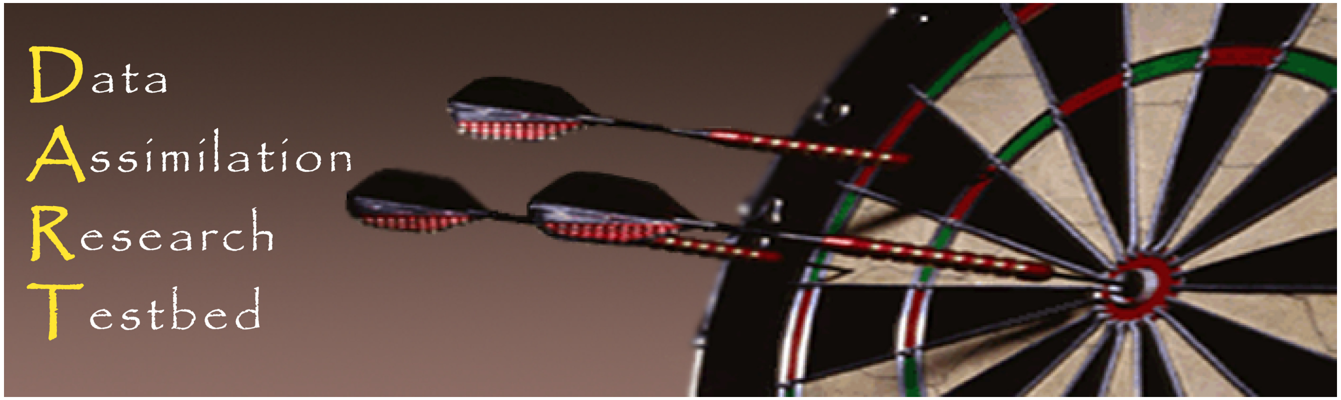
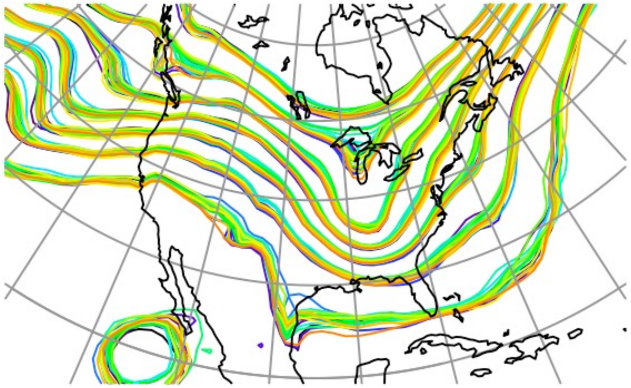


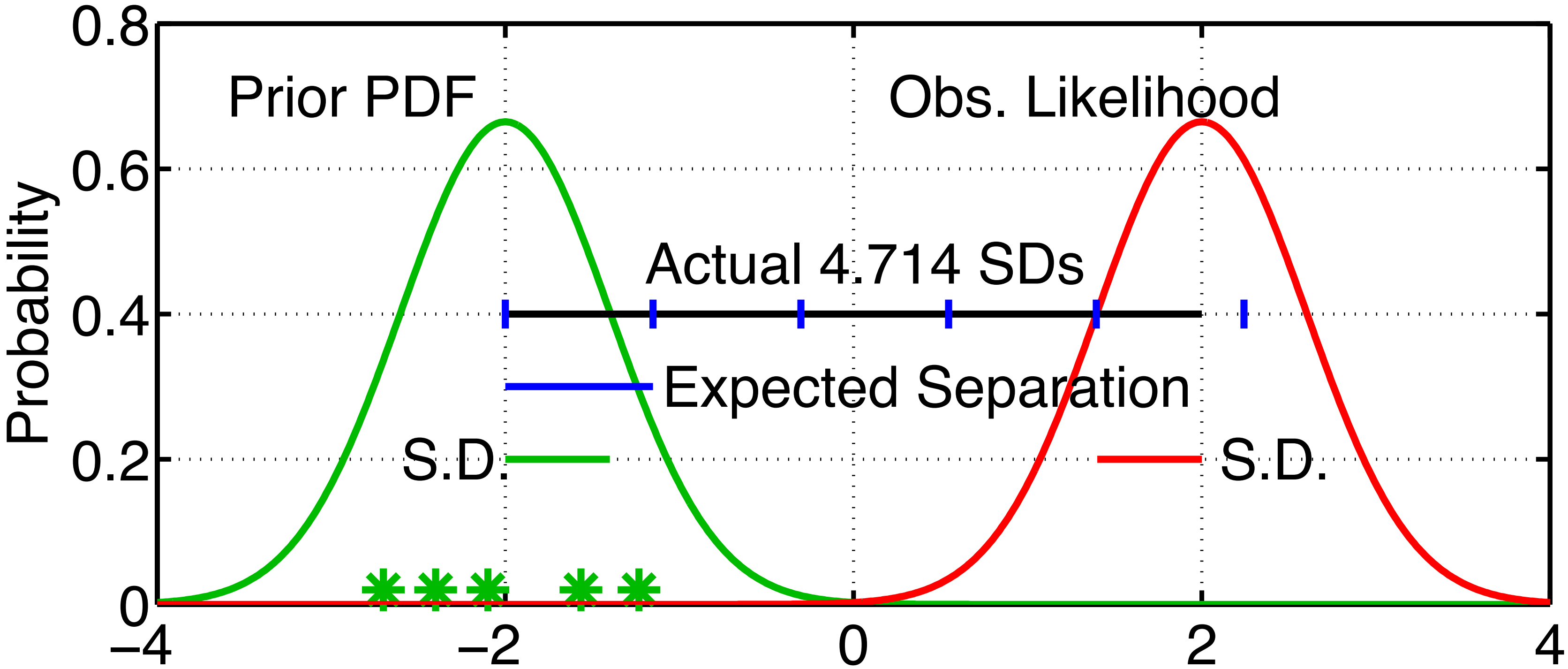
DART Tutorial Section 12:
Adaptive Inflation



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 UCAR | National Center for Atmospheric Research

Variance inflation for observations: An adaptive error tolerant filter

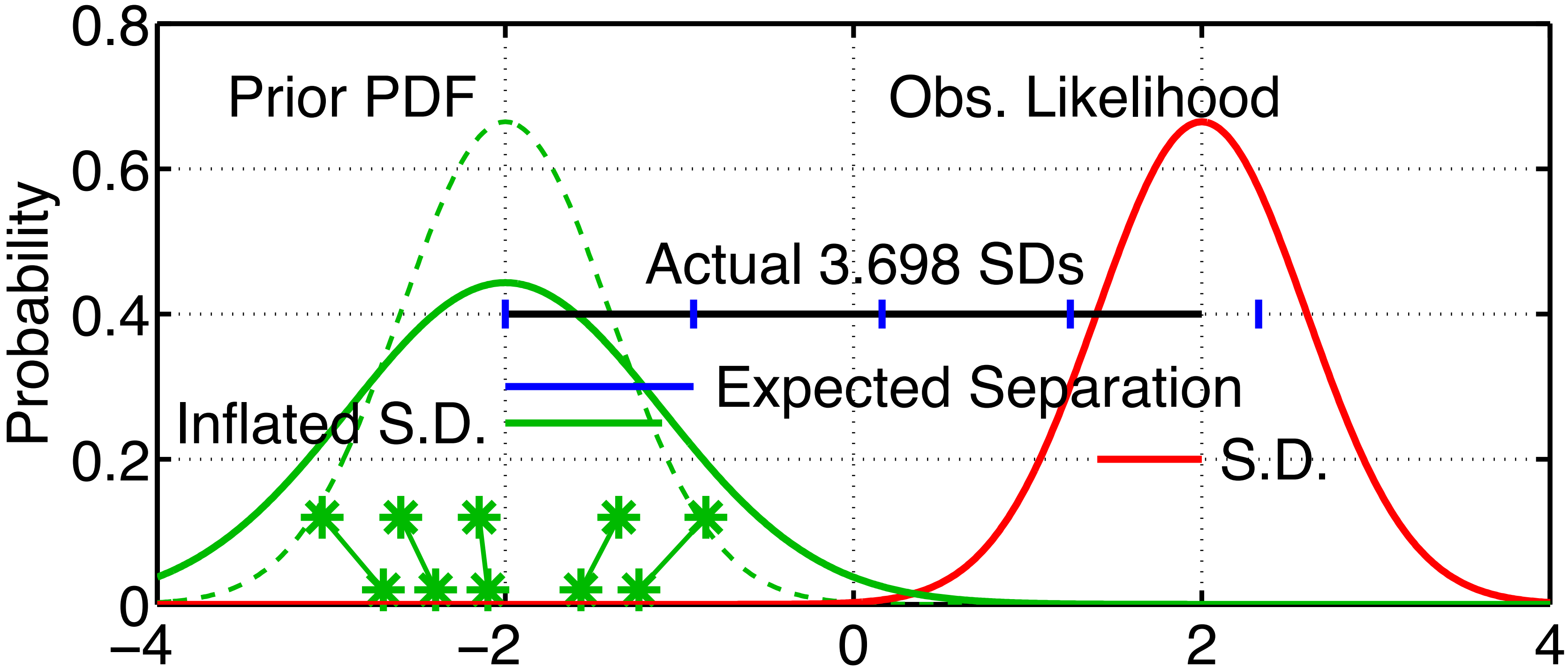


1. For observed variable, have estimate of prior-observed inconsistency.
2. Expected (prior_mean – observation) = $\sqrt{\sigma^2_{prior} + \sigma^2_{obs}}$

   Assumes that prior and observation are supposed to be unbiased.

   Is it model error or random chance?

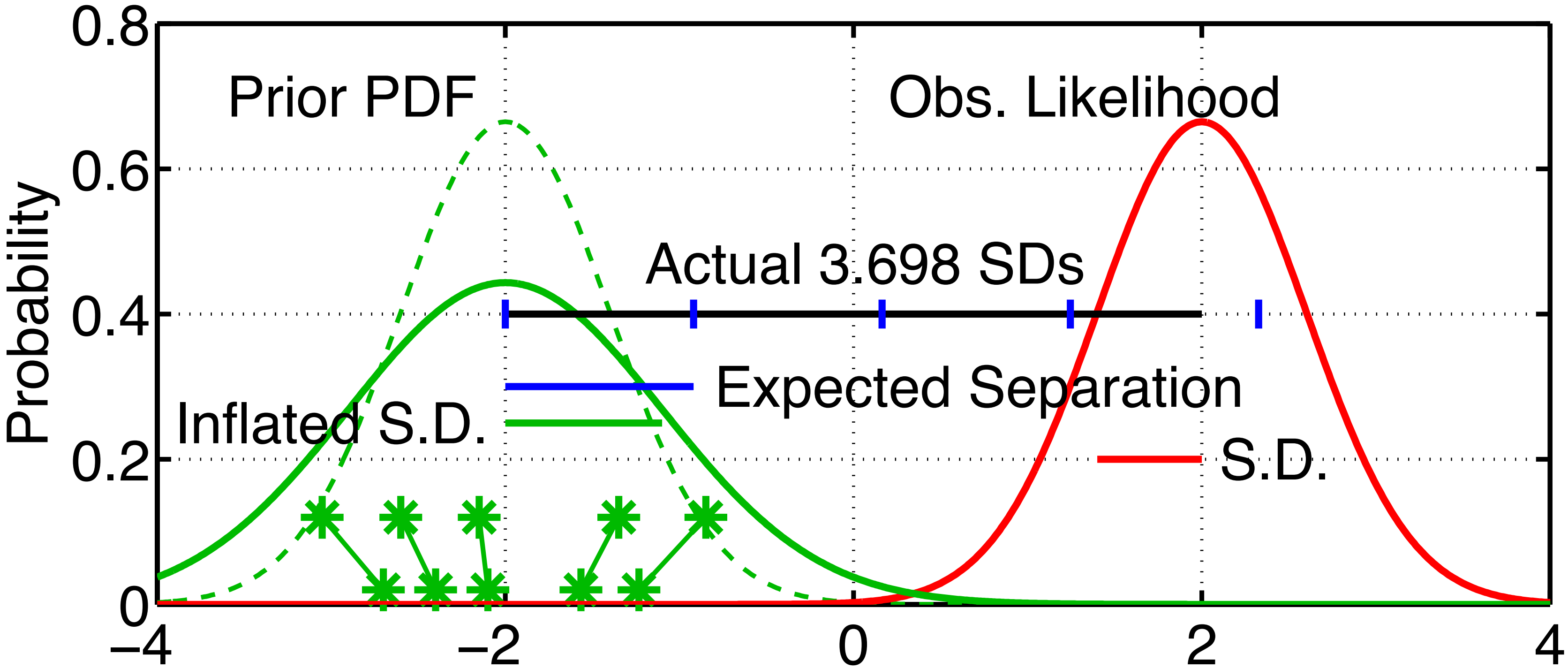
Variance inflation for observations: An adaptive error tolerant filter



1. For observed variable, have estimate of prior-observed inconsistency.
2. Expected (prior_mean – observation) = $\sqrt{\sigma_{prior}^2 + \sigma_{obs}^2}$
3. Inflating increases expected separation

   Increases 'apparent' consistency between prior and observation.

Variance inflation for observations: An adaptive error tolerant filter

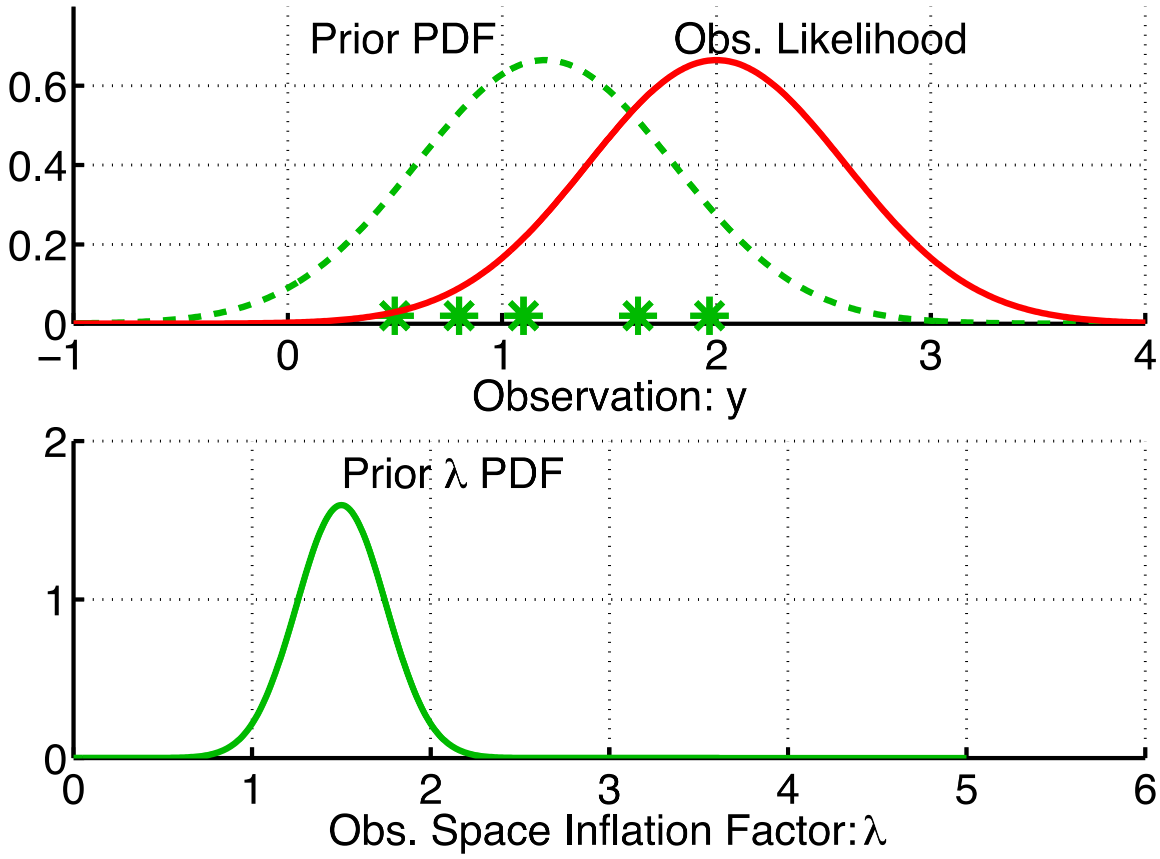


Distance D from prior mean y to obs is $N\left(0,\sqrt{\lambda\sigma_{prior}^{2}+\sigma_{obs}^{2}}\right)=N(0,\theta)$

Prob y_0 is observed given λ: $p(y_o \mid \lambda)=\left(2\pi\theta^2\right)^{-1/2}\exp\left(-D^2/2\theta^2\right)$

Variance inflation for observations: An adaptive error tolerant filter

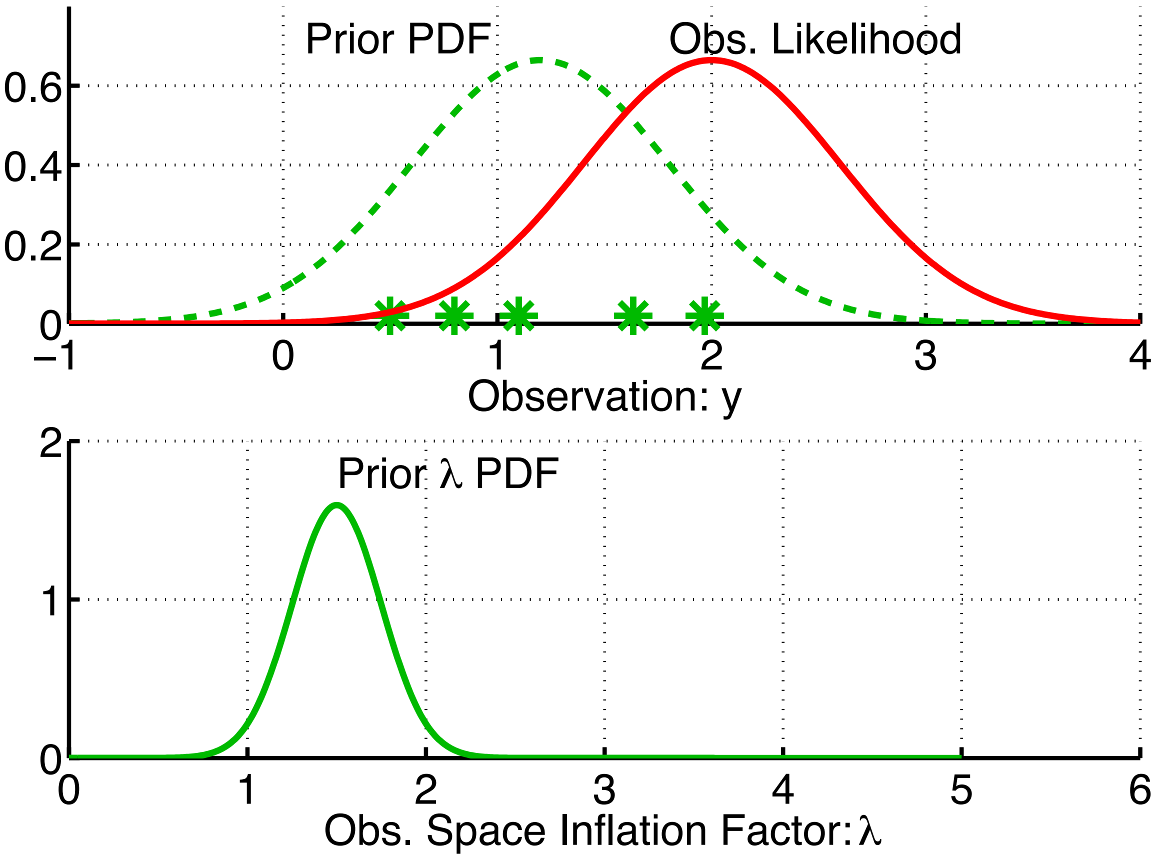
Use Bayesian statistics to get estimate of inflation factor λ.



Assume prior is Gaussian: $p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) = N\left(\bar{\lambda}_p, \sigma^2_{\lambda,p}\right)$

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.

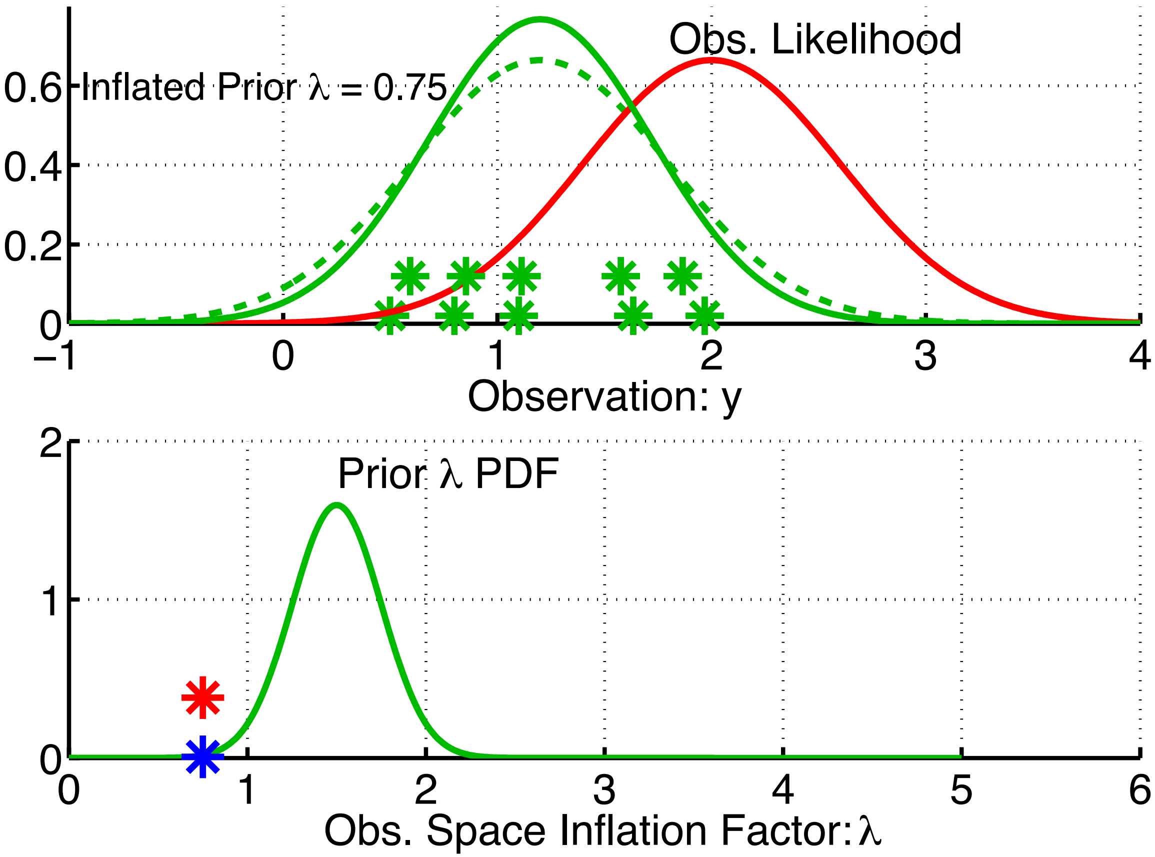
$p\left(\lambda, t_k \mid Y_{t_{k-1}}\right)$

Recall that $p\left(y_k \mid \lambda\right)$ can be evaluated From normal PDF.

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

Variance inflation for observations: An adaptive error tolerant filter

Use Bayesian statistics to get estimate of inflation factor λ.



Get $p(y_k \mid \lambda = 0.75)$ from normal PDF.

Multiply by

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

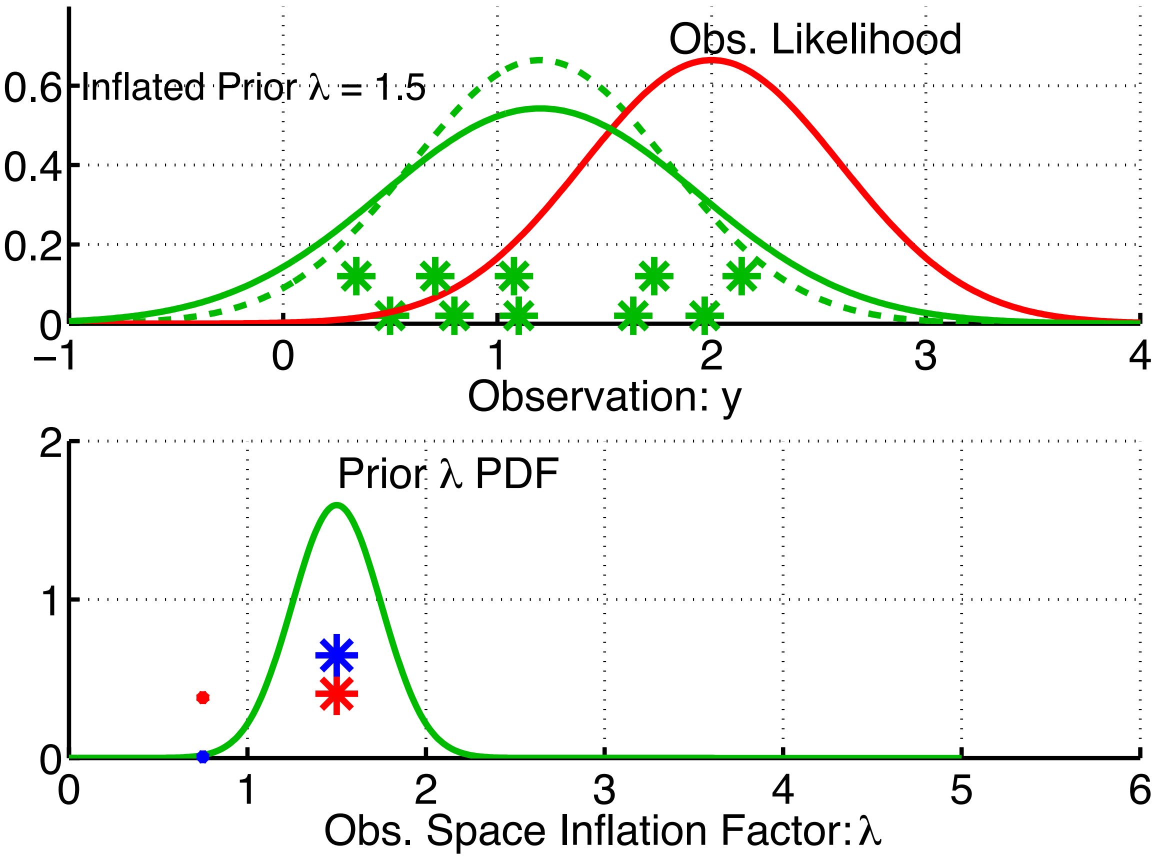
to get

$$p\left(\lambda = 0.75, t_k \mid Y_{t_k}\right)$$

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

Variance inflation for observations: An adaptive error tolerant filter

Use Bayesian statistics to get estimate of inflation factor λ.



Get $p(y_k \mid \lambda = 1.50)$ from normal PDF.

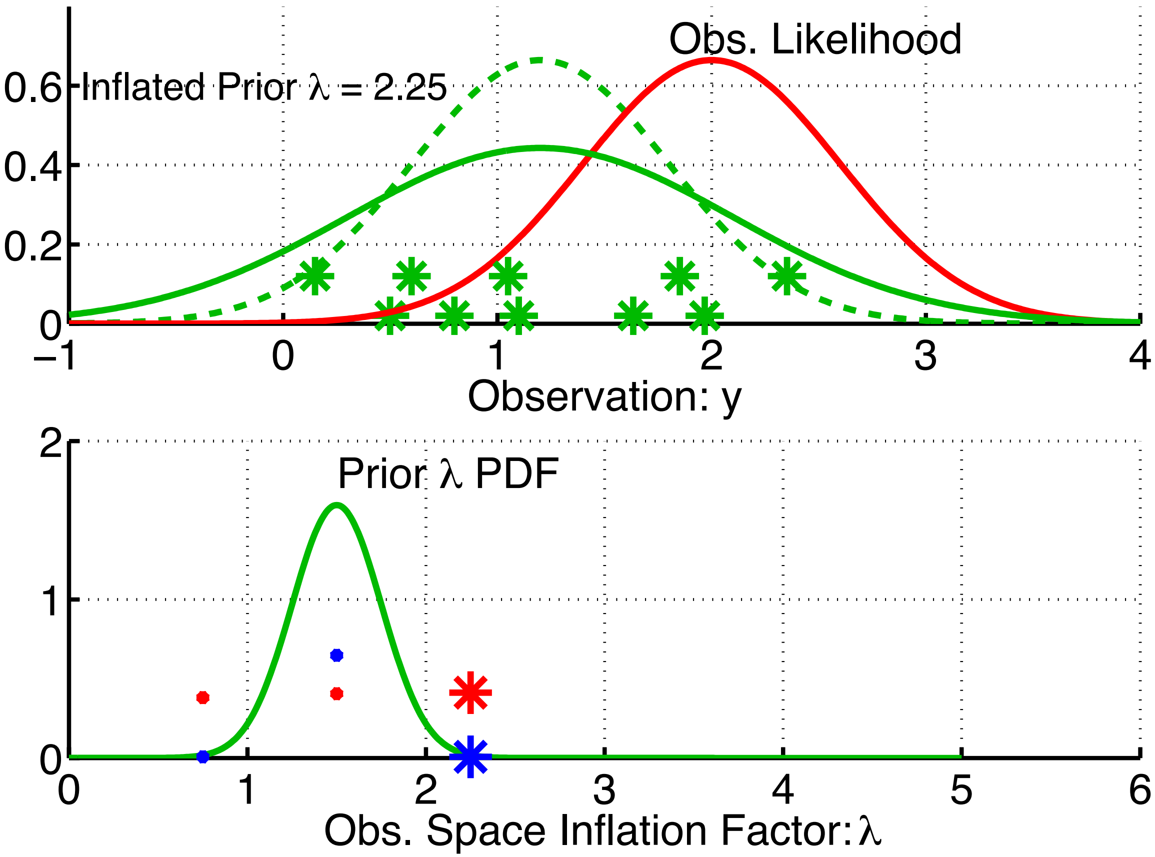
Multiply by $p(\lambda = 1.50, t_k \mid Y_{t_{k-1}})$

to get $p(\lambda = 1.50, t_k \mid Y_{t_k})$

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

Variance inflation for observations: An adaptive error tolerant filter

Use Bayesian statistics to get estimate of inflation factor λ.



Get $p(y_k \mid \lambda = 2.25)$ from normal PDF.

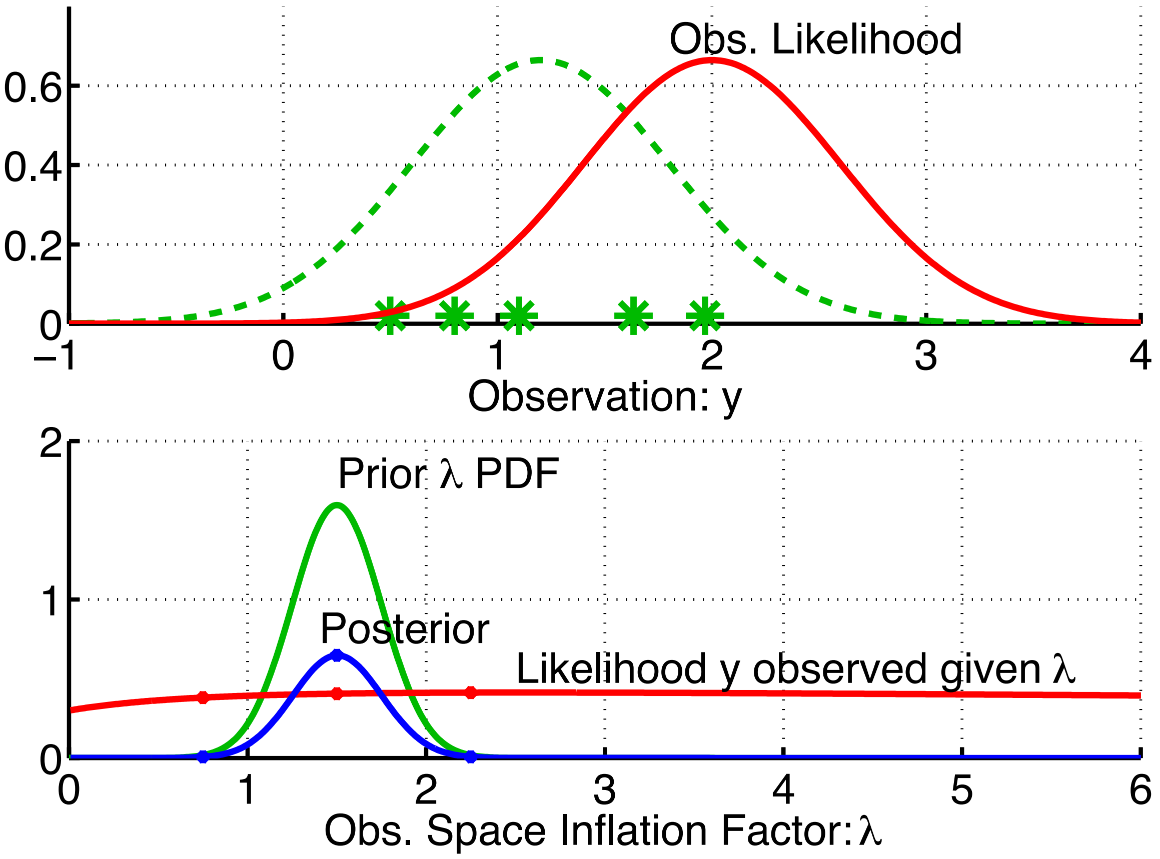
Multiply by $p(\lambda = 2.25, t_k \mid Y_{t_{k-1}})$

to get $p(\lambda = 2.25, t_k \mid Y_{t_k})$

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

Variance inflation for observations: An adaptive error tolerant filter

Use Bayesian statistics to get estimate of inflation factor λ.



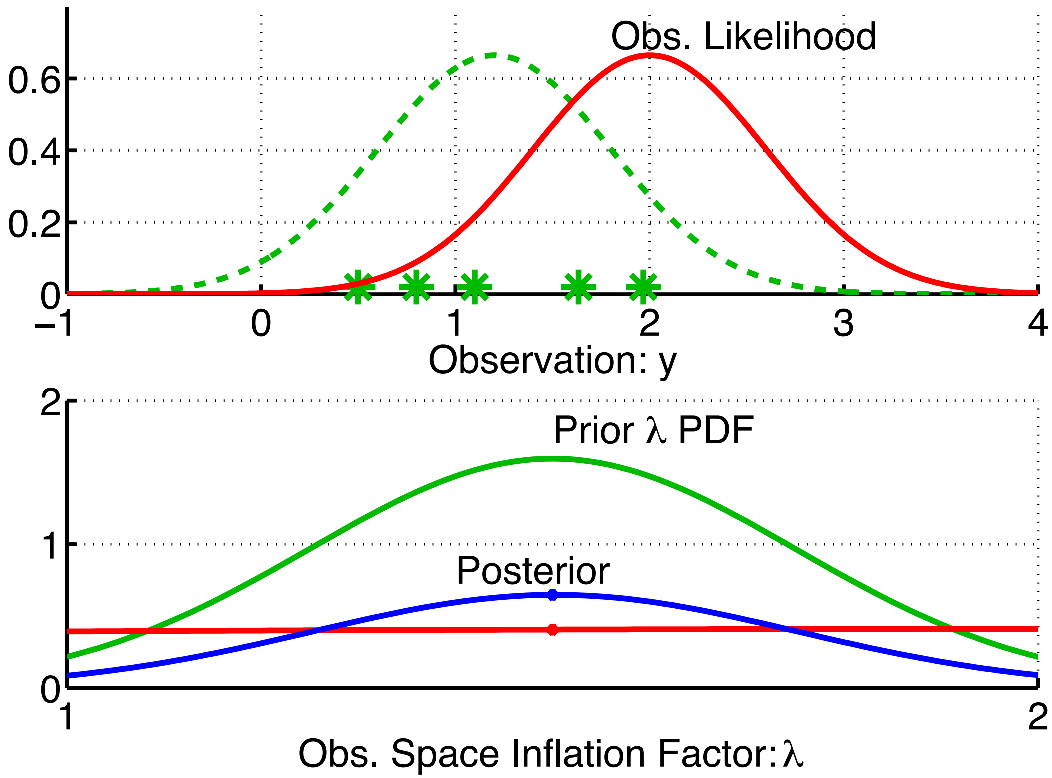
Repeat for a range of values of λ.

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

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

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.

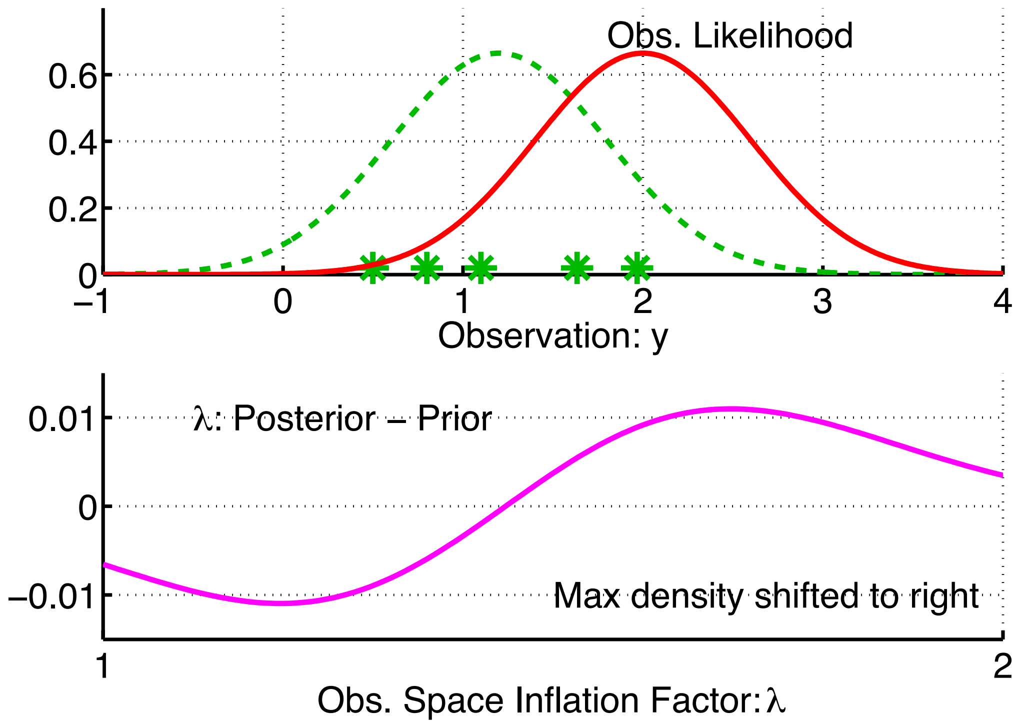
Posterior and prior are very similar.

Normalized posterior indistinguishable from prior.

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

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.

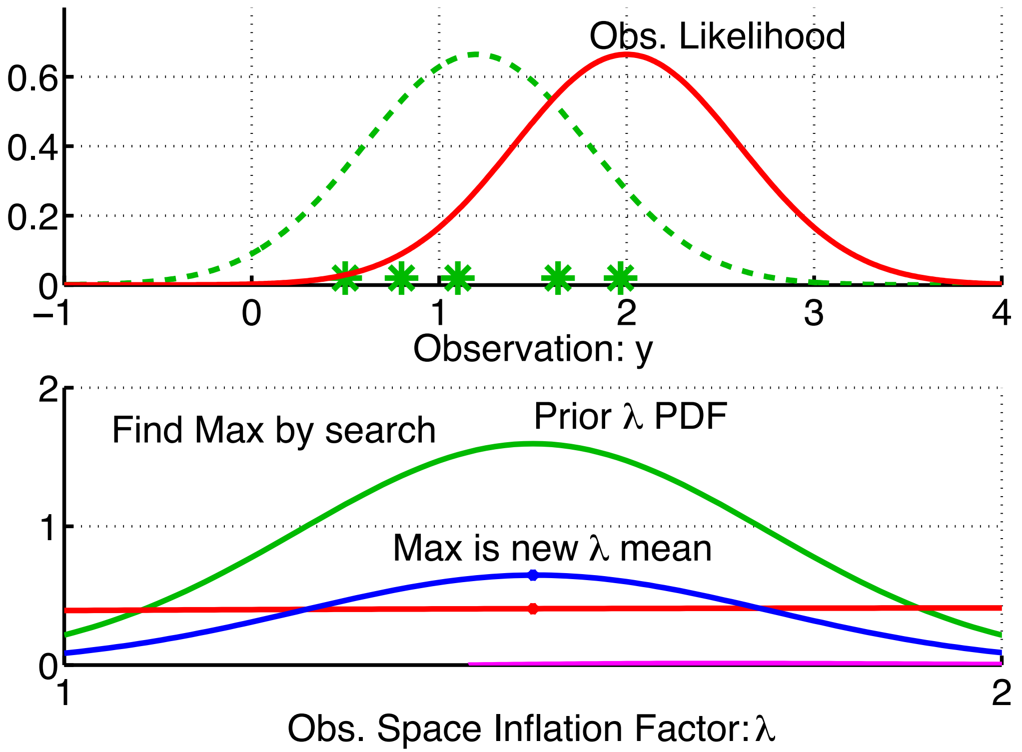
Posterior and prior are very similar.

Difference shows slight shift to larger values of λ.

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

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.

$$p\left(\lambda, t_k \mid Y_{t_k}\right) = p\left(y_k \mid \lambda\right) p\left(\lambda, t_k \mid Y_{t_{k-1}}\right) / normalization$$

Variance inflation for observations: An adaptive error tolerant filter

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

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

Solving $\partial\left[p(y_k \mid \lambda)\, p(\lambda, t_k \mid Y_{t_{k-1}})\right] / \partial y = 0$ leads to 6^{th} 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

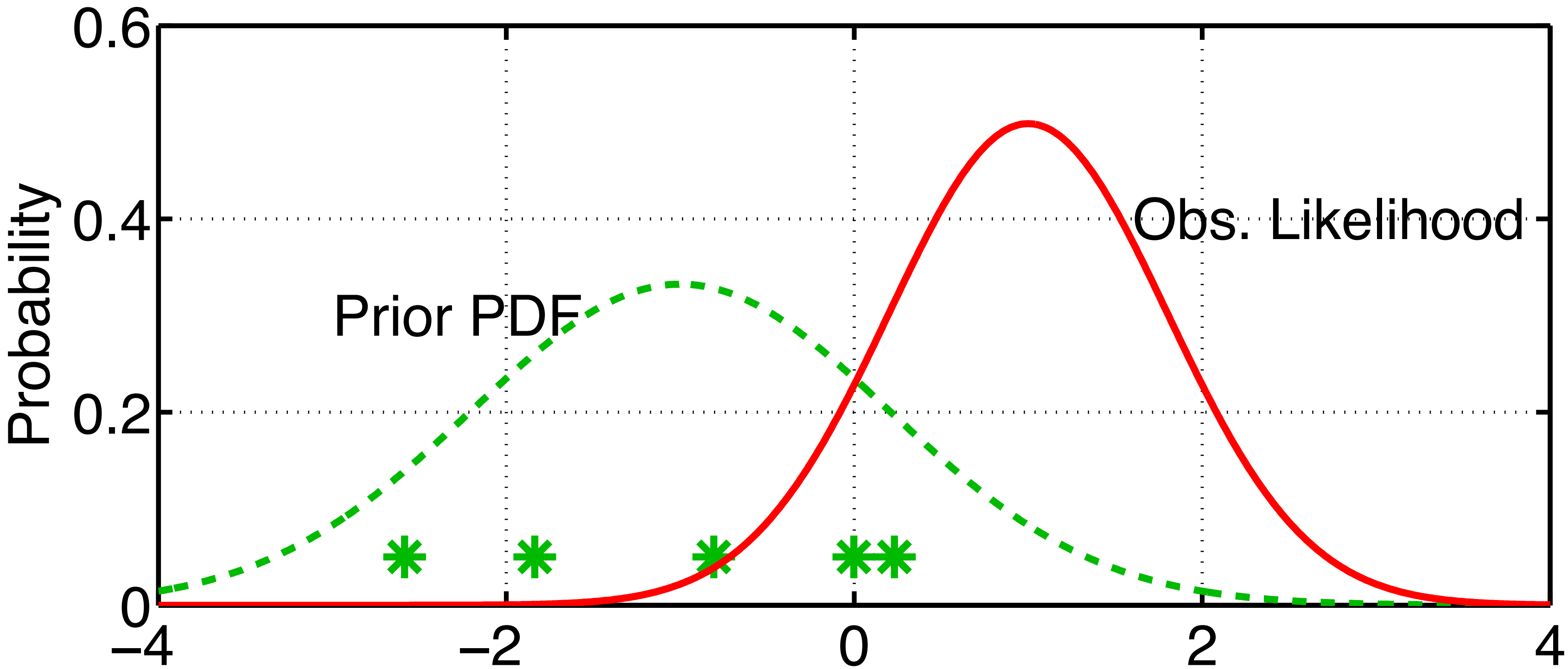
B. Computing updated inflation variance, $\sigma^2_{\lambda,u}$.

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

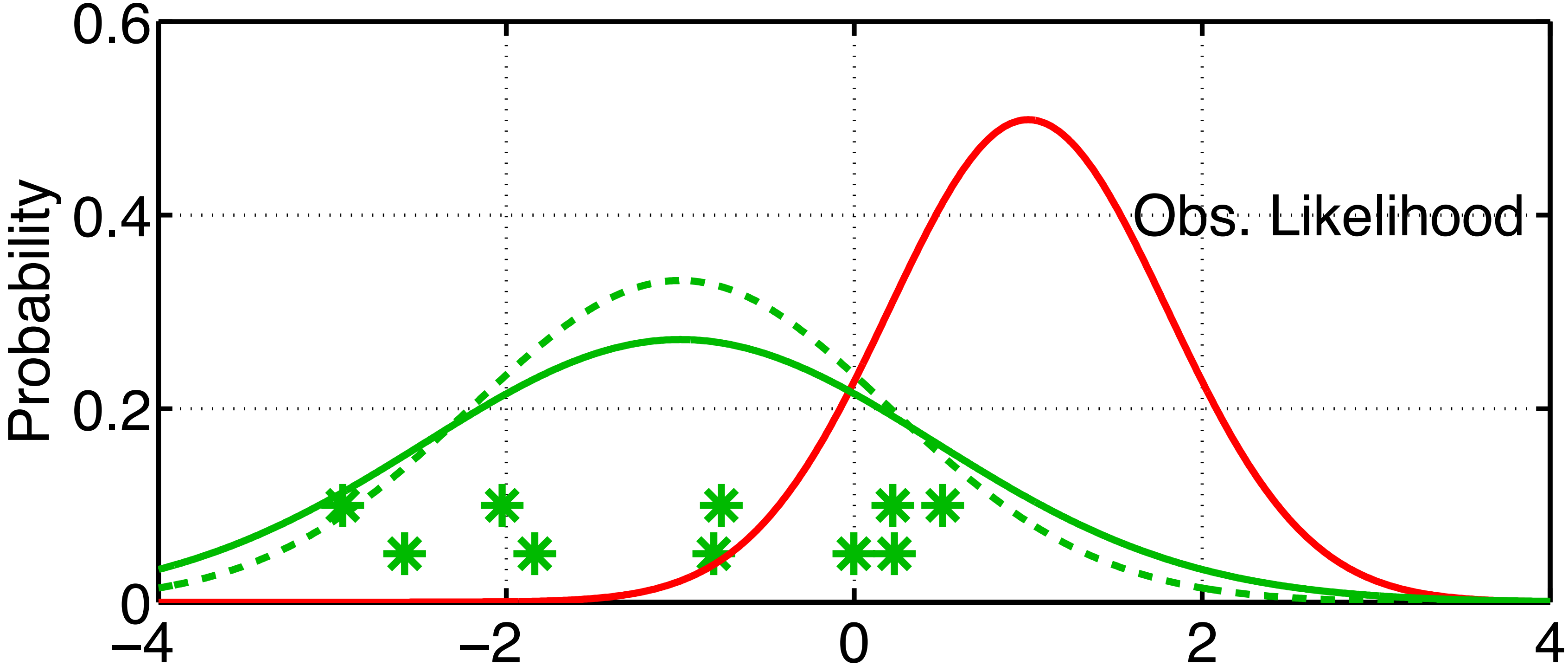
2. Find $\sigma^2_{\lambda,u}$ so $N\left(\bar{\lambda}_u, \sigma^2_{\lambda,u}\right)$ goes through $p\left(\bar{\lambda}_u\right)$ and $p\left(\bar{\lambda}_u + \sigma_{\lambda,p}\right)$

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

Observation Space Computations with Adaptive Error Correction

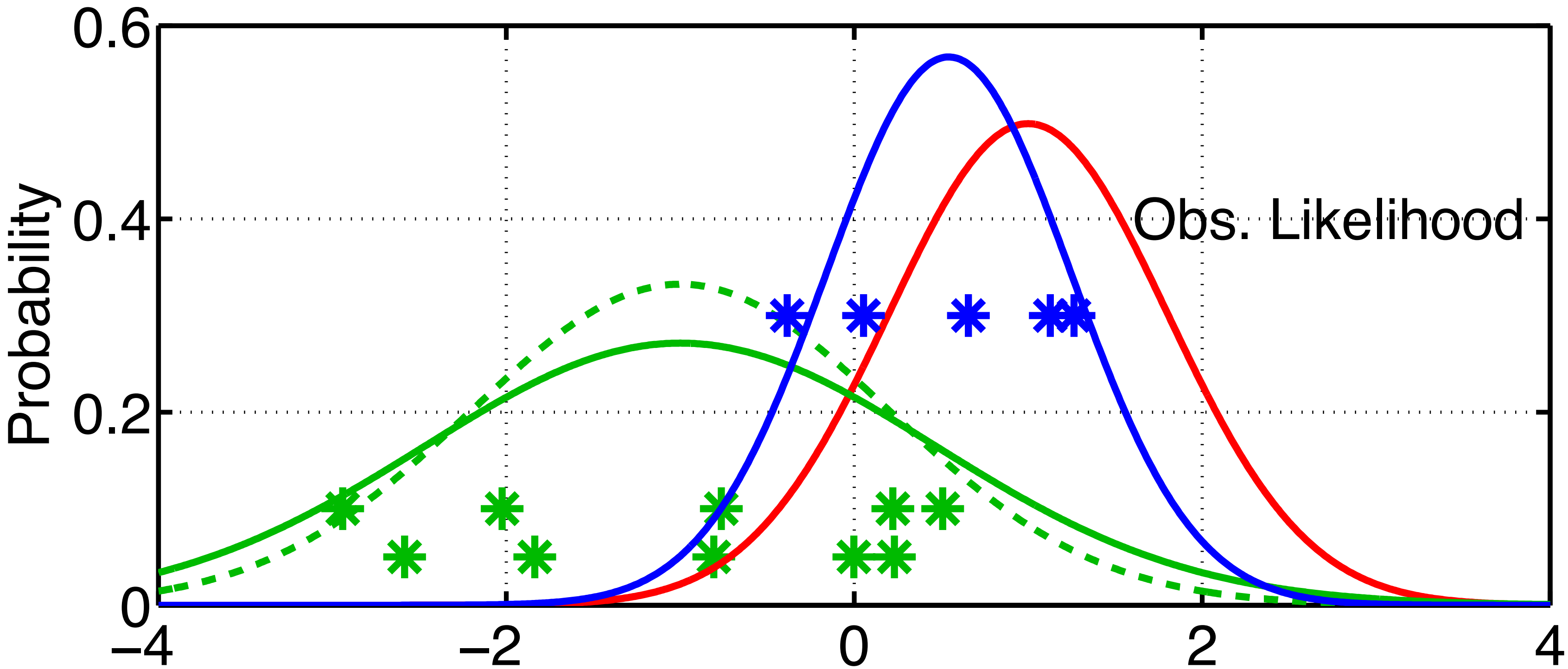


1. Compute updated inflation distribution, $p\left(\lambda, t_k \mid Y_{t_k}\right)$.



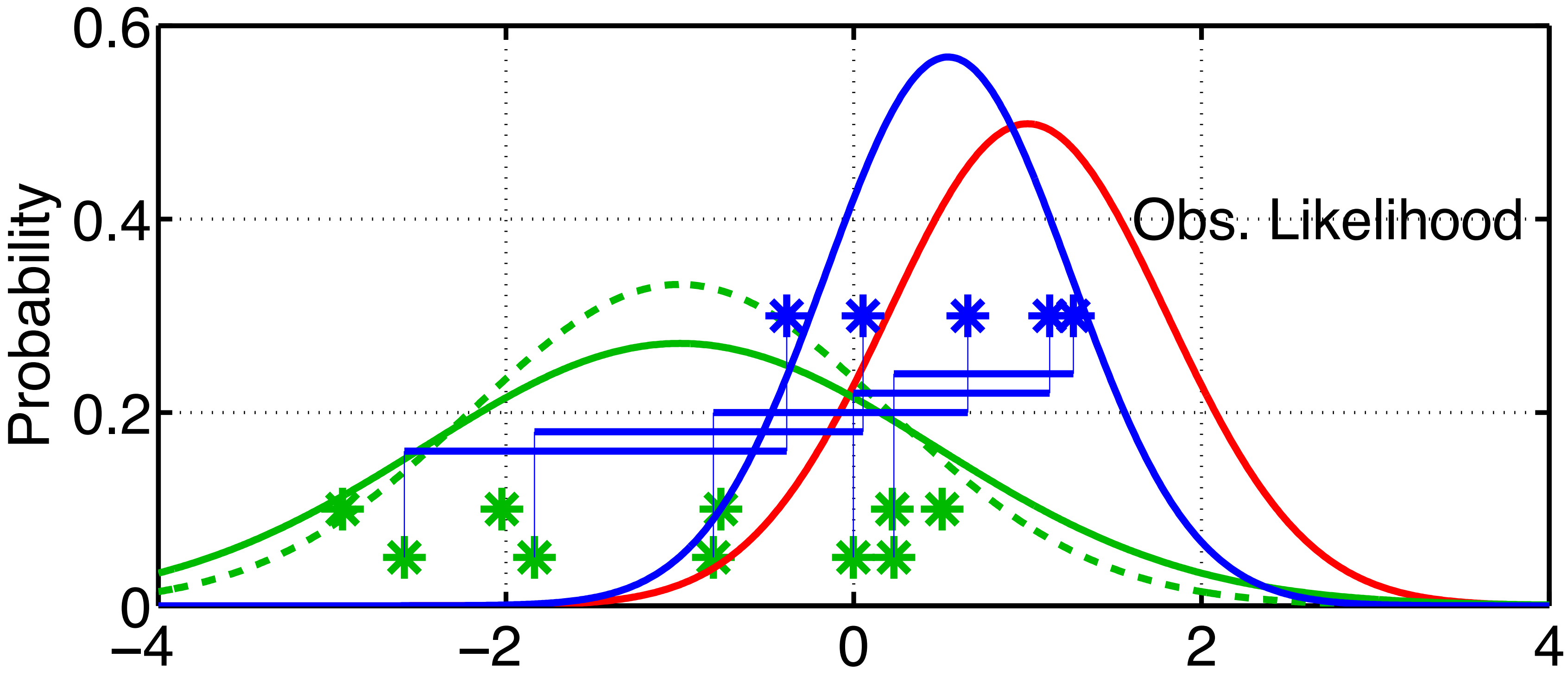
1. Compute updated inflation distribution, $p\left(\lambda, t_k \mid Y_{t_k}\right)$.
2. Inflate ensemble using mean of updated λ distribution.

Observation Space Computations with Adaptive Error Correction



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

Observation space adaptive inflation is not supported in DART Manhattan release

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)

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: $p(y_o \mid \lambda) = \left(2\pi\theta^2\right)^{-1/2} \exp\left(-D^2/2\theta^2\right)$

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

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.

Experimenting with spatially-constant state space inflation

	Before Assimilation	After Assimilation	
inf_flavor	= 3,	0,	
inf_initial_from_restart	= .false.,	.false.,	
inf_sd_initial_from_restart	= .false.,	.false.,	
inf_deterministic	= .true.,	.true.,	
inf_initial	= 1.00,	1.0,	Initial inflation value
inf_sd_initial	= 0.2,	0.0,	Initial standard deviation
inf_damping	= 1.0,	1.0,	
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.
	prior inflation	posterior inflation	

Flavor:
0=> NONE
2=> varying state space
3=> constant state space

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 spatially-constant state space inflation.

Experimenting with spatially-constant state space inflation

Run the filter:

Examine performance with *plot_total_err* in Matlab.

Time series of inflation mean and standard deviation are in *preassim.nc* file:

- This can be viewed with *ncview* (more on this later).
- Inflation adjusts with time.
- Inflation standard deviation is non-increasing with time.
- This file also has time series of the ensemble.

Final values of inflation for restart are in *filter_output.nc* file.

Adaptive Inflation Algorithmic Variants

1. Increase prior state 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 state space λ.

    Constant in space and time.

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

    We already tried this in section 9.

    Set *inflate_sd_initial* and *inf_sd_lower_bound* to 0.

    Set *inf_initial* to desired inflation value.

Adaptive Inflation 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 = 0.10 to 0.60 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 0.20,
   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.

Adaptive Inflation Algorithmic Variants

4. Inflation damping

Inflation mean damped towards 1.0 every assimilation time.

Set by namelist entry *inf_damping*.

inf_damping = 0.9: 90% of the inflation difference from 1.0 is retained.

Can be useful in models with heterogeneous observations in time.
For instance, a well-observed hurricane crosses a model domain.
Adaptive inflation increases along hurricane trace.
After hurricane, fewer observations, no longer need so much inflation.

For large earth system models, following values may work:

inf_sd_initial = 0.6,
inf_damping = 0.9,
inf_sd_lower_bound = 0.6.

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.

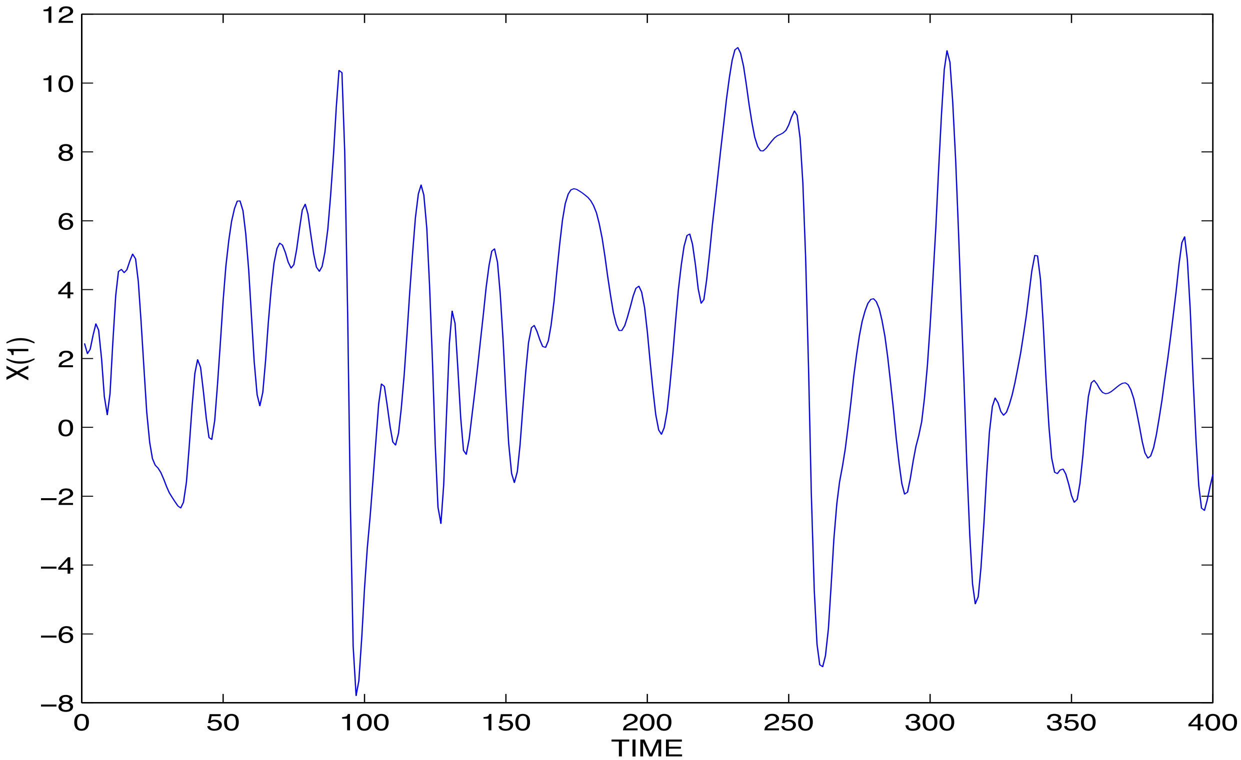
Simulating Model Error in 40-Variable Lorenz 96 Model

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 Lorenz 96 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

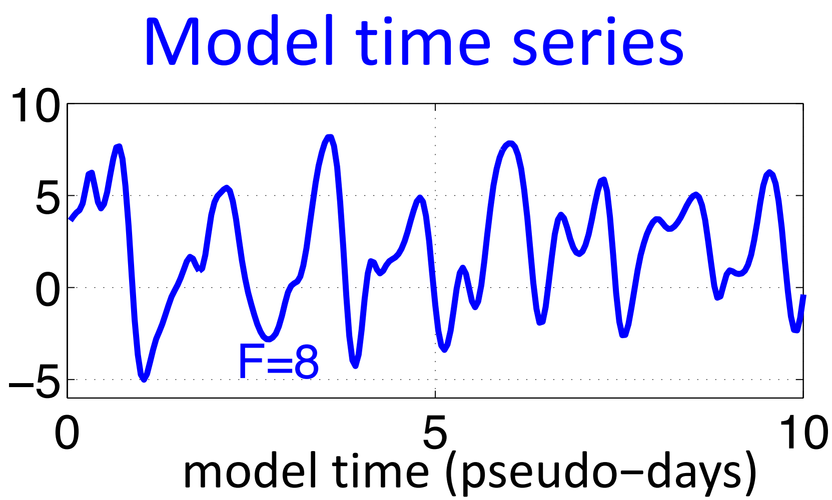
4 groups of 20 members each (80 ensemble members total)

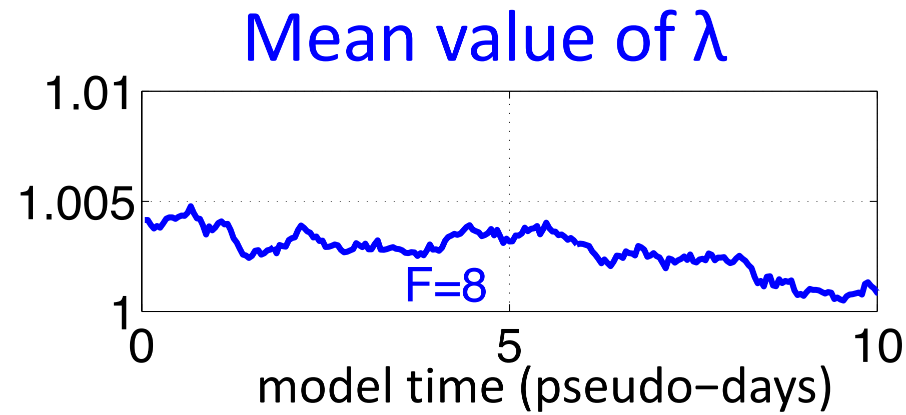
Results from 10 days after 40 day spin-up

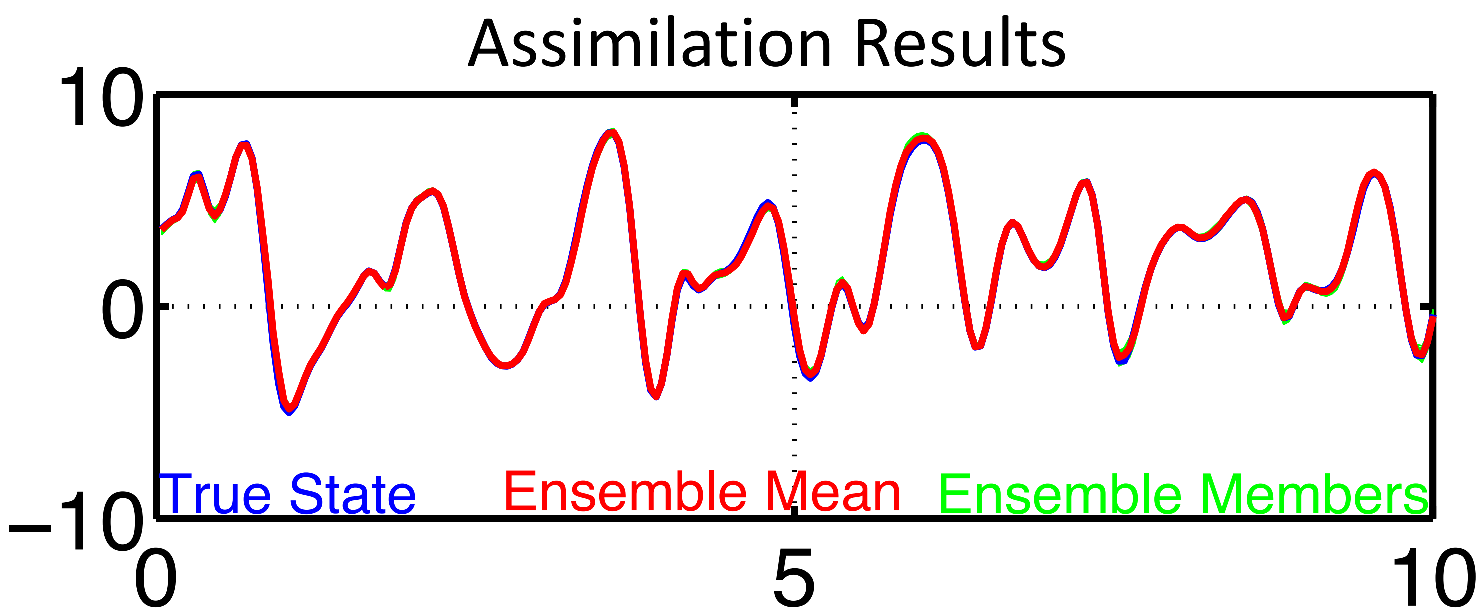
Vary assimilating model forcing: F=8, 6, 3, 0

Simulates increasing model error

Assimilating F=8 Truth with F=8 Ensemble

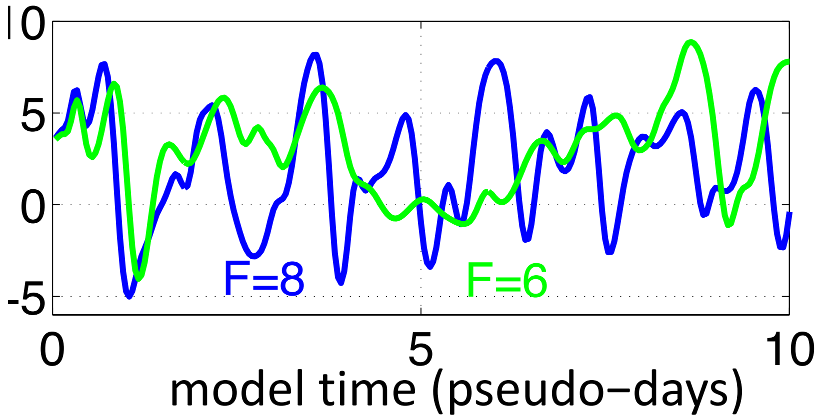




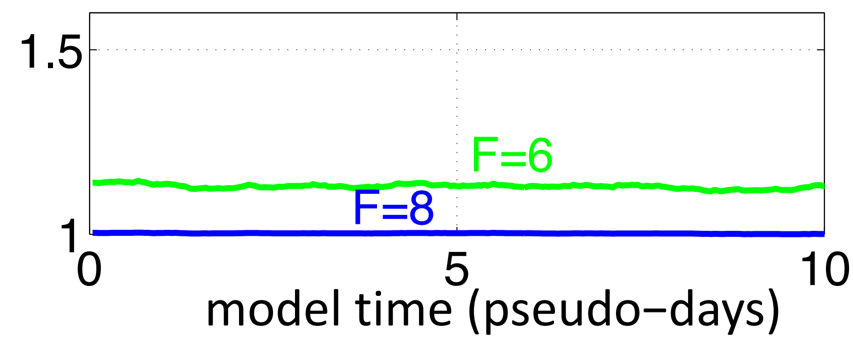


Assimilating F=8 Truth with F=6 Ensemble

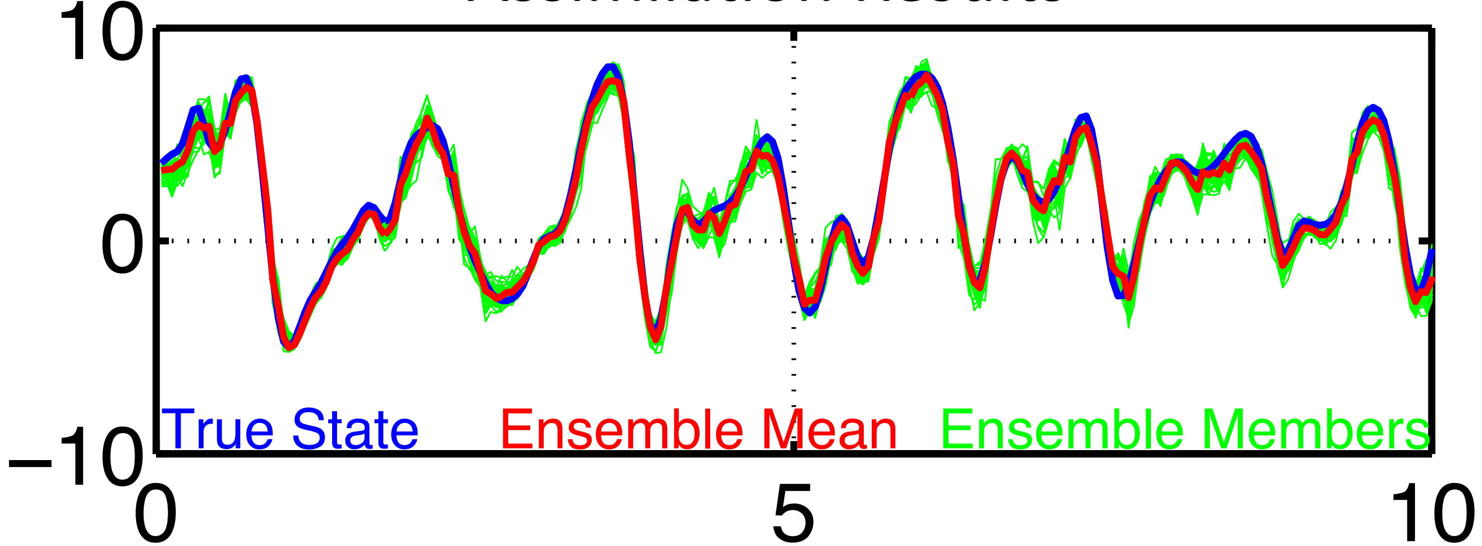
Model time series



Mean value of λ

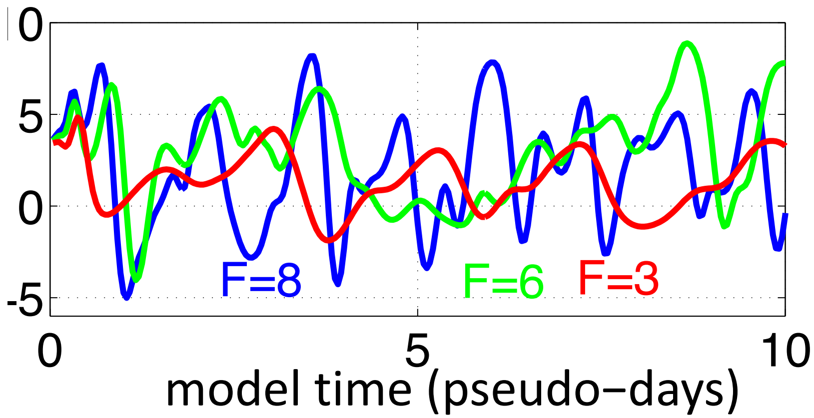


Assimilation Results

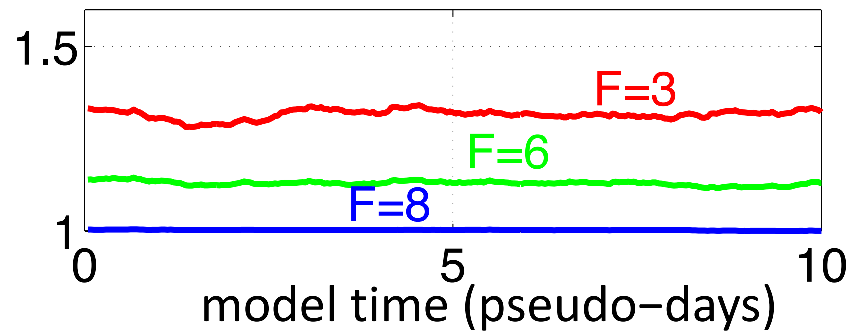


Assimilating F=8 Truth with F=3 Ensemble

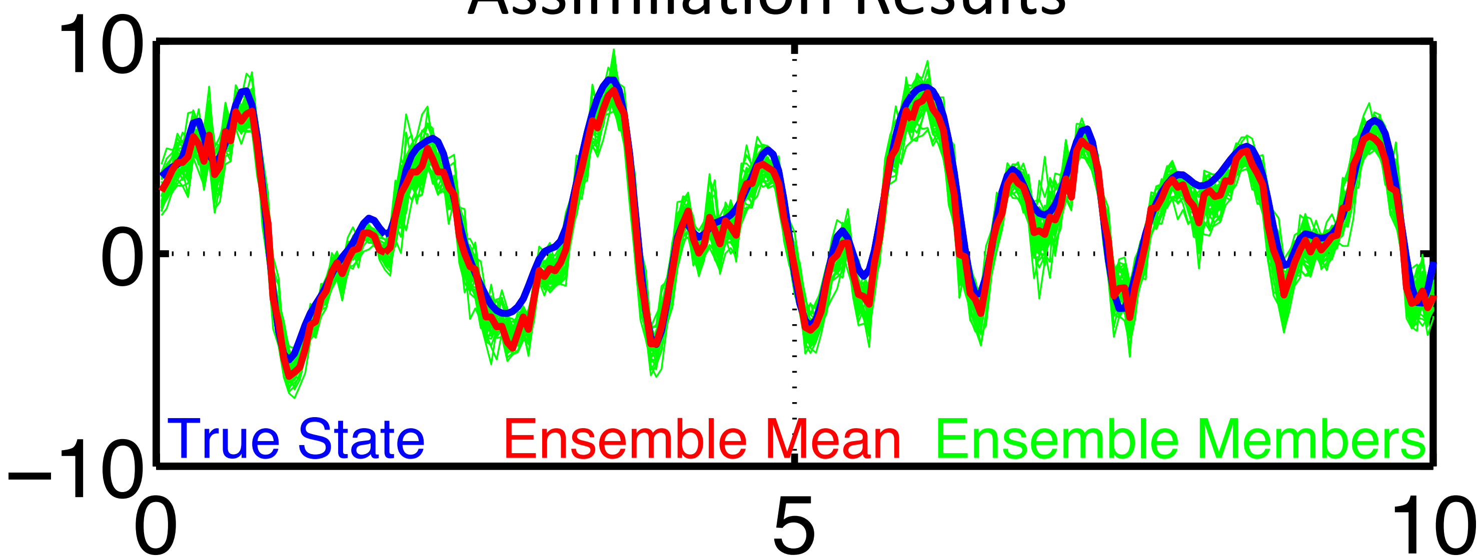
Model time series



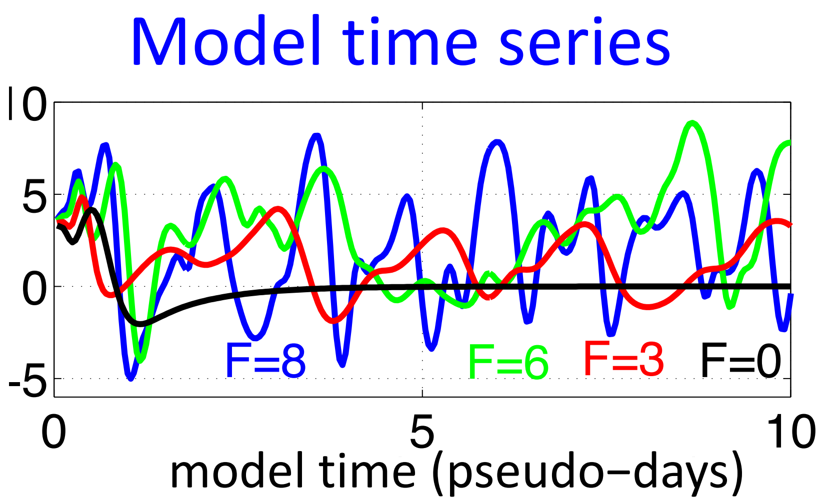
Mean value of λ

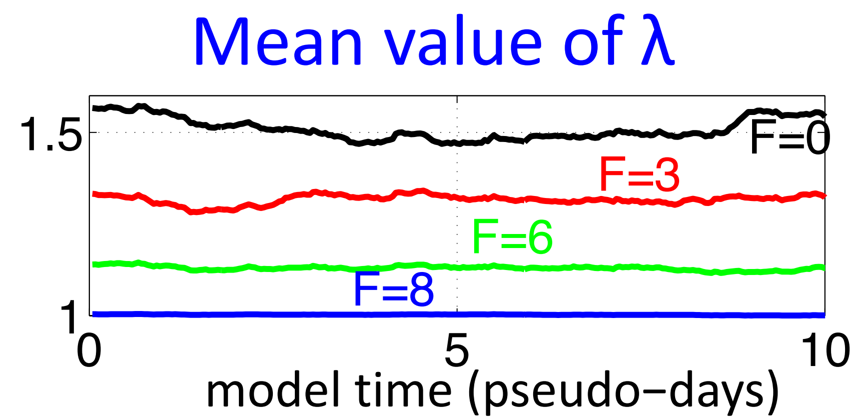


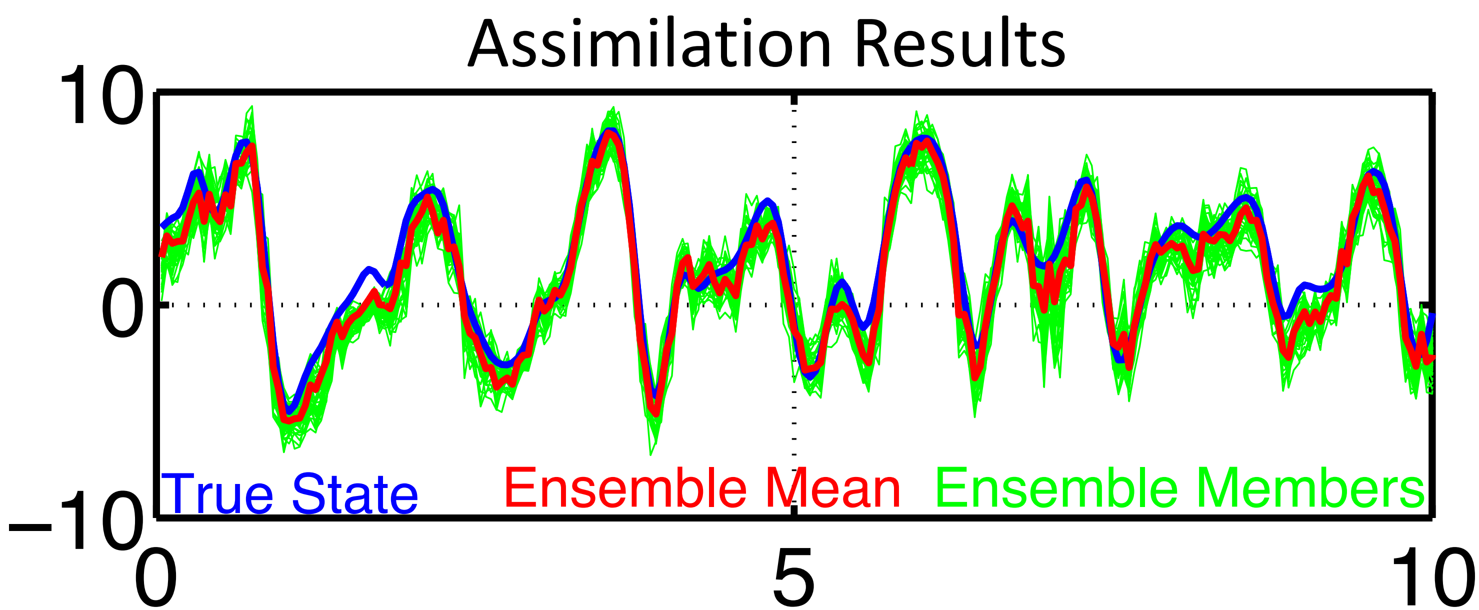
Assimilation Results



Assimilating F=8 Truth with F=0 Ensemble

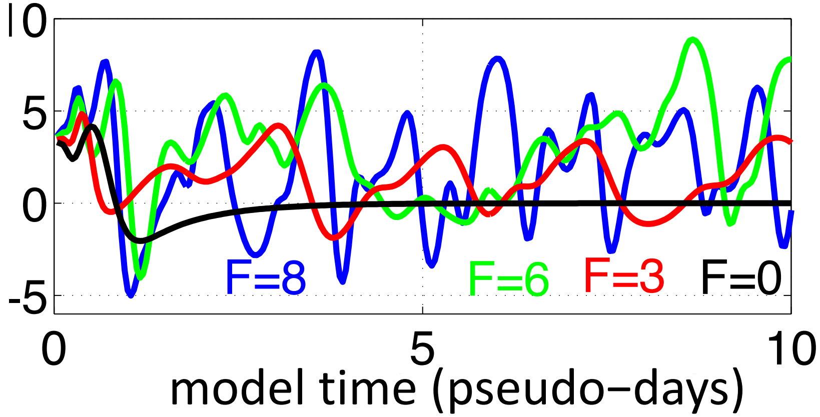




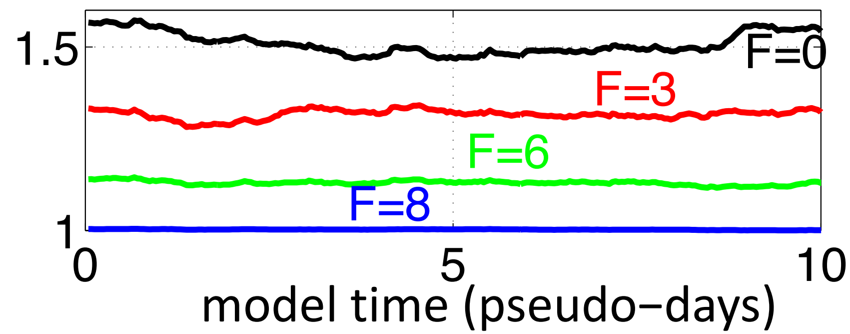


Assimilating F=8 Truth with F=0 Ensemble

Model time series

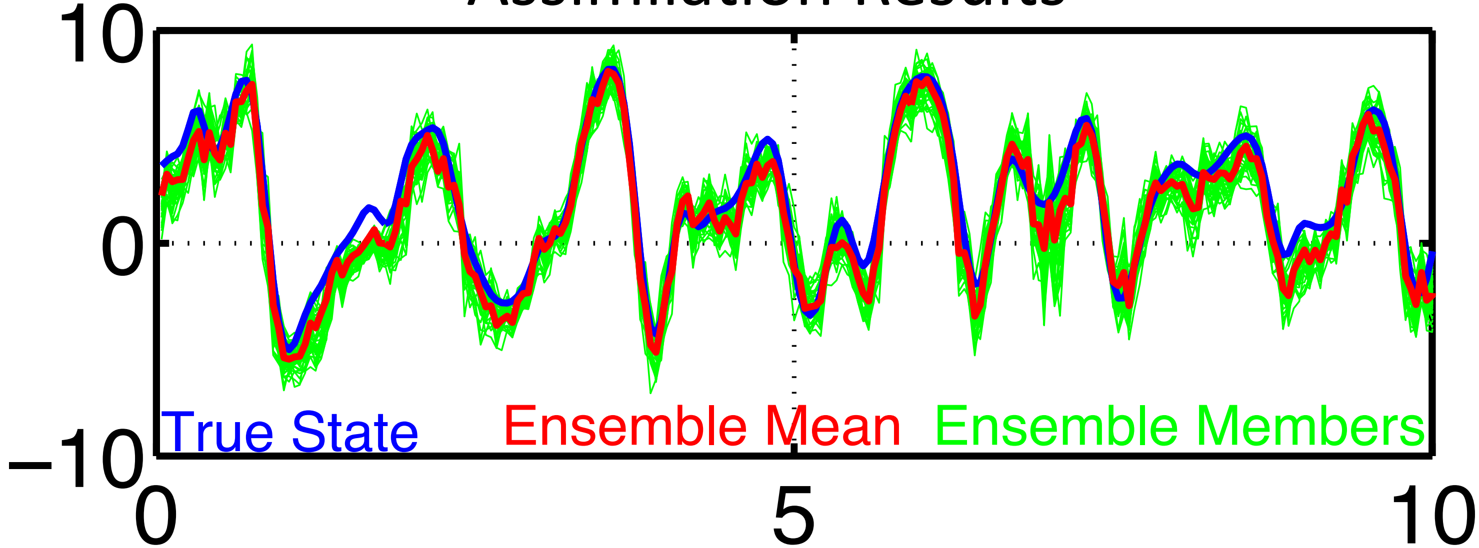


Mean value of λ

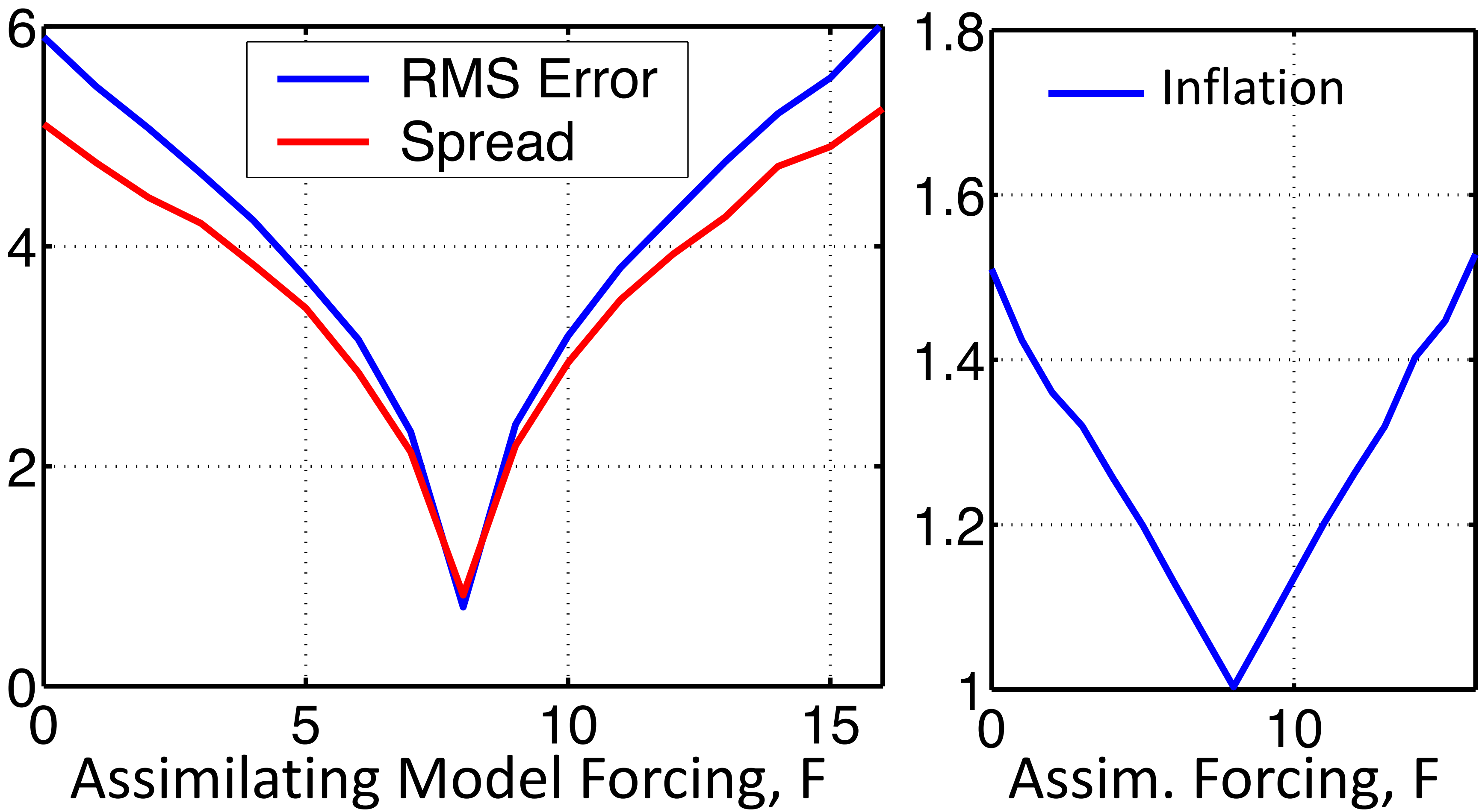


Prior RMS Error, Spread, and λ Grow as Model Error Grows

Assimilation Results



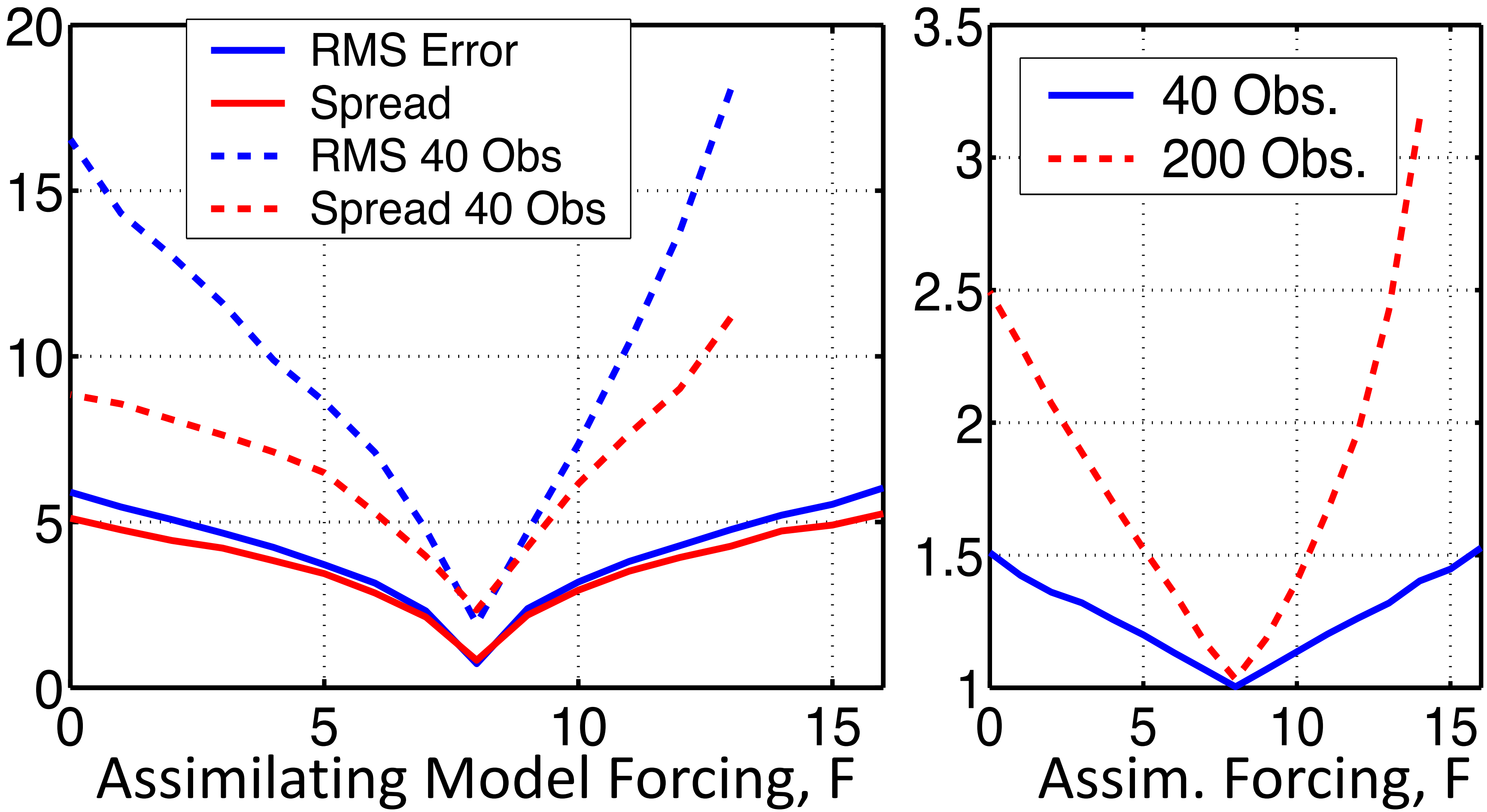
Base case: 200 randomly located observations per time



(Error saturation is approximately 30.0)

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

Less well observed case, 40 randomly located obs per time



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

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

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-varying state space inflation

	Before Assimilation	After Assimilation
inf_flavor	= 2,	0,
inf_initial_from_restart	= .false.,	.false.,
inf_sd_initial_from_restart	= .false.,	.false.,
inf_deterministic	= .true.,	.true.,
inf_initial	= 1.00,	1.0,
inf_sd_initial	= 0.2,	0.0,
inf_damping	= 1.0,	1.0,
inf_lower_bound	= 1.0,	1.0,
inf_upper_bound	= 1000000.0,	1000000.0,
inf_sd_lower_bound	= 0.0,	0.0,

Flavor: 2=> varying state space
3=> constant state space
0=> NONE

Initial inflation value

Initial standard deviation

Lower bound on s.d.

models/lorenz_96/work/

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 spatially-varying state space inflation.

Experimenting with spatially-varying state space inflation

Can try this with the other algorithmic variants.

Spatially-varying adaptive inflation is the most common choice in DART for large earth system models.

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.

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'.

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**