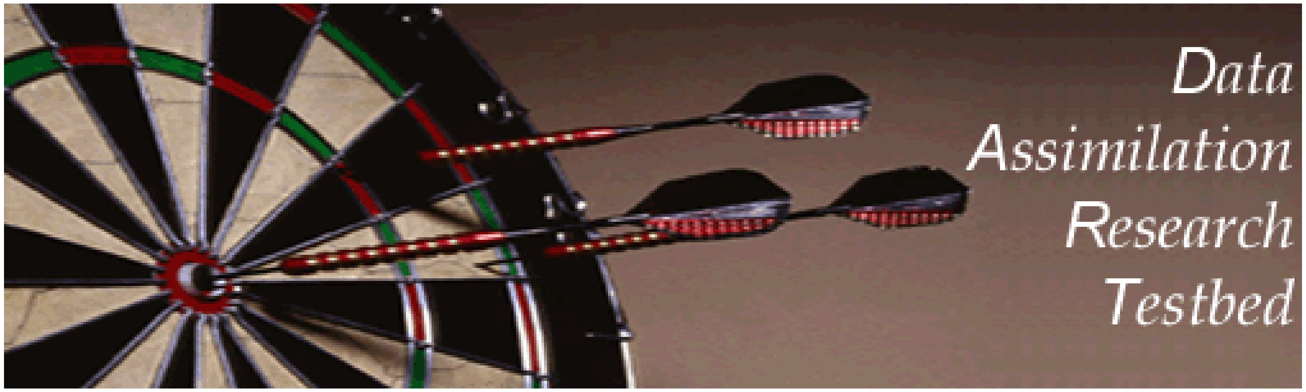
Data Assimilation Research Testbed Tutorial

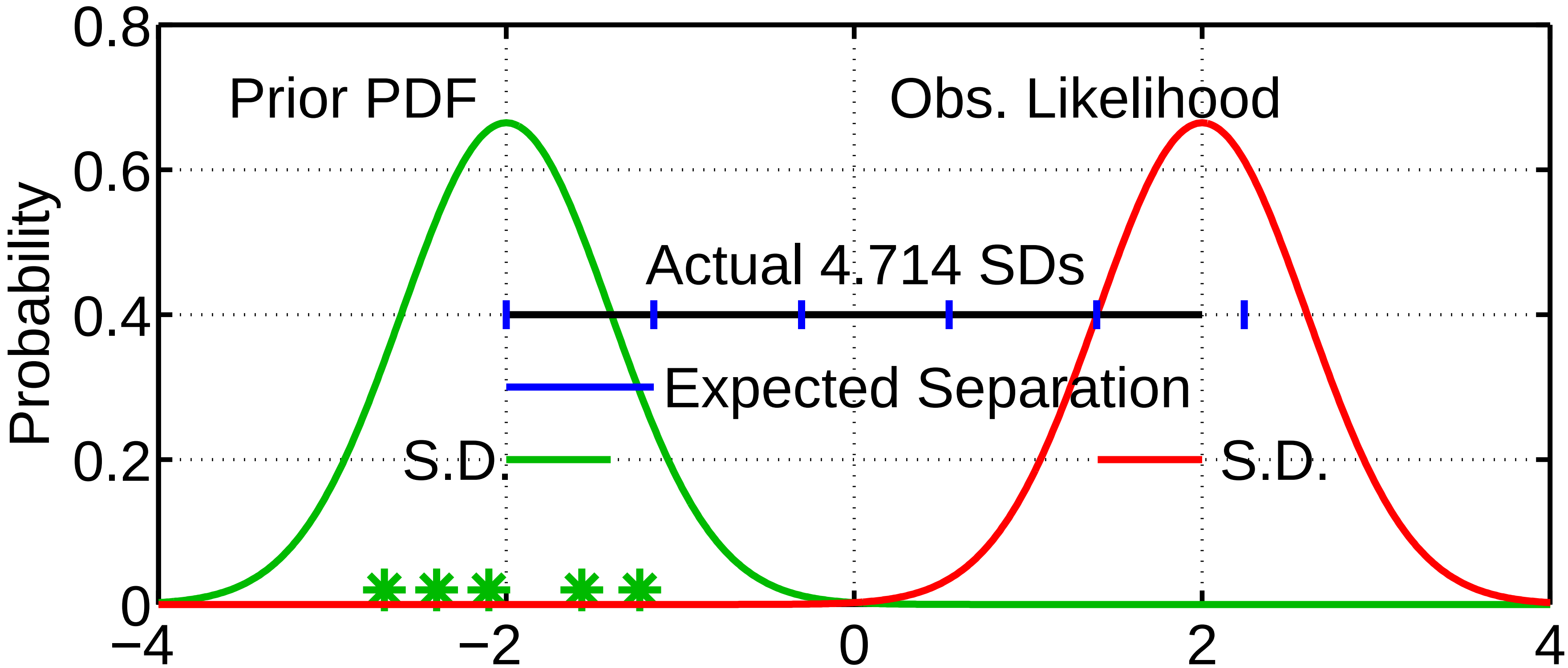


Section 12: Adaptive Inflation

Version 2.0: September, 2006

Variance inflation for Observations: An Adaptive Error Tolerant Filter

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



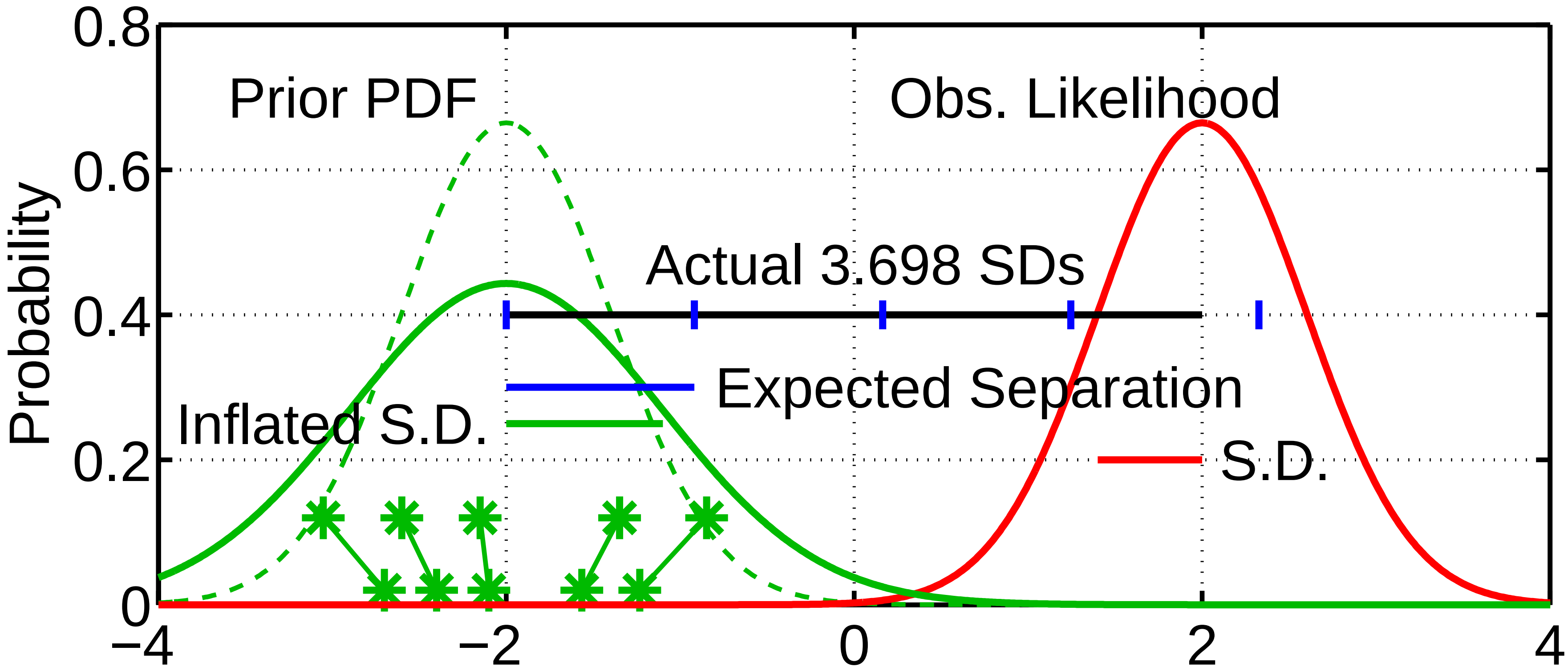
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

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



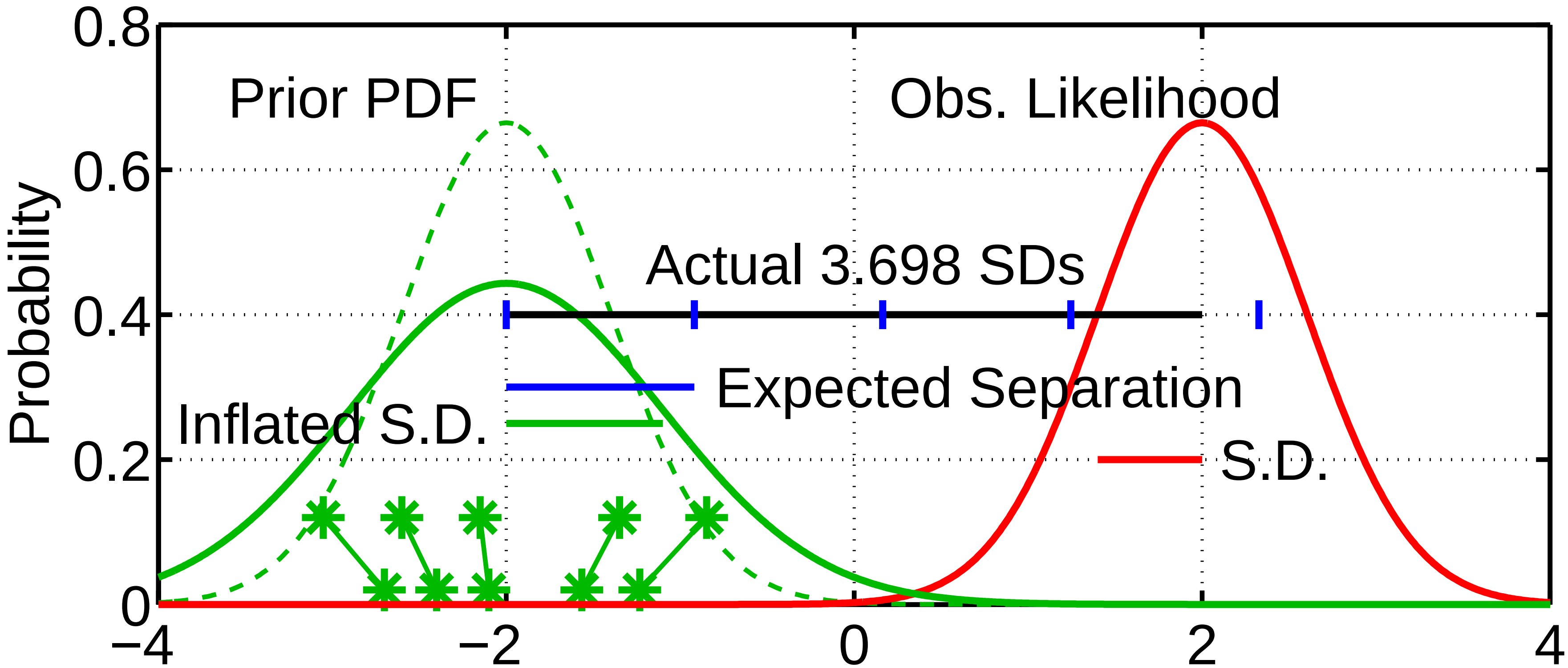
2. Expected(prior mean - observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$.

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

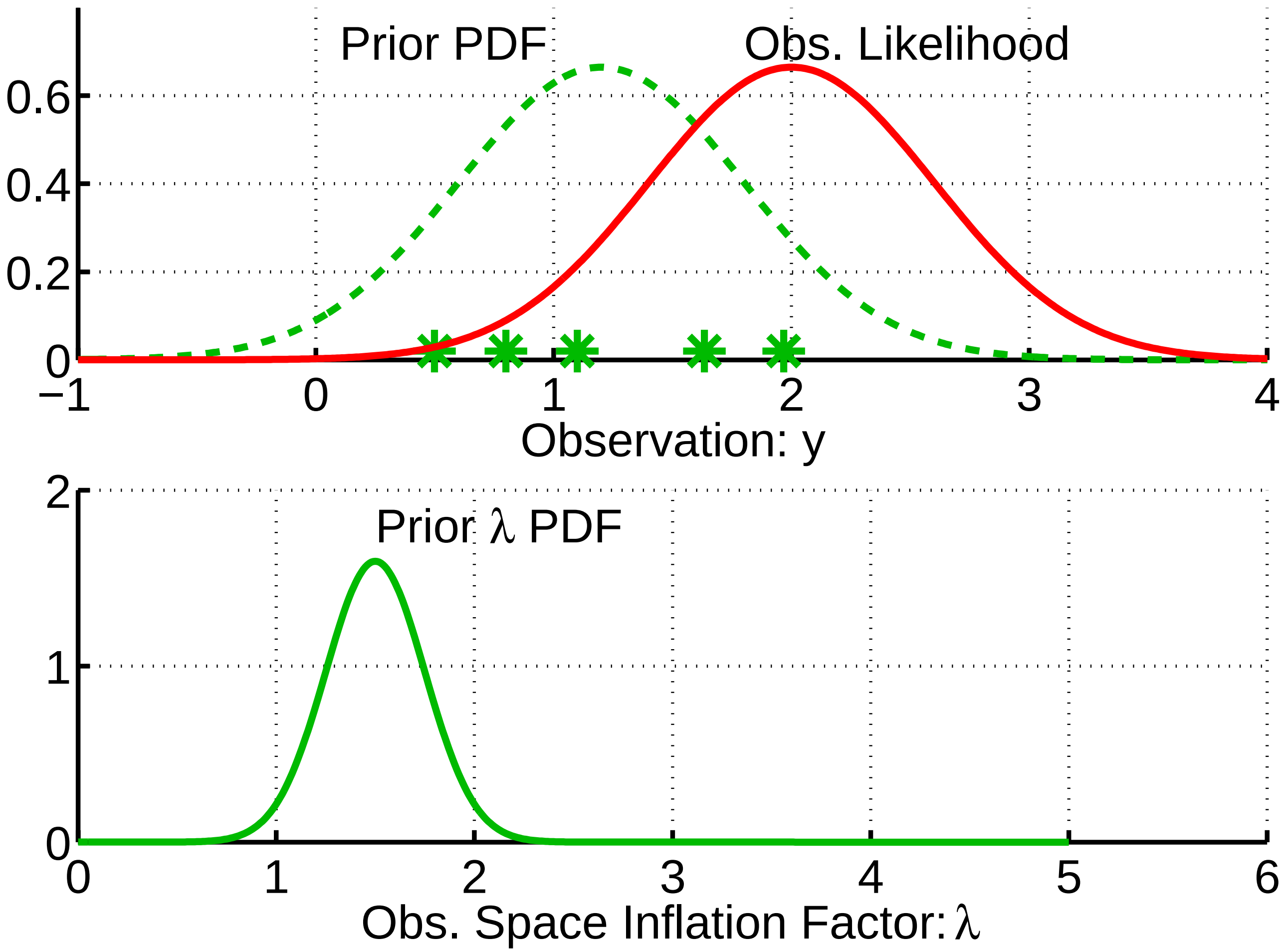


Distance, D, from prior mean y to obs. is $N\left(0, \sqrt{\lambda\sigma_{prior}^2 + \sigma_{obs}^2}\right) = N(0, \theta)$

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

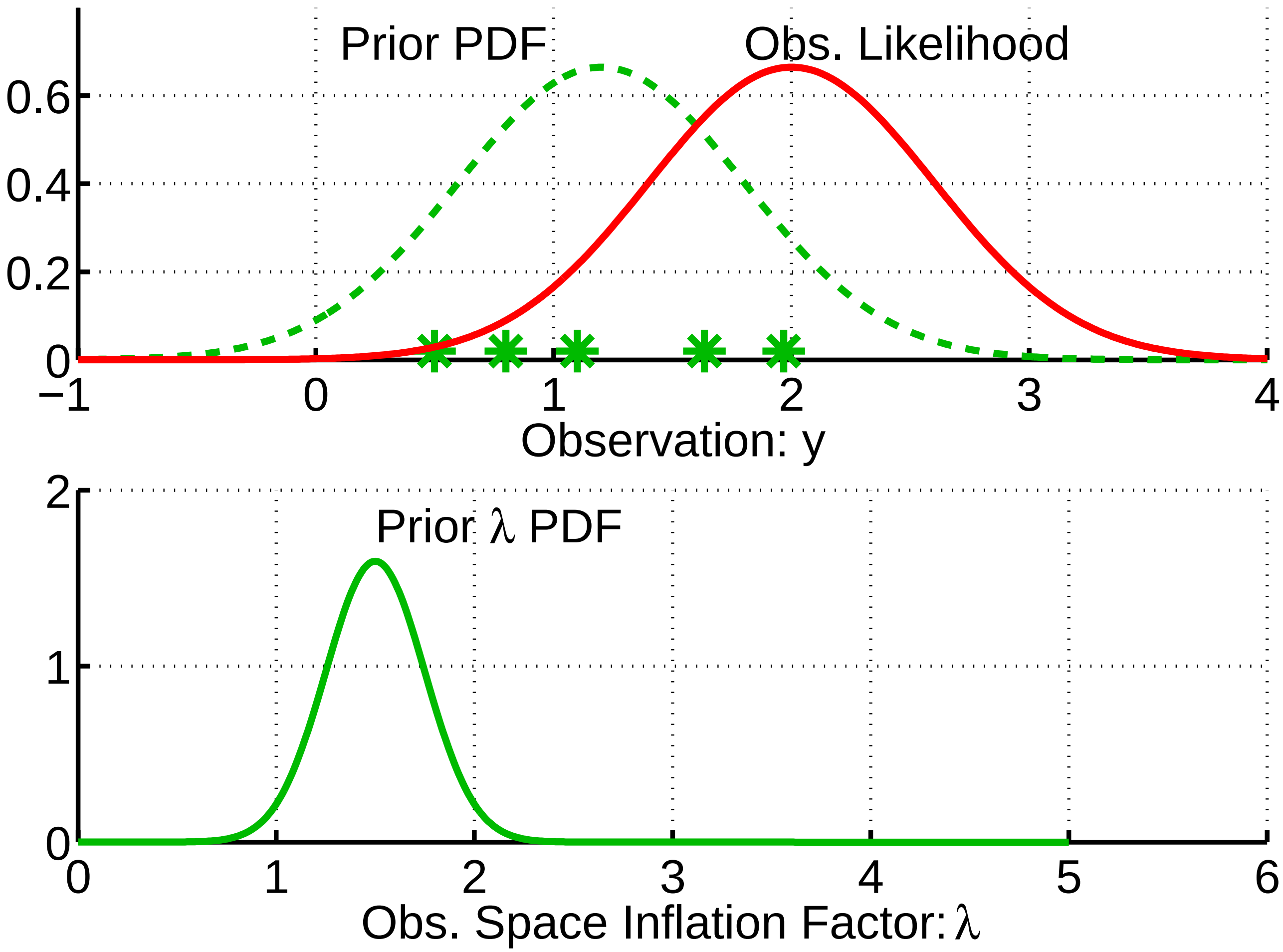
Use Bayesian statistics to get estimate of inflation factor, λ.



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.



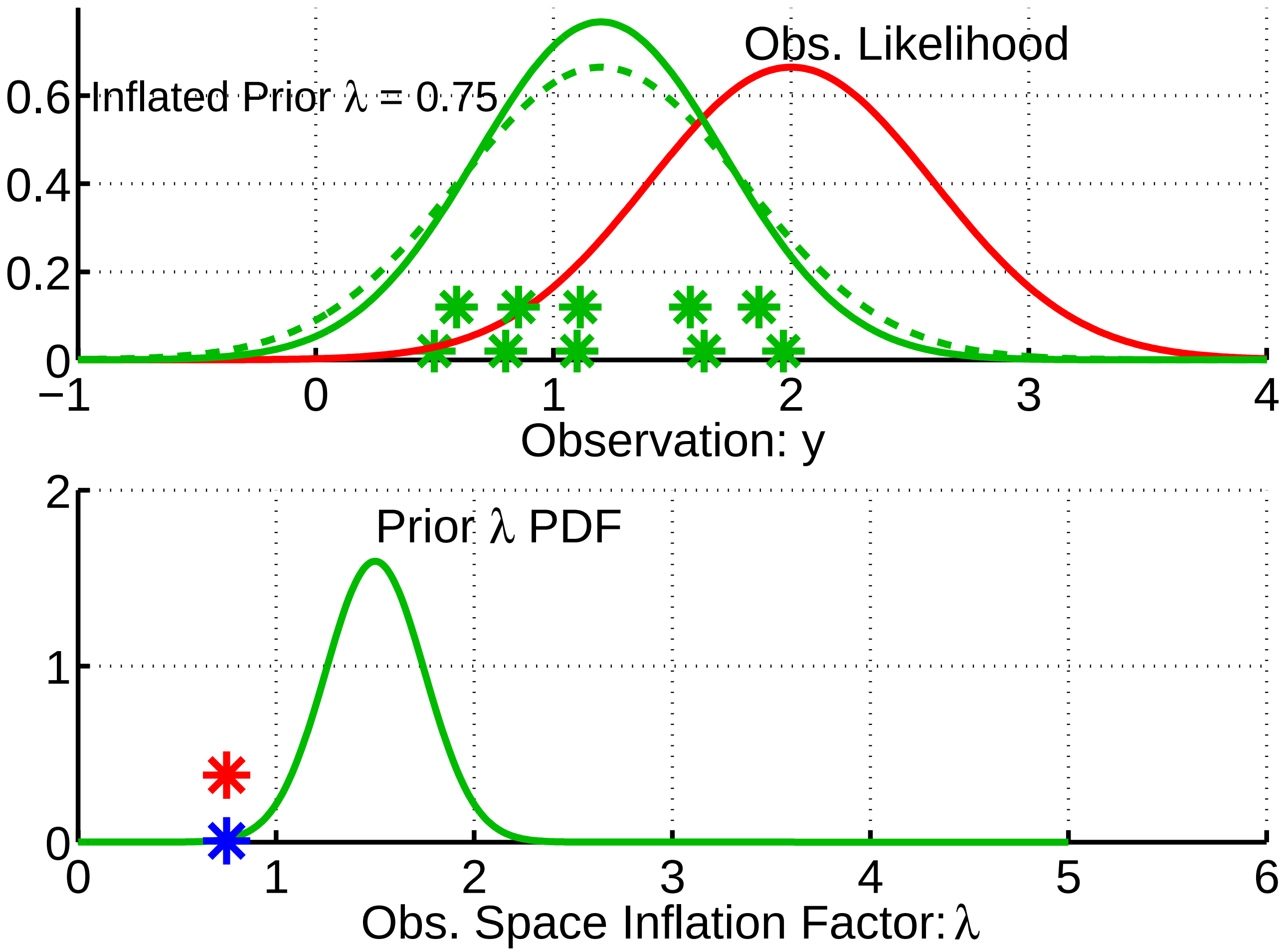
We've assumed a gaussian for prior $p(\lambda, t_k | Y_{t_{k-1}})$.

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

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

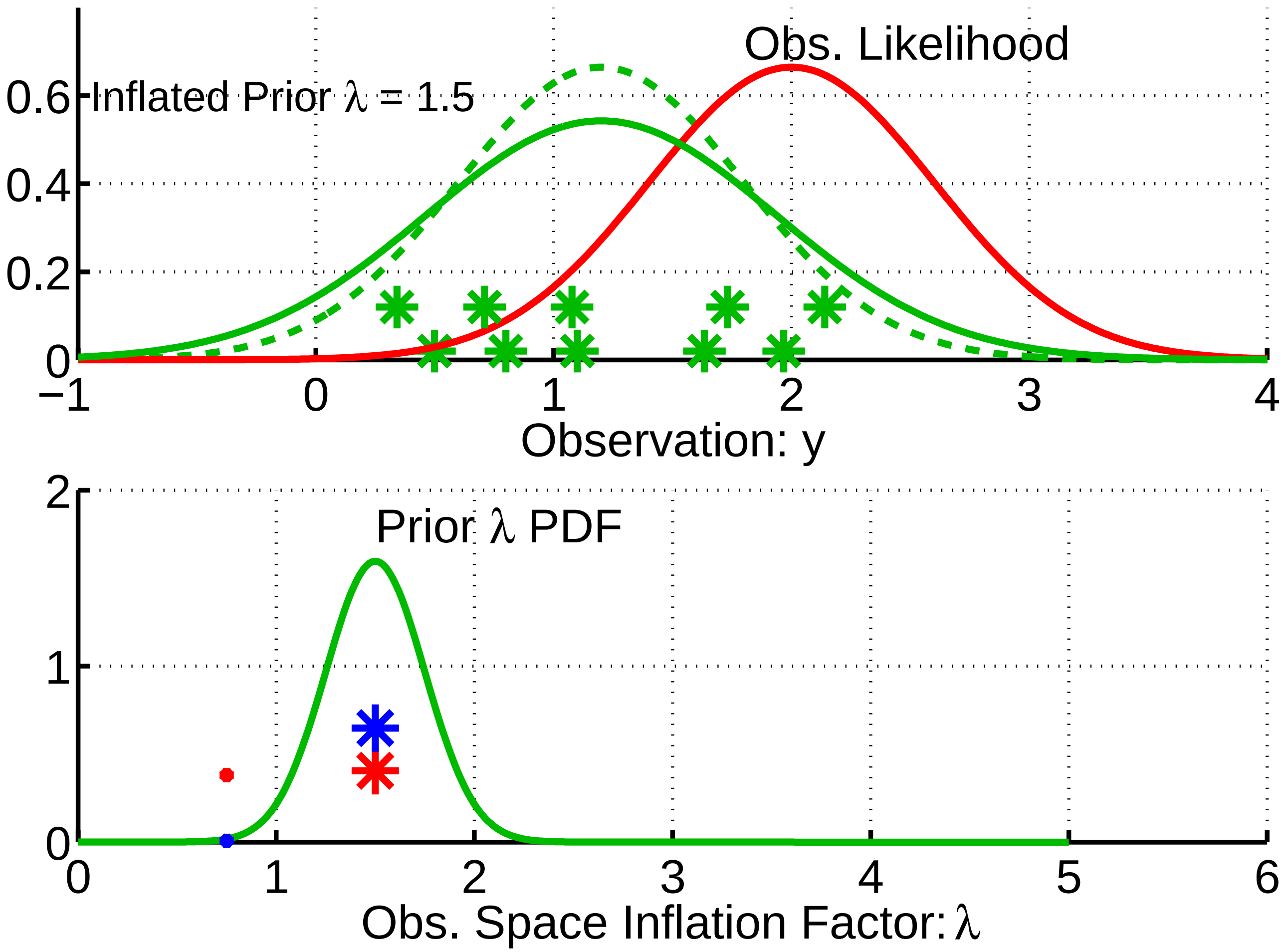
to get

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

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

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

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

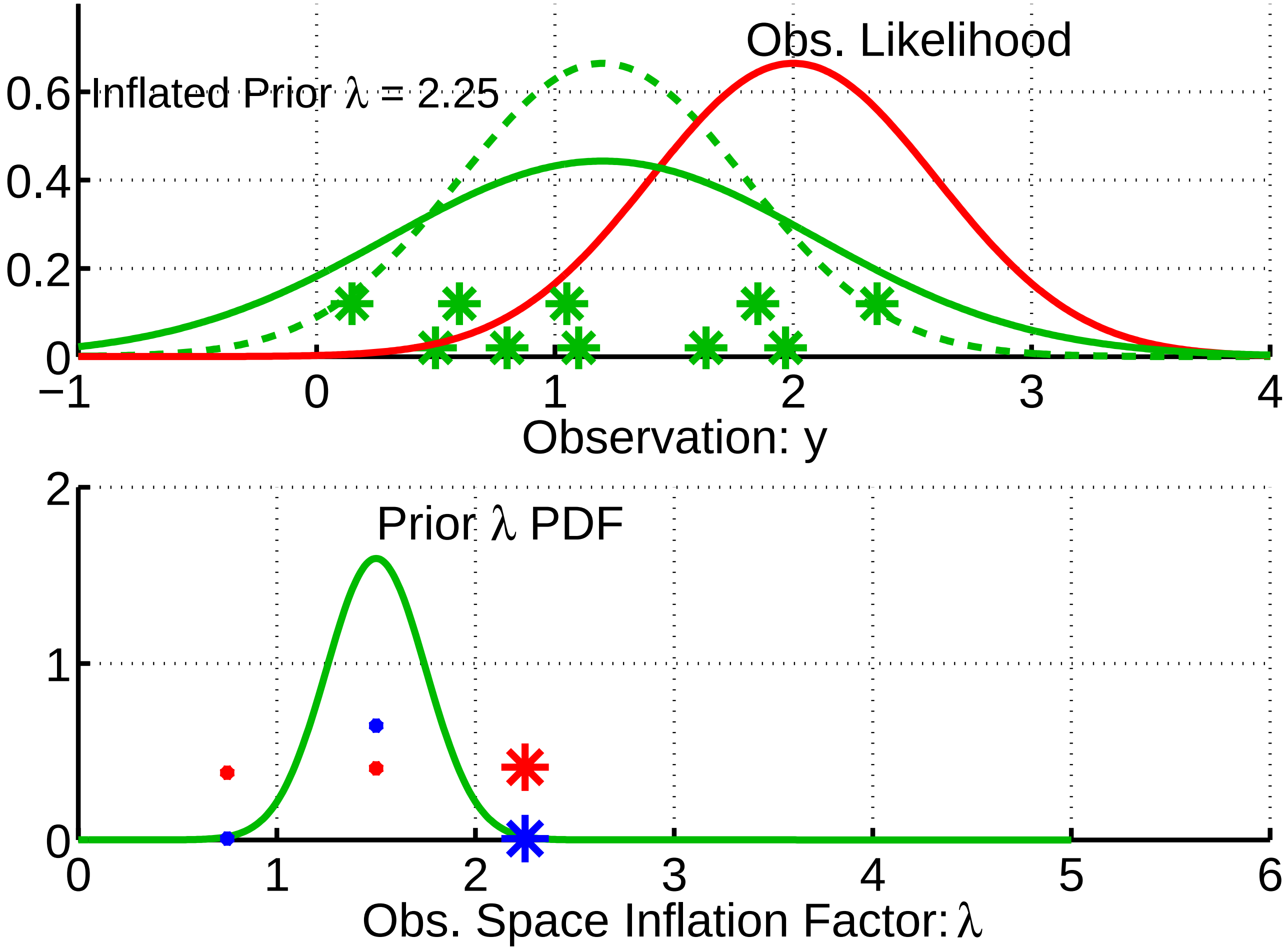
to get

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

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

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.

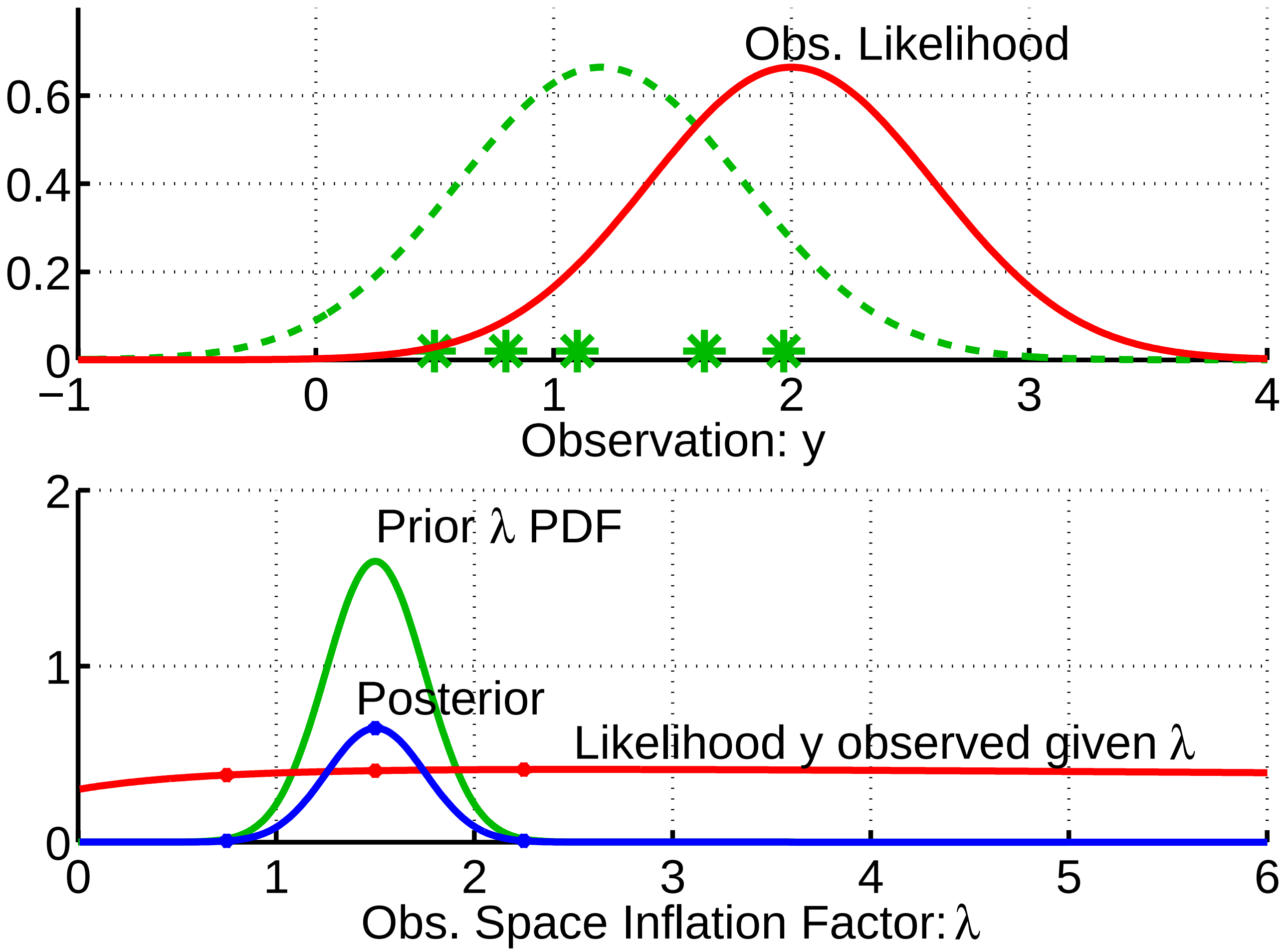
Multiply by $p(\lambda = 2.25, t_k | Y_{t_{k-1}})$

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.



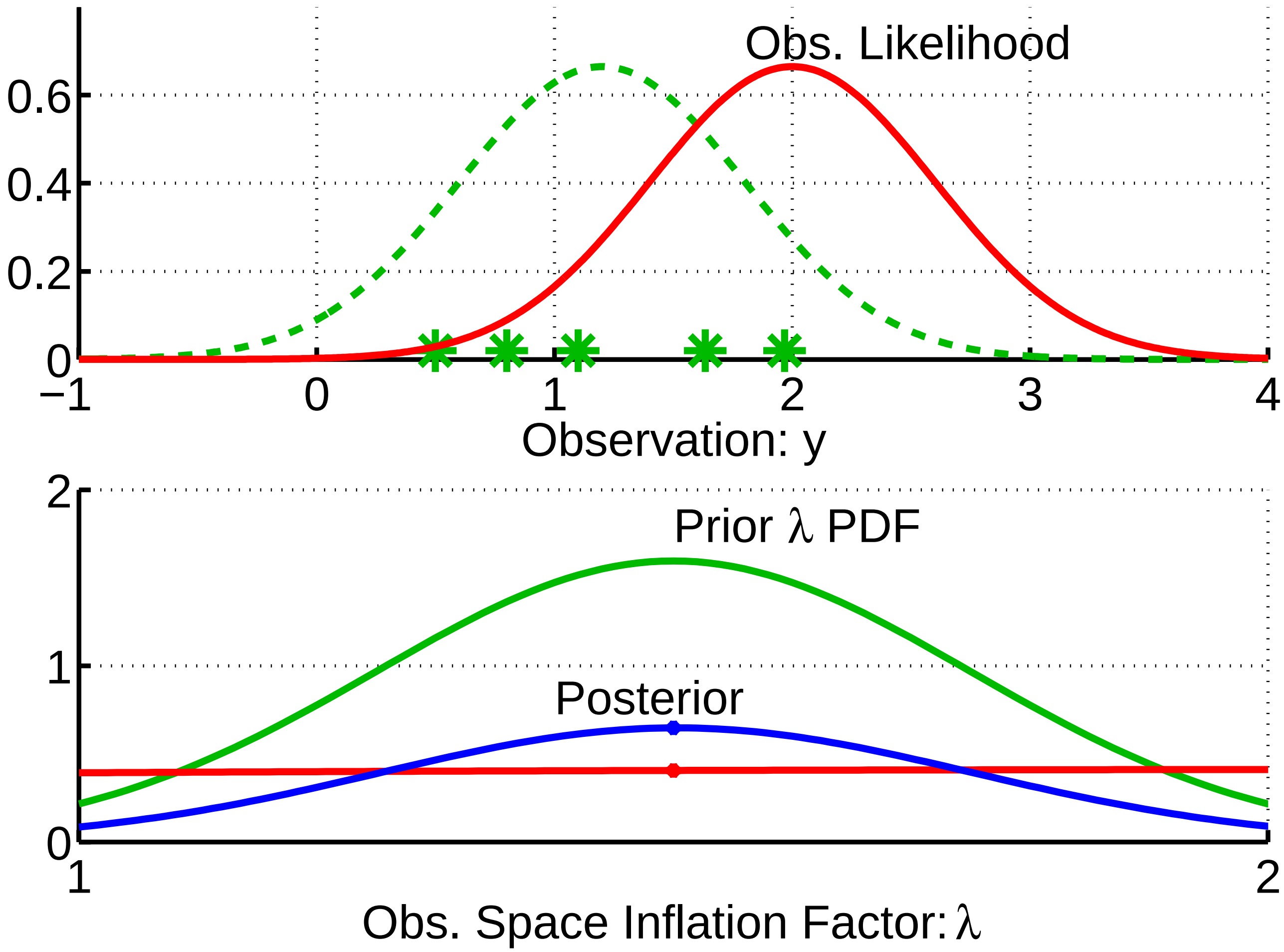
Repeat for a range of values of λ.

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.

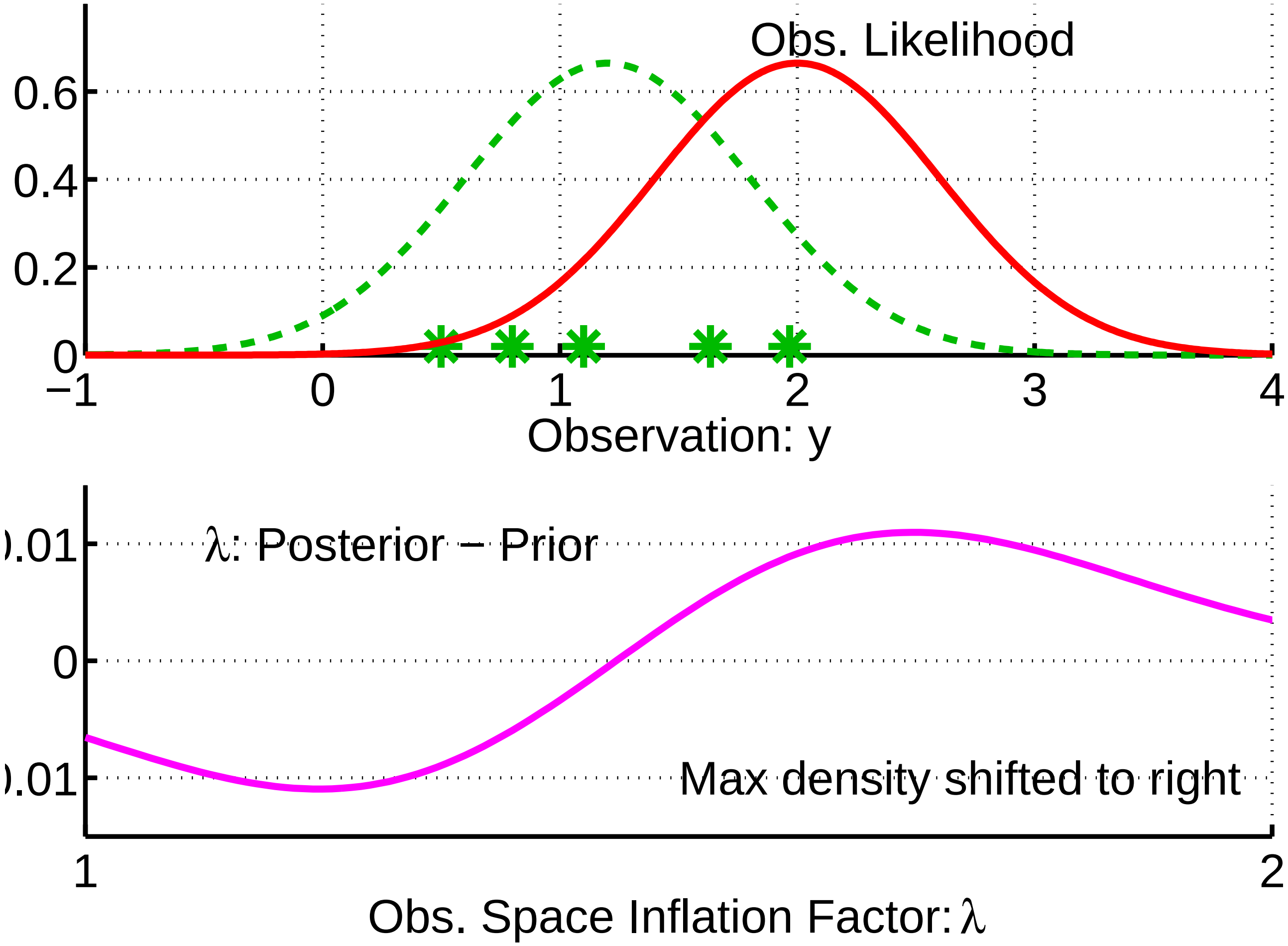
Posterior and prior are very similar.

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.

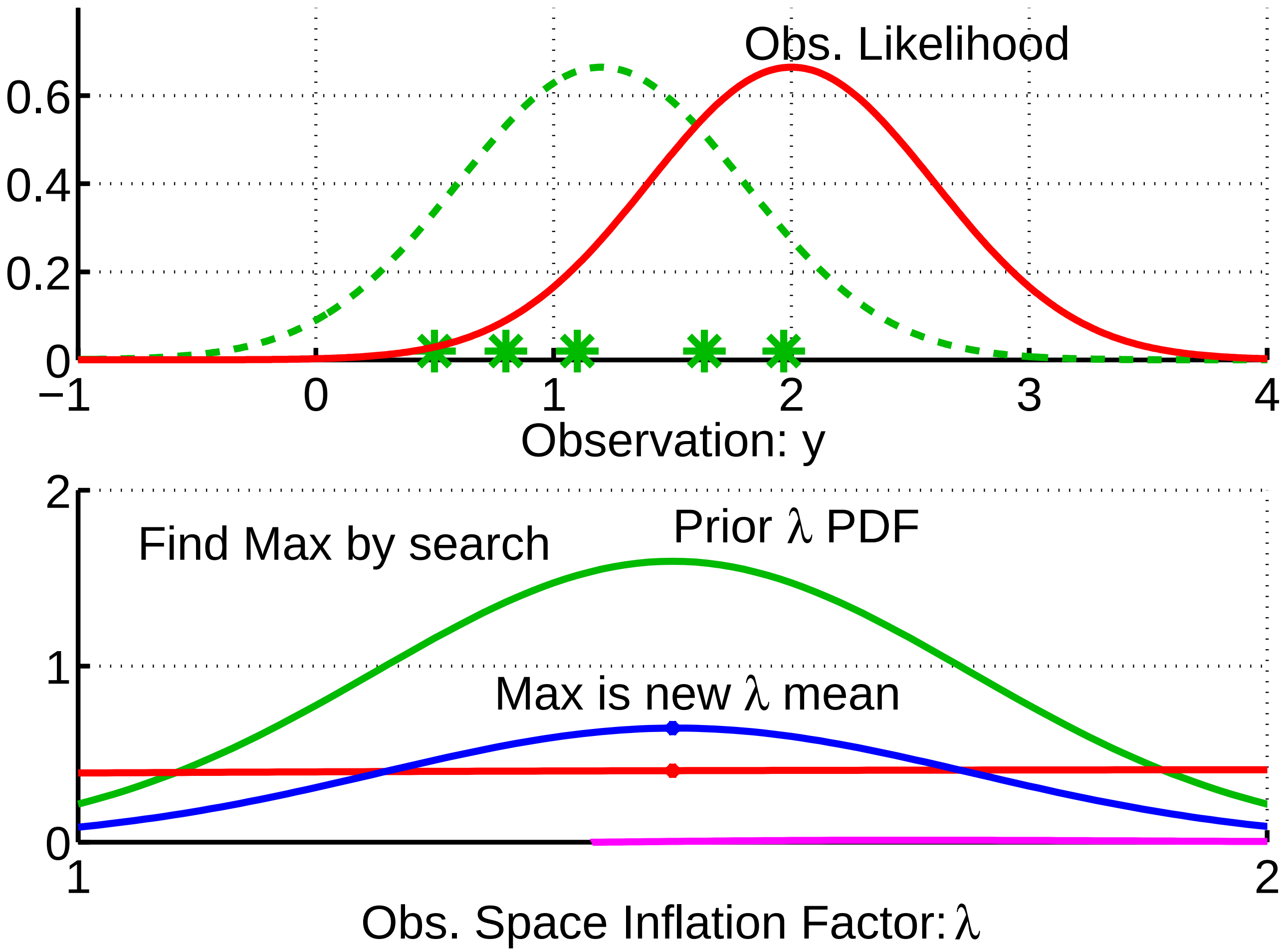
Posterior and prior are very similar.

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

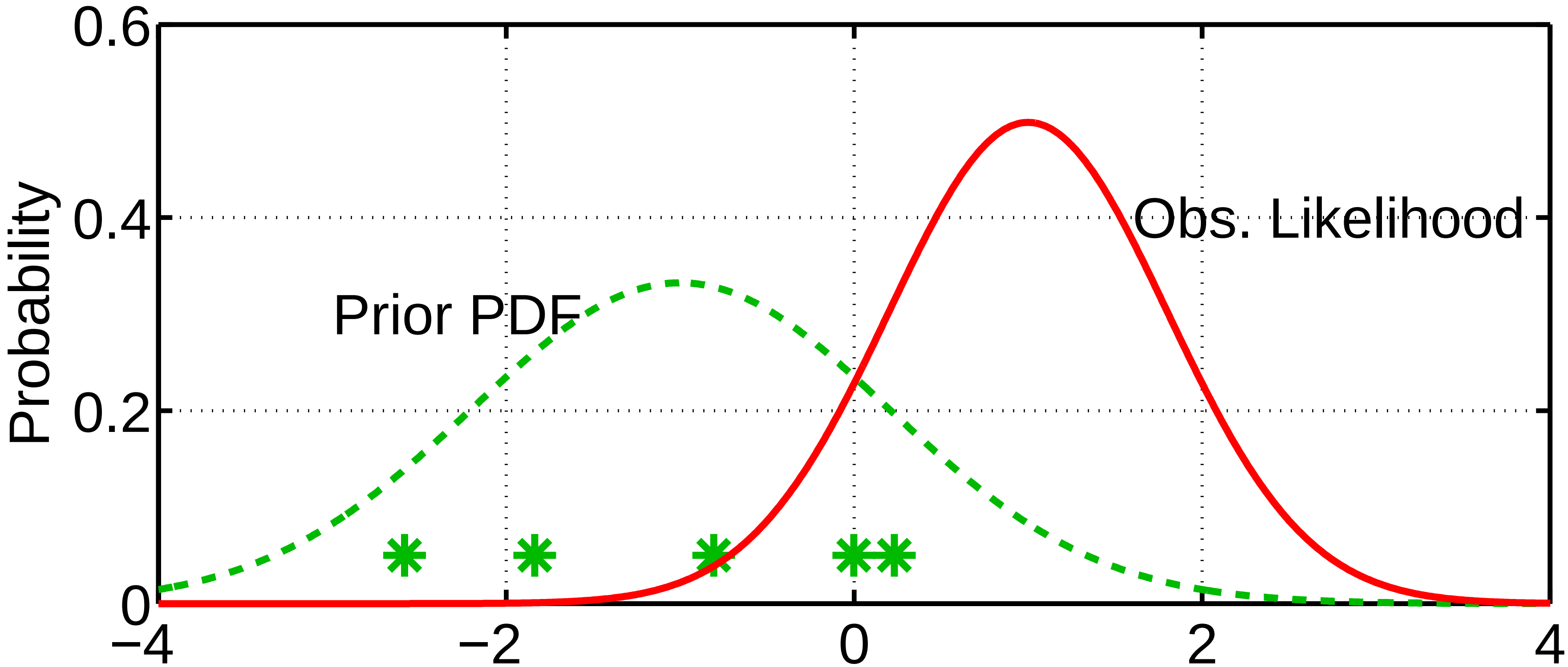
A. Computing updated inflation variance, $\sigma^2_{\lambda,\, u}$

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

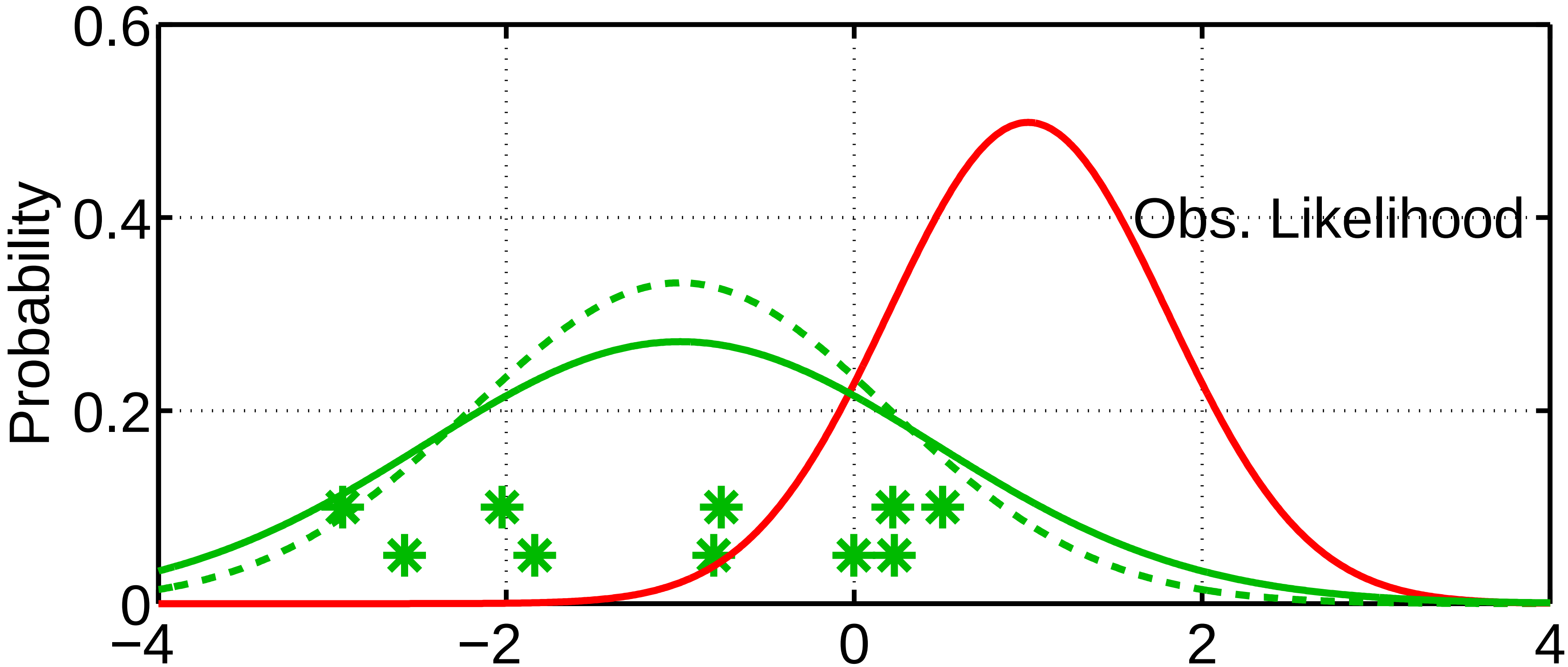
3. Compute as $\sigma^2_{\lambda,\, u} = -\sigma^2_{\lambda,\, p} / 2\, ln r$ where $r = p(\bar{\lambda}_u + \sigma_{\lambda,\, p}) / p(\bar{\lambda}_u)$

Observation Space Computations with Adaptive Error Correction



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

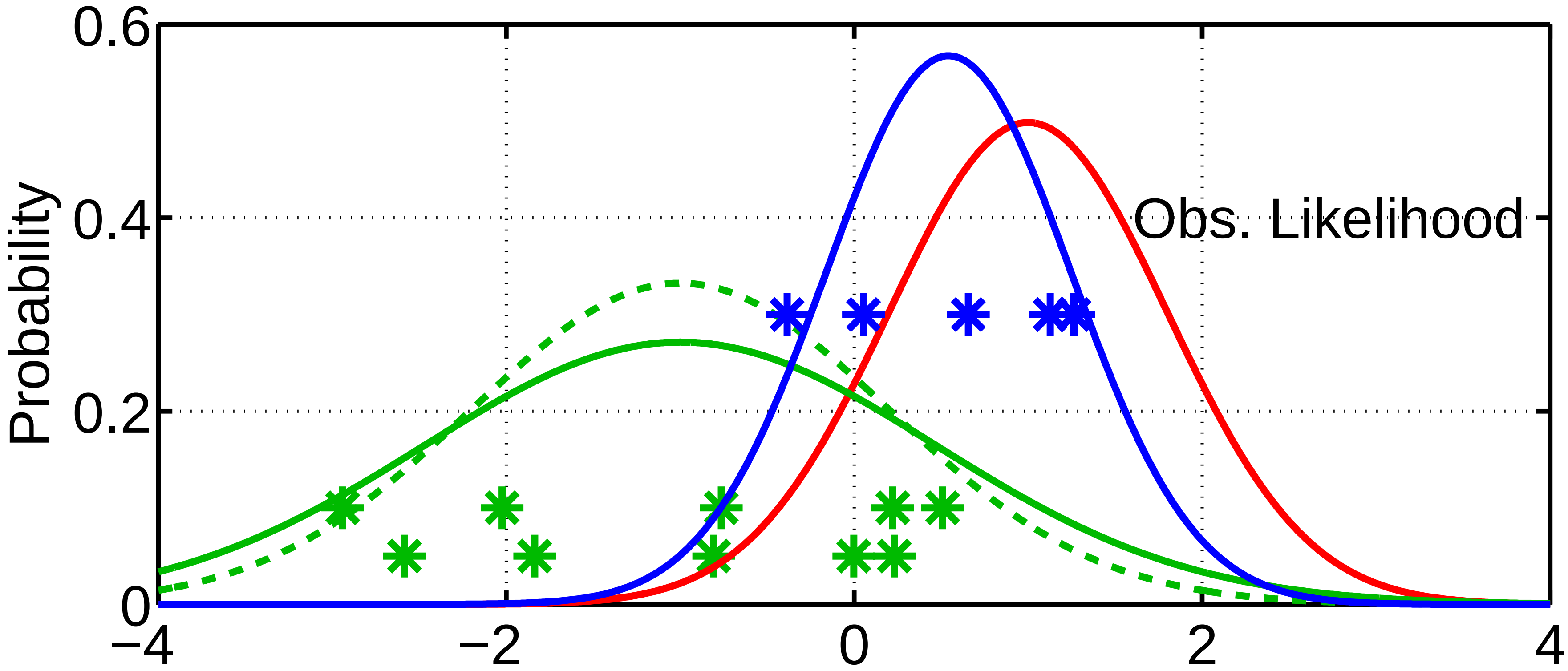
Observation Space Computations with Adaptive Error Correction



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

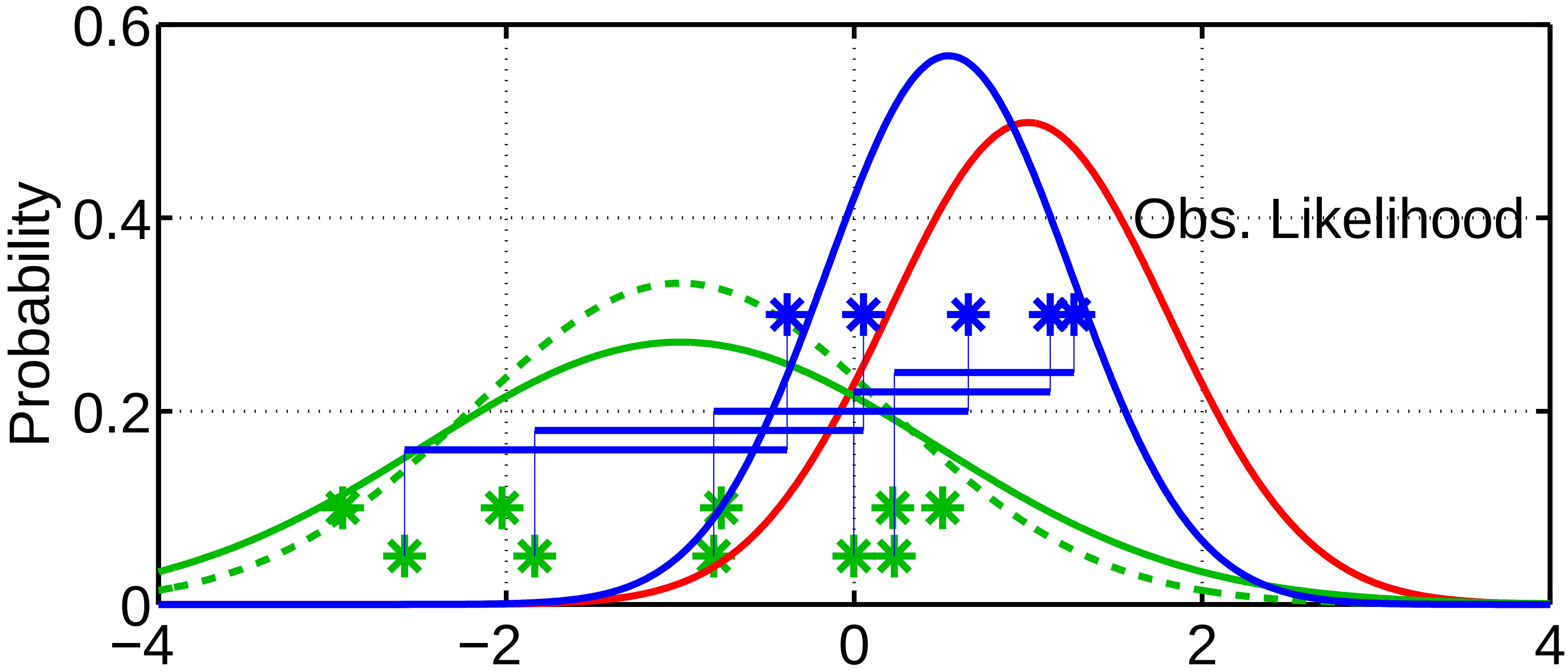


1. Compute updated inflation distribution, $p(\lambda, t_k | Y_{t_k})$.

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 Inflation in DART

	Before Assimilation	After Assimilation		
inf_flavor	= 1,	0,	Flavor:	1=> obs. space
inf_start_from_restart	= .false.,	.false.,		3=>physical space
inf_output_restart	= .true.,	.true.,		0=>NONE
inf_deterministic	= .true.,	.true.,		
inf_in_file_name	= 'prior_inflate_ics',	'post_inflate_ics',		
inf_out_file_name	= 'prior_inflate_restart',	'post_inflate_restart',		
inf_diag_file_name	= 'prior_inflate_diag',	'post_inflate_diag',		
inf_initial	= 1.00,	1.00,	Initial inflation value	
inf_sd_initial	= 0.2,	0.0,	Initial standard deviation	
inf_lower_bound	= 1.0,	1.0,		
inf_upper_bound	= 1000000.0,	1000000.0,		
inf_sd_lower_bound	= 0.0,	0.0	Lower bound on s.d.	

Try this in Lorenz-96 (verify other aspects of input.nml).

Use 40 member ensemble. (set *ens_size = 40* in *filter_nml*).

Set red values as above for adaptive observation space inflation.

Adaptive Observation Space Inflation in Lorenz-96

Run the filter

Examine performance with *plot_total_err* in matlab

Time series of inflation and standard deviation are in *prior_inflate_diag*

Inflation adjusts with time

Standard deviation is non-increasing

Algorithmic variants:

1. Increase prior y variance by adding random gaussian noise.

   As opposed to ‘deterministic’ linear inflating.

   Set *inf_deterministic* in first column to .false.

   Change it back to .true. after checking this out.

2. Just have a fixed value for obs. space λ

   Cheap, handles blow up of state vars unconstrained by obs.

   We already tried this in section 9.

Algorithmic variants:

3. Fix value of λ standard deviation, σ_λ.

Reduces cost, computation of σ_λ can sometimes be tricky.

Avoids σ_λ getting small (error model filter divergence, Yikes!).

Have to have some intuition about the value for σ_λ.

This appears to be most viable option for large models.

Values of σ_λ = 0.05 to 0.10 work for very broad range of problems.

This is a sampling error closure problem (akin to turbulence).

To fix σ_λ:

Set *inflate_sd_initial* to fixed value, for instance 0.10,

Set *inflate_sd_lower_bound* to same value.

(s.d. can't get any smaller).

Try this in lorenz-96. Look at how the inflation varies.

Potential problems with observation space adaptive inflation

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)

   Try turning this on in 9var model.

   Fixed 0.05 for *inf_sd_initial*, *sd_lower_bound*.

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* in *model_nml* 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.

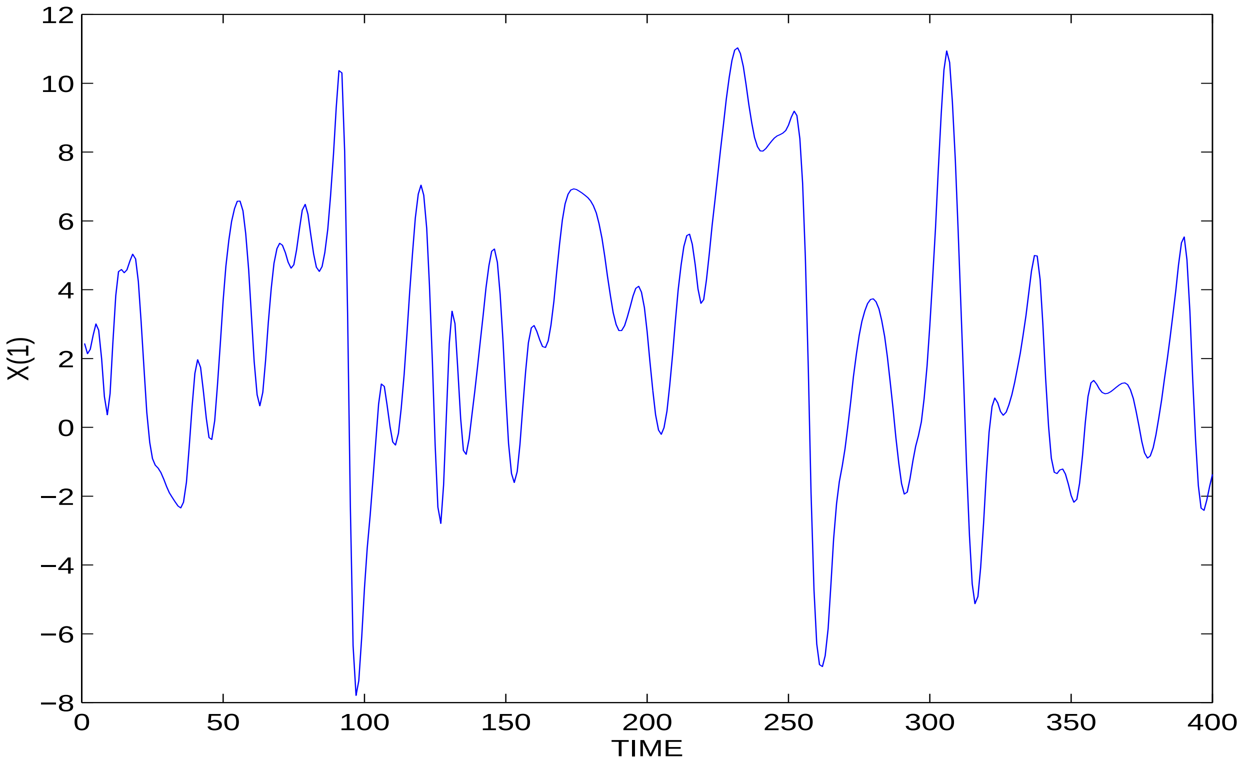
Simulating Model Error in 40-Variable Lorenz-96 Model

40 state variables: $X_1, X_2, \ldots, X_N$

$dX_i / dt = (X_{i+1} - X_{i-2})X_{i-1} - X_i + F;$

$i = 1, \ldots, 40$ with cyclic indices

Use $F = 8.0$, 4th-order Runge-Kutta with $dt=0.05$



Time series of state variable from free L96 integration

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)

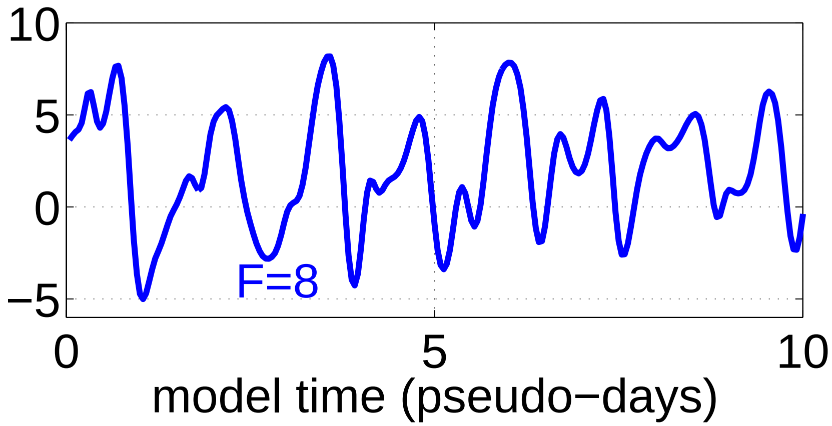
Results from 10 days after 40 day spin-up

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

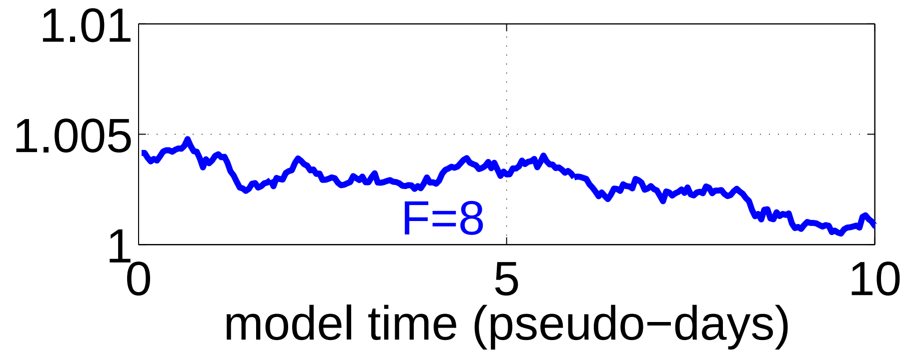
Simulates increasing model error

Assimilating F=8 Truth with F=8 Ensemble

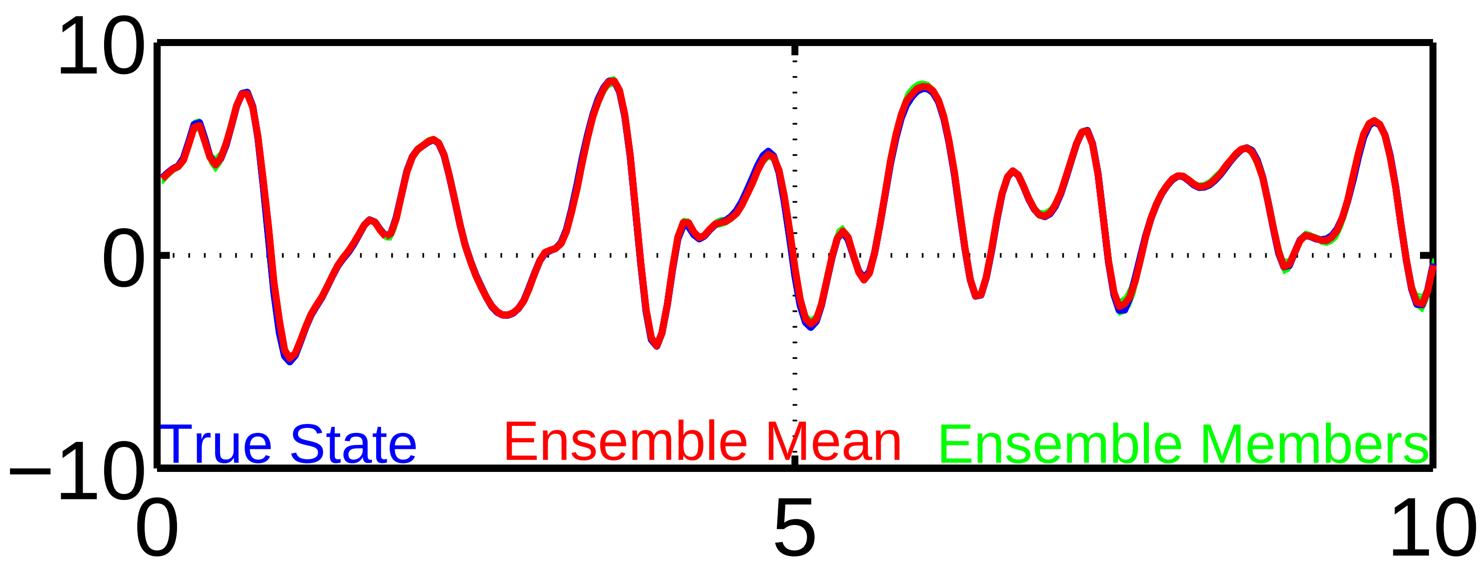
Model time series



Mean value of λ

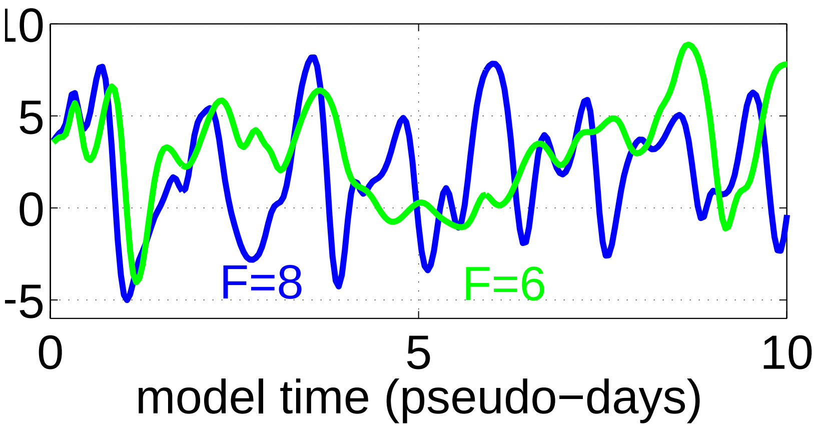


Assimilation Results

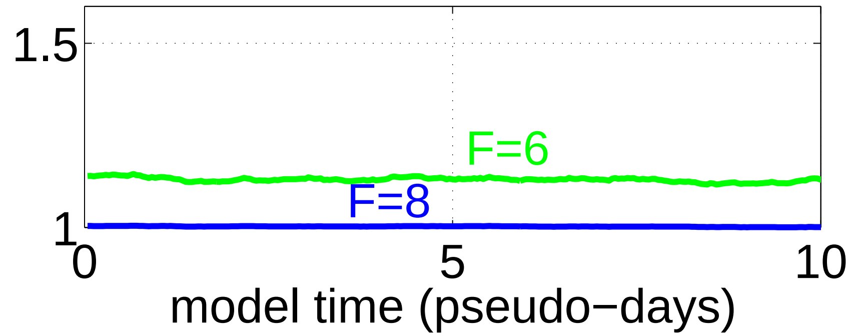


Assimilating F=8 Truth with F=6 Ensemble

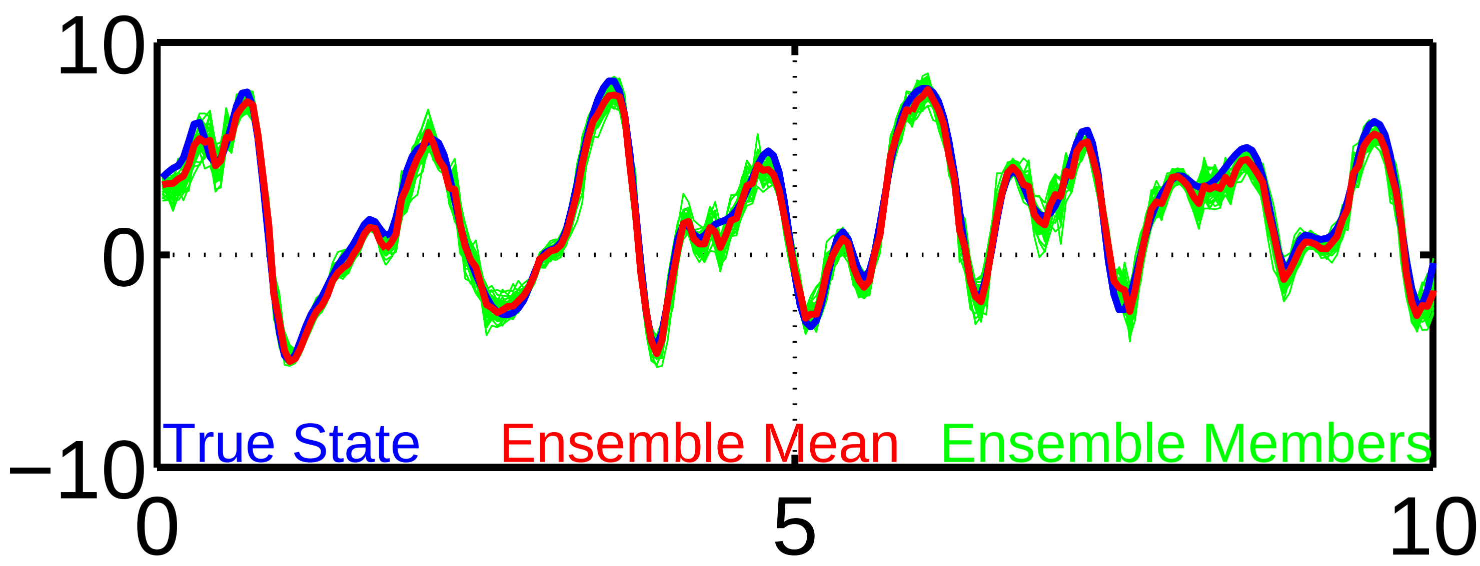
Model time series



Mean value of λ

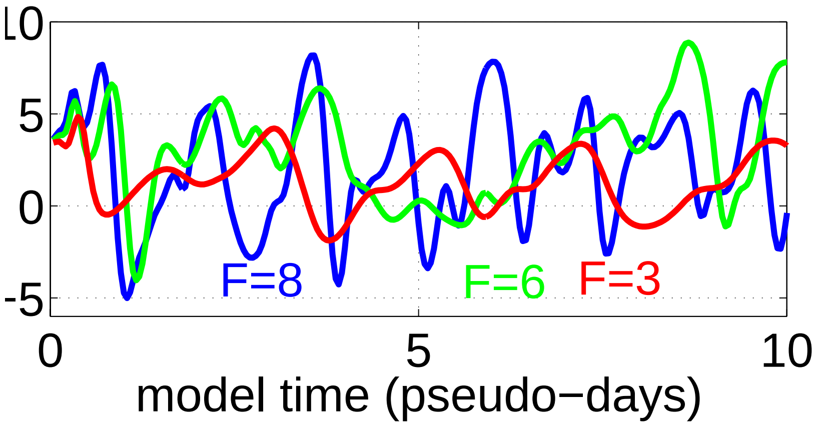


Assimilation Results

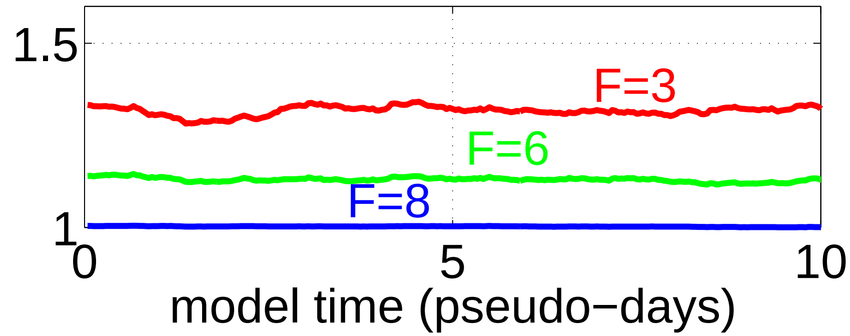


Assimilating F=8 Truth with F=3 Ensemble

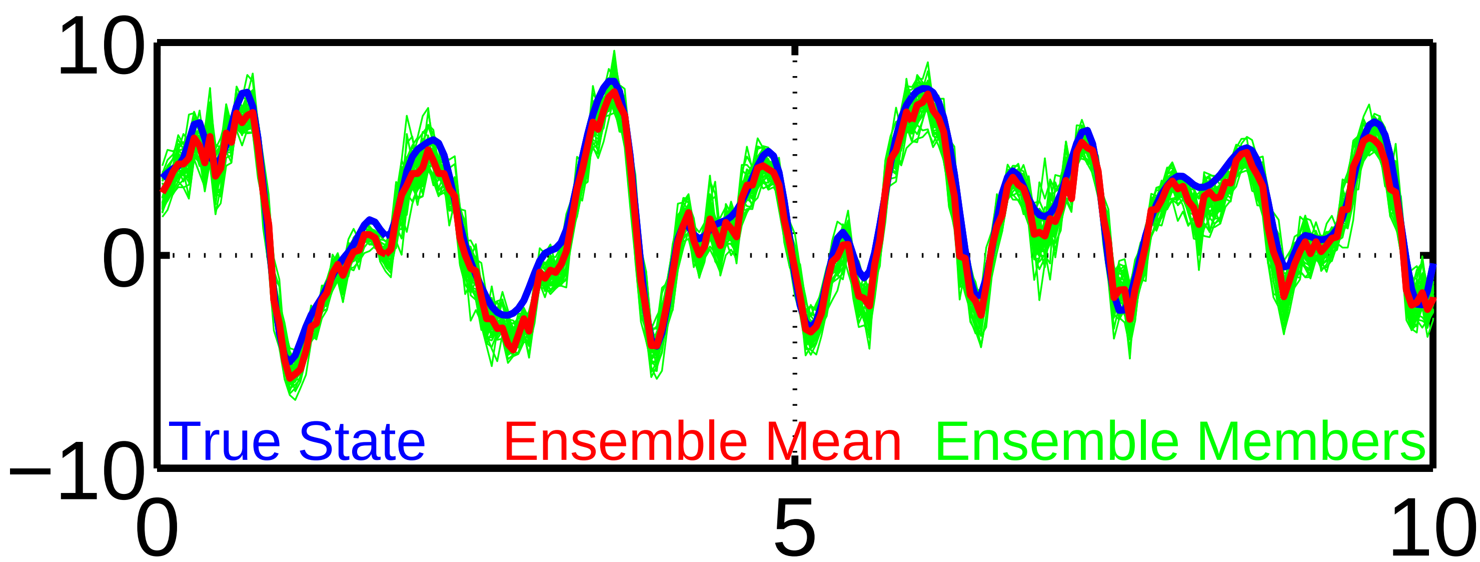
Model time series



Mean value of λ

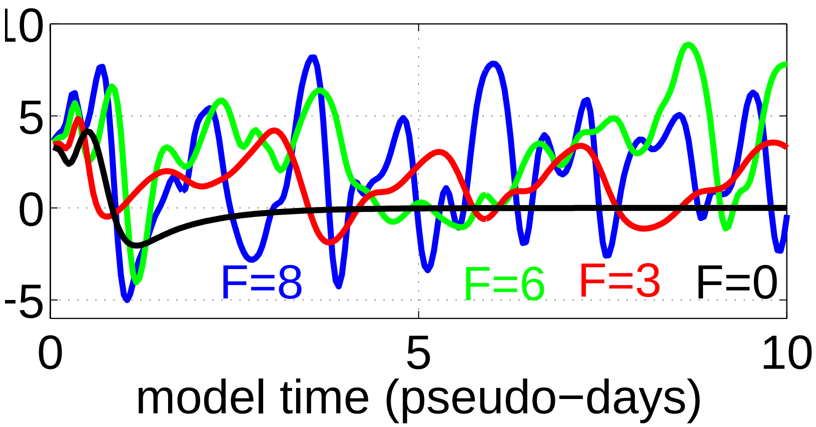


Assimilation Results

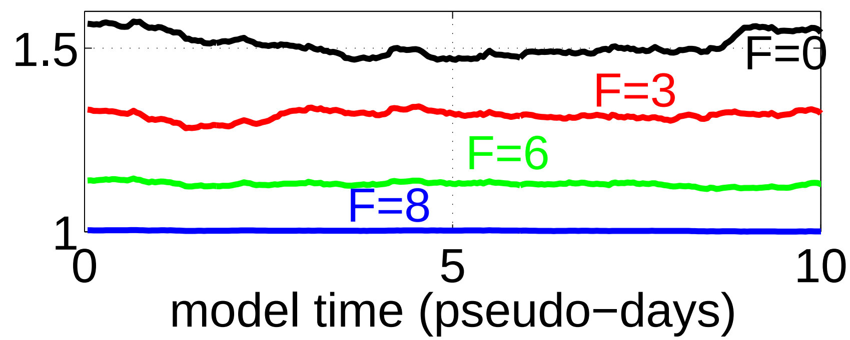


Assimilating F=8 Truth with F=0 Ensemble

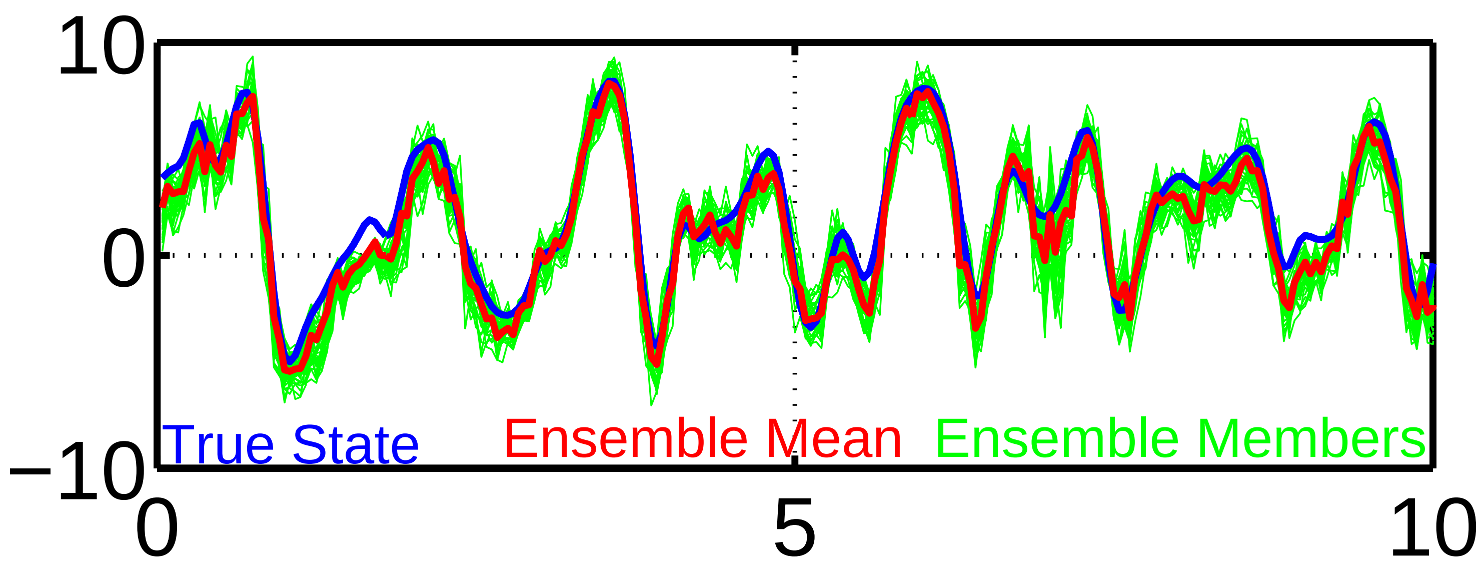
Model time series



Mean value of λ

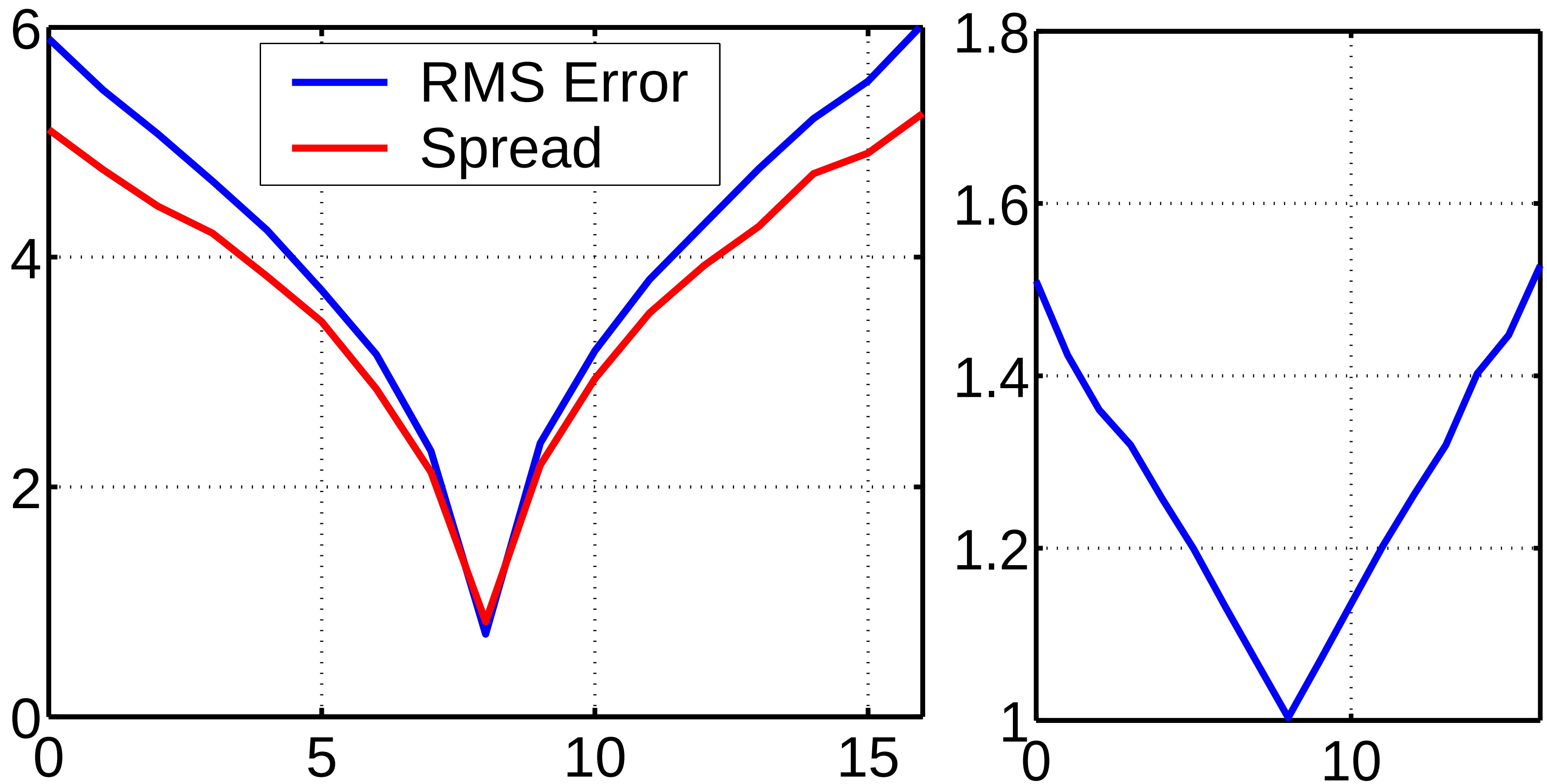


Assimilation Results



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

Base case: 200 randomly located observations per time



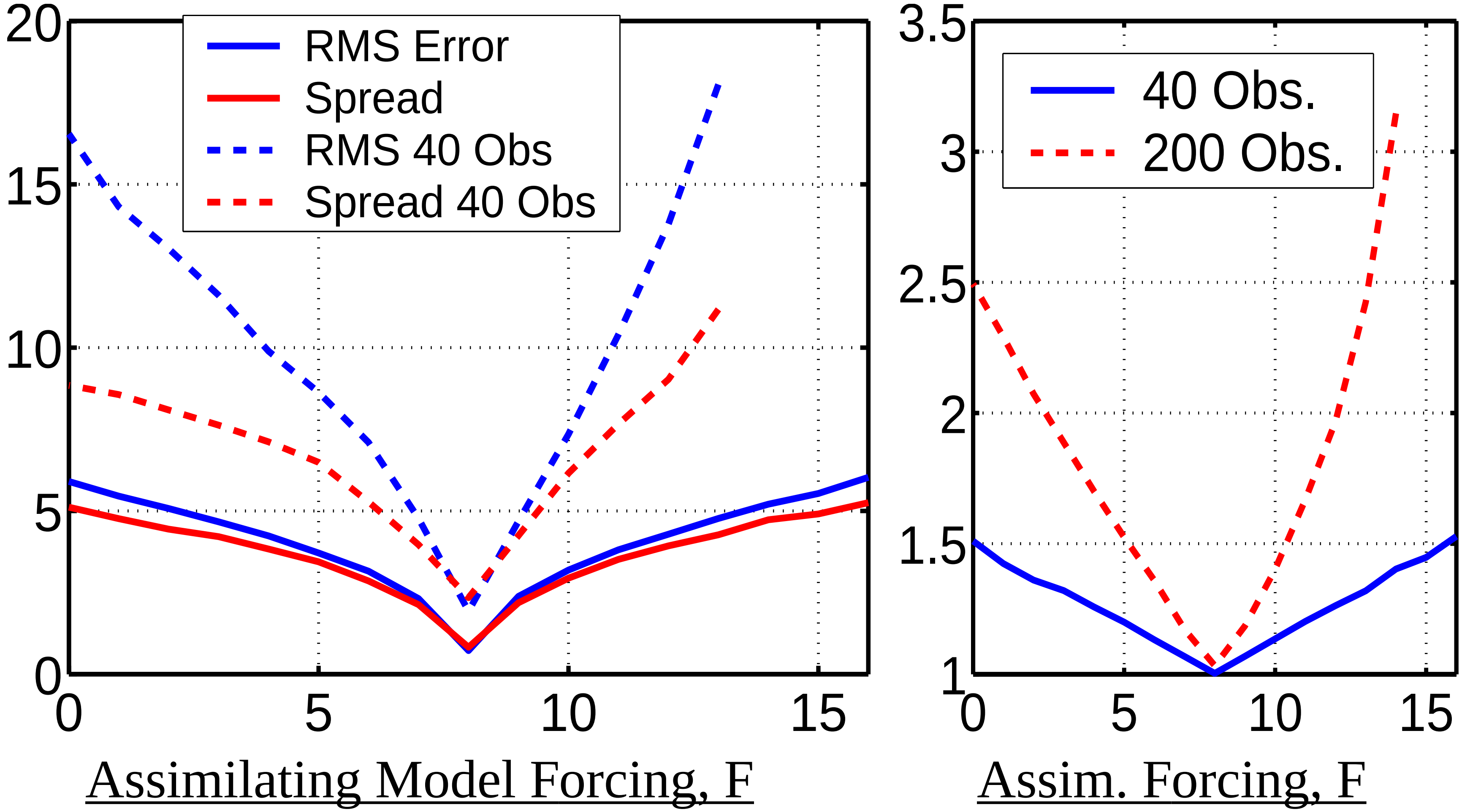
Assimilating Model Forcing, F

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: $p(y_o|\lambda) = (2\Pi\theta^{\mathbf{2}})^{-1/2}\exp(-D^2/2\theta^2)$

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

Compute updated distribution for λ_s exactly as for observation space.

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?

Experimenting with spatially-constant state space inflation

To try adaptive state inflation, set *inf_flavor=3* in first column.
May help to increase initial value, *inf_initial*
Diagnostics are in *Prior_Diag.nc* file;
    This can be viewed with *ncview* (more on this later)
Final values are in *prior_inflate_restart file*

Spatially varying adaptive inflation algorithm:

Have a distribution for λ 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!

Experimenting with spatially-constant state space inflation

To try adaptive state inflation, set *inf_flavor=2* in first column.
May help to increase initial value, *inf_initial*
Diagnostics are in *Prior_Diag.nc* file;
    This can be viewed with *ncview*
Final values are in *prior_inflate_restart file*

Posterior Inflation

So far, we’ve always used the first column of the inflation namelist
Inflation is performed after model advances but before assimilation
Can also do posterior inflation using second column
This does inflation after assimilation but before model advance
Technically, this is cheating except in the limit of many observations
Assumption that observations and error are independent is lost
Helps to increase variance in forecasts

Can also do both prior and posterior inflation (use both columns)
Diagnostics are in same files with ‘post’ instead of ‘prior’

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)

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

Different λ's see different observations

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