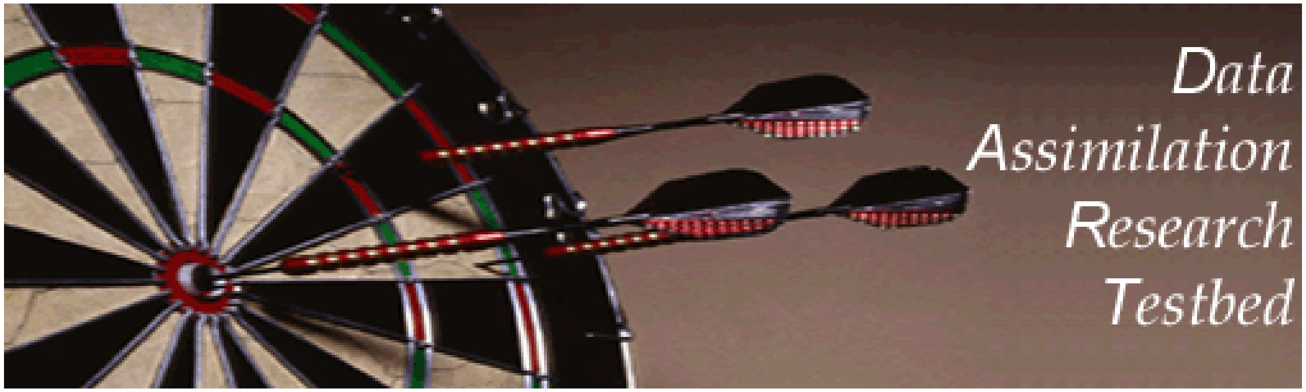
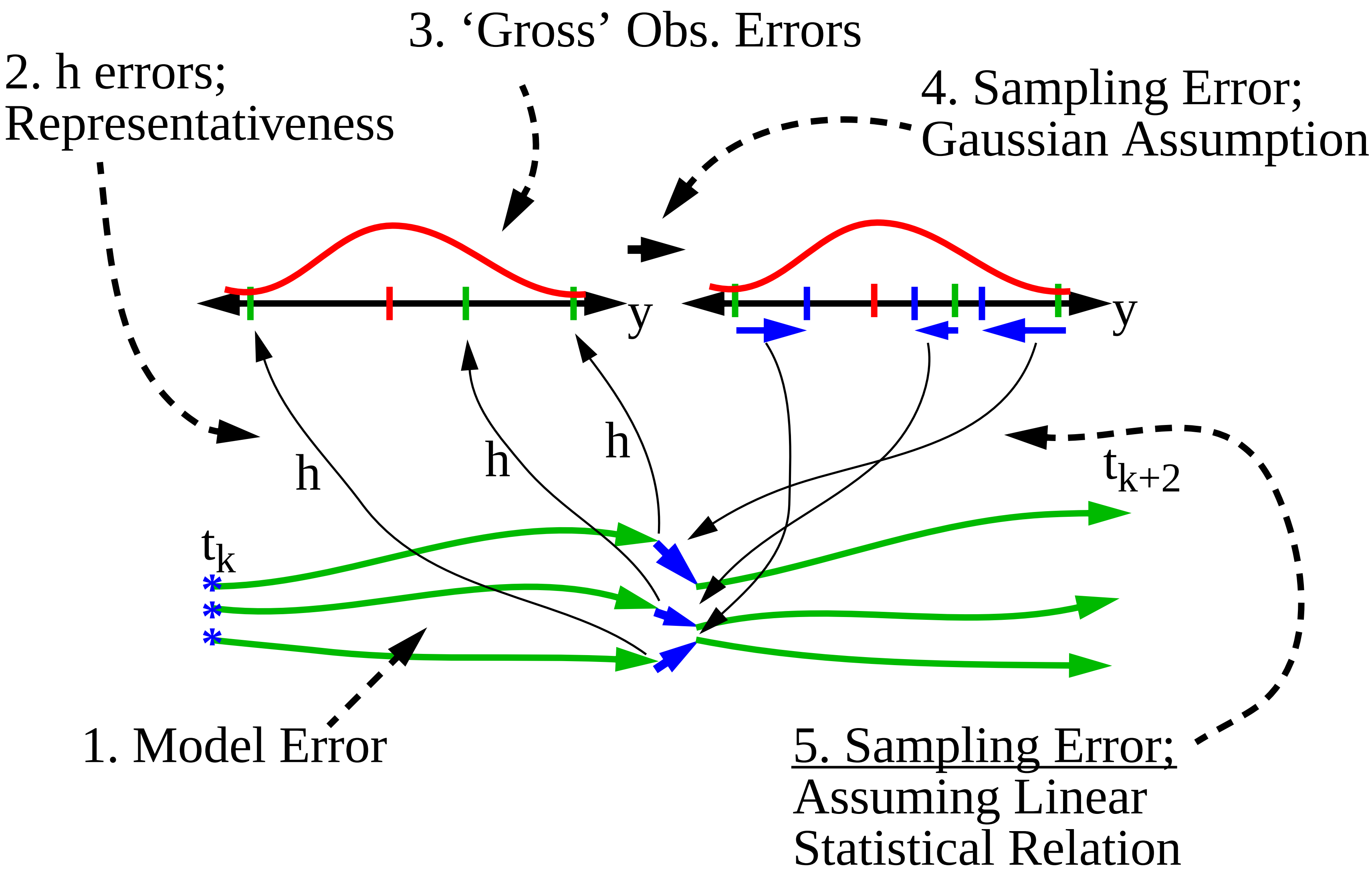
Adaptive Inflation for Ensemble Assimilation



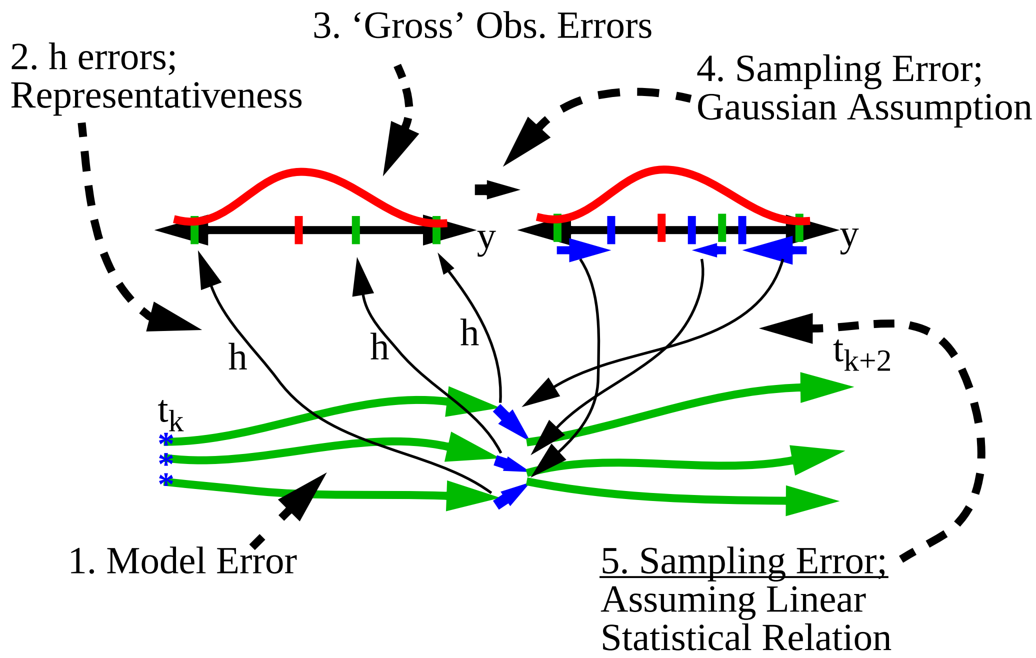
Jeffrey Anderson
IMAGe Data Assimilation Research Section (DAReS)

Thanks to Ryan Torn, Nancy Collins, Tim Hoar, Hui Liu,
Kevin Raeder, Xuguang Wang

Some Error Sources in Ensemble Filters



All Error Sources are Expected to Lead to Too Little Variance



Increase variance by linearly inflating ensemble around mean.

Can do this to PRIOR before computing obs. operator or,

In OBS. SPACE after computing obs. operator or,

To POSTERIOR after doing regression.

Pretty Much the Simplest Low-Order Model

1. Single state variable, linear growth around 0.

$$x_{t+1} = \alpha x_t$$ α=2 in first examples.

2. For perfect model, truth is perpetually 0.

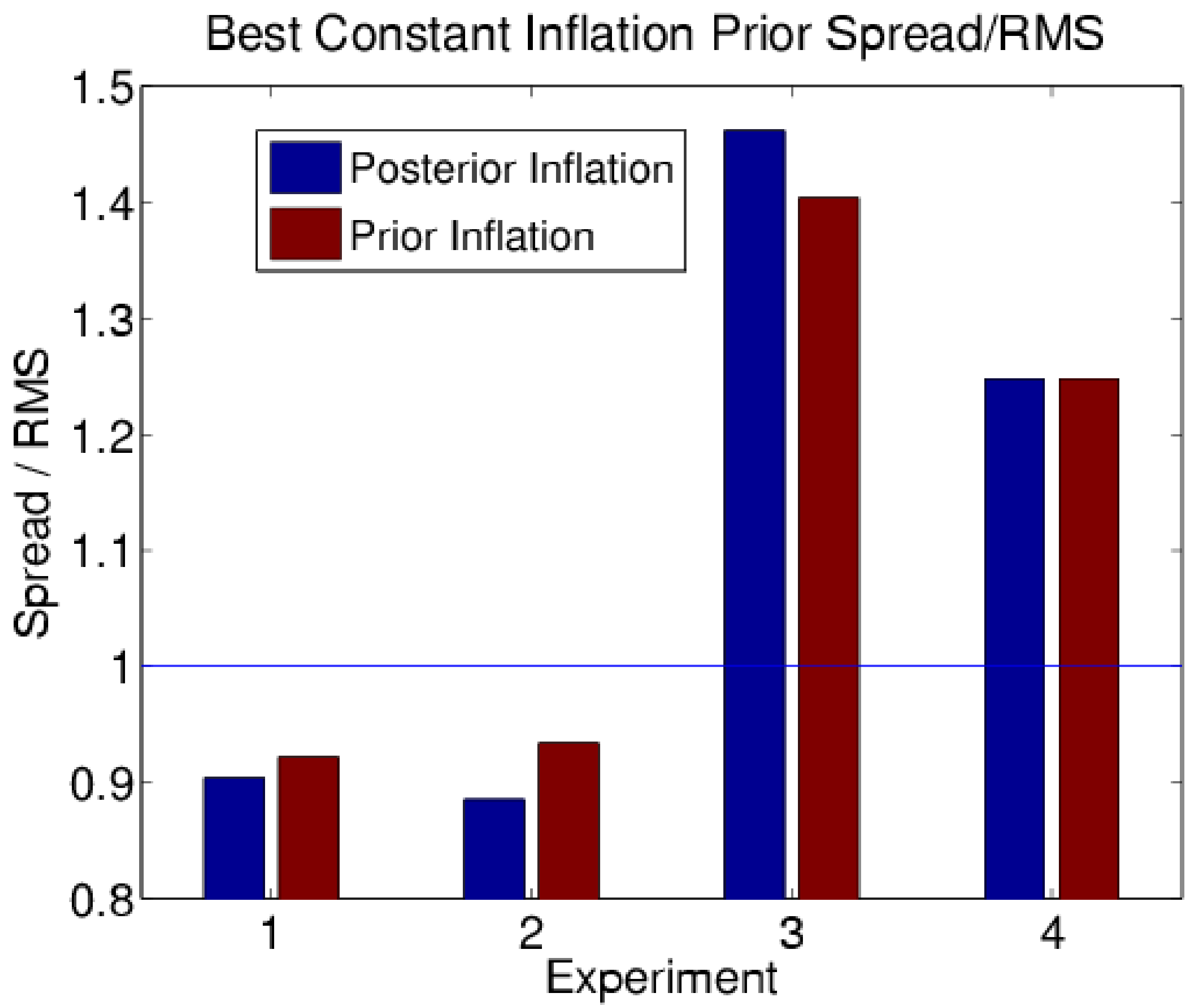
3. Variable observed directly with synthetic noise.

4. Model error in assimilating model is additive.

$$x_{t+1} = \alpha x_t + \gamma$$ γ=0.1 in first examples.

Case 1: obserr variance 1.0, 5 member EnKF.
Case 2: same as 1). but obserr variance 10.0 with 10 obs per time.
Case 3: same as 1). but observed every 12th step.
Case 4: same as 1). but 20 member EAKF, γ=0.1.

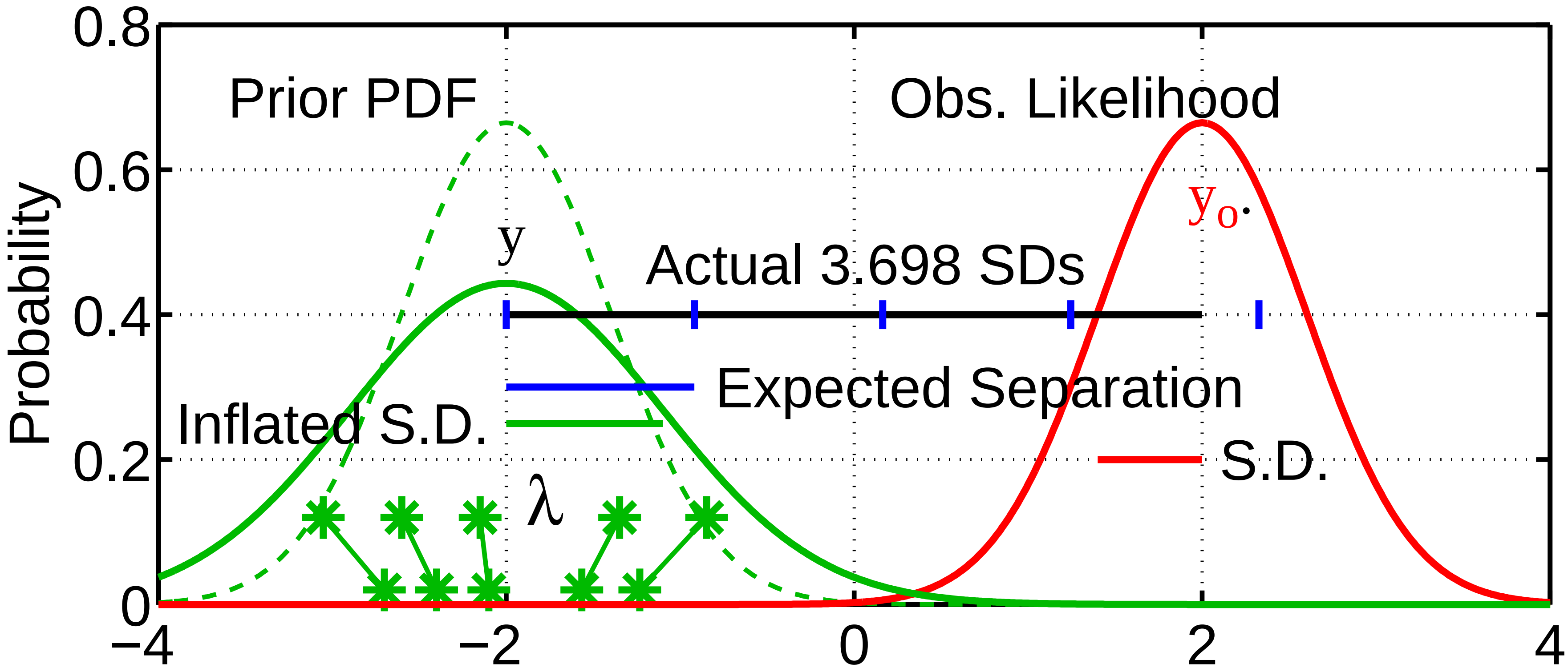
Results for BEST time constant inflation; Prior and Posterior



Lowest Prior RMS does NOT mean spread/RMS is unity.

Adaptive Inflation for Ensemble Filtering

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

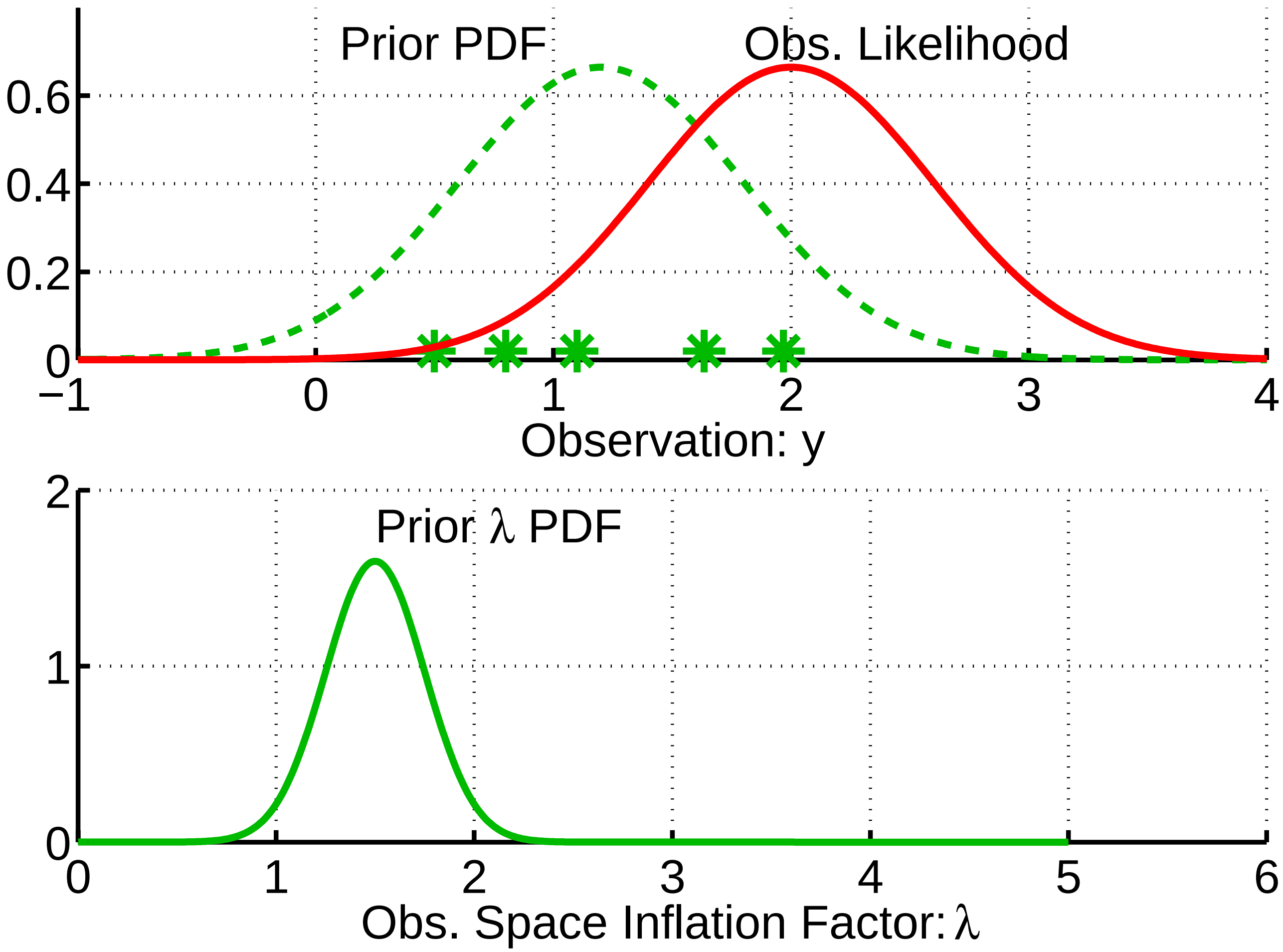


Distance, D, from prior mean y to obs. is $N\left(0, \sqrt{\lambda\sigma_{prior}^2 + \sigma_{obs}^2}\right) = N(0, \theta)$

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

Adaptive Inflation for Ensemble Filtering

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



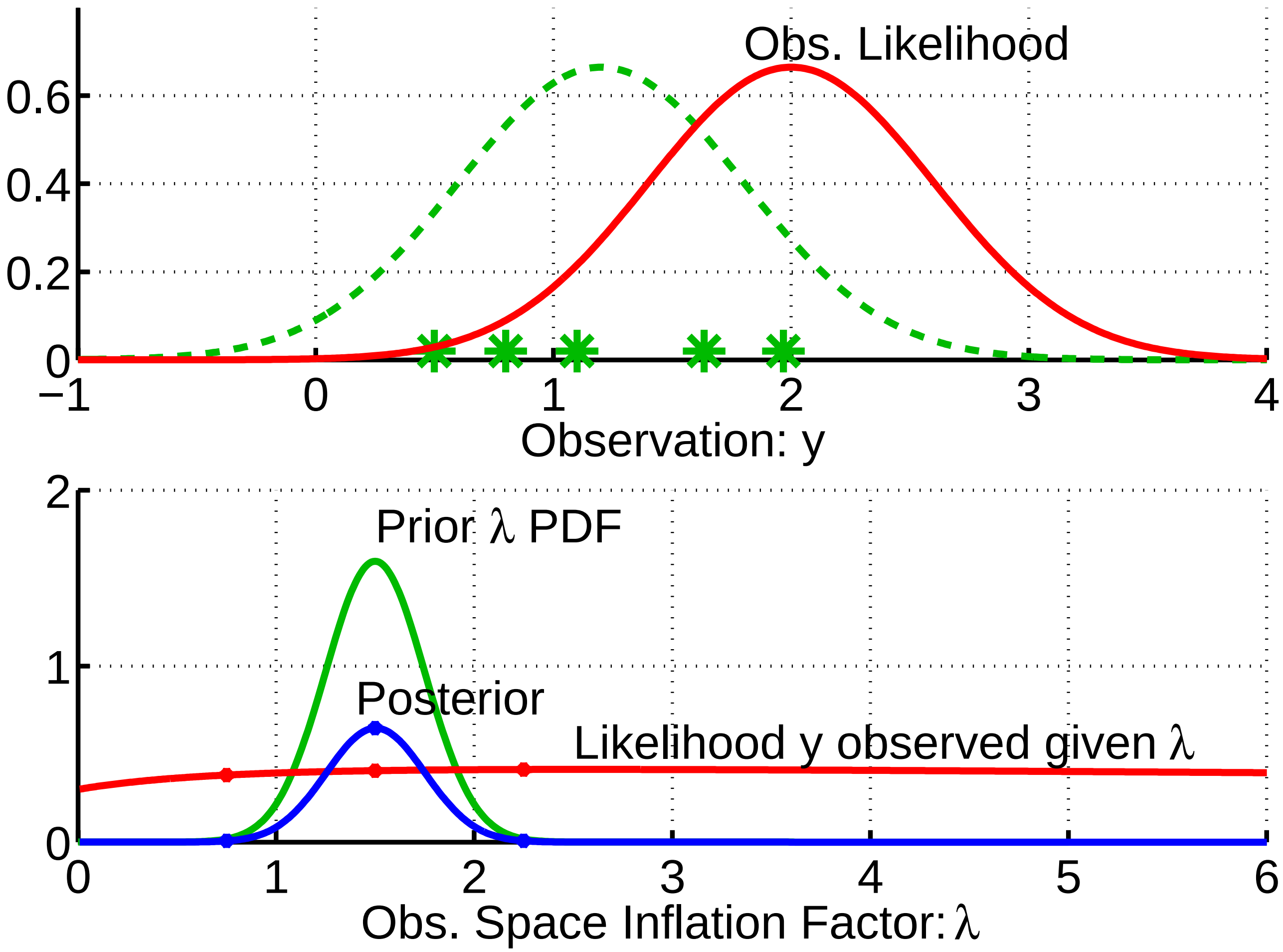
We've assumed a gaussian for prior $p(\lambda|Y_{prev})$.

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

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization.$$

Adaptive Inflation for Ensemble Filtering

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).

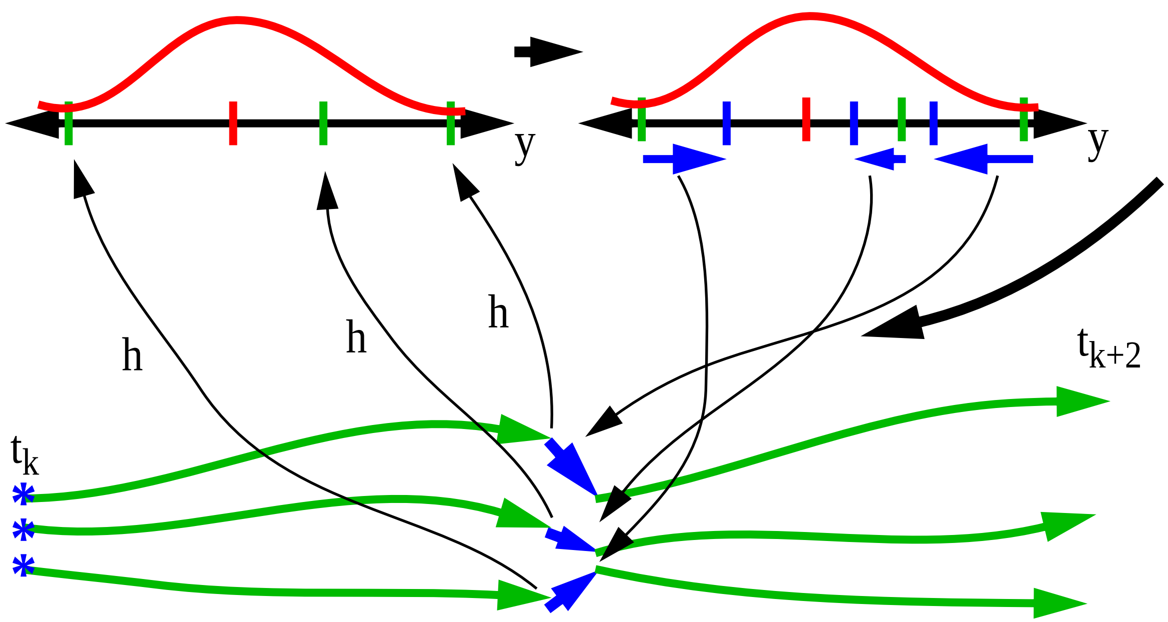
$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization\ .$$

State Space Adaptive Inflation

Computations so far adapt inflation for observation space.

What is relation between observation and state space inflation?

Have to use prior ensemble observation/state joint distribution.



Regress changes in inflation onto state variable inflation.

Spatially varying adaptive inflation algorithm:

Have a distribution for λ at each time 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

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

Equation for finding mode of posterior is now full 12th order:

Analytic solution appears unlikely.

Can do Taylor expansion of θ around $\lambda_{s,i}$.

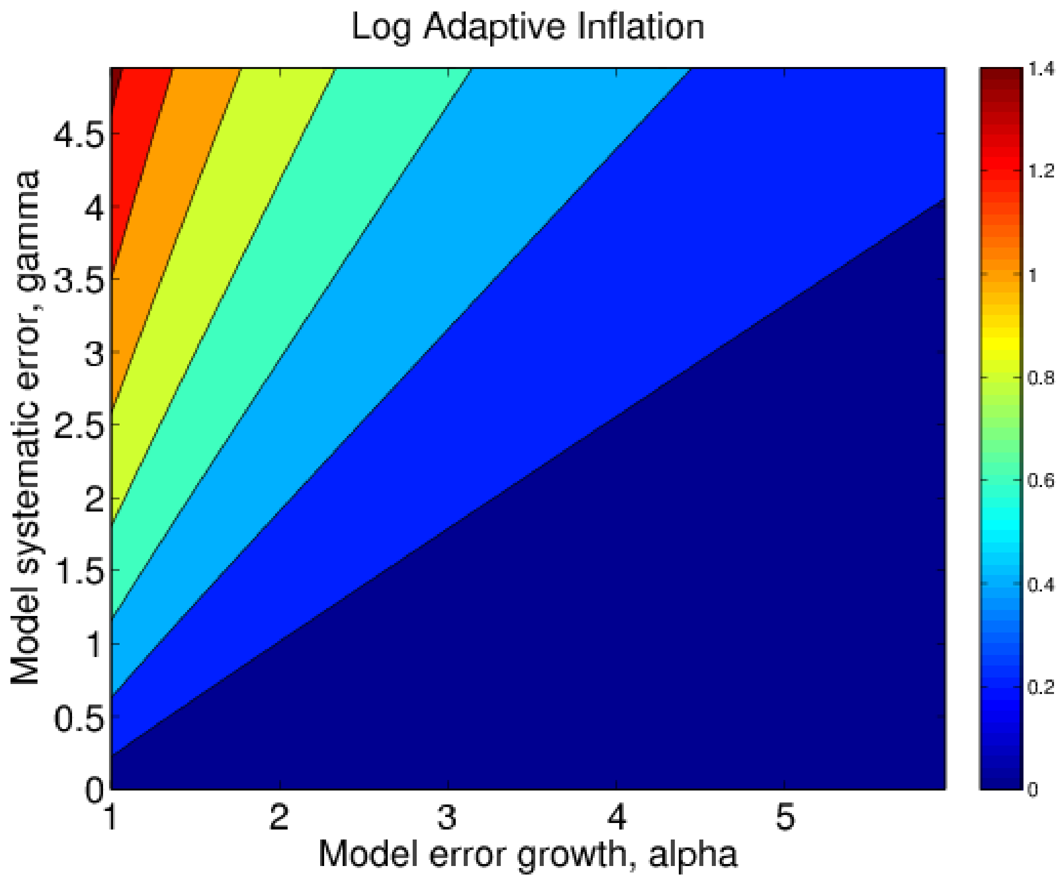
Retaining linear term is normally quite accurate.

There is an analytic solution to find mode of product in this case!

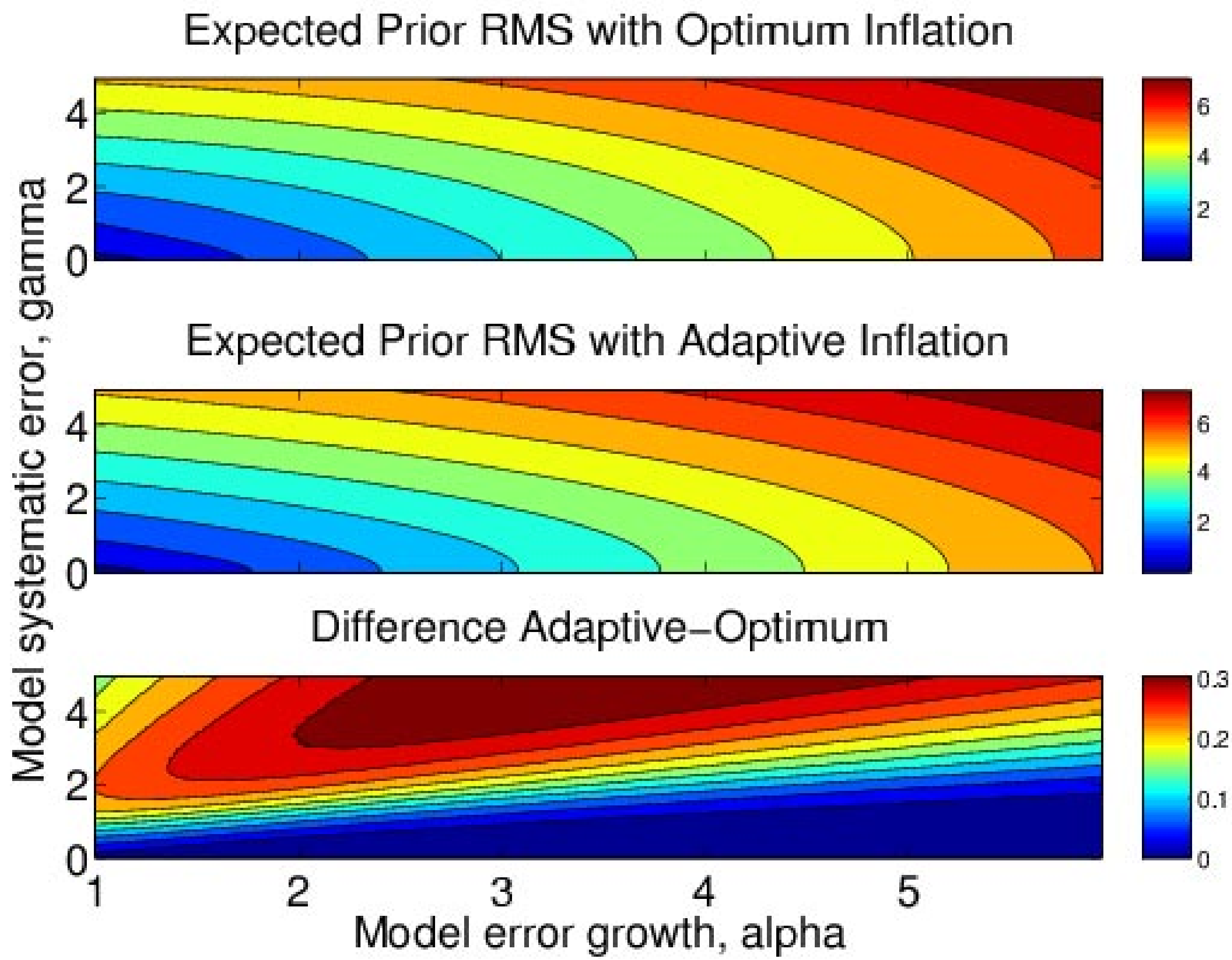
Prior Adaptive Inflation in Simplest Model with Model Error

There is analytic solution for the inflation that minimizes prior RMS.

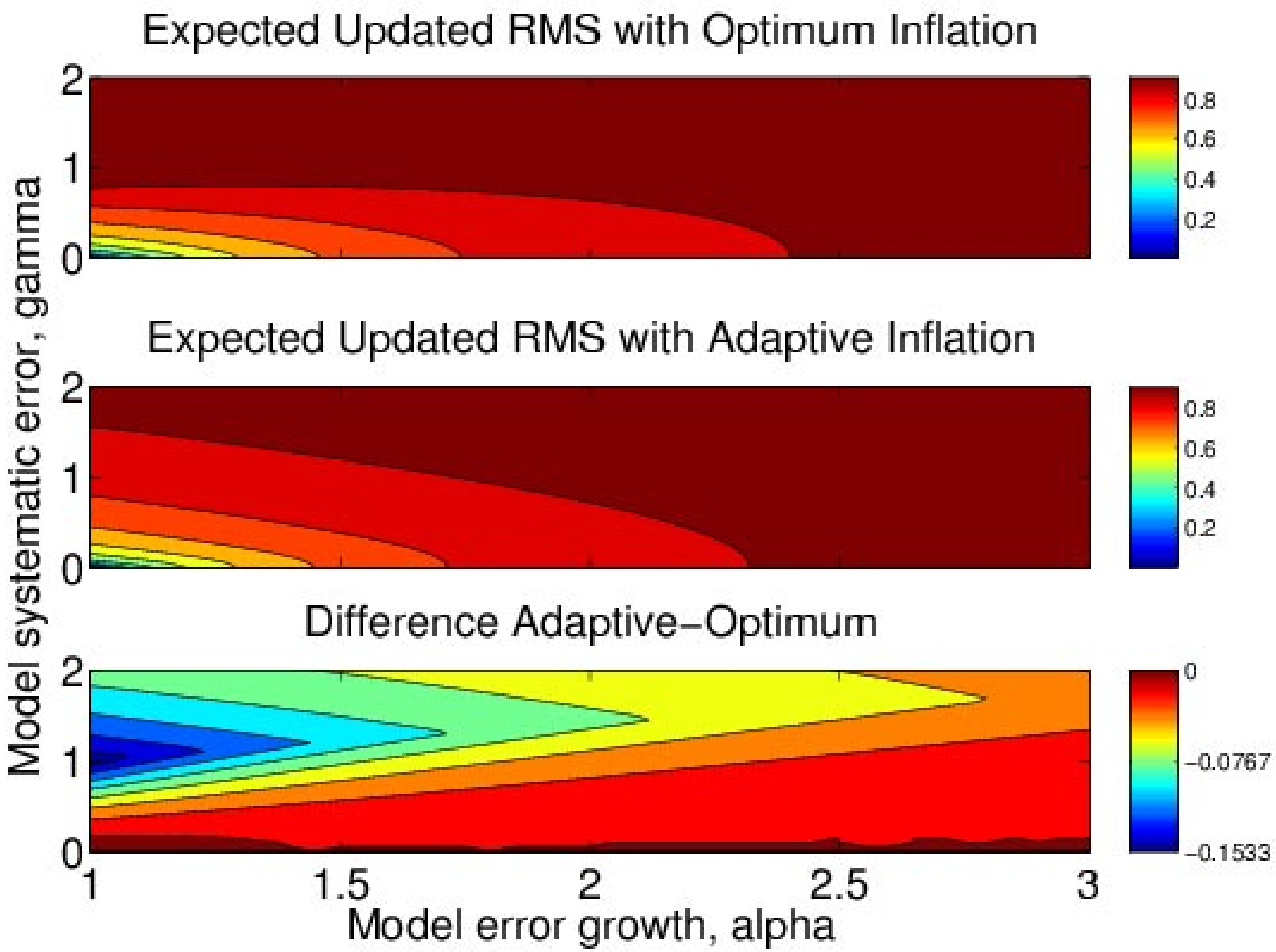
Adaptive is trying to make spread and RMS agree.



Prior Adaptive Inflation in Simplest Model with Model Error



Prior Adaptive Inflation in Simplest Model with Model Error



Adaptive Prior Works Marginally in Simplest Model.

Why do we need spatially-varying inflation?

1. Observation density can be spatially-varying.
   - S.H. vs. N.H. for global.
   - Near hurricane vortex vs. oceanic environment.
   - In squall line vs. non-convecting environment.

   Large enough inflation to do well in one may blow-up in the other.

2. Sampling error will be larger in densely observed regions.

3. Model error may also be spatially-varying.

Adaptive Inflation in Global NWP

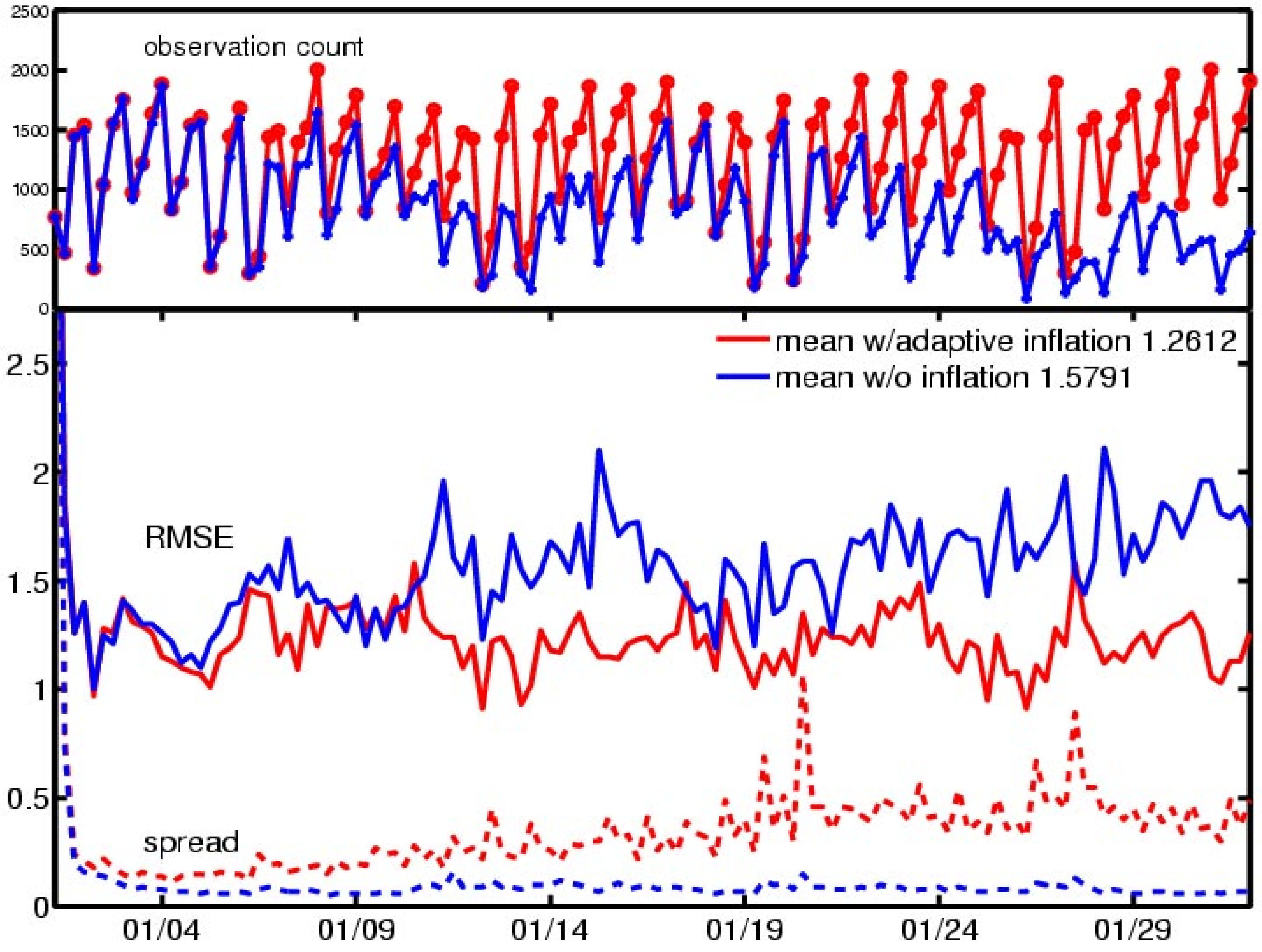
Model: CAM 3.1 T85L26.

Initialized from a climatological distribution (huge spread).

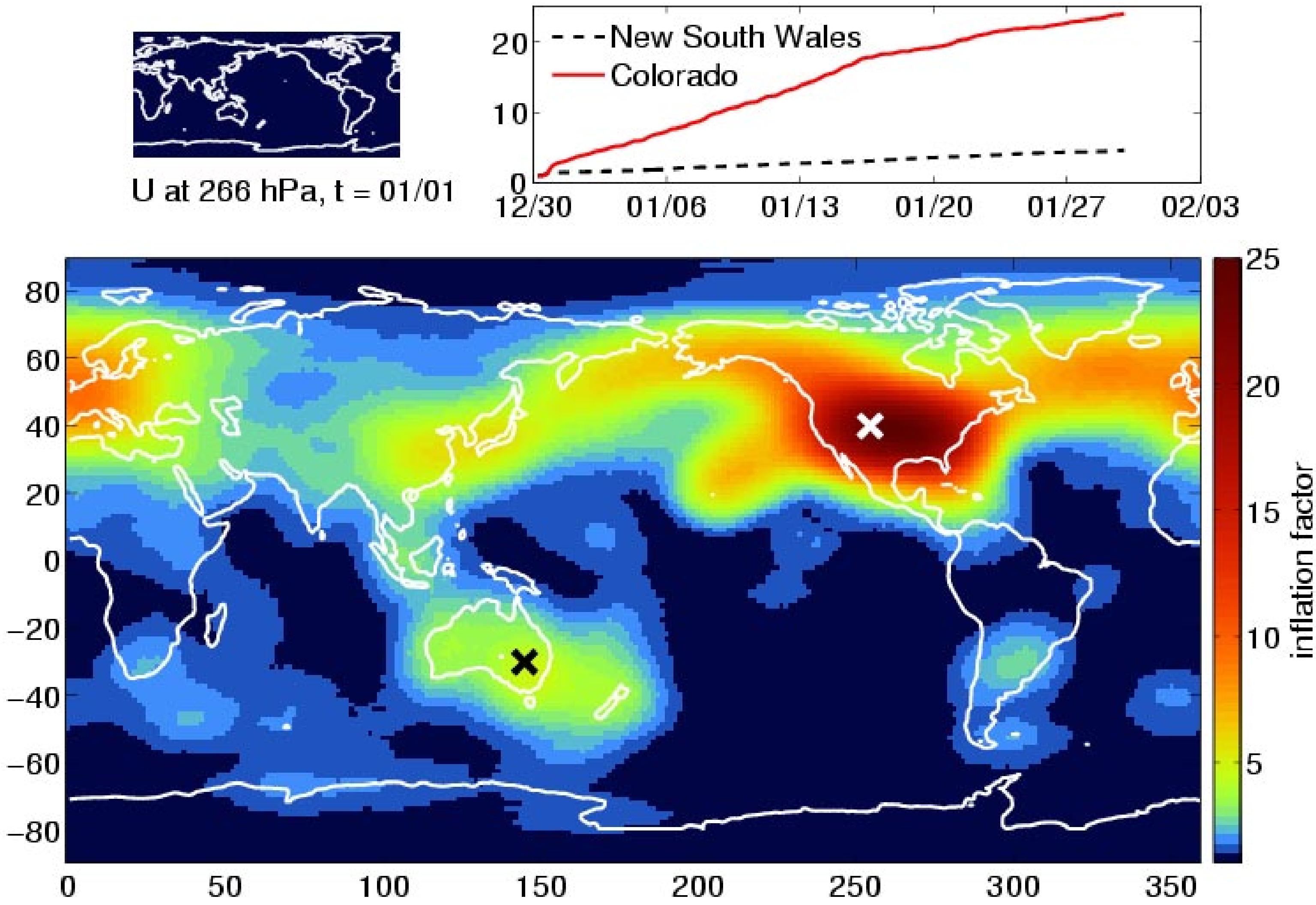
Observations: Radiosondes, ACARS, Satellite Winds.

Subset of observations used in NCAR/NCEP reanalysis.

Adaptive Inflation in CAM; 500 hPa T Obs. Space Prior RMS, Spread



Adaptive Inflation in CAM after 1-Month; 266 hPa U



Adaptive Inflation in CAM

1. Largest inflation caused by model error in densely observed regions.

2. RMS reduced, spread increased.

3. Fewer observations rejected.
    Note: really need to verify on set of trusted obs.

4. Inflation distribution has no time tendency.
    Can still cause blow-ups if inflation gets ‘orphaned’.

    Example: Inflation is larger near Antarctica in summer.
    When winter comes, obs. disappear.
    Inflation will just stay at previous values.

Really Need a Model of Time Evolution of Inflation Estimate

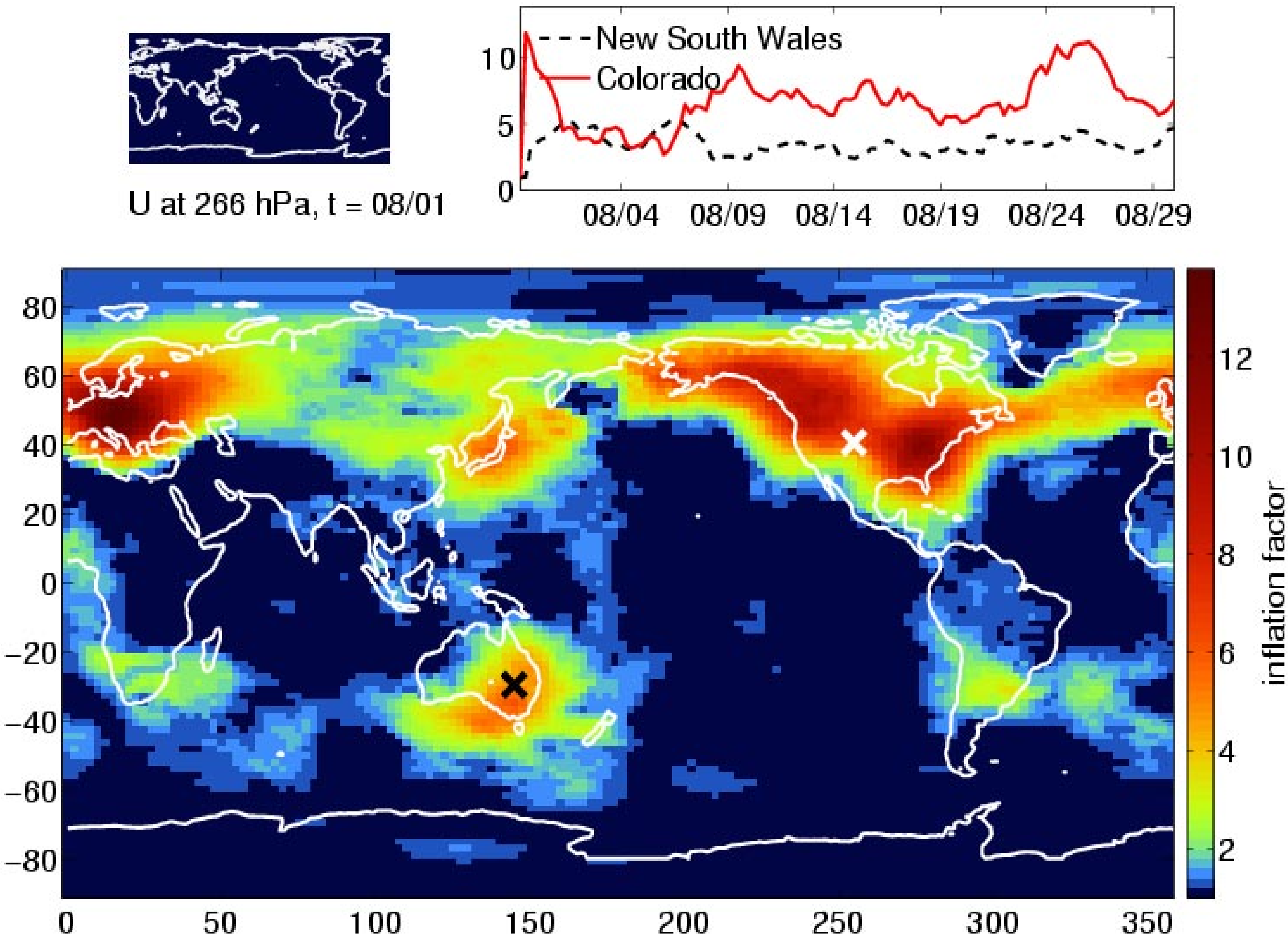
Simple and Naive:

1. Mean inflation value is damped towards 1 as function of time.

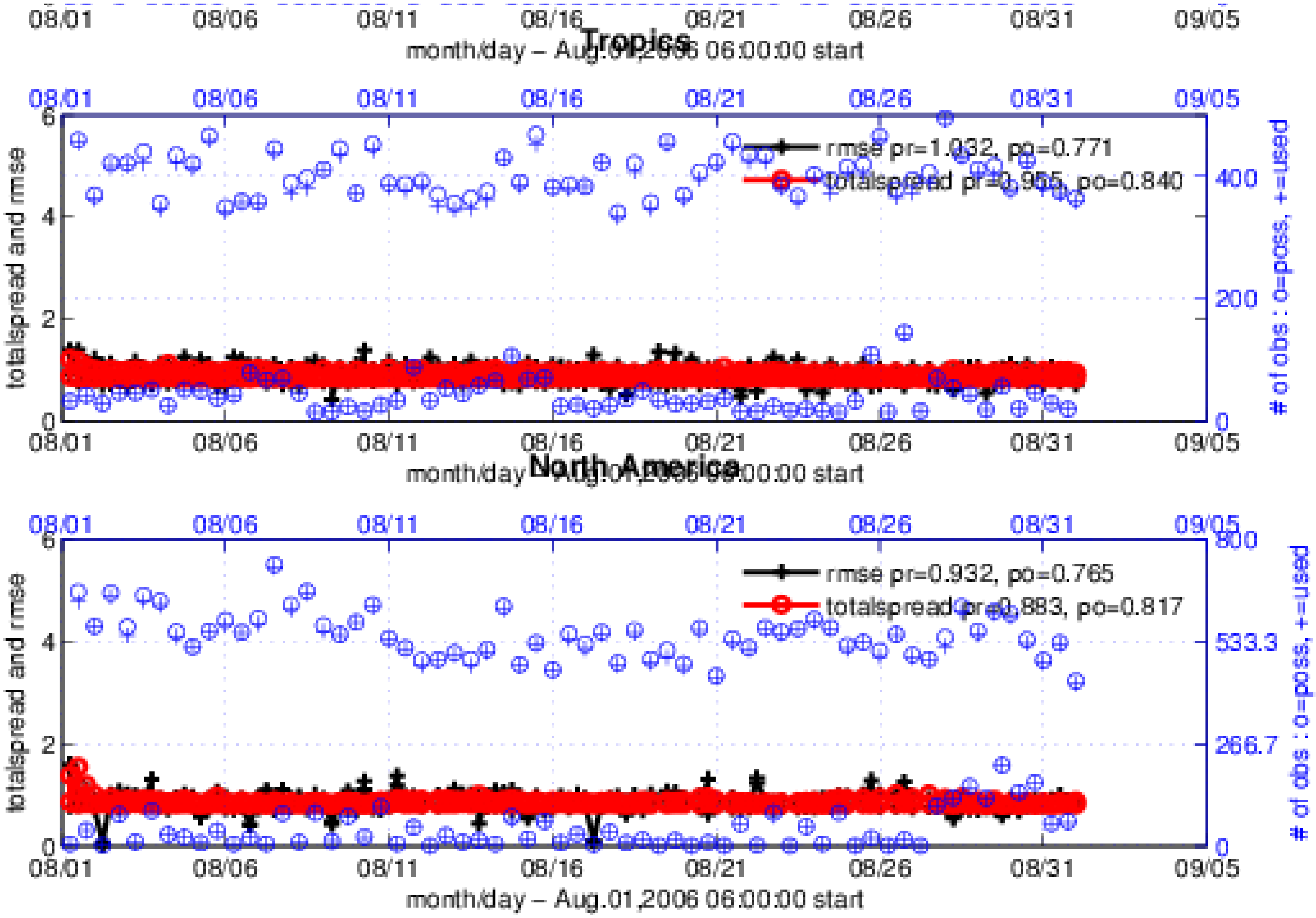
2. Inflation variance is damped towards some climatological value.

3. Example in CAM:
   Mean is damped 10% every 6 hours.
   Variance is held fixed at 0.6.

Damped Adaptive Inflation in CAM after 1-Month; 266 hPa U



Damped Adaptive Inflation in CAM: Obs. Space 500 hPa T



Damped Adaptive Inflation

1. Works well in CAM until...
   Reanalysis BUFR files had most obs. missing on 8 Jan., 2007...
   Then, it blew up.

2. Can this work with a hurricane or thunderstorm?
   Observations may be quite variable in time/space.
   Model error may also be quite variable.
   People have been trying this with some success.

General Thoughts

1. Don't expect ensembles to have consistent spread and RMS,
   Unless you want increased RMS error.
   Adaptive aims for consistency, so not optimal.
   Ensembles that look like random samples won't come from EnKF.

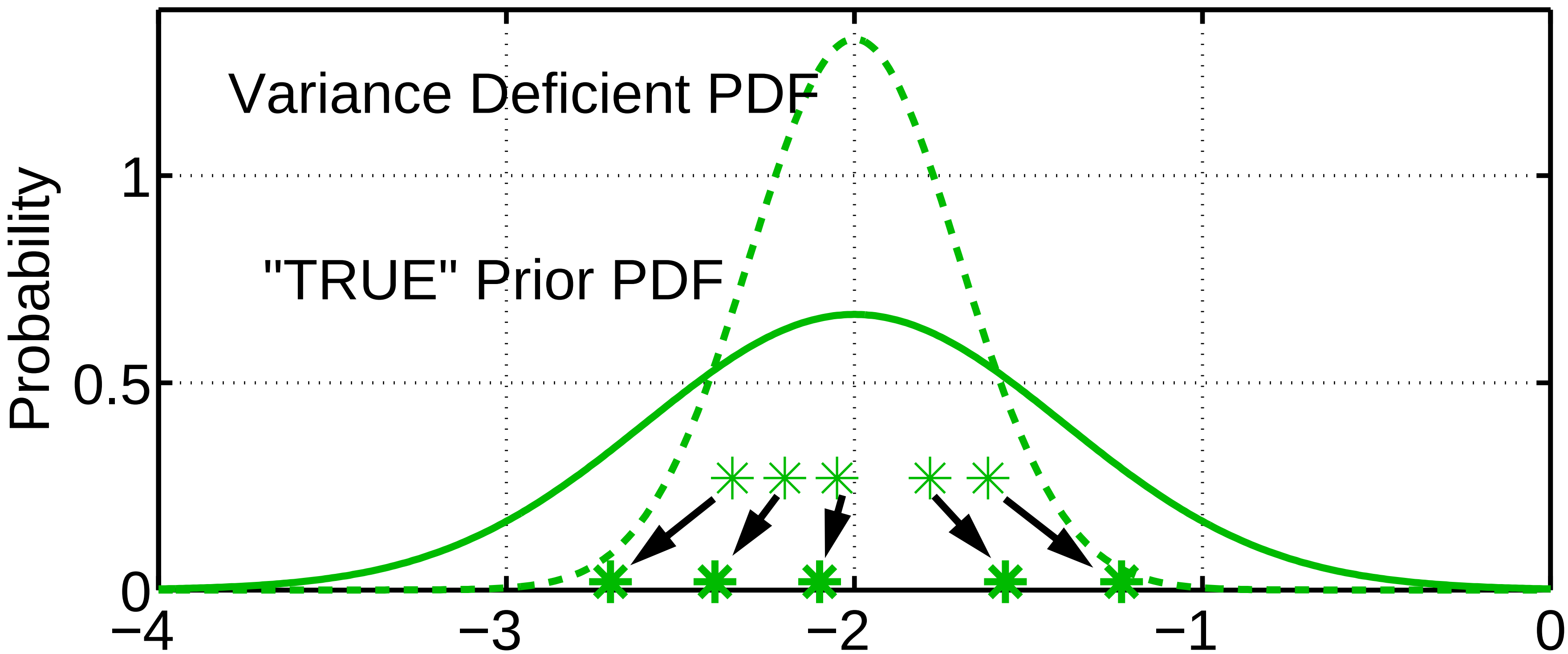
2. When model error dominates lack of spread,
   Need to change the general model of inflation.
   Work with model error and sampling error separately.
   For model error, model as additive in time?

3. Need trusted observation sets for comparing assimilations.

Everything shown here is available for play in DART:
   www.image.ucar.edu/DAReS/DART

Model/Filter Error; Filter Divergence and Variance Inflation

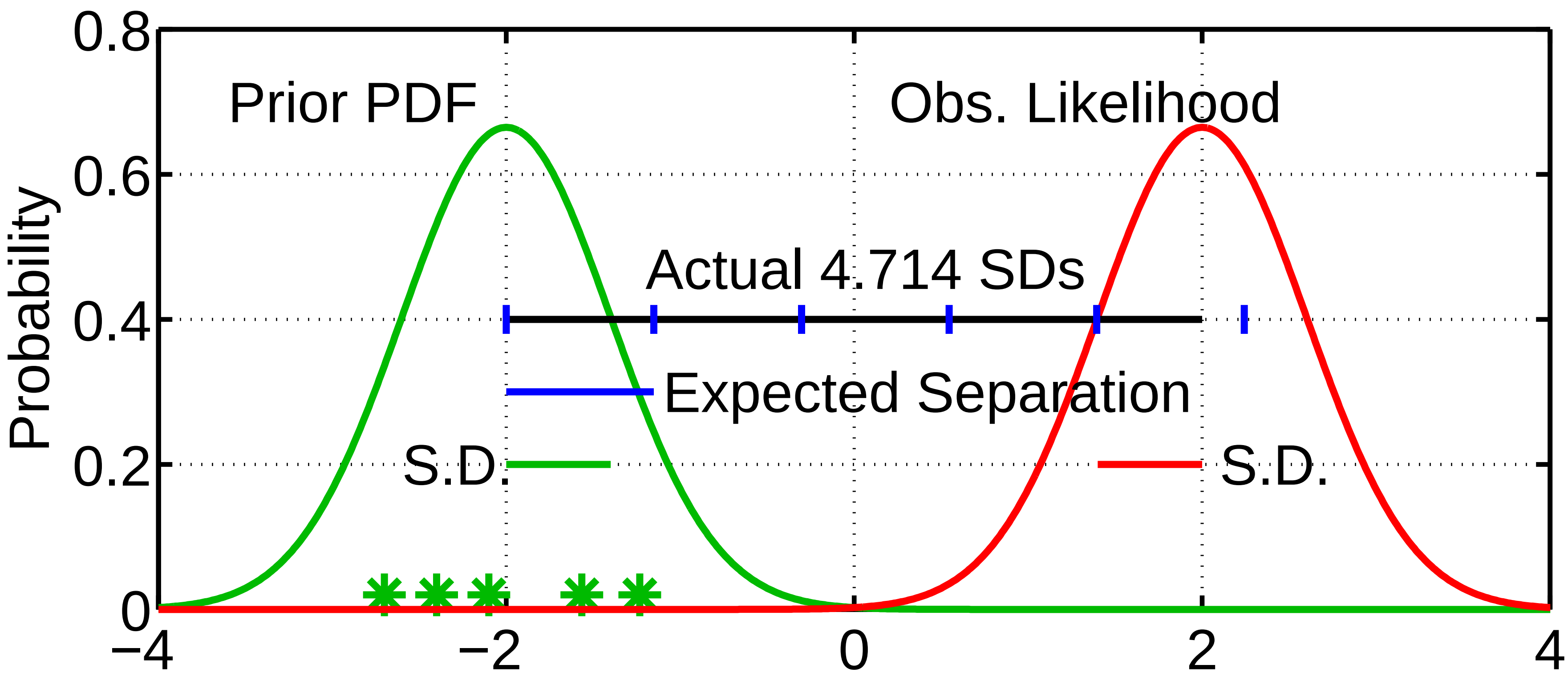
1. History of observations and physical system => ‘true’ distribution.
2. Sampling error, some model errors lead to insufficient prior variance.



3. Naive solution is Variance inflation: just increase spread of prior
4. For ensemble member i, $inflate(x_i) = \sqrt{\lambda}(x_i - \bar{x}) + \bar{x}$.

Adaptive Inflation for Ensemble Filtering

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



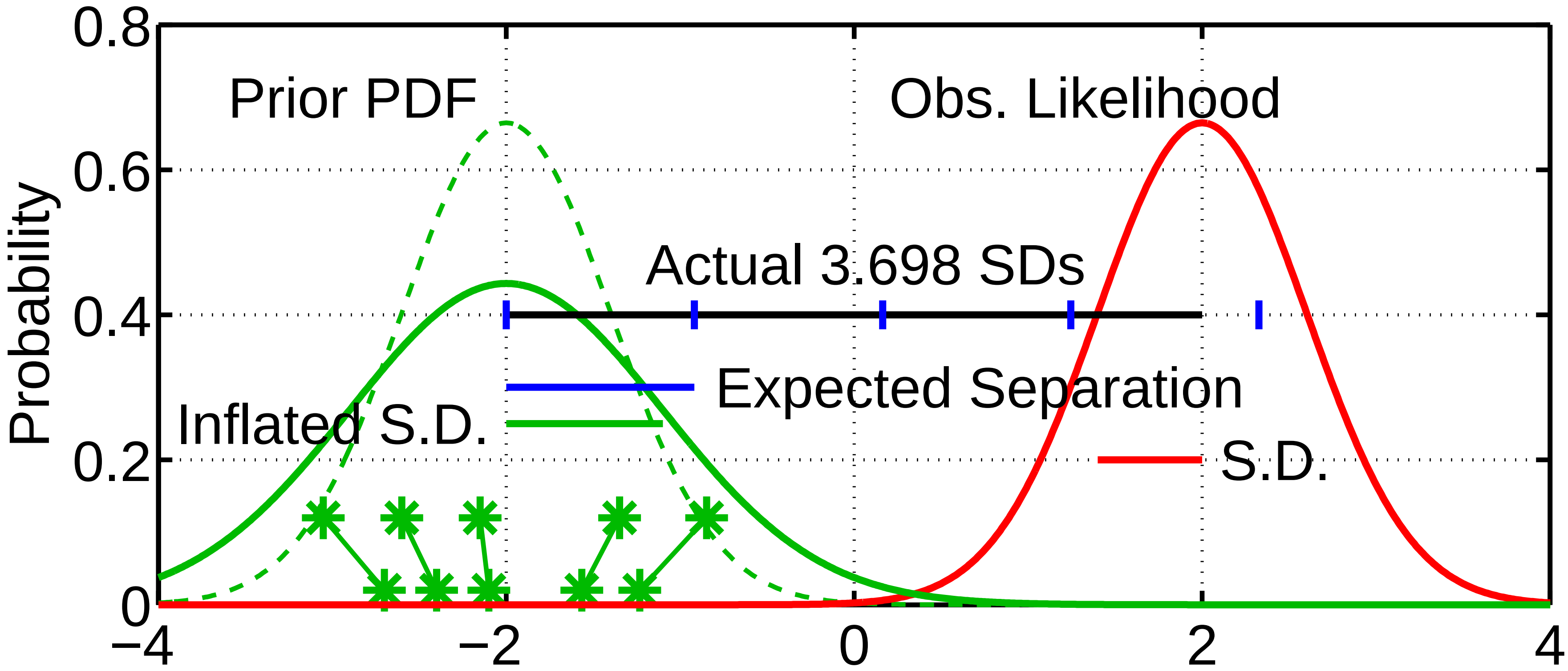
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

Assumes that prior and observation are supposed to be unbiased.

Is it model error or random chance?

Adaptive Inflation for Ensemble Filtering

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



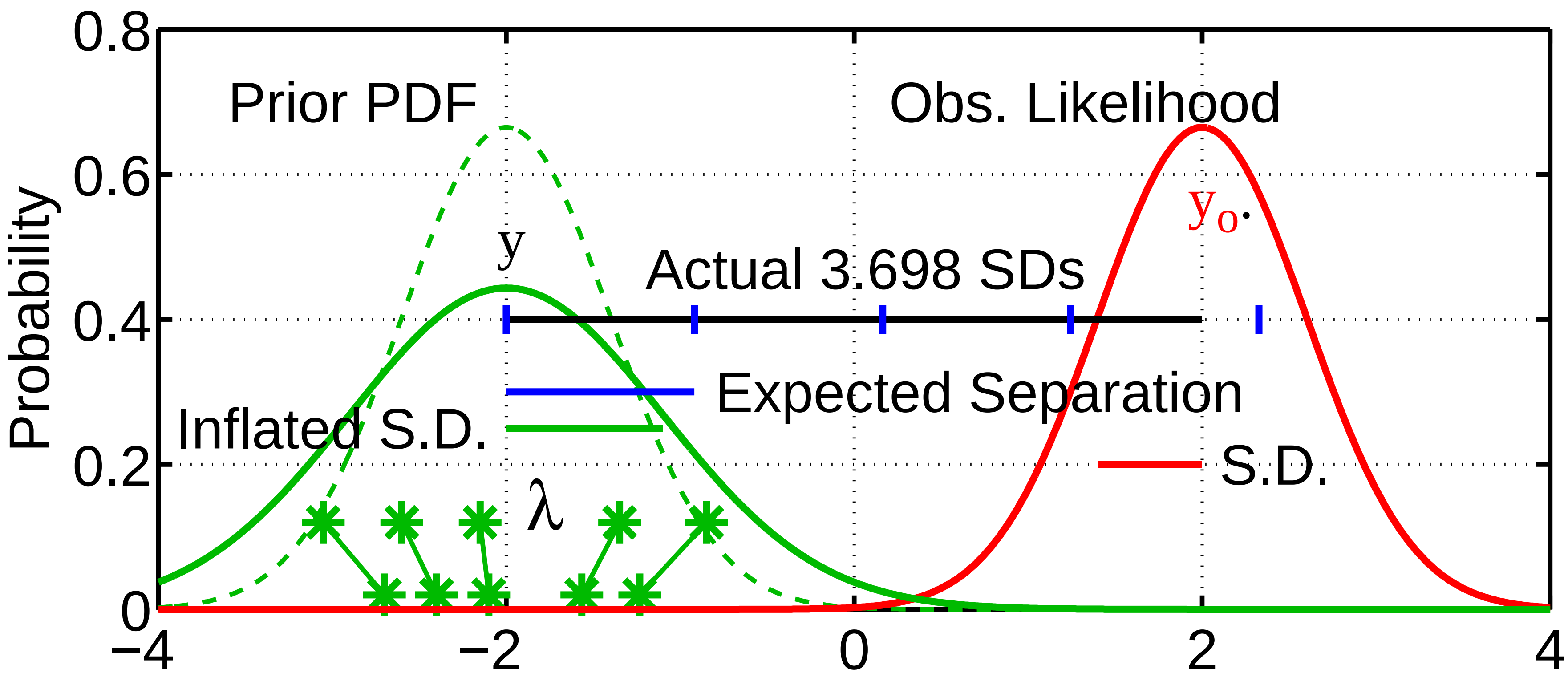
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

3. Inflating increases expected separation.

   Increases ‘apparent’ consistency between prior and observation.

Adaptive Inflation for Ensemble Filtering

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

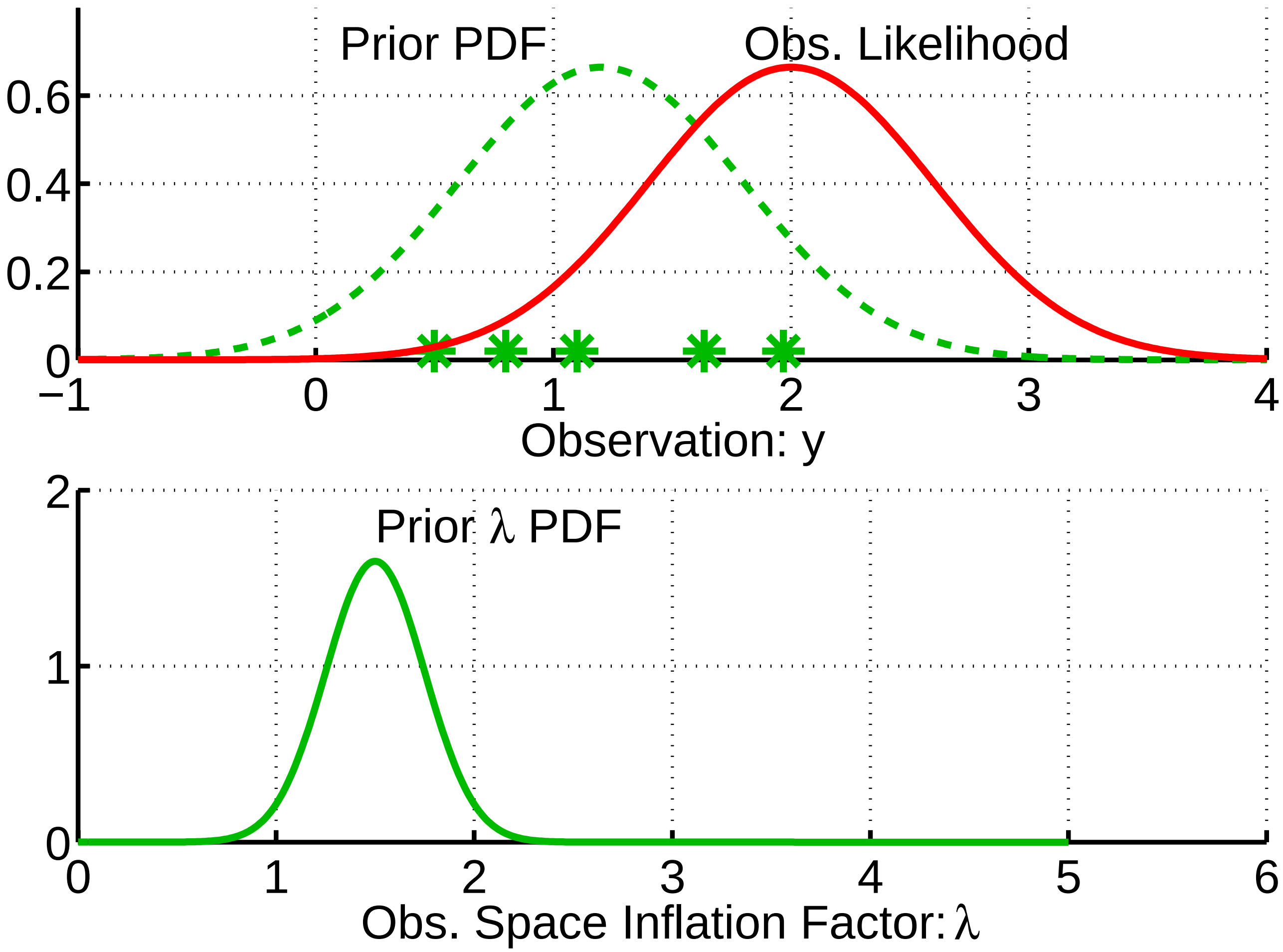


Distance, D, from prior mean y to obs. is $N\left(0, \sqrt{\lambda\sigma_{prior}^2 + \sigma_{obs}^2}\right) = N(0, \theta)$

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

Adaptive Inflation for Ensemble Filtering

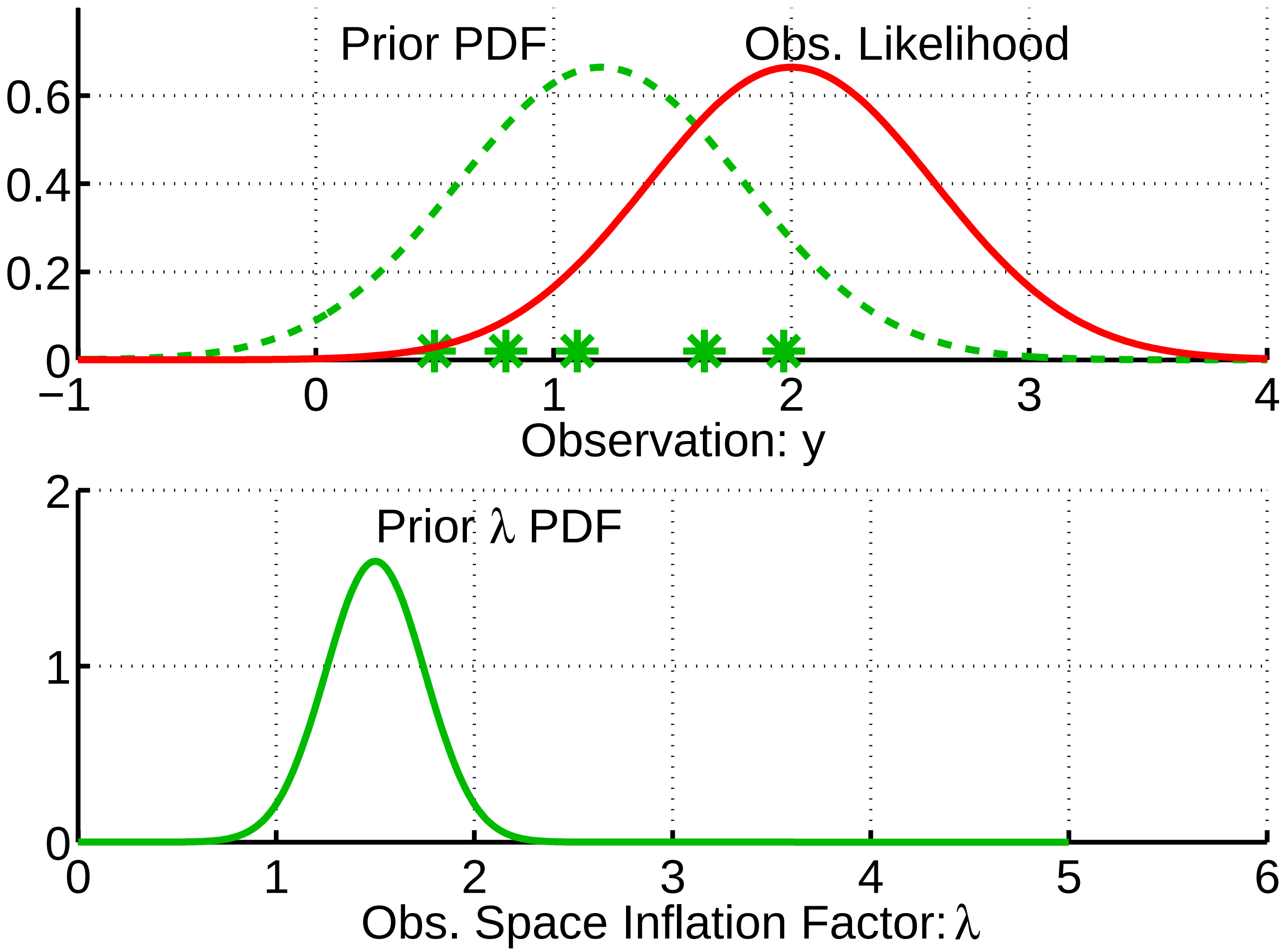
Use Bayesian statistics to get estimate of inflation factor, λ.



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

Adaptive Inflation for Ensemble Filtering

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



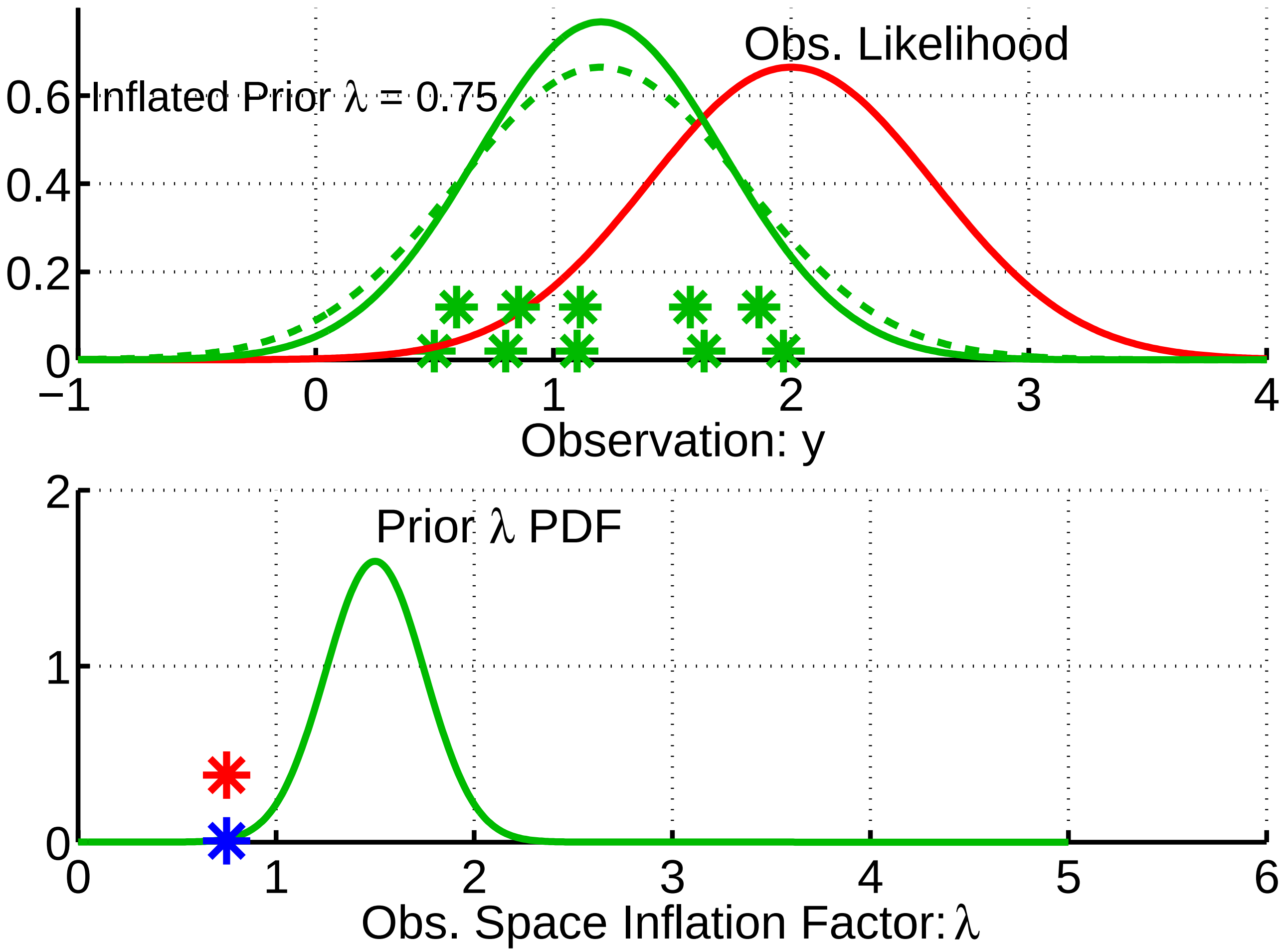
We've assumed a gaussian for prior $p(\lambda|Y_{prev})$.

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

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization.$$

Adaptive Inflation for Ensemble Filtering

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



Get $p(y_o|\lambda = 0.75)$ from normal PDF.

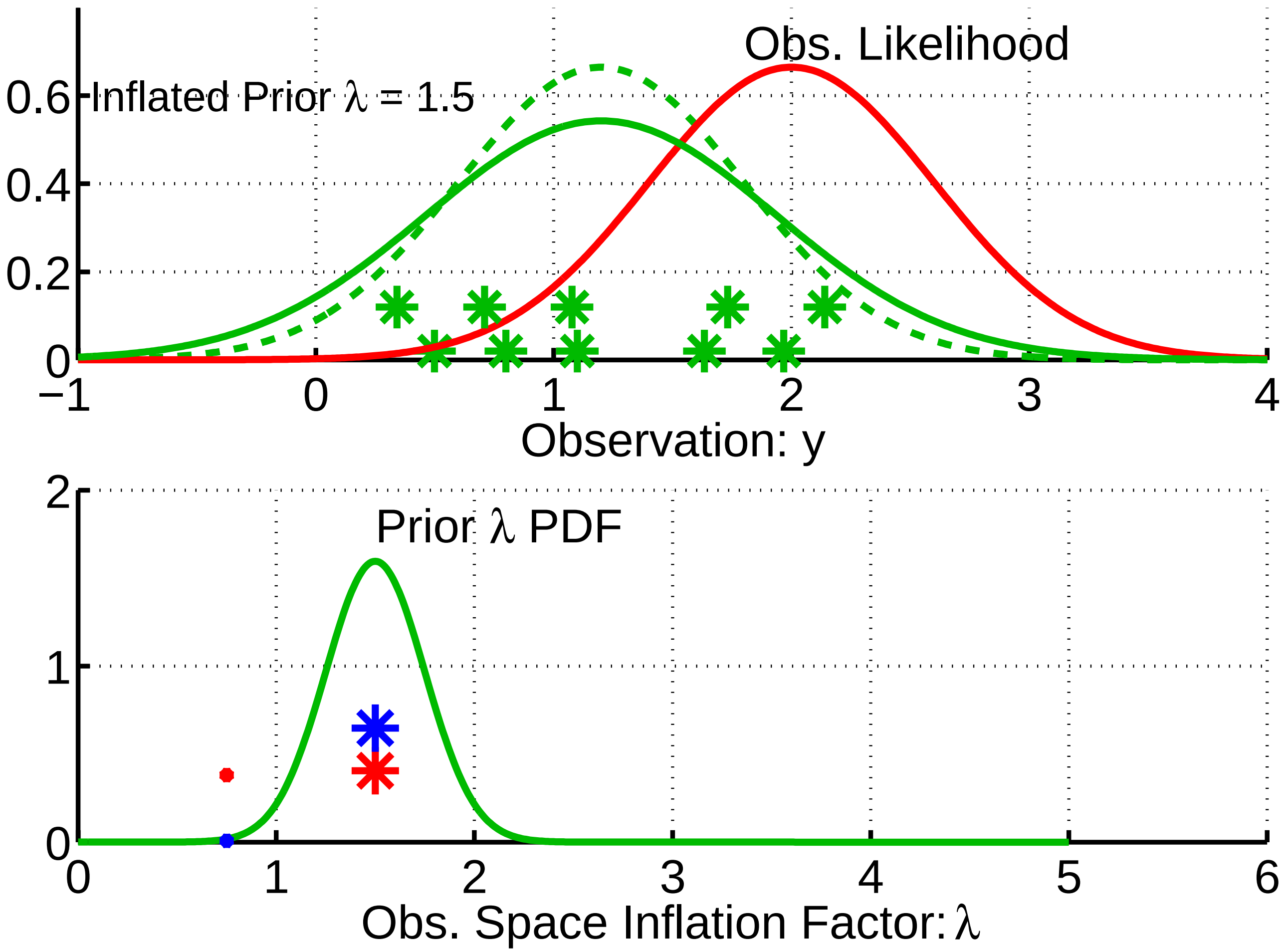
Multiply by $p(\lambda = 0.75|Y_{prev})$ to get

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

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization .$$

Adaptive Inflation for Ensemble Filtering

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



Get $p(y_o|\lambda = 1.5)$

from normal PDF.

Multiply by

$p(\lambda = 1.5|Y_{prev})$

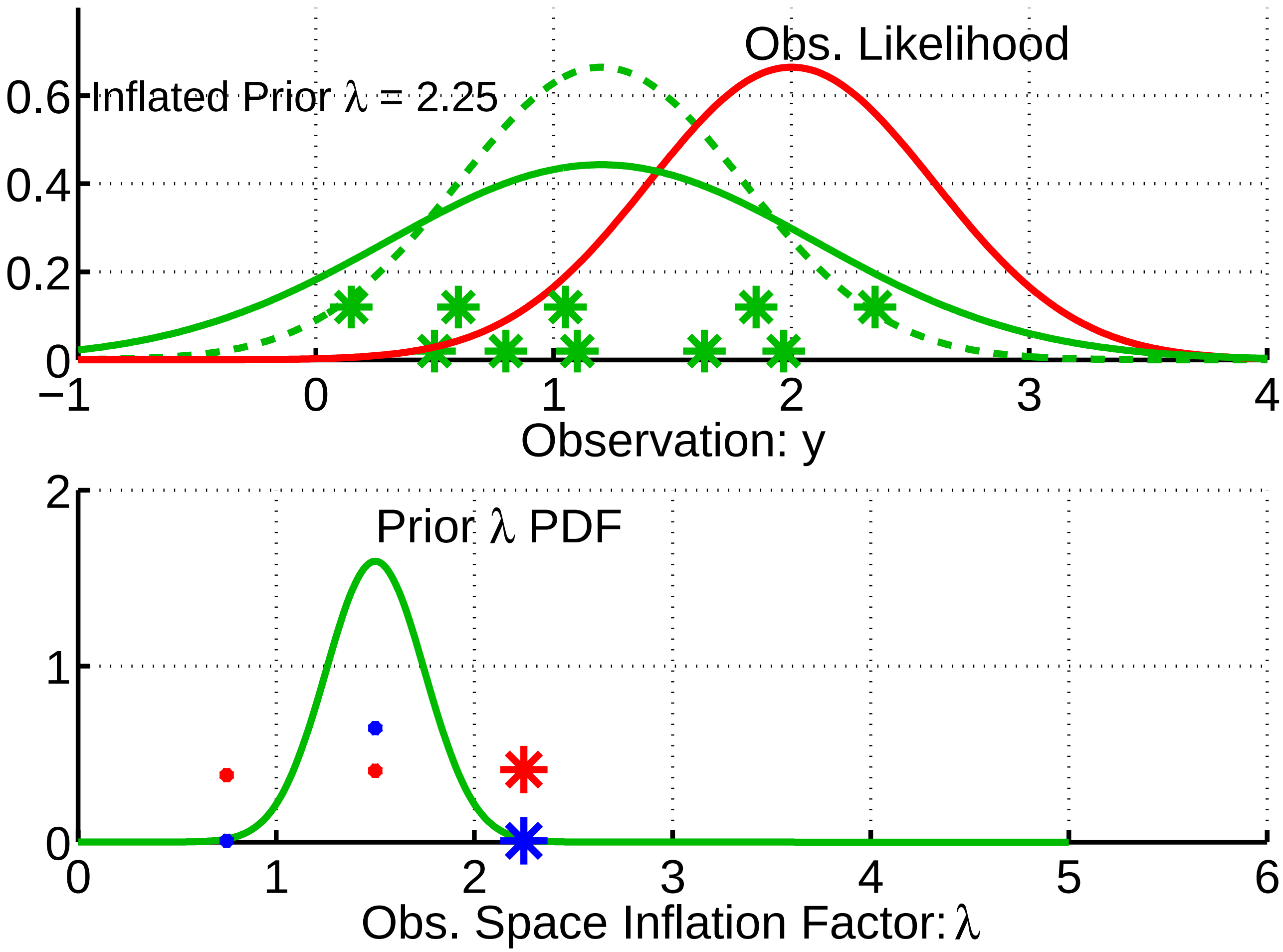
to get

$p(\lambda = 1.5|Y_{prev}, y_o)$

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization .$$

Adaptive Inflation for Ensemble Filtering

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



Get $p(y_o|\lambda = 2.2)$ from normal PDF.

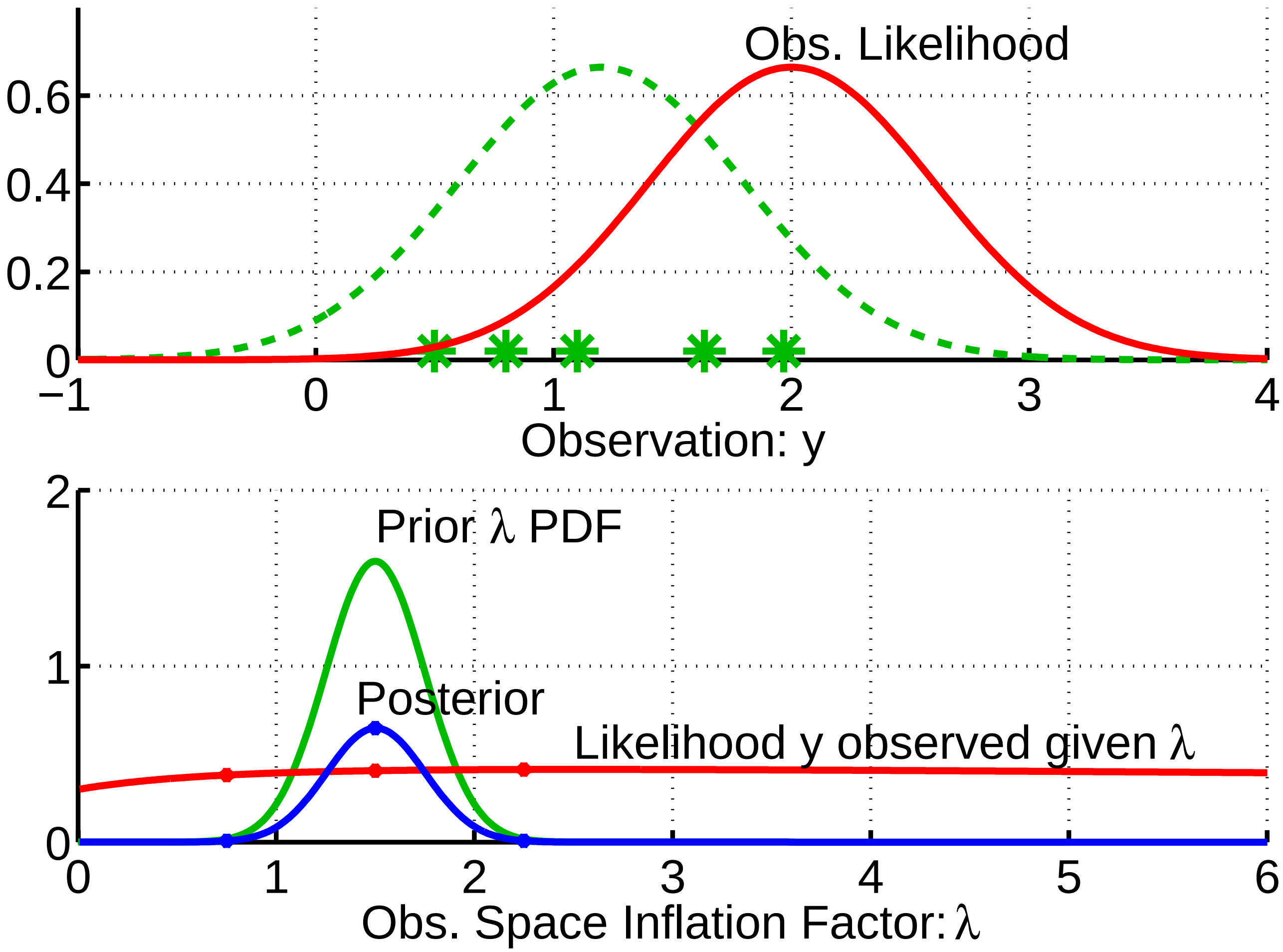
Multiply by $p(\lambda = 2.2|Y_{prev})$

to get $p(\lambda = 2.2|Y_{prev}, y_o)$

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization.$$

Adaptive Inflation for Ensemble Filtering

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



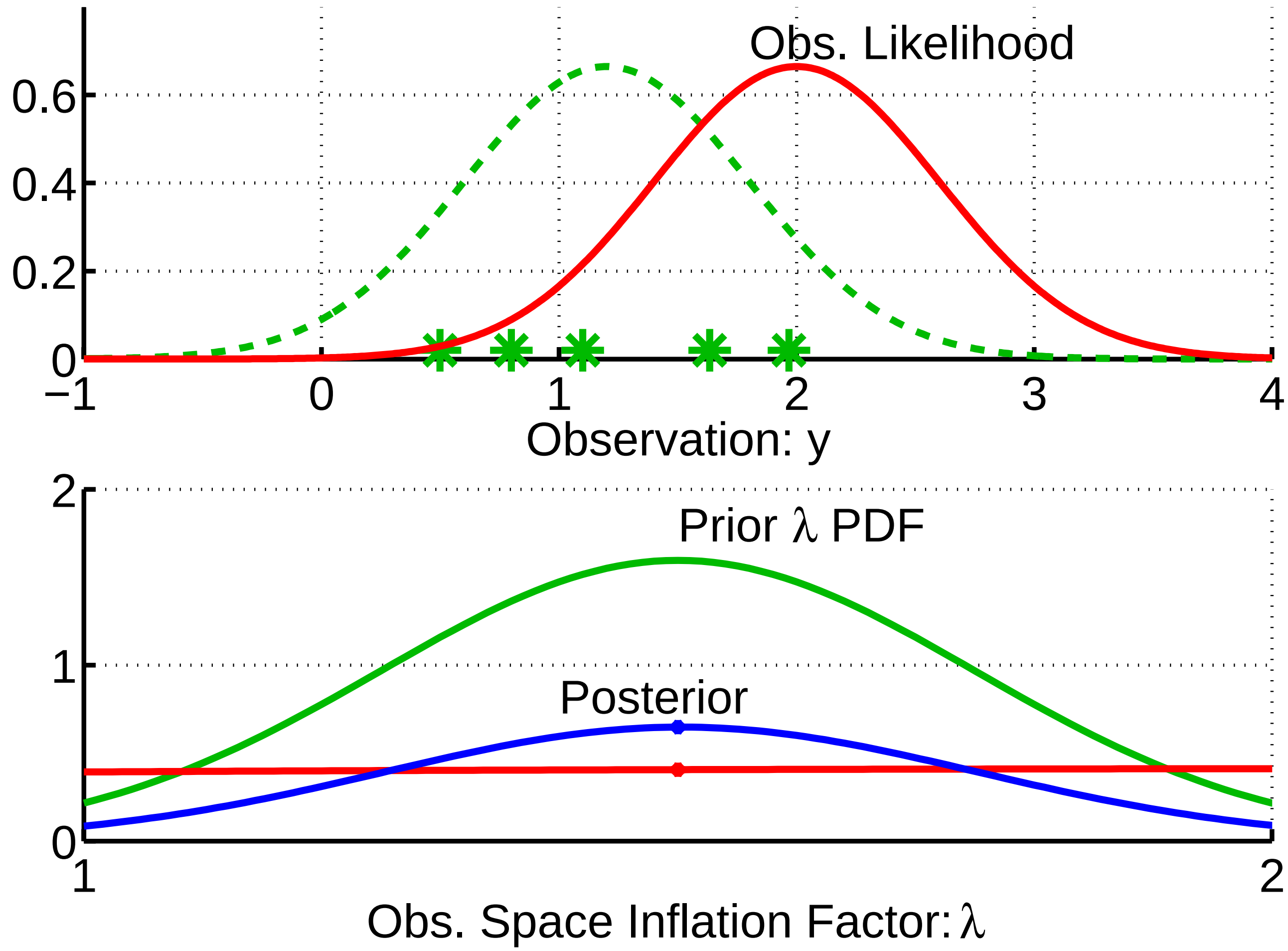
Repeat for a range of values of λ.

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

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization\ .$$

Adaptive Inflation for Ensemble Filtering

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



Very little information about λ in a single observation.

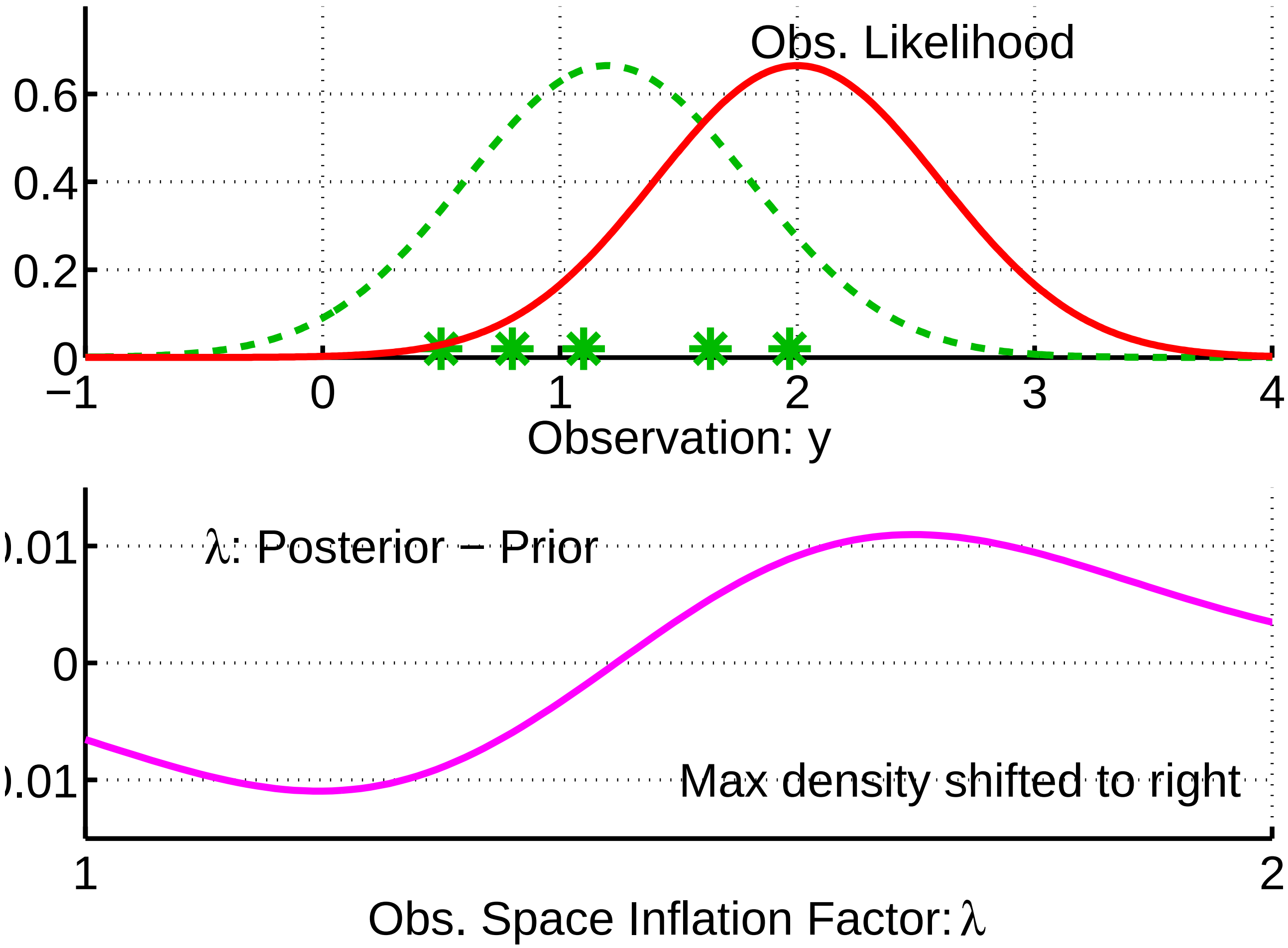
Posterior and prior are very similar.

Normalized posterior indistinguishable from prior.

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization\ .$$

Adaptive Inflation for Ensemble Filtering

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



Very little information about λ in a single observation.

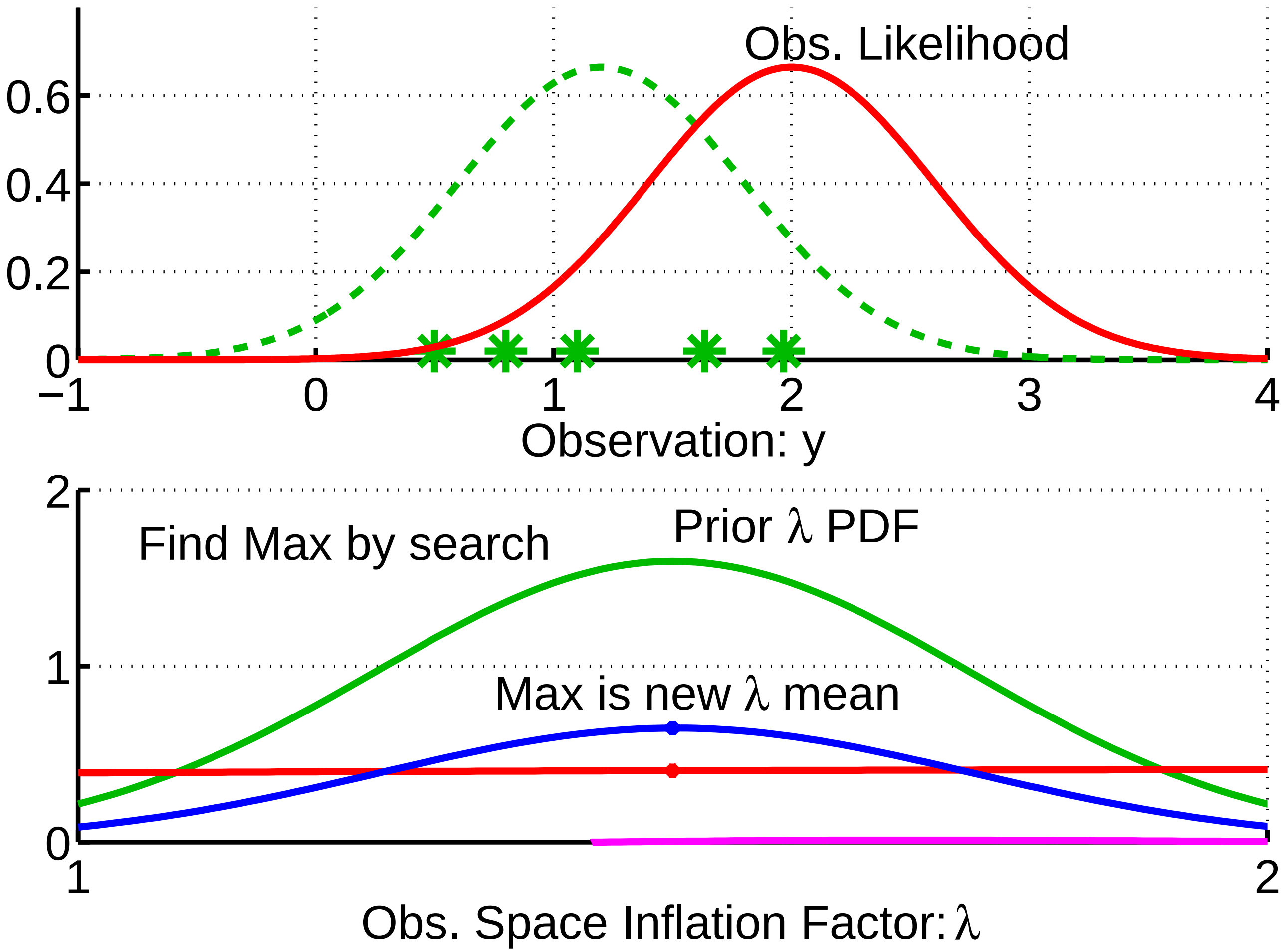
Posterior and prior are very similar.

Difference shows slight shift to larger values of λ.

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization\ .$$

Adaptive Inflation for Ensemble Filtering

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.

$$p(\lambda|Y_{prev}, y_o) = p(y_o|\lambda)p(\lambda|Y_{prev})/normalization\ .$$

Adaptive Inflation for Ensemble Filtering

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

Mode of $p(y_o|\lambda)p(\lambda|Y_{prev})$ can be found analytically!

Solving $\partial[p(y_o|\lambda)p(\lambda|Y_{prev})]/\partial\lambda = 0$ 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.

Adaptive Inflation for Ensemble Filtering

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 $p(\bar{\lambda}_u + \sigma_{\lambda,p})$.

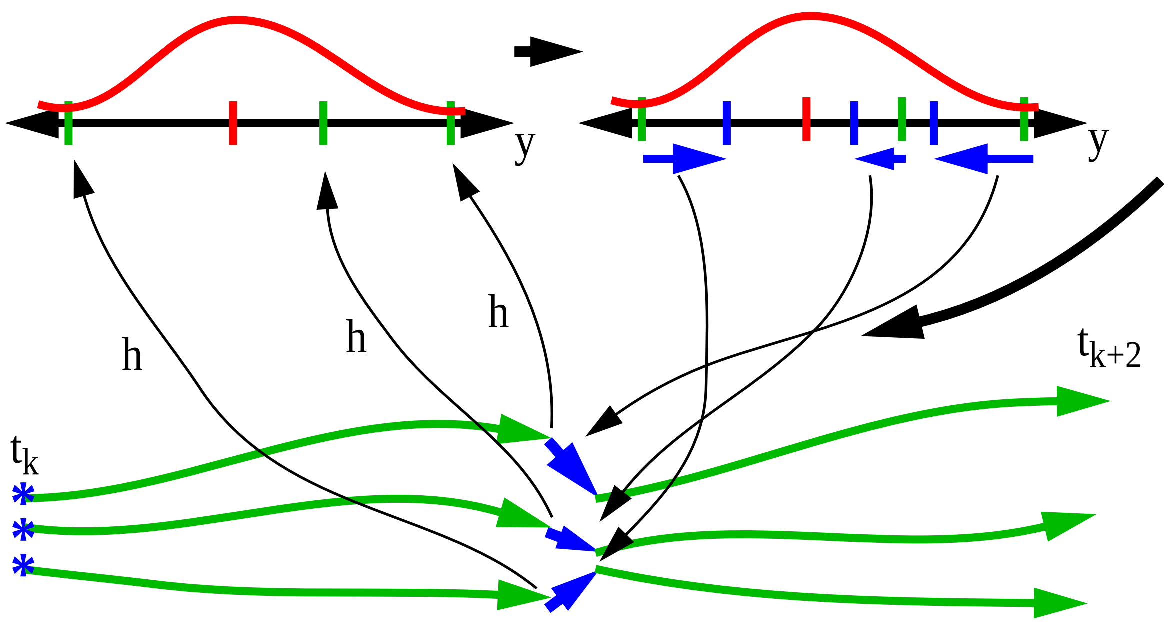
3. Compute as $\sigma^2_{\lambda,u} = -\sigma^2_{\lambda,p}/2\, ln r$ where $r = p(\bar{\lambda}_u + \sigma_{\lambda,p})/p(\bar{\lambda}_u)$.

State Space Adaptive Inflation

Computations so far adapt inflation for observation space.

What is relation between observation and state space inflation?

Have to use prior ensemble observation/state joint distribution.



Regress changes in inflation onto state variable inflation.

Spatially varying adaptive inflation algorithm:

Have a distribution for λ at each time 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

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

Equation for finding mode of posterior is now full 12th order:

Analytic solution appears unlikely.

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!

Hierarchical Bayesian Methods for Adaptive Filters: Summary

1. Localization:
   Run an ensemble of ensembles.
   Use regression coefficient signal-to-noise ratio to minimize error.

2. Inflation:
   Use each observation twice.
      Once to adjust parameter (inflation) of filter system.
      Second time to adjust mean and variance of estimate.