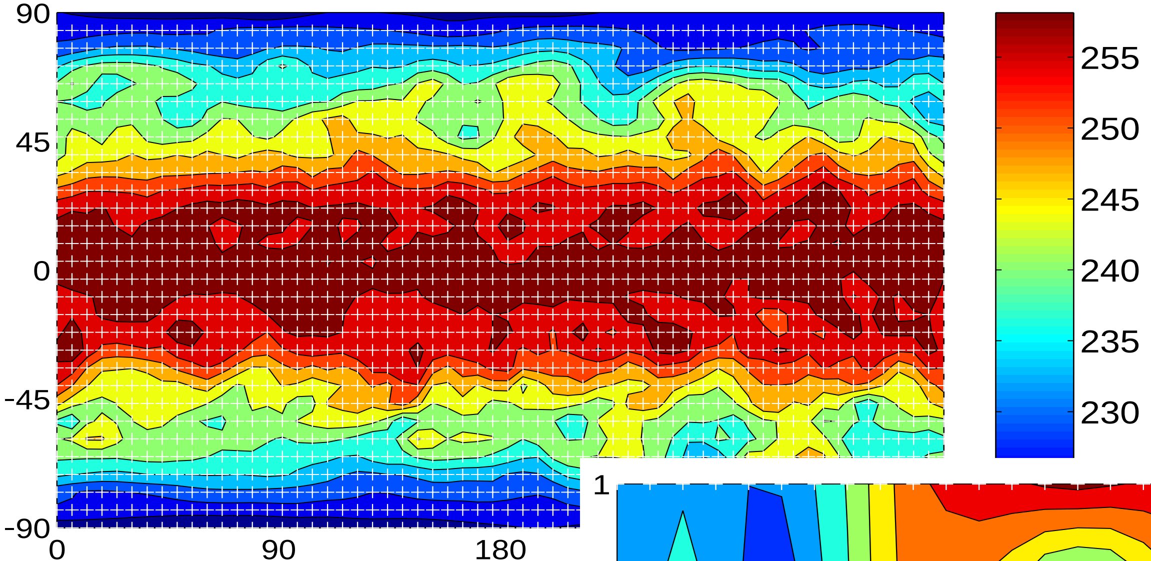


Moves with group velocity (approximately); dies off with lead

Assimilation in Idealized AGCM: GFDL FMS B-Grid Dynamical Core (Havana)

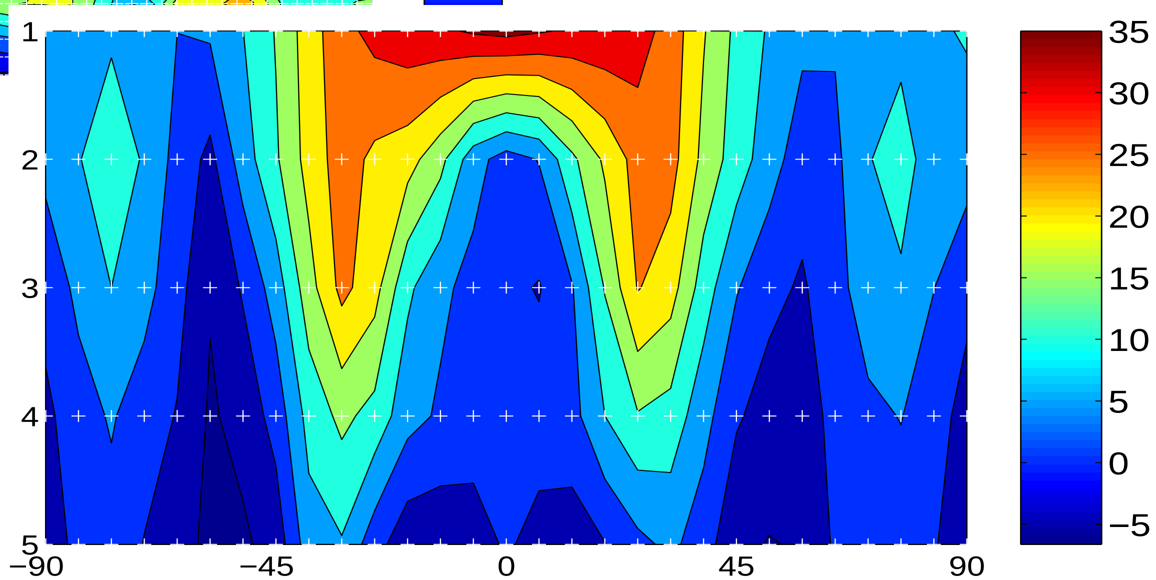
Held-Suarez Configuration (no zonal variation, fixed forcing)

Low-Resolution (60 lons, 30 lats, 5 levels); Timestep 1 hour



T at Level 3
(Middle of Atmosphere)

Cross Section of
Zonal Mean U



Has Baroclinic Instability

Results for 4x20 group filter

Assimilation for 400 days; starting from climatological distribution

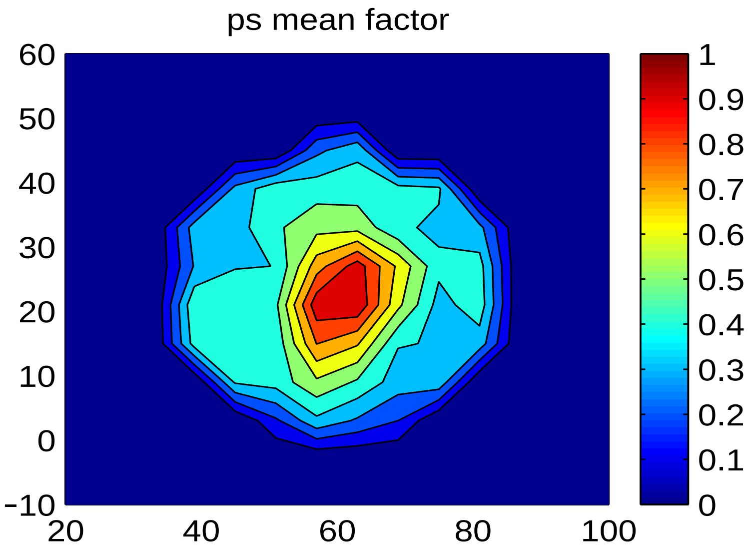
Summary results are from last 200 days

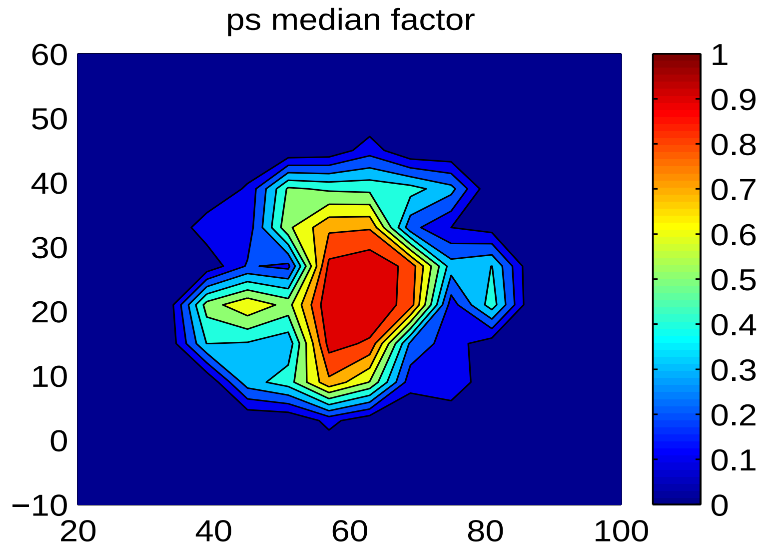
No covariance inflation

1800 randomly located surface pressure stations observe once every 24 hours

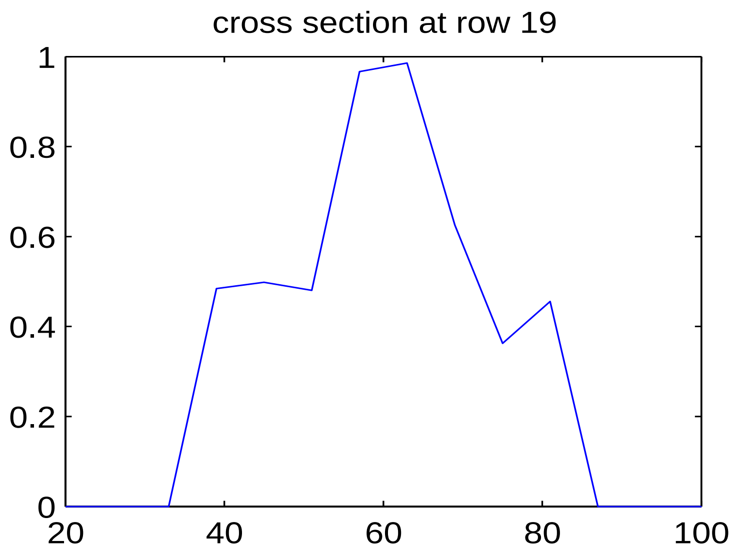
Observational error variance is 1 mb

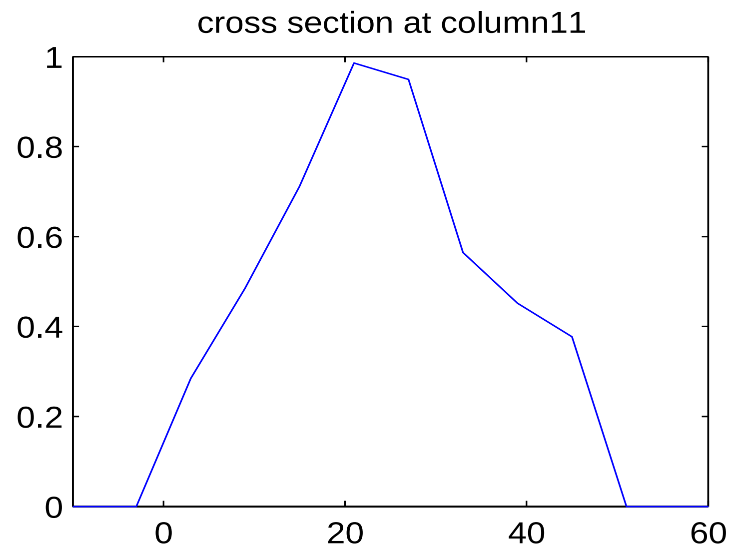
Hierarchical Filter Regression Confidence Factors: PS Obs. at 20N, 60E



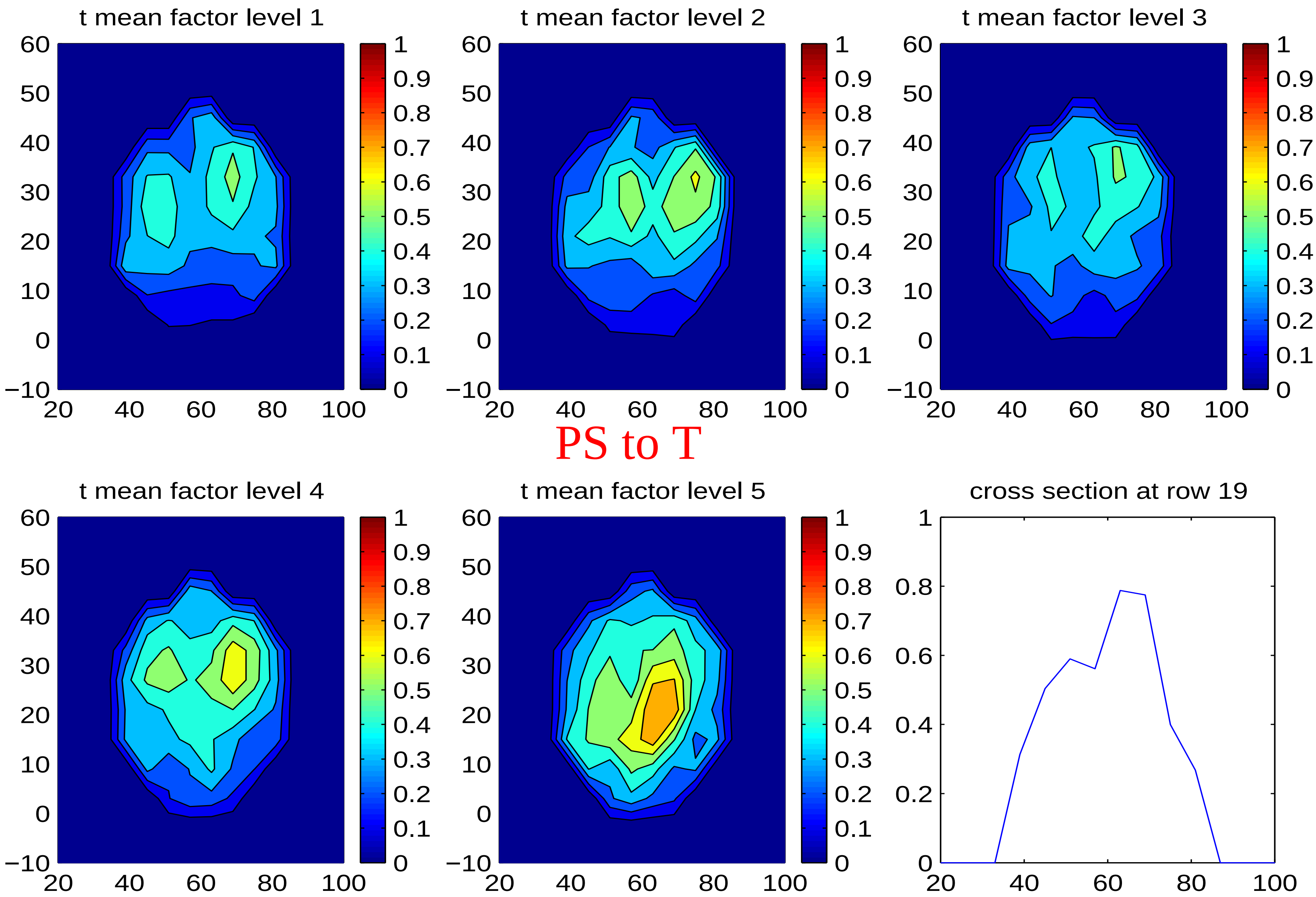


PS to PS

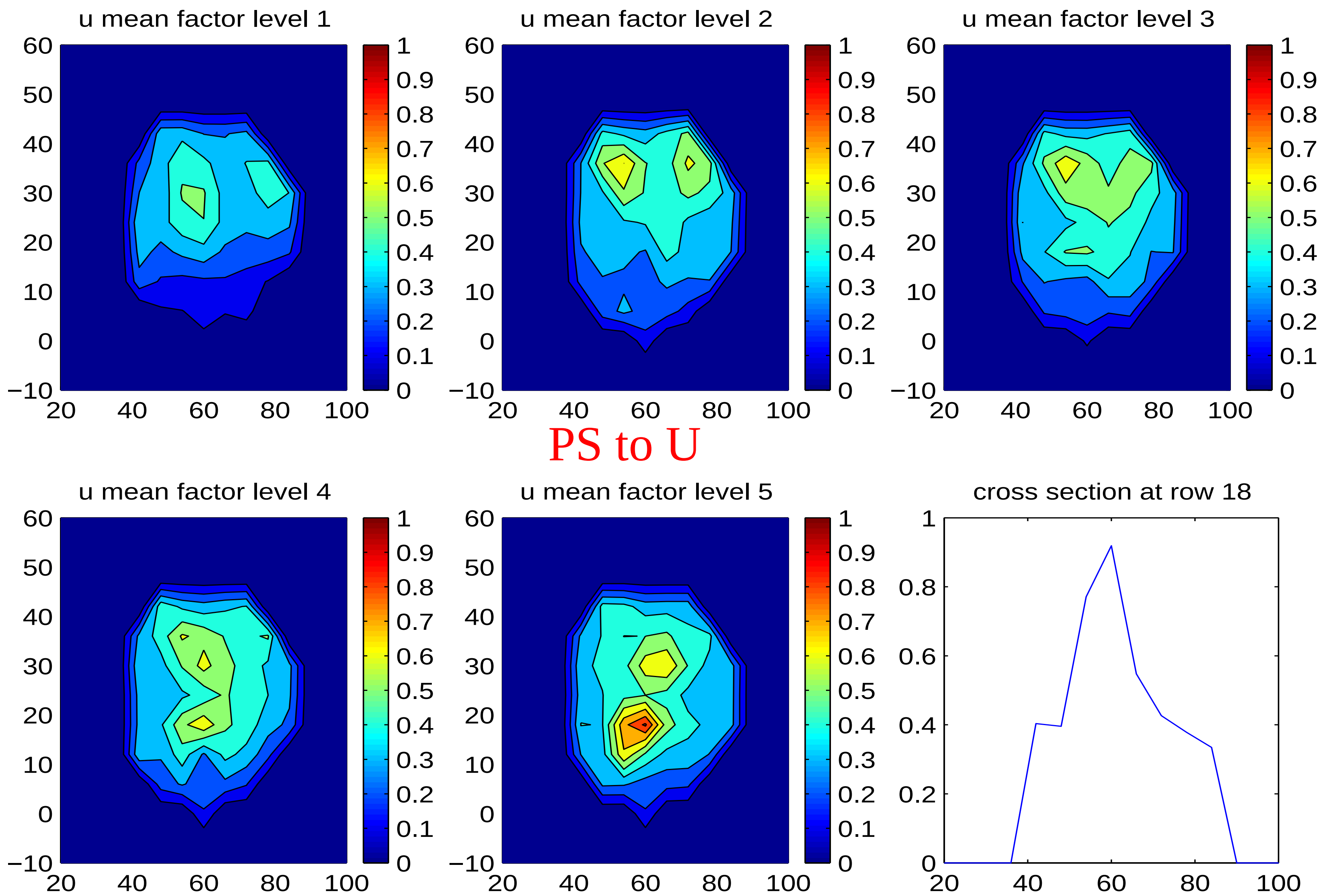




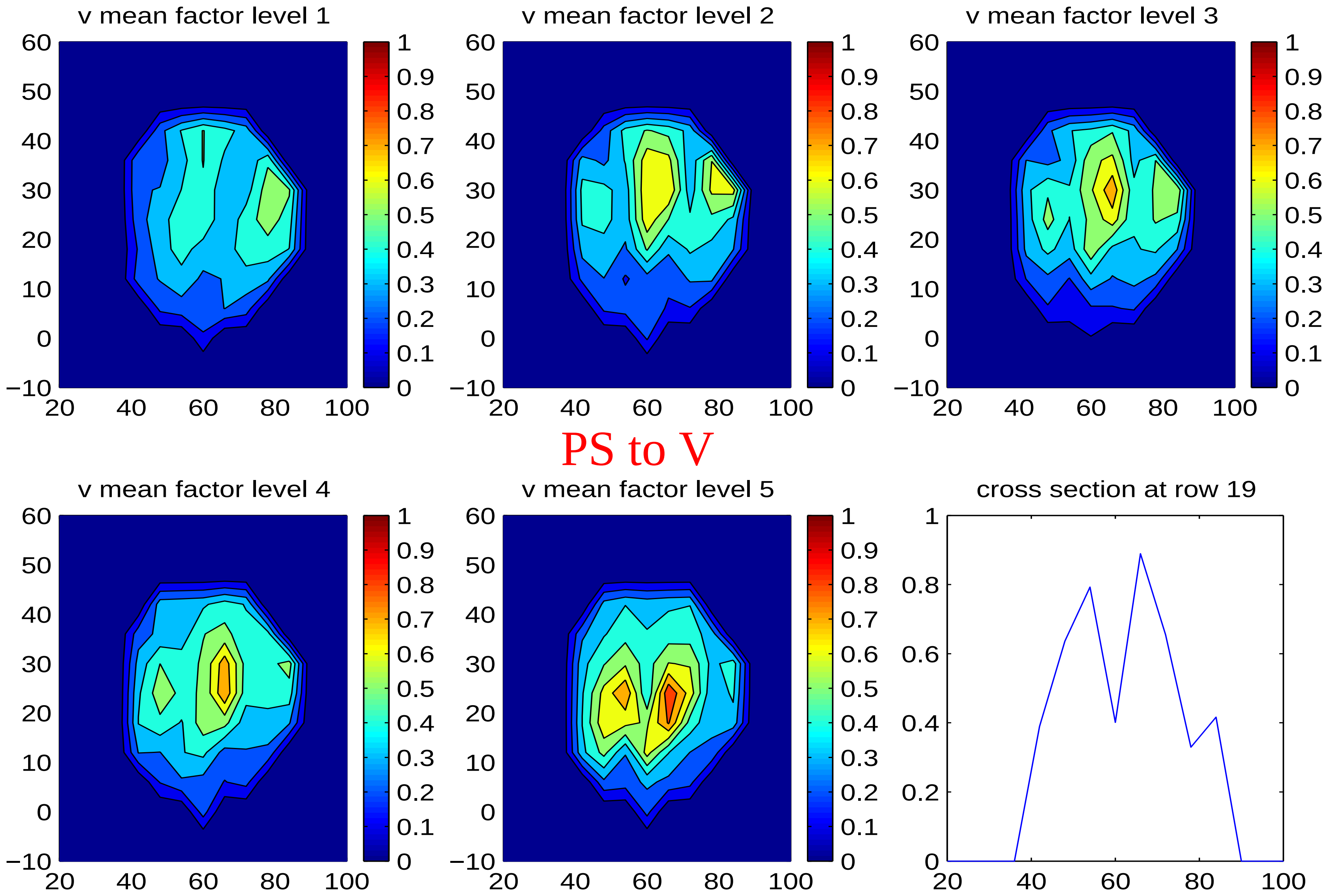
Hierarchical Filter Regression Confidence Factors: PS Obs. at 20N, 60E



Hierarchical Filter Regression Confidence Factors: PS Obs. at 20N, 60E



Hierarchical Filter Regression Confidence Factors: PS Obs. at 20N, 60E



Hierarchical Bayesian Methods for Assimilation Parameter Estimation

1. Is it worth it?
    - A. Added cost needs to lead to reduced error and/or...
    - B. Wider applicability for parameters.
    - C. In other words, don't want to have to tune.

2. Ensemble / ensemble hierarchies are very expensive.
    - A. Probably only for exploratory work.
    - B. Extract information on parameter distributions.

3. Ensemble / continuous may not be expensive.
    - A. Running these in real time not out of the question.
    - B. Could look at more sophisticated error correction models.