

Ensemble Filters for Atmosphere and Ocean Data Assimilation

Jeffrey L. Anderson

3 June, 2003

I. Overview of 'Assimilation' Problem

II. A Bayesian context

III. A Monte Carlo approximation

IV. Several heuristic simplifications

V. An example from a simple atmospheric model

VI. Challenges and opportunities

The Data Assimilation Problem

Given:

1. A physical system (atmosphere, ocean...)

2. Observations of the physical system

Usually sparse and irregular in time and space

Instruments have error of which we have a (poor) estimate

Observations may be of 'non-state' quantities

Many observations may have very low information content

3. A model of the physical system

Usually thought of as approximating time evolution

Could also be just a model of balance (attractor) relations

Truncated representation of 'continuous' physical system

Often quasi-regular discretization in space and/or time

Generally characterized by 'large' systematic errors

May be ergodic with some sort of 'attractor'

The Data Assimilation Problem (cont.)

We want to increase our information about all three pieces:

1. Get an improved estimate of state of physical system

Includes time evolution and ‘balances’

Initial conditions for forecasts

High quality analyses (re-analyses)

2. Get better estimates of observing system error characteristics

Estimate value of existing observations

Design observing systems that provide increased information

3. Improve model of physical system

Evaluate model systematic errors

Select appropriate values for model parameters

Evaluate relative characteristics of different models

Examples:

A. Numerical Weather Prediction

Model: Global troposphere / stratosphere $O(1 \text{ degree by } 50 \text{ levels})$

Observations: radiosondes twice daily, surface observations, satellite winds, aircraft reports, etc.

B. Tropical Upper Ocean State Estimation (ENSO prediction)

Model: Global (or Pacific Basin) Ocean $O(1 \text{ degree by } 50 \text{ levels})$

Observations: Surface winds (possibly from atmospheric assimilation), TAO buoys, XBTs, satellite sea surface altimetry

C. Mesoscale simulation and prediction

Model: Regional mesoscale model (WRF), $O(1 \text{ km resolution})$

Observations: Radial velocity from Doppler radar returns

D. Global Carbon Sources and Sinks

Nonlinear Filtering

Dynamical system governed by (stochastic) DE

$$dx_t = f(x_t, t) + G(x_t, t)d\beta_t, \quad t \geq 0 \quad (1)$$

Observations at discrete times

$$y_k = h(x_k, t_k) + v_k; \quad k = 1, 2, \dots; \quad t_{k+1} > t_k \geq t_0 \quad (2)$$

Observational error is white in time and Gaussian

$$v_k \rightarrow N(0, R_k) \quad (3)$$

Complete history of observations is

$$Y_\tau = \{y_l; \quad t_l \leq \tau\} \quad (4)$$

Goal: Find probability distribution for state at time t

$$p(x, t | Y_t) \quad (5)$$

Nonlinear Filtering (cont.)

State between observation times obtained from DE

Need to update state given new observation

$$p(x, t_k | Y_{t_k}) = p(x, t_k | y_k, Y_{t_{k-1}}) \quad (6)$$

Apply Bayes' rule

$$p(x, t_k | Y_{t_k}) = \frac{p(y_k | x_k, Y_{t_{k-1}}) p(x, t_k | Y_{t_{k-1}})}{p(y_k | Y_{t_{k-1}})} \quad (7)$$

Noise is white in time (3) so

$$p(y_k | x_k, Y_{t_{k-1}}) = p(y_k | x_k) \quad (8)$$

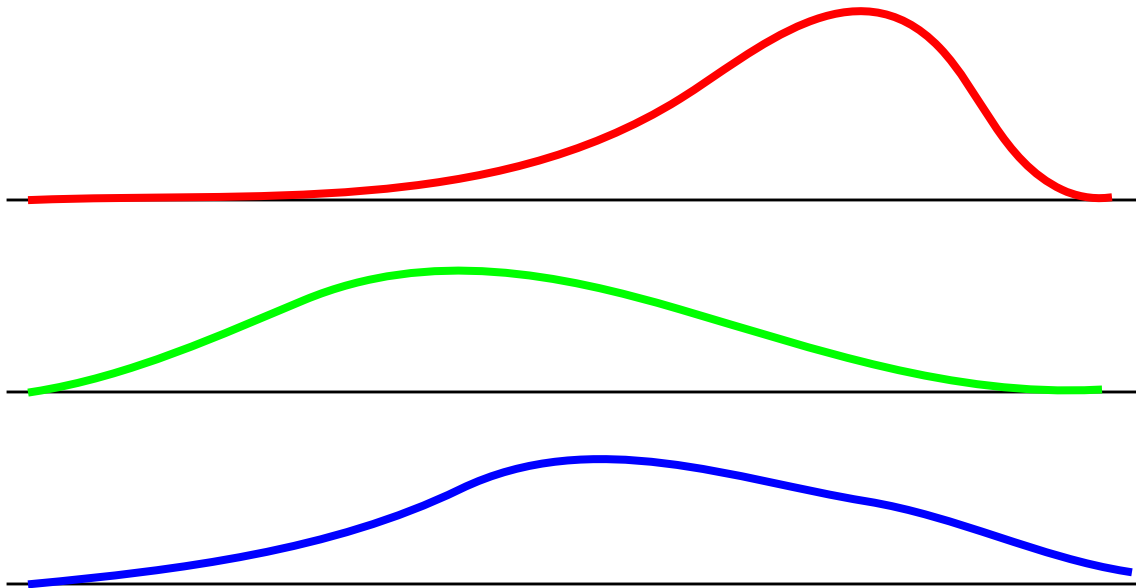
Also have

$$p(y_k | Y_{t_{k-1}}) = \int p(y_k | x) p(x, t_k | Y_{t_{k-1}}) dx \quad (9)$$

Nonlinear Filtering (cont.)

Probability after new observation

$$p(x, t_k | Y_{t_k}) = \frac{p(y_k | x) p(x, t_k | Y_{t_{k-1}})}{\int p(y_k | \xi) p(\xi, t_k | Y_{t_{k-1}}) d\xi} \quad (10)$$



Second term in numerator, denominator comes from DE
First term comes from distribution of observational error

General methods for solving the filter equations are known:

1. Advancing state estimate in time
2. Taking product of two distributions

But, these methods are far too expensive for problems of interest

1. Huge model state spaces (10 is big!), NWP models at $O(10 \text{ million})$
2. Need truncated representations of probabilistic state to avoid exponential solution time and storage

The ART of Data Assimilation:

Find heuristic simplifications that make approximate solution affordable

1. Localization (spatial or other truncated basis)
2. Linearization of models, represent time evolution as linear (around a control non-linear trajectory)
3. Represent distributions as Gaussian (or sum of Gaussians)
4. Monte Carlo methods
5. Application of simple balance relations

Ensemble Kalman Filters:

Simplifications:

1. Monte Carlo approximation to probability distributions
2. Localization in space, avoids degeneracy from samples smaller than state space
3. Gaussian representation of probability distributions generally used for computing update

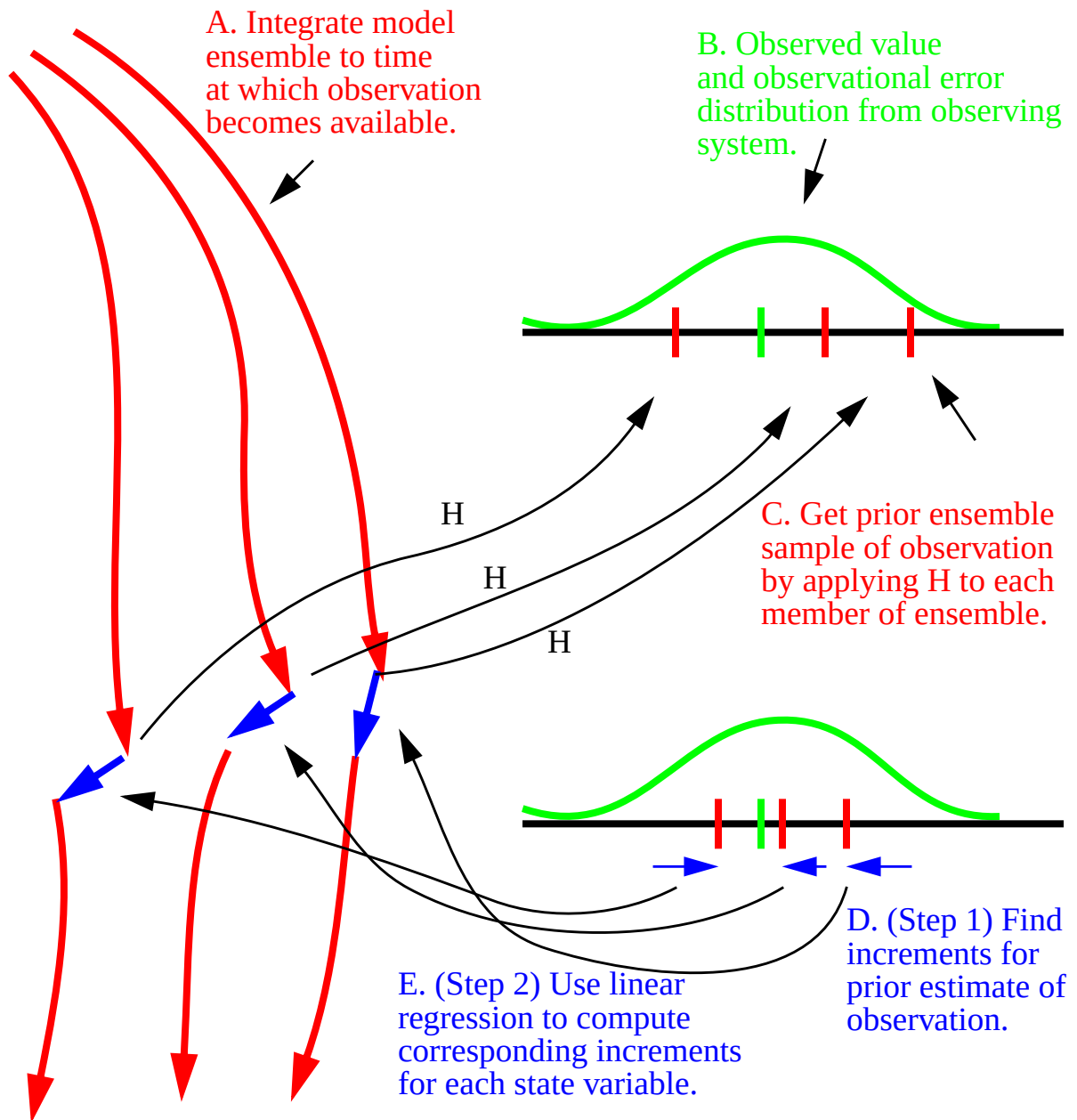
Problems:

1. Selecting initial samples for ensembles (Monte Carlo samples)
2. Determining degree of spatial localization
3. Maintaining appropriate model 'balances' in ensemble members

BUT, UNPRECEDENTED EASE OF INITIAL APPLICATION

How an Ensemble Filter Works

Theory: Impact of observations can be handled sequentially
Impact of observation on each state variable can be handled sequentially



Details of Step 1: Updating Observation Variable Ensemble

Scalar Problem: Wide variety of options available and affordable

Begin with two previously documented methods:

1. Perturbed Observation Ensemble Kalman Filter
 2. Ensemble Adjustment Kalman Filter
-

Both make use of following (key to Kalman filter...)

Given prior ensemble with sample mean $\bar{z}^p$ and covariance Σ^p

Observation y^o with observational error variance matrix R

Note: Product of Gaussians is Gaussian

$$\Sigma^u = \left\{ (\Sigma^p)^{-1} + H^T R^{-1} H \right\}^{-1} \quad (9)$$

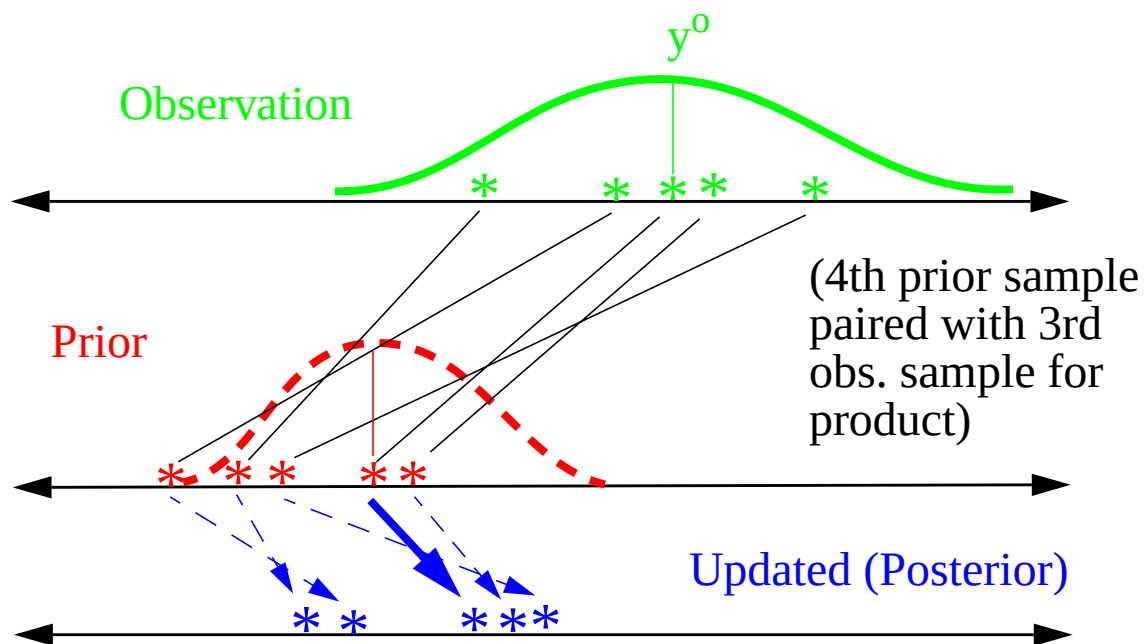
and mean:

$$\bar{z}^u = \Sigma^u \left\{ (\Sigma^p)^{-1} \bar{z}^p + H^T R^{-1} y^o \right\} \quad (10)$$

Details of Step 1: Perturbed Obs. Ensemble Kalman Filter

1. Compute prior sample variance and mean, Σ^p and $\bar{z}^p$
2. Apply (9) once to compute updated covariance, Σ^u
3. Create an N-member random sample of observation distribution by adding samples of obs. error to y^o
4. Apply (10) N times to compute updated ensemble members
Replace $\bar{z}^p$ with value from prior ensemble, y_i^p
Replace y^o with value from random sample, y_i^o
Updated ensemble value is y_i^u ($= \bar{z}^u$ from 10)

NOTE: When combined with linear regression for step 2, this gives identical results to EnKF's described in literature!



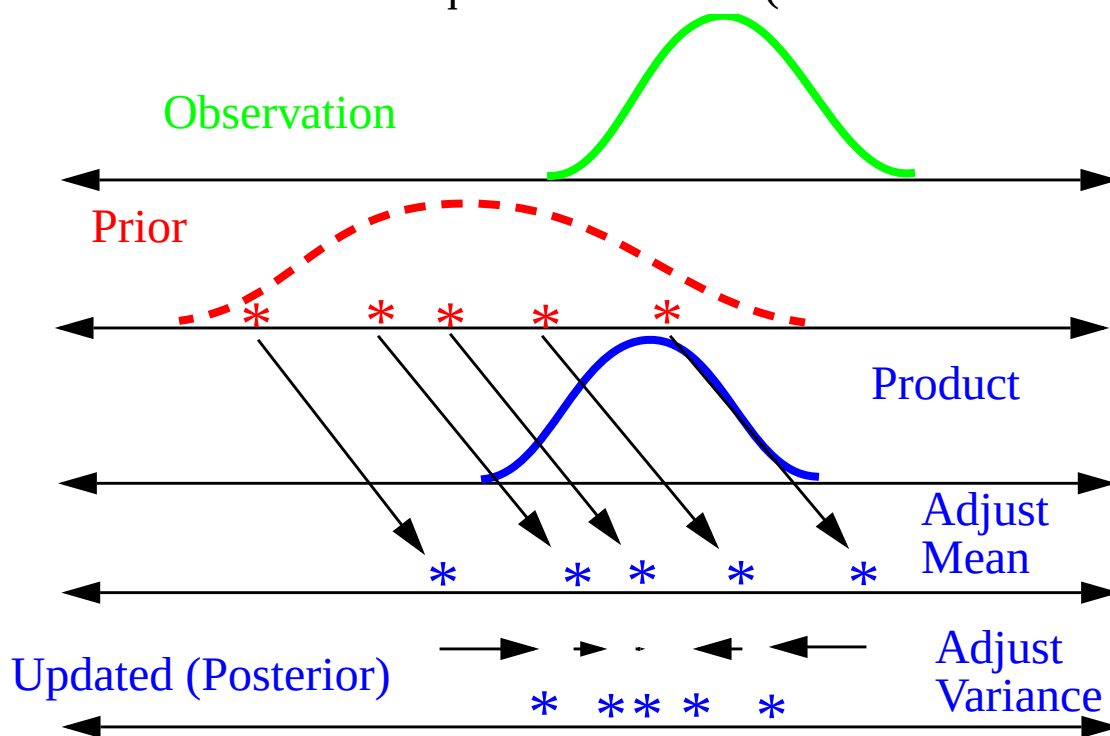
Details of Step 1: Ensemble Adjustment Kalman Filter

1. Compute prior sample variance and mean, Σ^p and $\bar{z}^p$
2. Apply (9) once to compute updated covariance, Σ^u
3. Apply (10) to compute updated mean, $\bar{z}^u$
4. Adjust prior ensemble of y so that mean and variance are exactly $\bar{z}^u$ and Σ^u

$$y_i^u = \left(y_i^p - \bar{y}^p \right) \sqrt{\sigma^u / \sigma^p} + \bar{y}^u, \quad i = 1, \dots, N$$

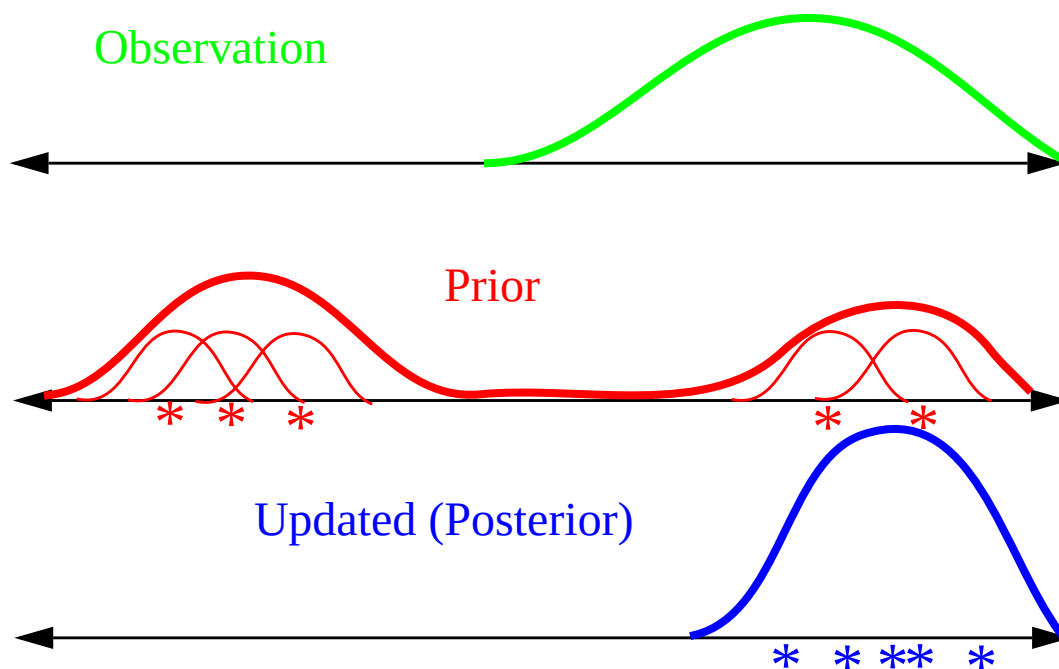
σ is variance of y

Similar methods called square root filters (for obvious reasons)



Details of Step 1: Quadrature kernel filter

1. Compute prior sample variance Σ^p
2. Use a Gaussian (or other) kernel approximation to get continuous approximation to $p(y^p)$
3. Use quadrature to take product in (6) directly
Can do individual Gaussians kernels if Gaussian
4. Create an N-member random sample of $p(z^u)$
5. An interesting variant uses boxcar kernels
6. Only useful for non-Gaussian structure in prior;
very powerful for Lorenz-63 model



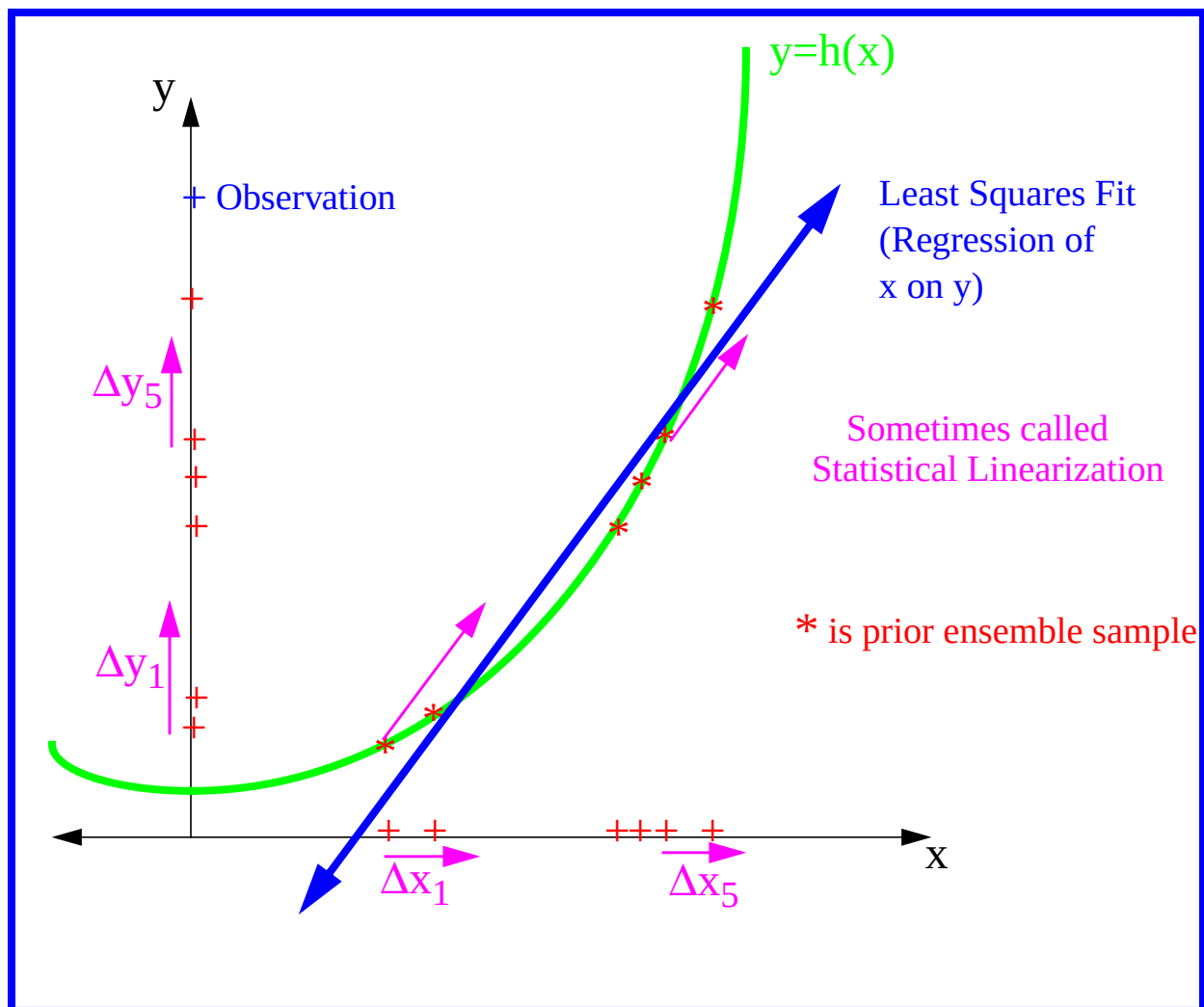
Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y , find increments for state variable ensembles

Simple idea: Do linear regression of x^p on y^p

Equivalent to doing: Least squares fit
 Assuming Gaussian prior relation

(Doing previously documented ensemble Kalman filters)

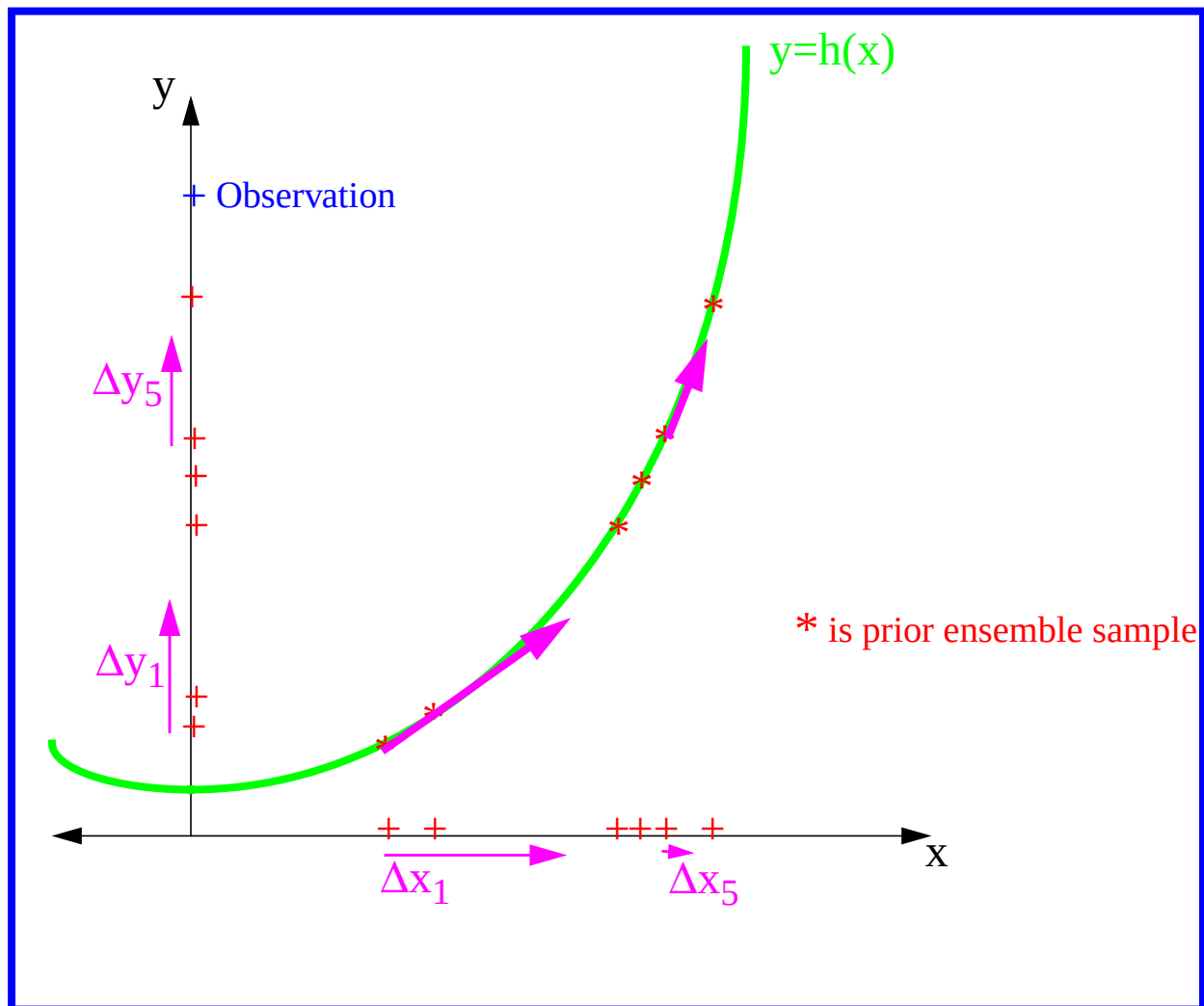


Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y , find increments for state variables

Idea: Do linear regression of x^p on y^p

Could also do local linearizations:
(Related to Gaussian Kernel approximation)

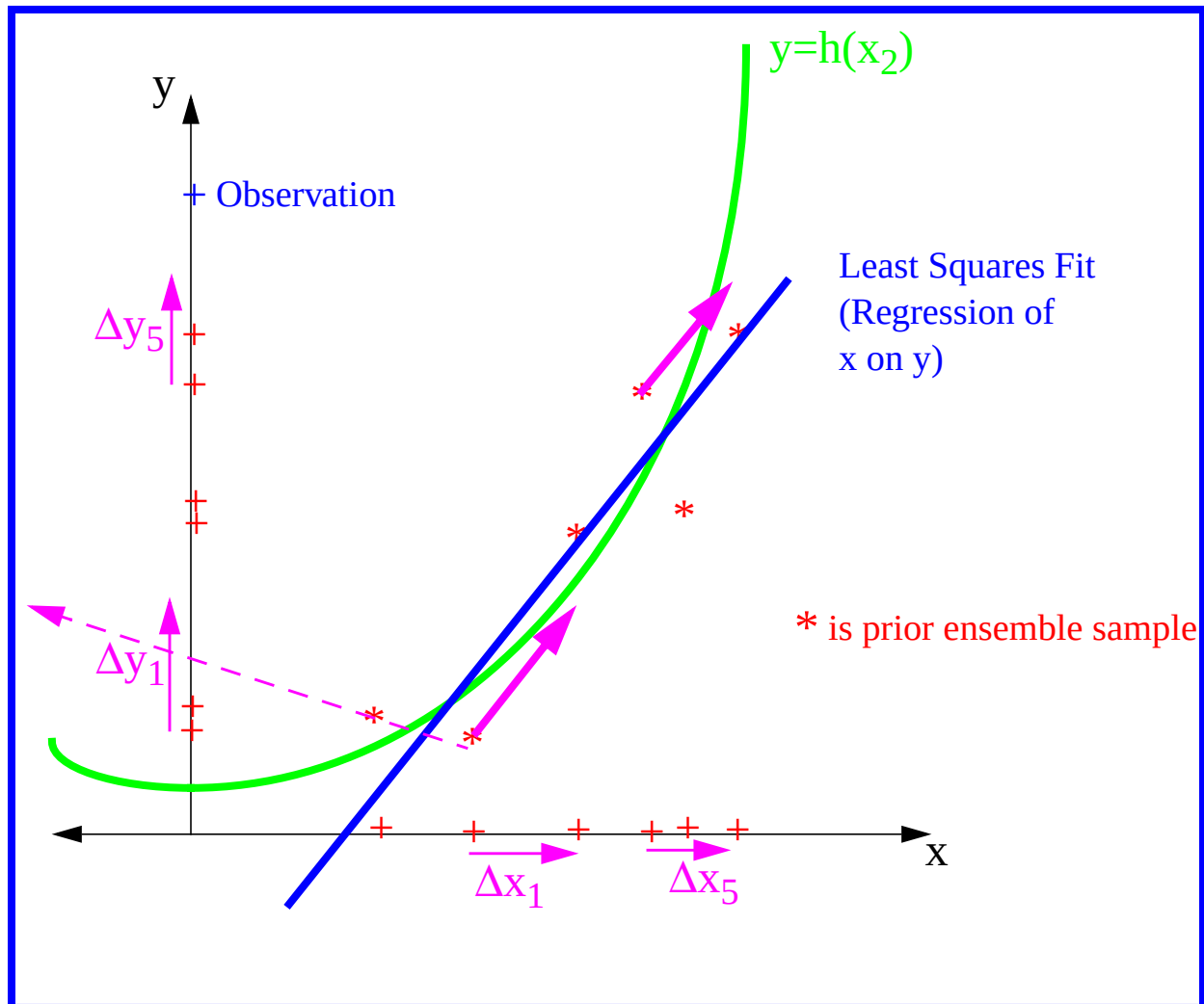


Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y , find increments for state variables

More challenging when obs and state are not functionally related

Example: $y = h(x_2)$, x and x_2 strongly correlated

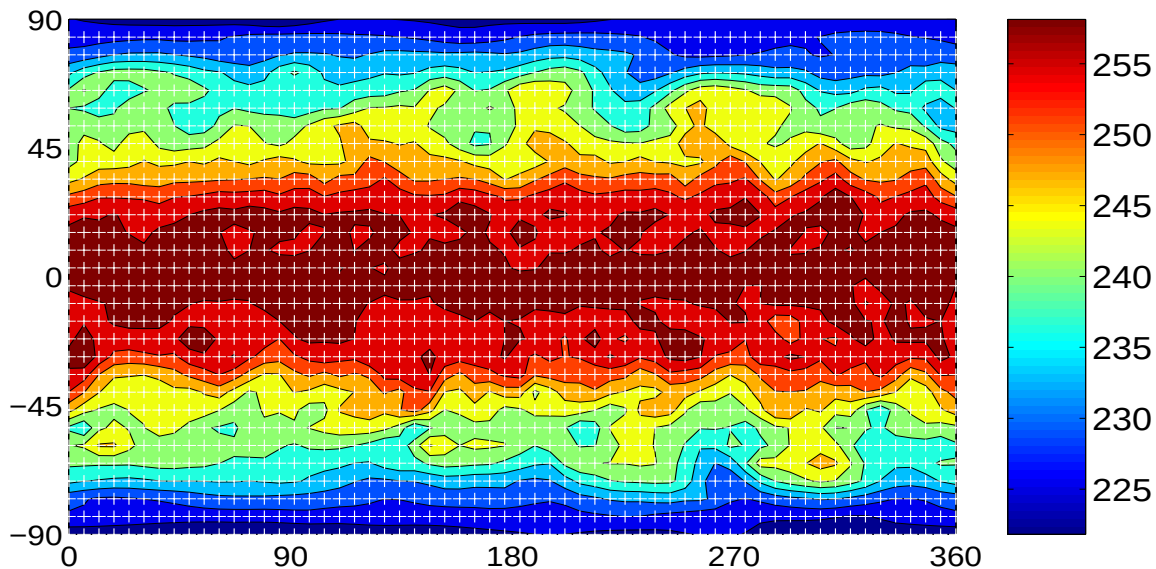


Large sample size needed to 'remove' noise

Trade-offs with local linearization (dotted magenta)

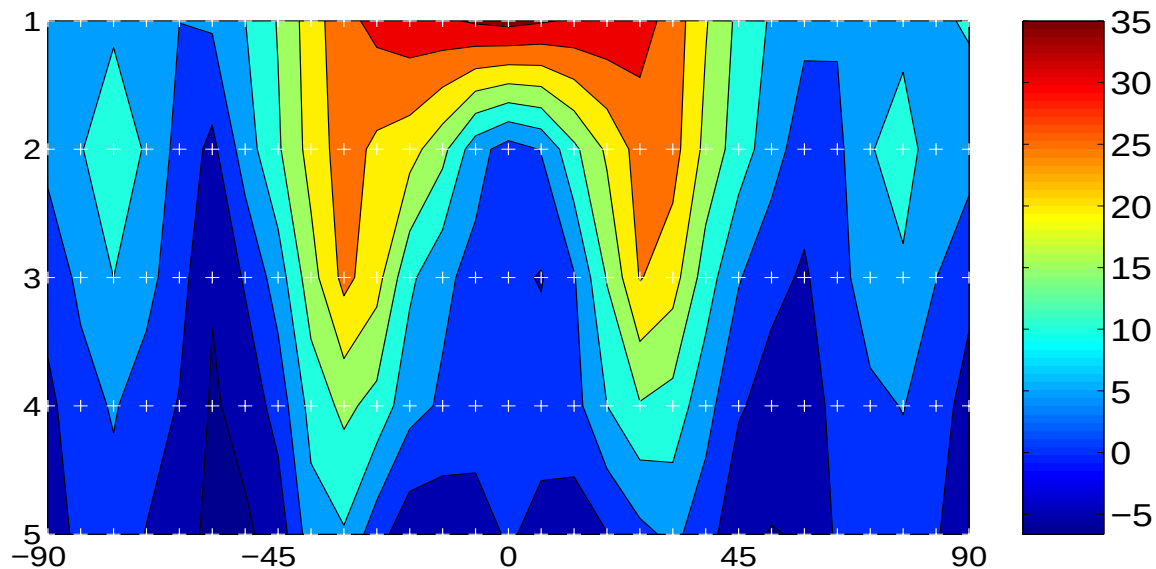
GFDL FMS B-Grid Dynamical Core (Havana)
Held-Suarez Configuration (no zonal variation, fixed forcing)

T at Level 3 (Middle of Atmosphere)

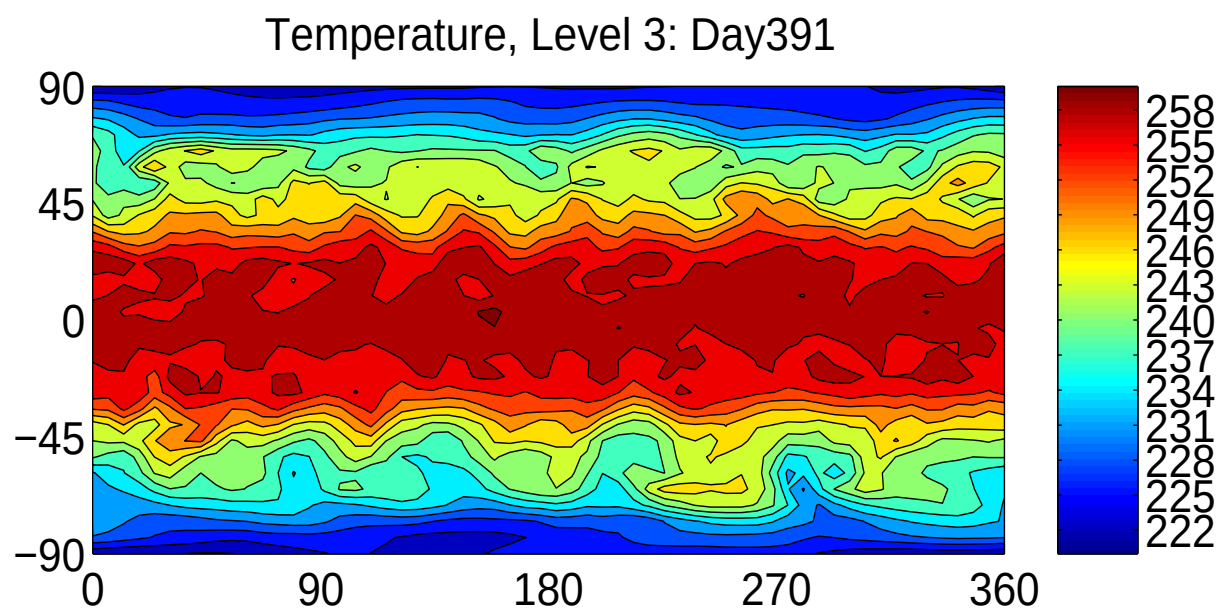
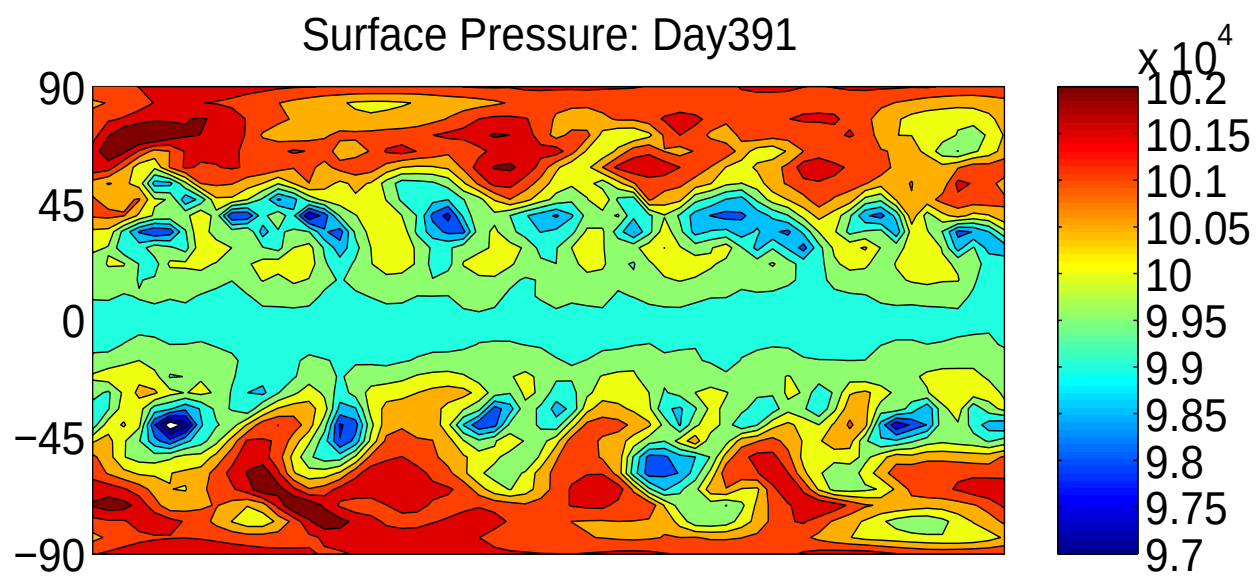


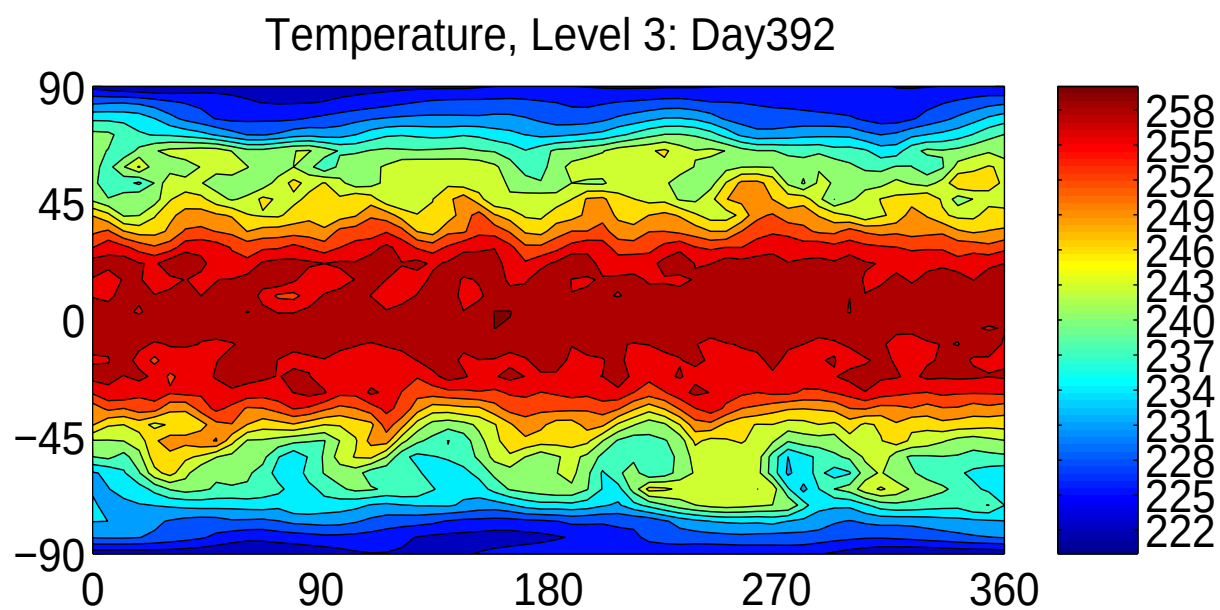
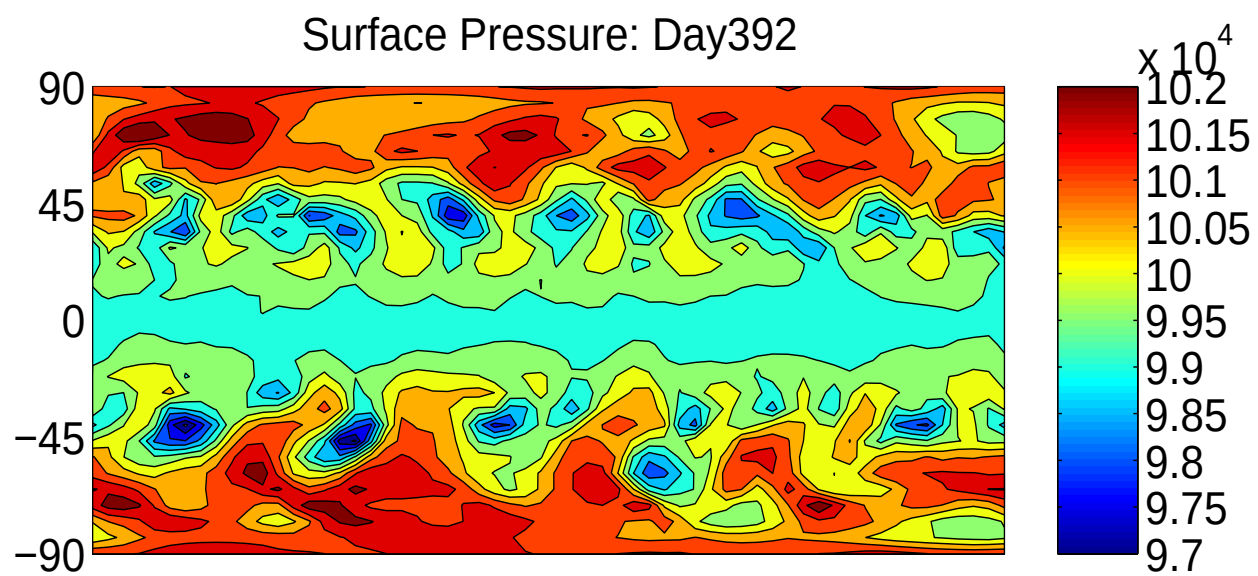
Low-Resolution (60 longitudes, 30 latitudes, 5 levels)
Damping coefficients reduced to 0.10 for error growth
Timestep 1 hour (or less for frequent observations)

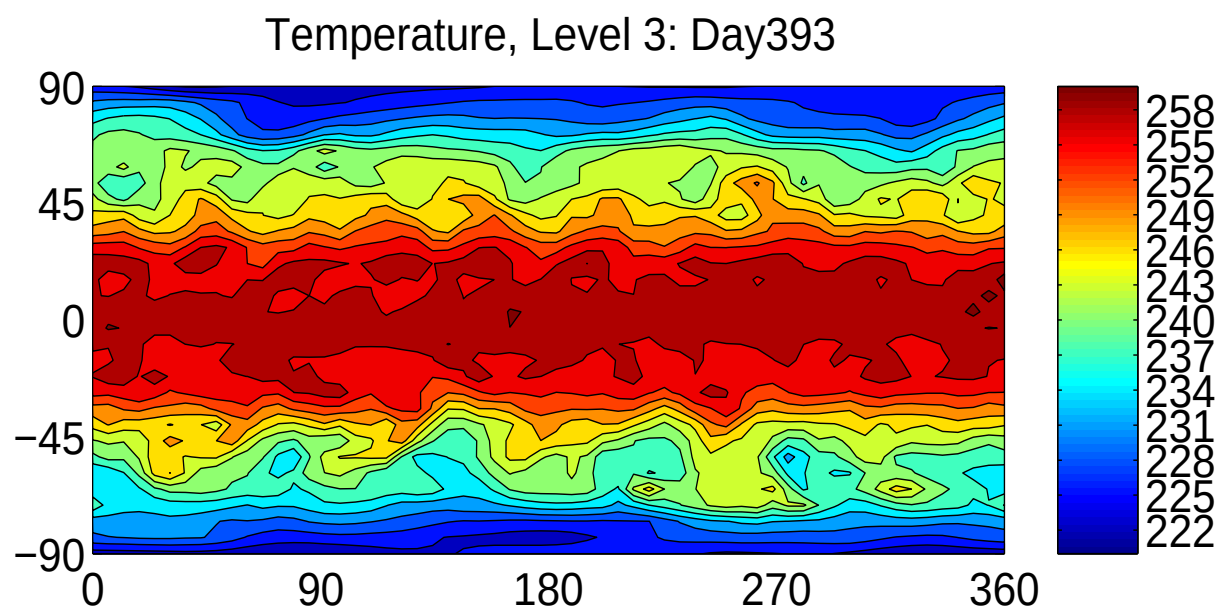
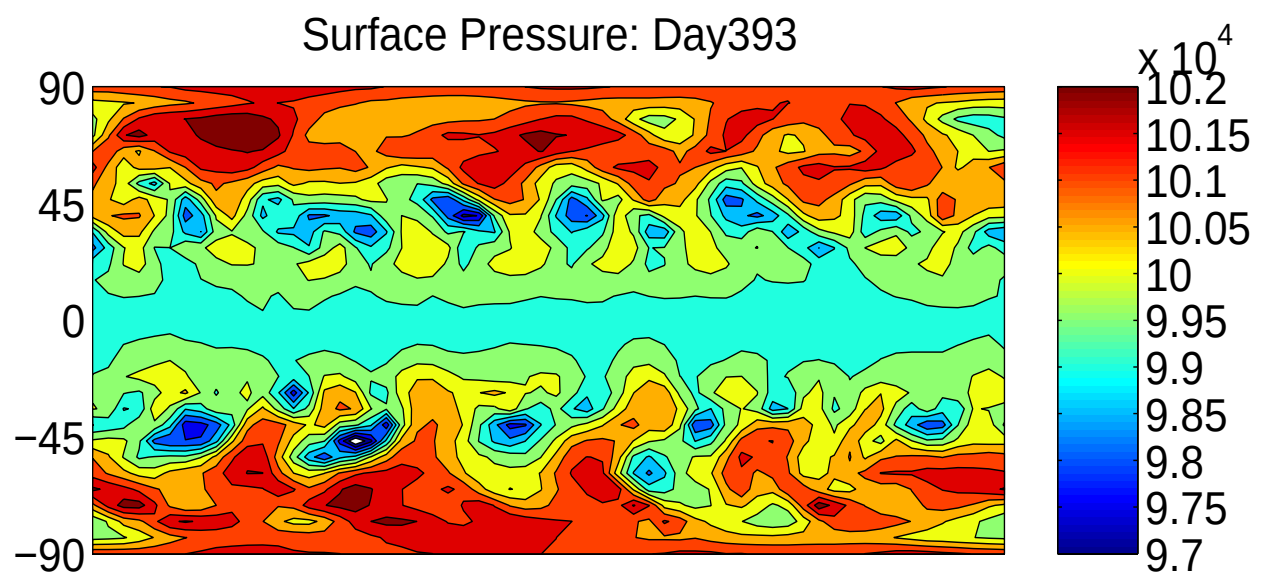
Cross section of Zonal mean U

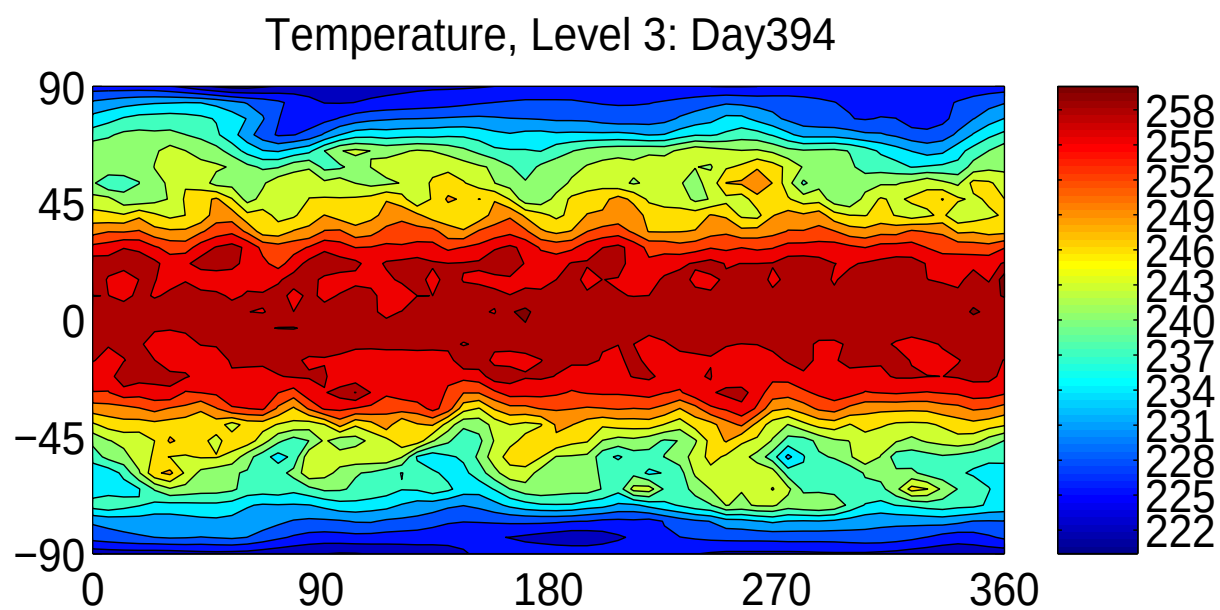
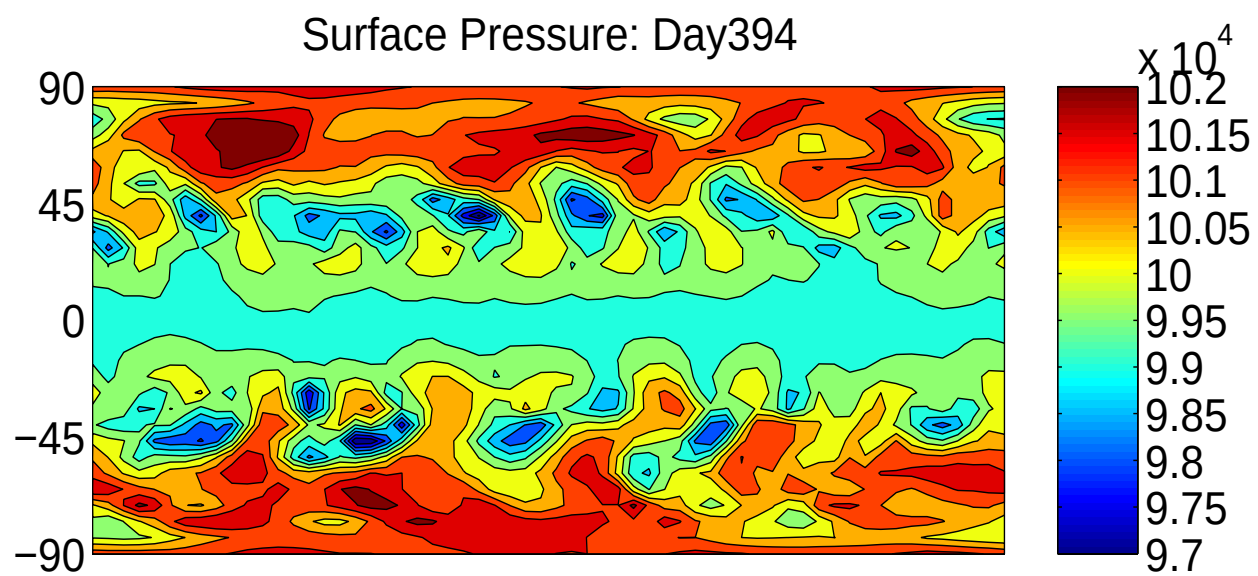


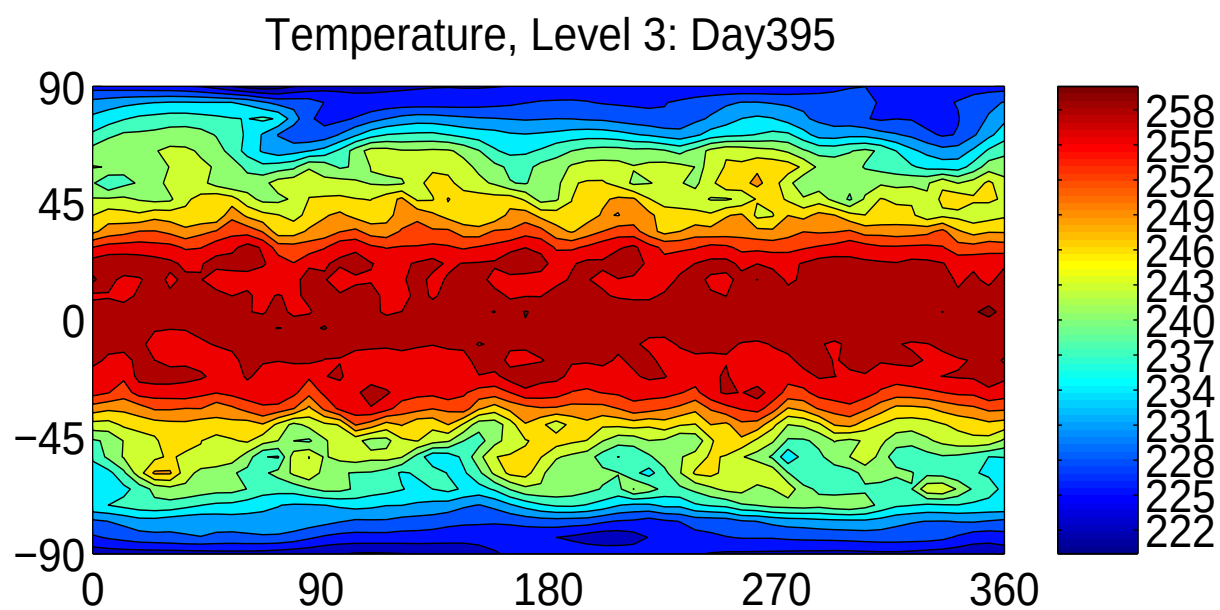
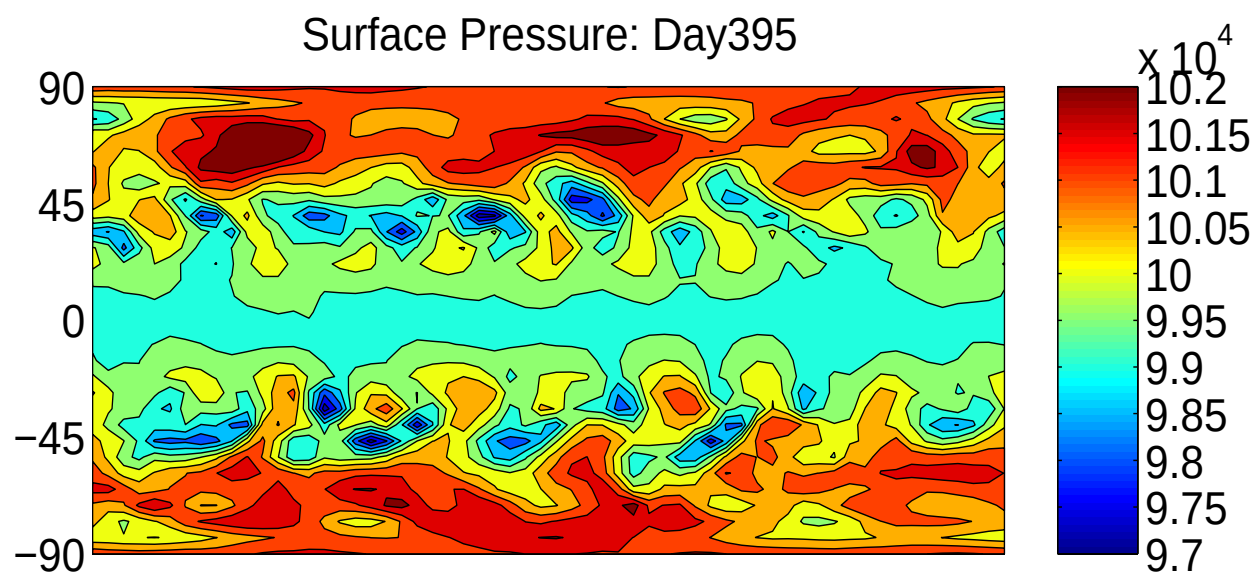
Has baroclinic instability

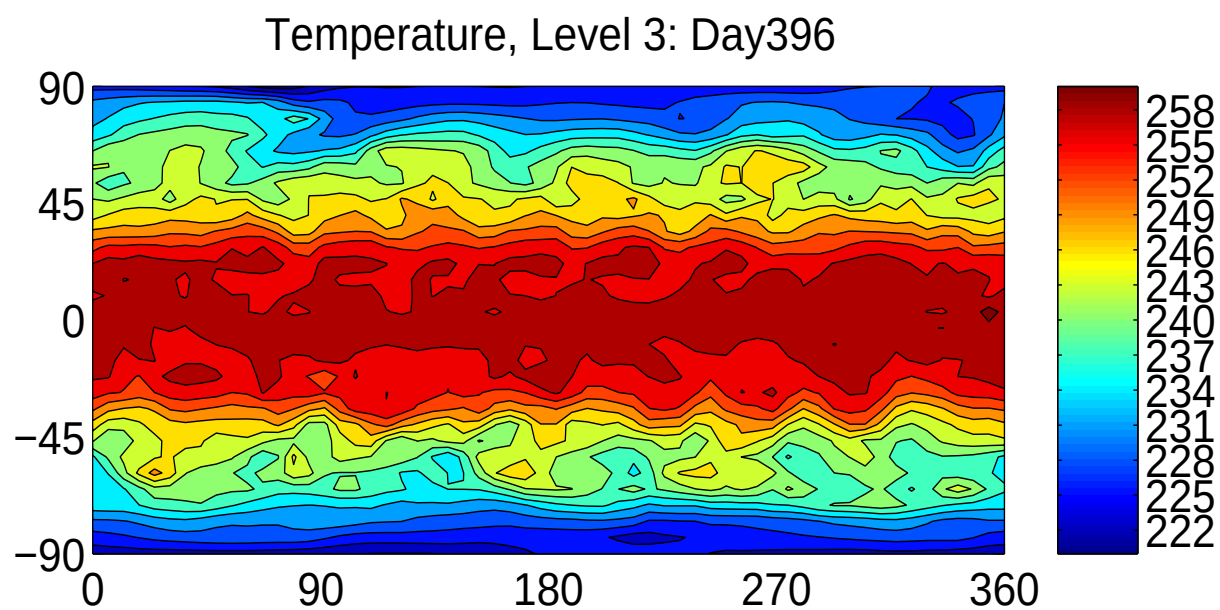
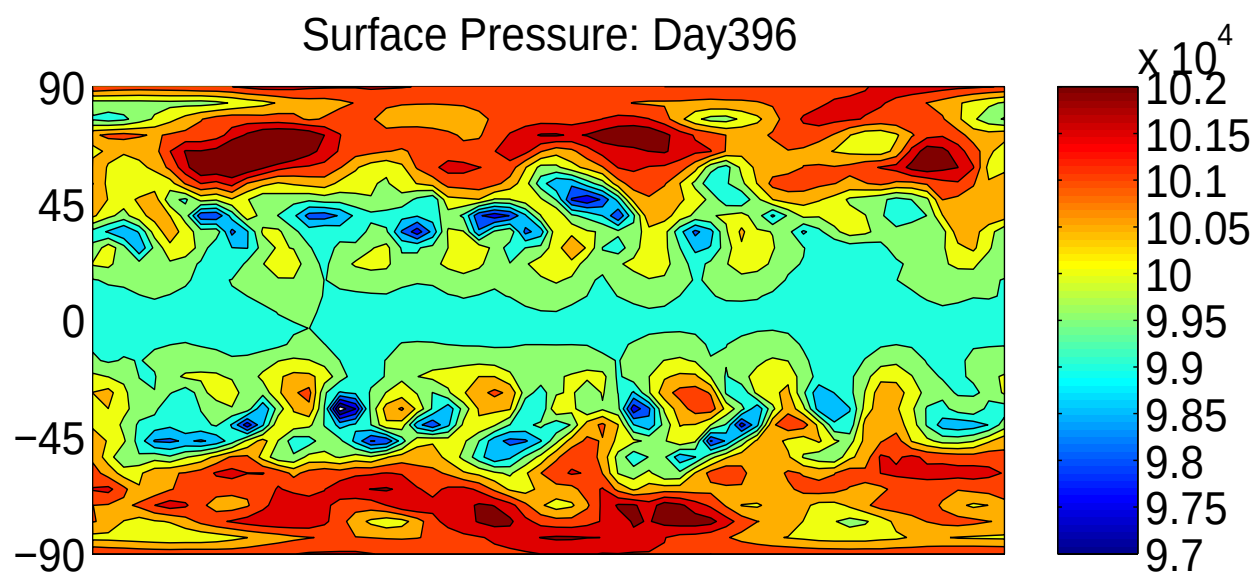


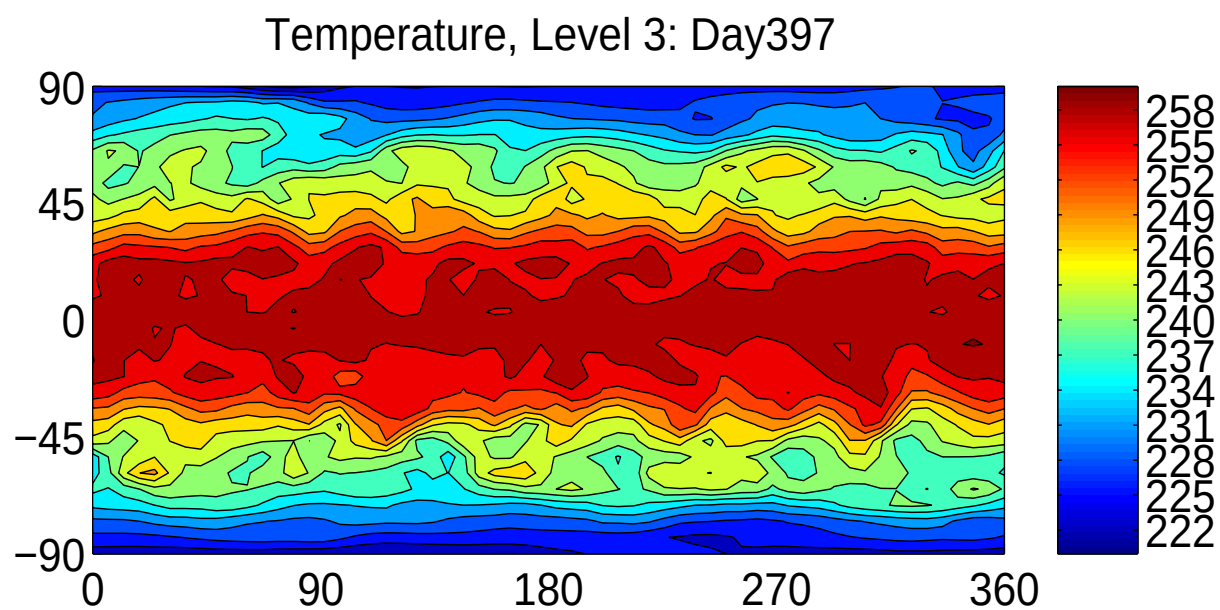
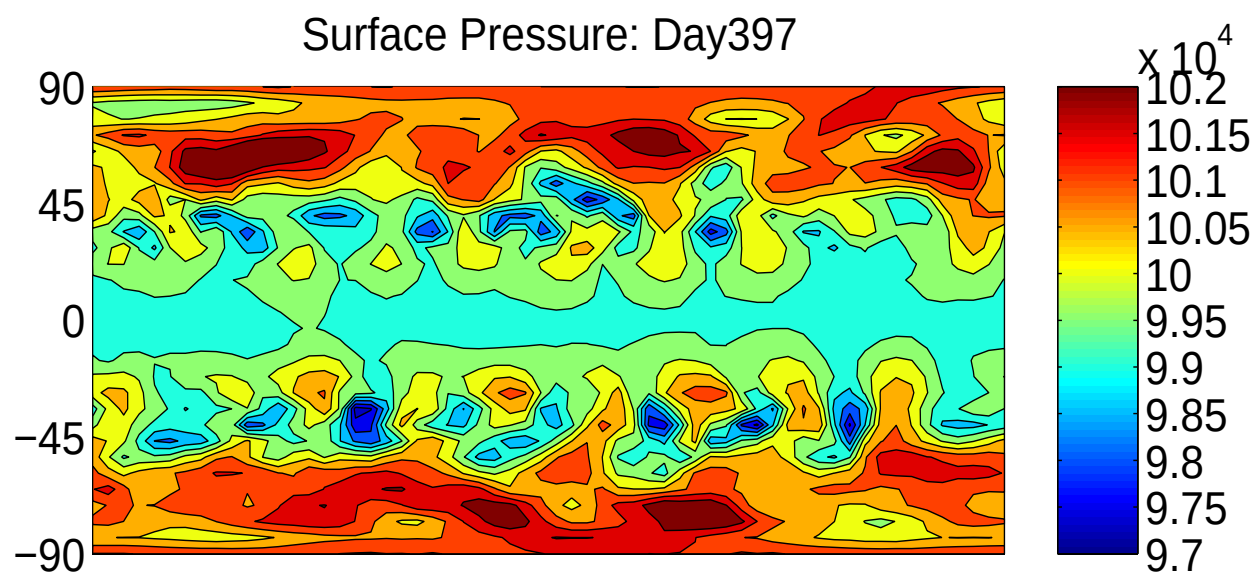


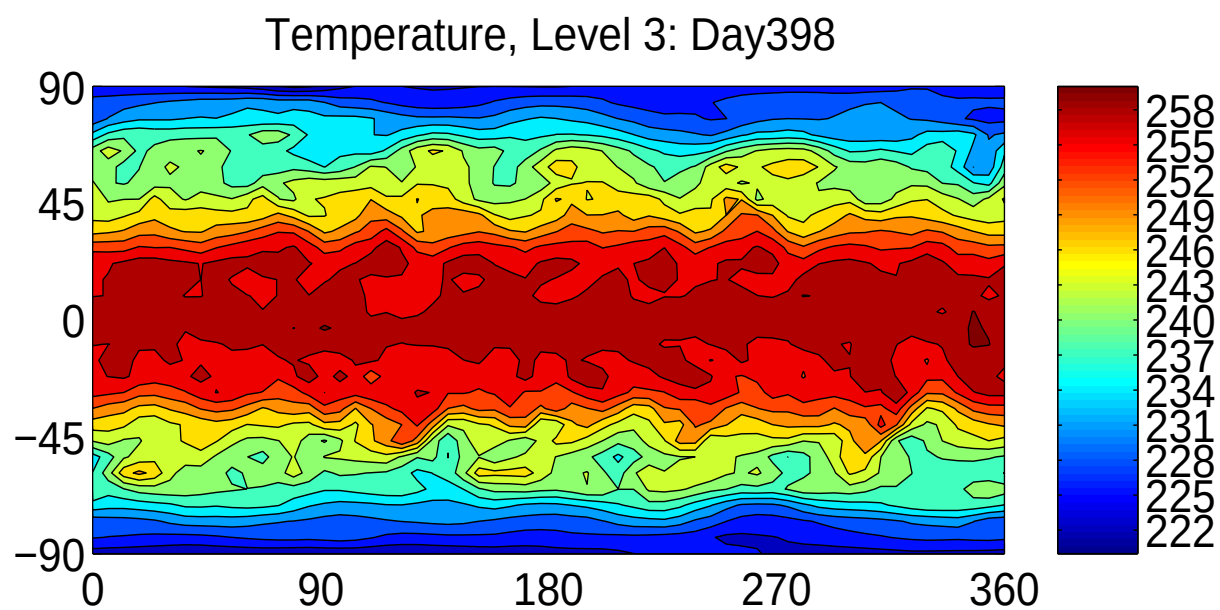
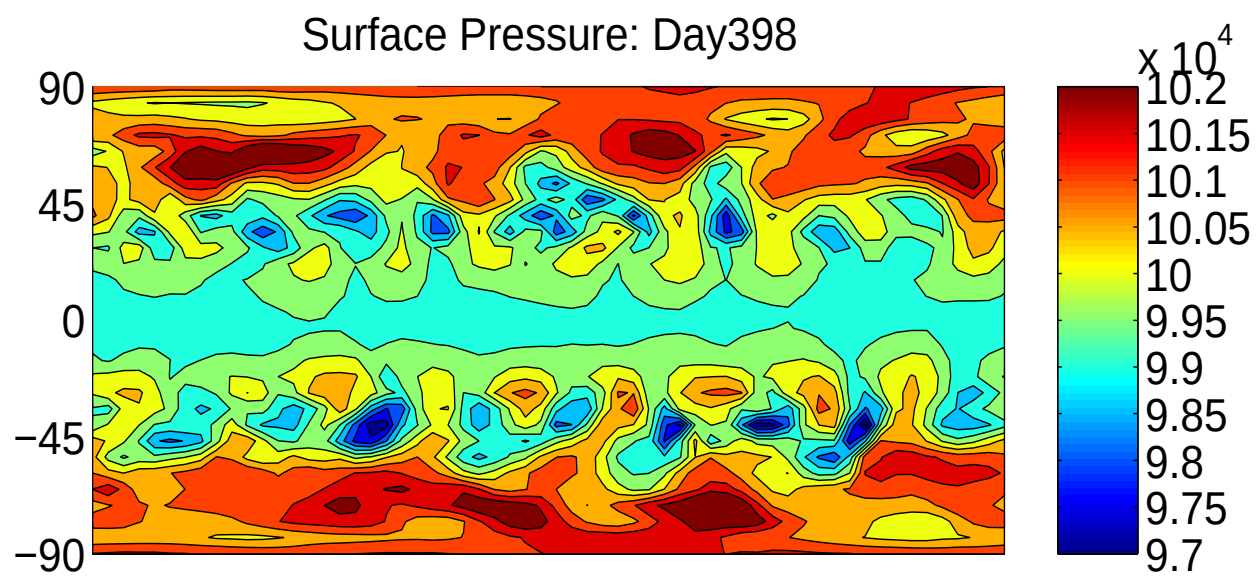


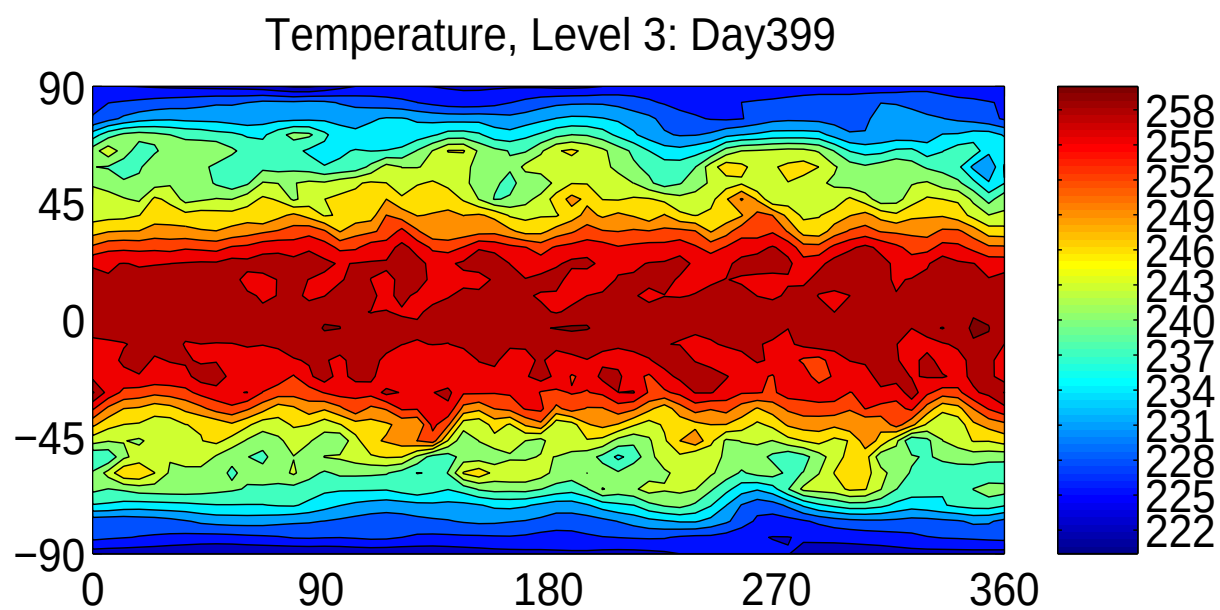
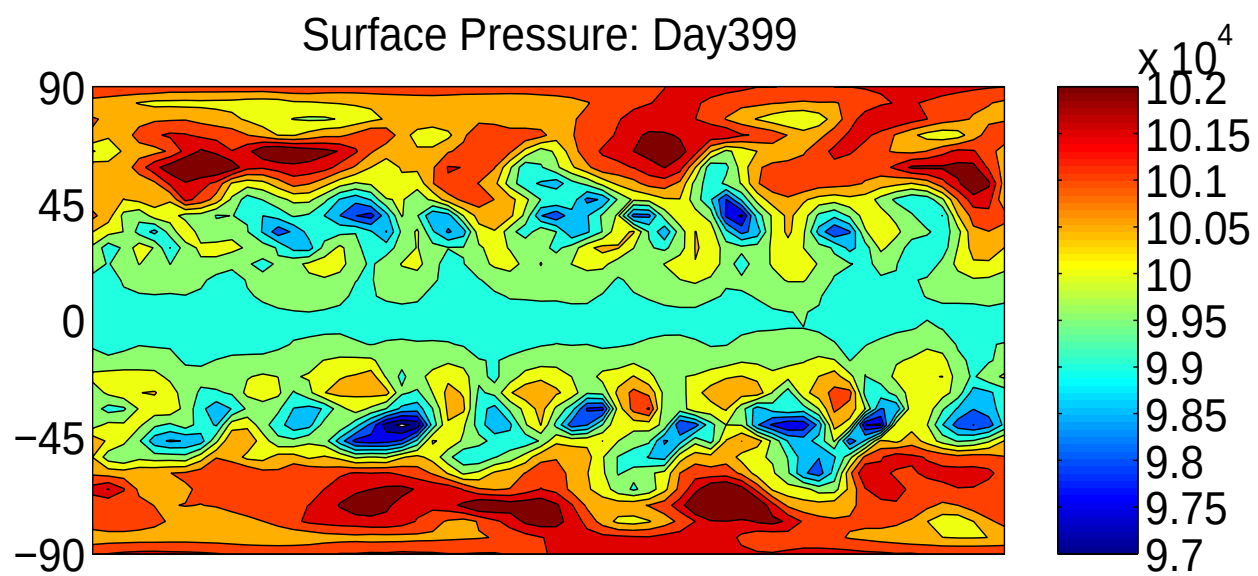


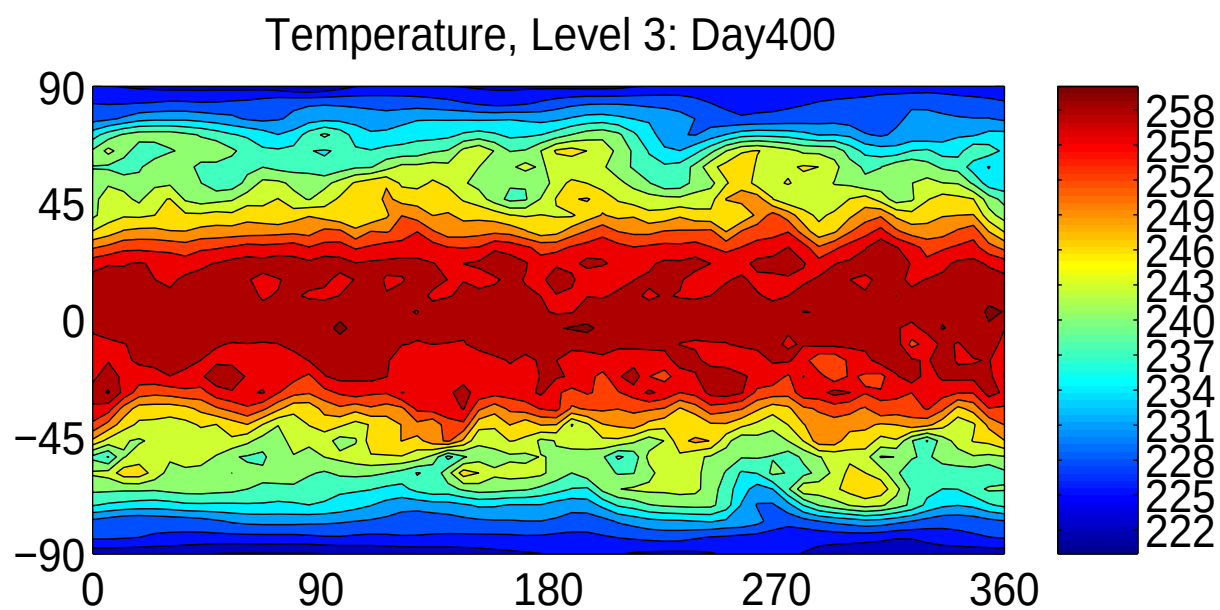
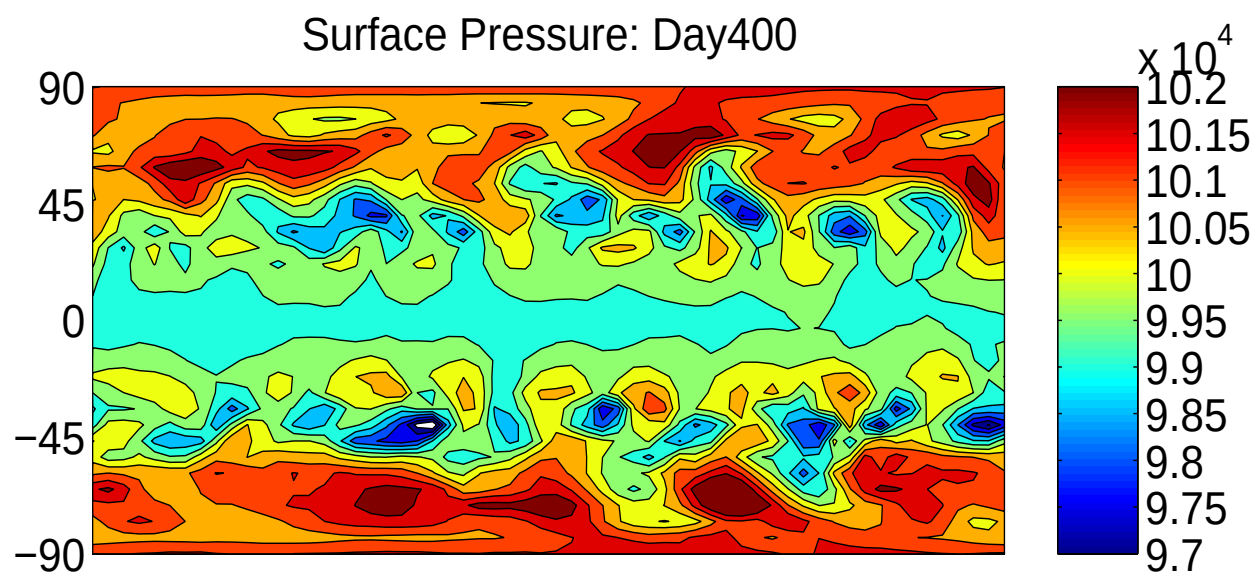








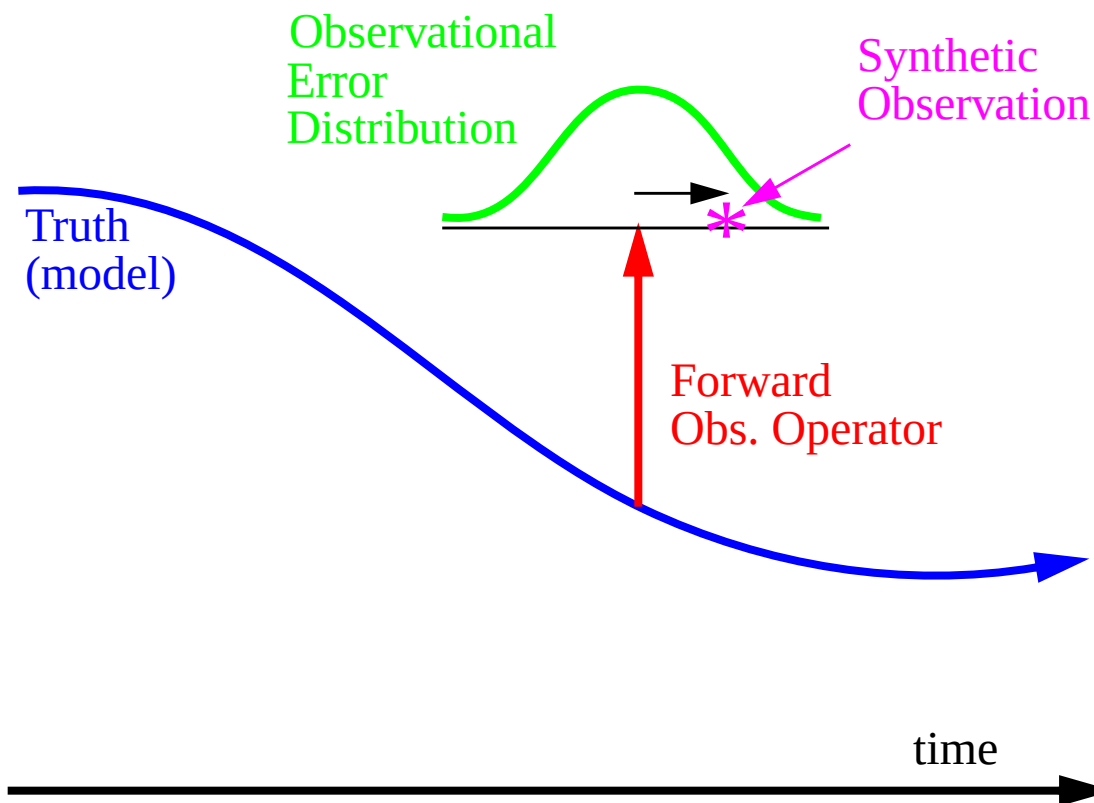




Perfect model experiments

‘Truth’ is generated by integrating model

B-grid, integrated for 100 years from state of rest before starting
(Multi-year spin-up for upper level temperatures)



‘Synthetic’ obs. by applying ‘forward observation’ operator to truth
(Here, this is just interpolating to a random horizontal location)

Instrument error simulated by adding random draw from a specified
Gaussian distribution to the interpolated observation

All the assimilation algorithm ever sees is these simulated observations

Result of assimilation can be compared to ‘truth’

Experimental Design Details

Base case assimilation starts from 'climatological' ensemble

Add tiny perturbations to control integration (truth)

Integrate this ensemble for several years

Ensemble size is 20 for ALL cases here

Each assimilation case is run for 400 days

Summary results are from last 200 days

No bias correction steps taken (no covariance inflation)

Single tuning parameter controls distance dependent correlation mask

Gives less weight to distant observations

This was tuned to give best RMS results in base case

Not changed for any other experiments

Note: Level 1 temperature in Held-Suarez configuration has very low frequency adjustment,

Experimental Sets

1. Impact of spatial density of observations:
150, 300, 450, 900, 1800, 3600, 7200, 14400, 28800 PS obs
Every 24 hours
PS observational error standard deviation 1.0 mb
2. Impact of frequency of observations
1800 PS observations
Every 24, 12, 6, 4, 3, 2, and 1 hours, 30, 15, and 5 minutes
PS observational error standard deviation 1.0 mb
3. Information content of different observation types
1800 observations of PS, or low-level T, or low-level U/V
Every 24 hours
PS observational error SD 2.0 and 1.0 mb
T observational error SD 1.0 and 0.5 K
U/V observational error SD 2.0 and 1.0 m/s, U, V errors independent
4. What happens if observations are confined to limited spatial domain
450 PS obs, only in N. Hemisphere between 90 and 270 deg. longitude
Every 24 hours
PS observational error standard deviation 1.0 mb
5. Impact of increased vertical resolution
1800 PS obs
Every 24 hours
PS observational error standard deviation 1.0 mb
5 and 18 vertical levels
6. Impact of adding stochastic 'sub-grid scale' noise
1800 PS obs, Every 24 hours
PS observational error standard deviation 1.0
Temperature time tendency noise standard deviation 0, 10%, 40%

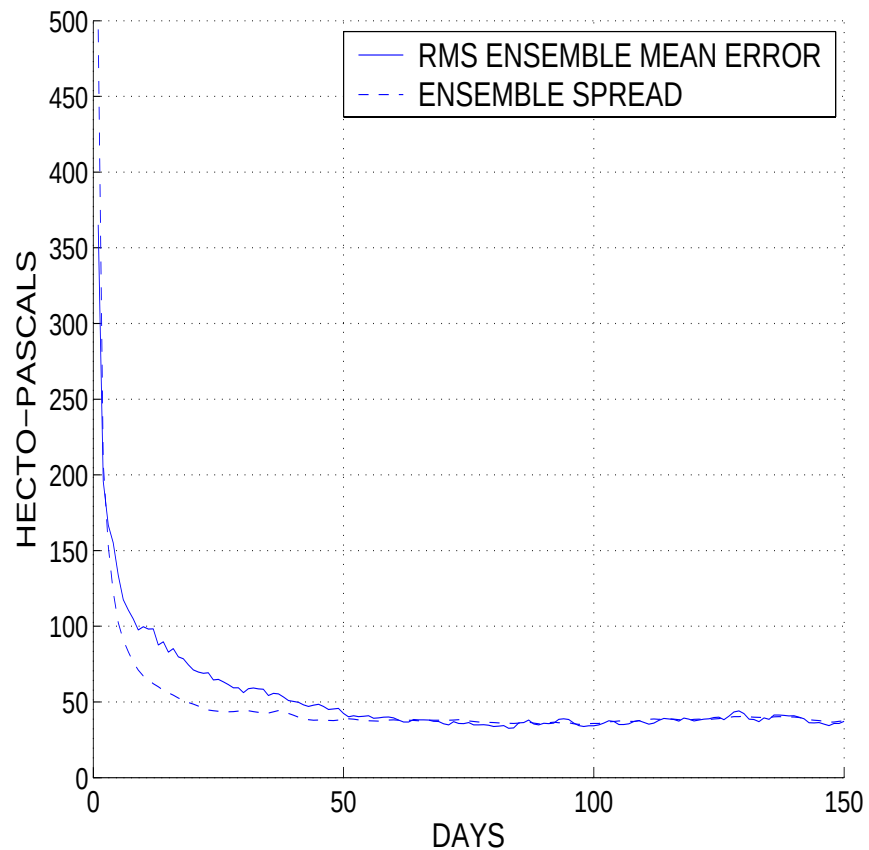
Baseline Case: 1800 PS Obs every 24 hours

PS Error Reduces
by about factor of
10

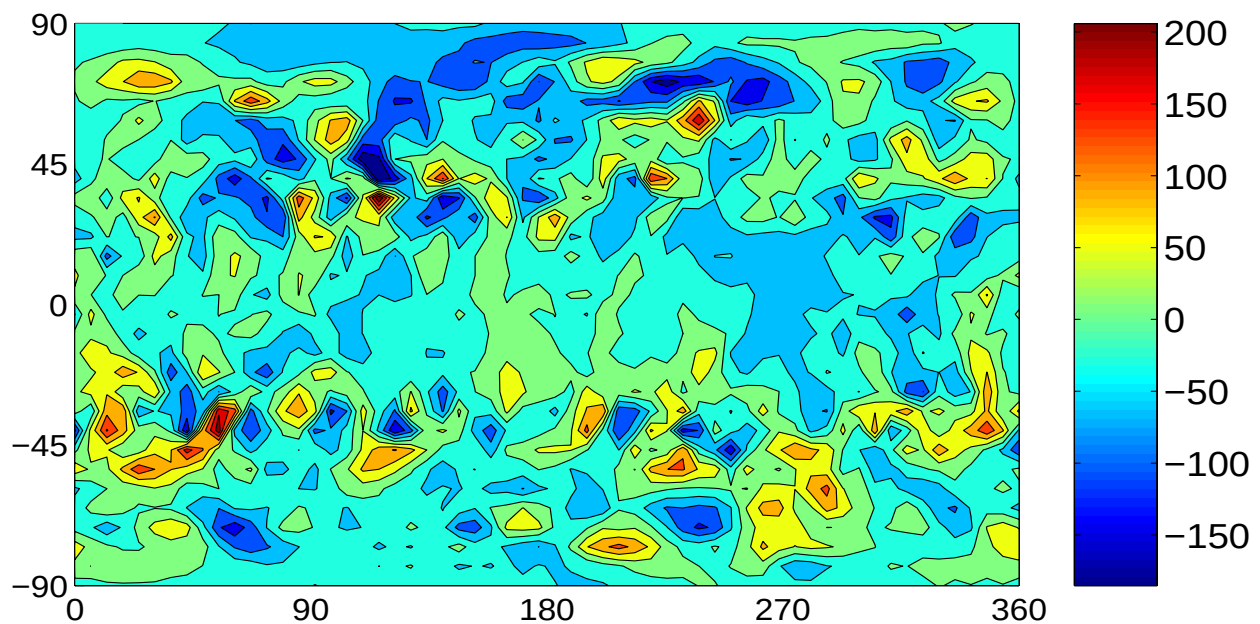
Asymptotes after
about 50 days

Ensemble spread
is approximately
correlated with
RMS error

Final Error is
about 0.4 mb



Largest error in mid-latitudes, 'synoptic' scales after 400 days

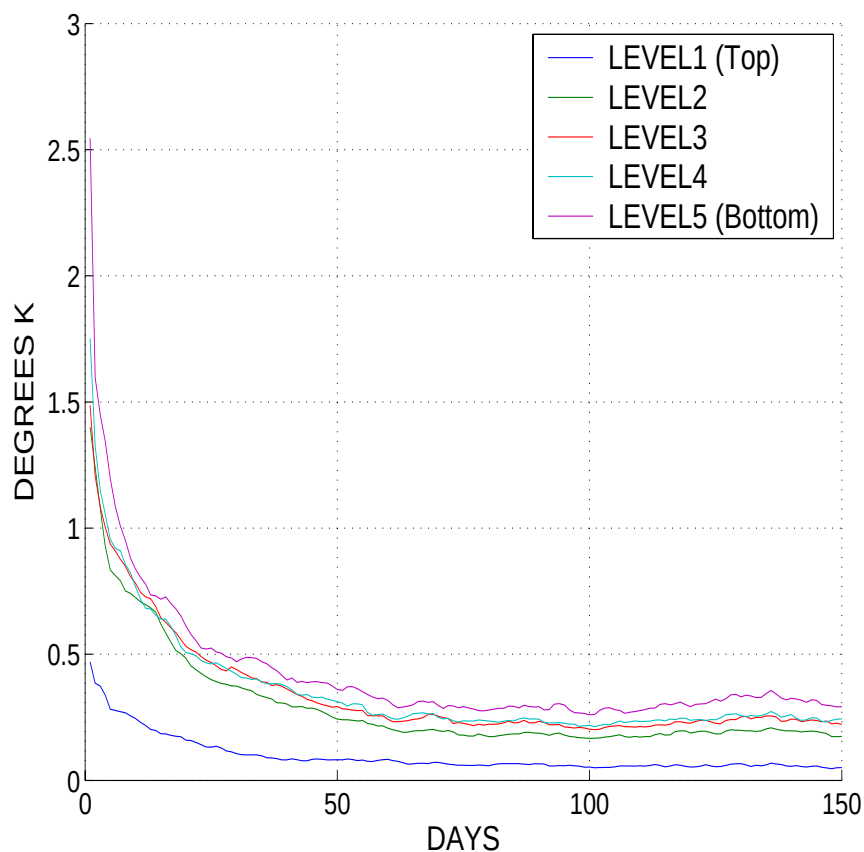


Baseline Case: 1800 PS Obs every 24 hours

T Error also reduced
by about factor of 10

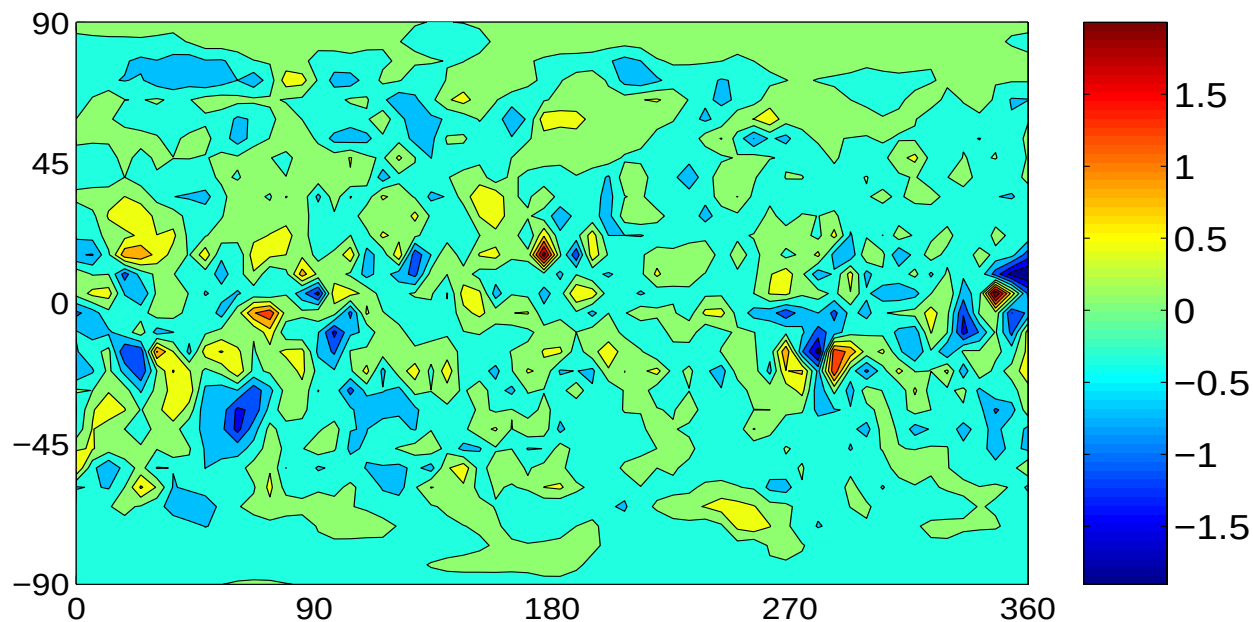
Asymptotes about
day 70

Final error about
0.25 K for interior
levels



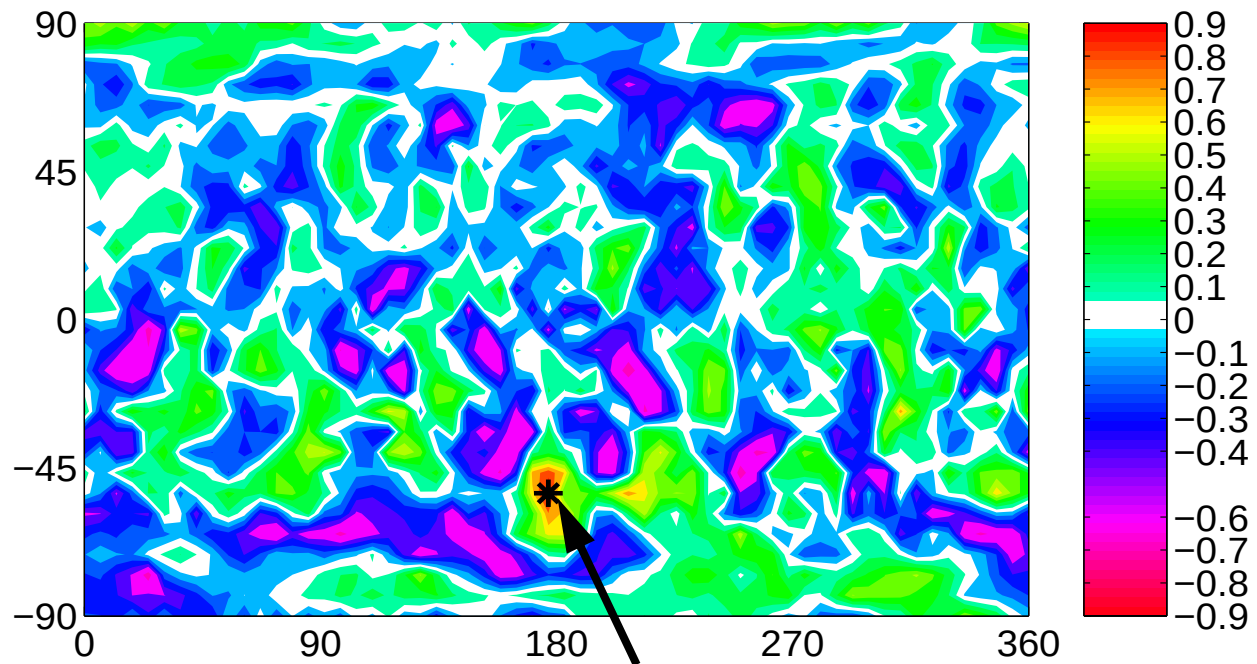
Largest T error in tropics for interior levels (level 3, day 400 shown)

Min = -1.9012 Max = 2.3893 RMS ERROR = 0.24305

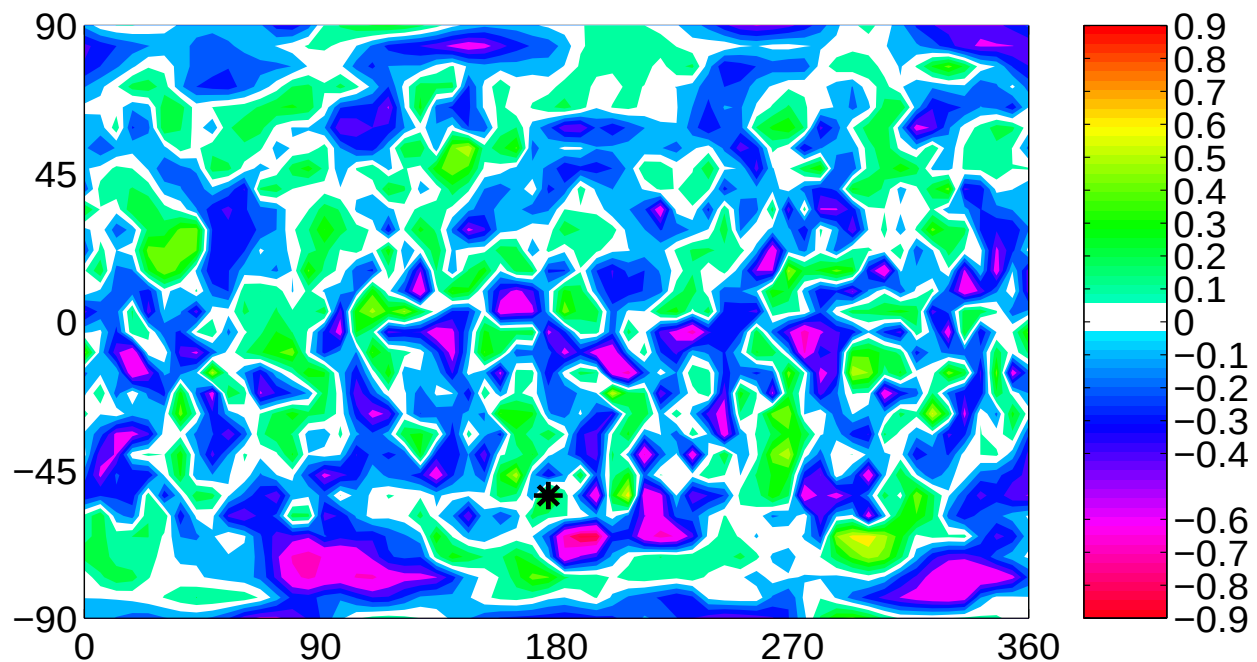


Sample Correlation: Baseline Case

Sample correlations reflect how observations can impact state variables:



Correlation of PS with PS at (180, 50S): largest values local but noisy



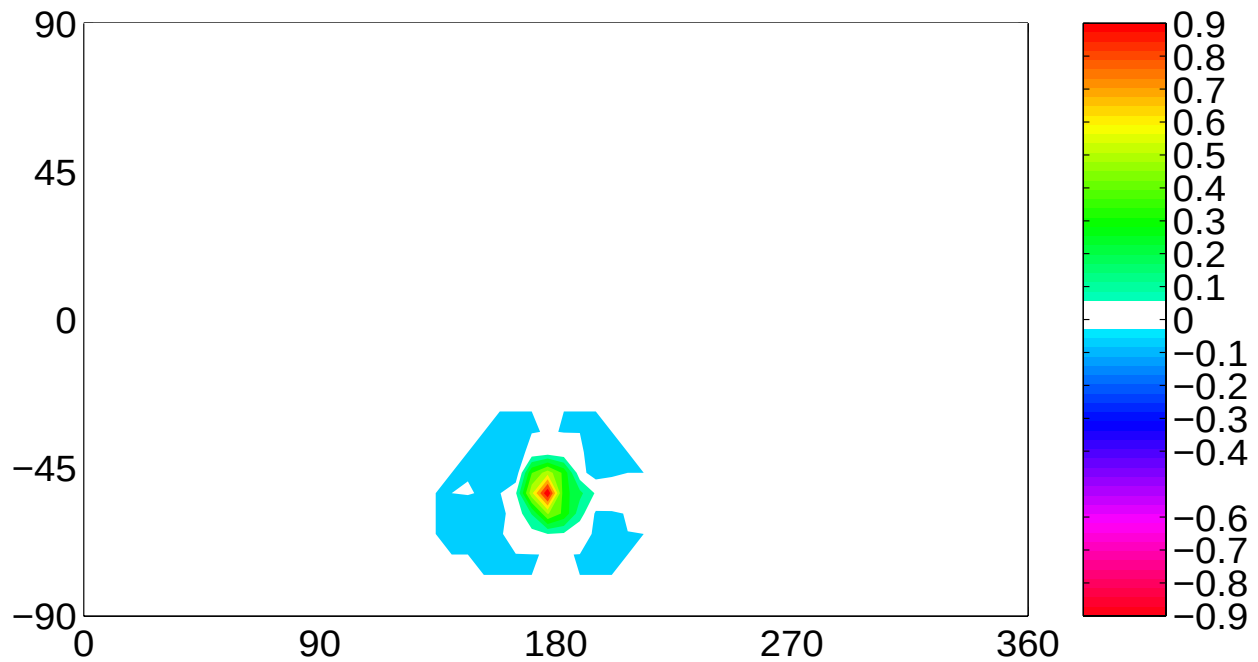
Correlation of T at level 3 with PS at (180, 50S);

Lots of noise, limited local signal

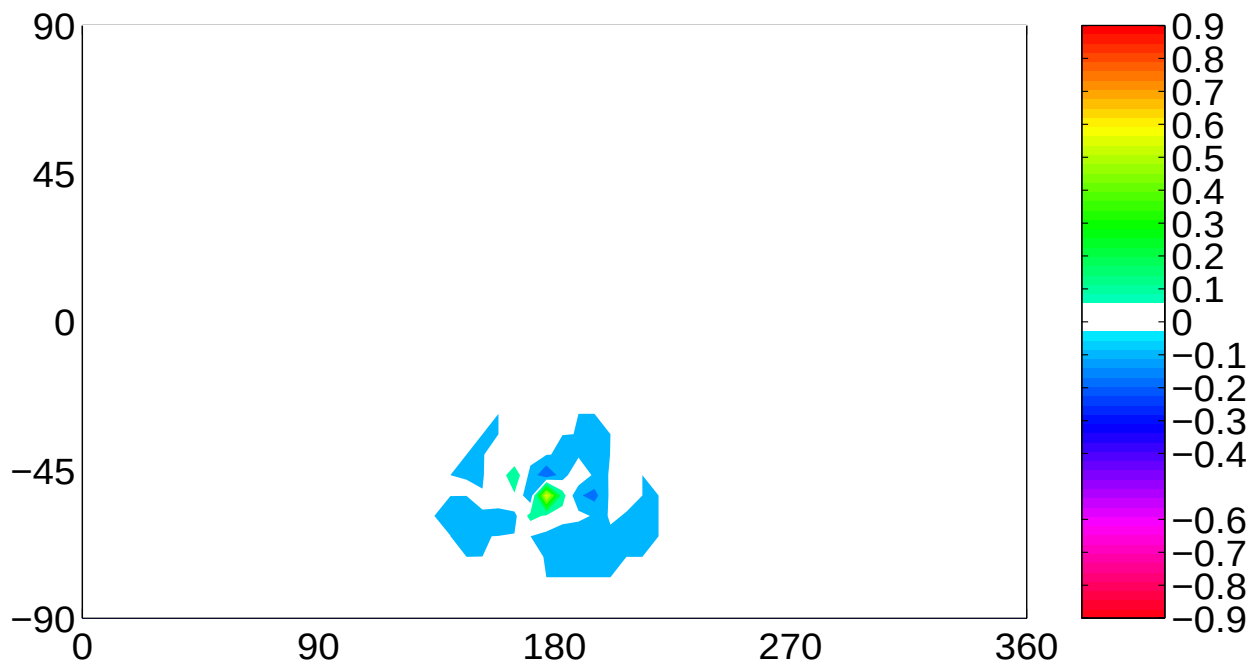
Filter must be able to extract limited signal from lots of noise

Sample Correlation with Envelope: Baseline Case

Sample correlations reflect how observations can impact state variables:

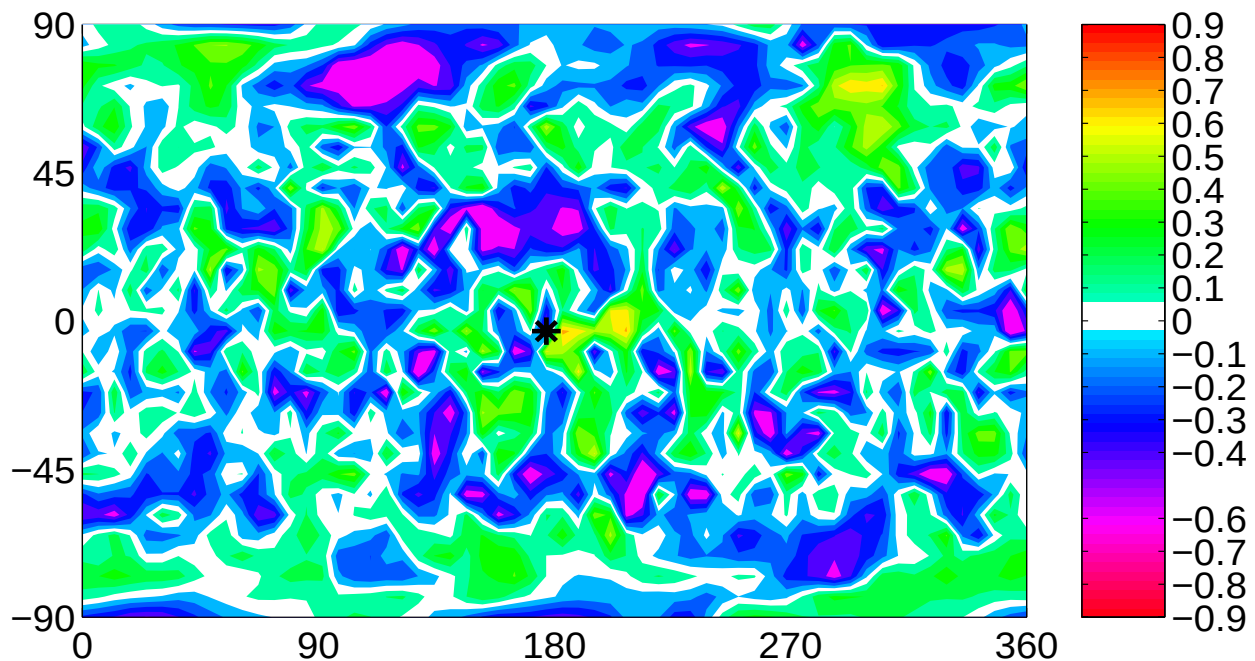
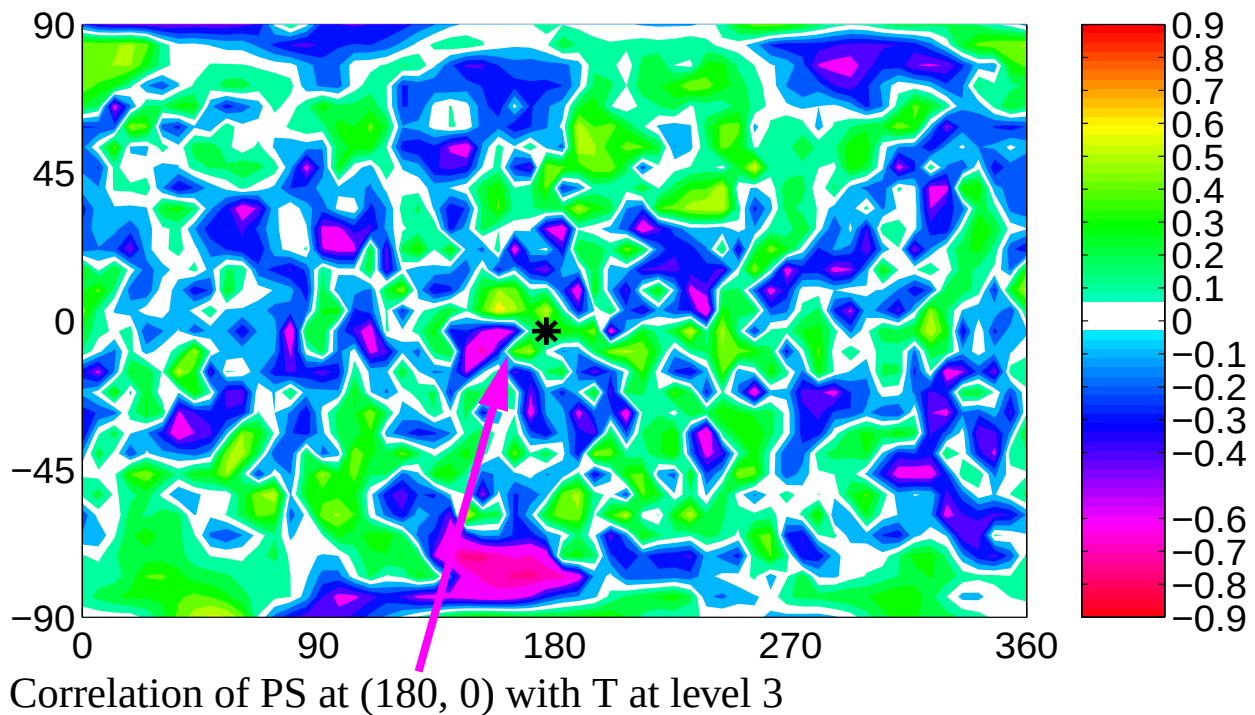


Correlation of PS with PS at (180, 50S):



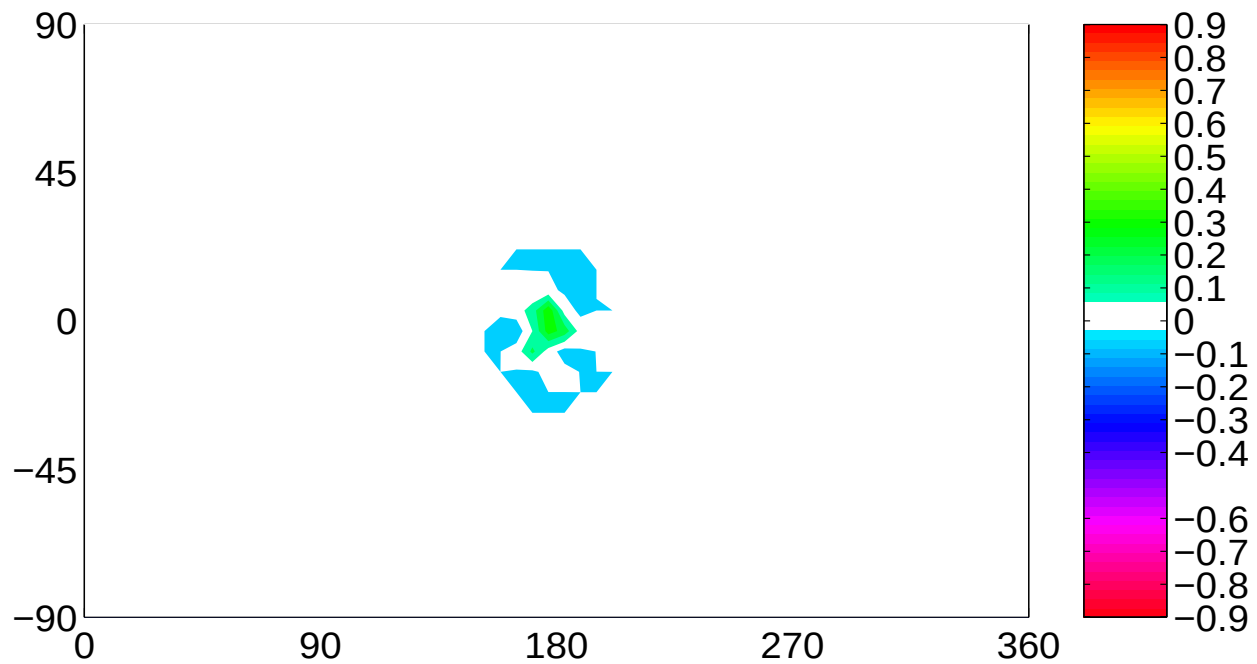
Correlation of T at level 3 with PS at (180, 50S);

Sample correlations vary significantly in time and space

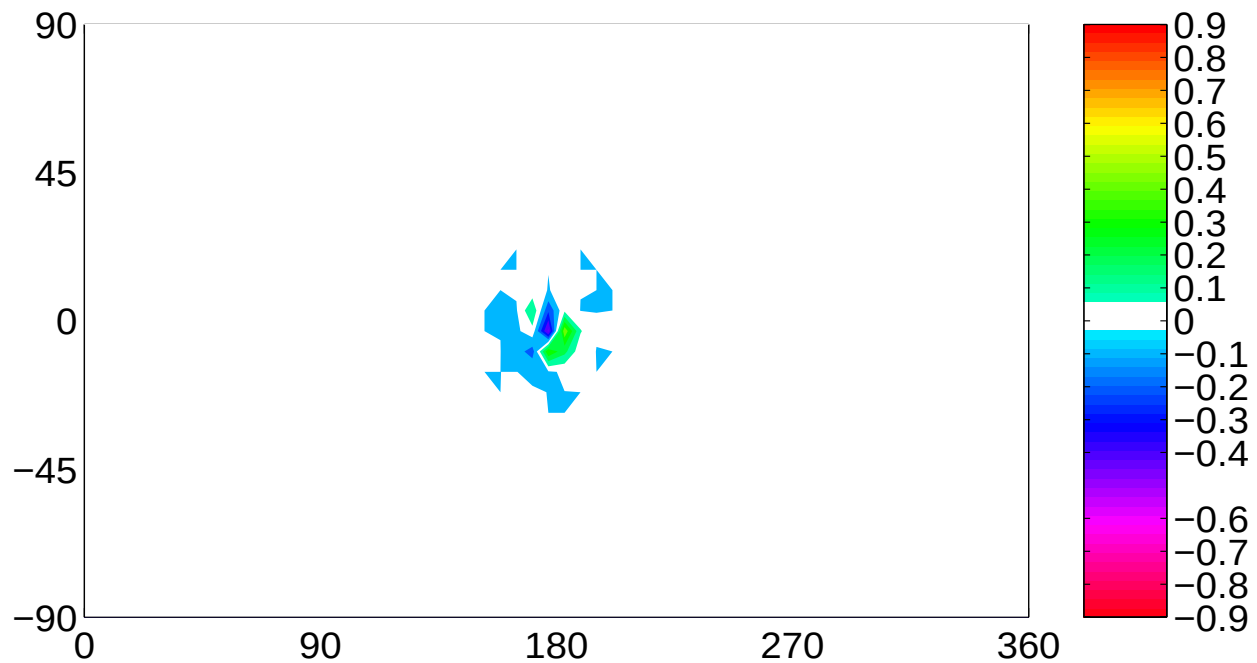


Must take actions to avoid impact from remote noise

Sample correlations vary significantly in time and space



Correlation of PS at (180, 0) with T at level 3



Same field, but 10 days later

Impacts of spatial density of PS obs

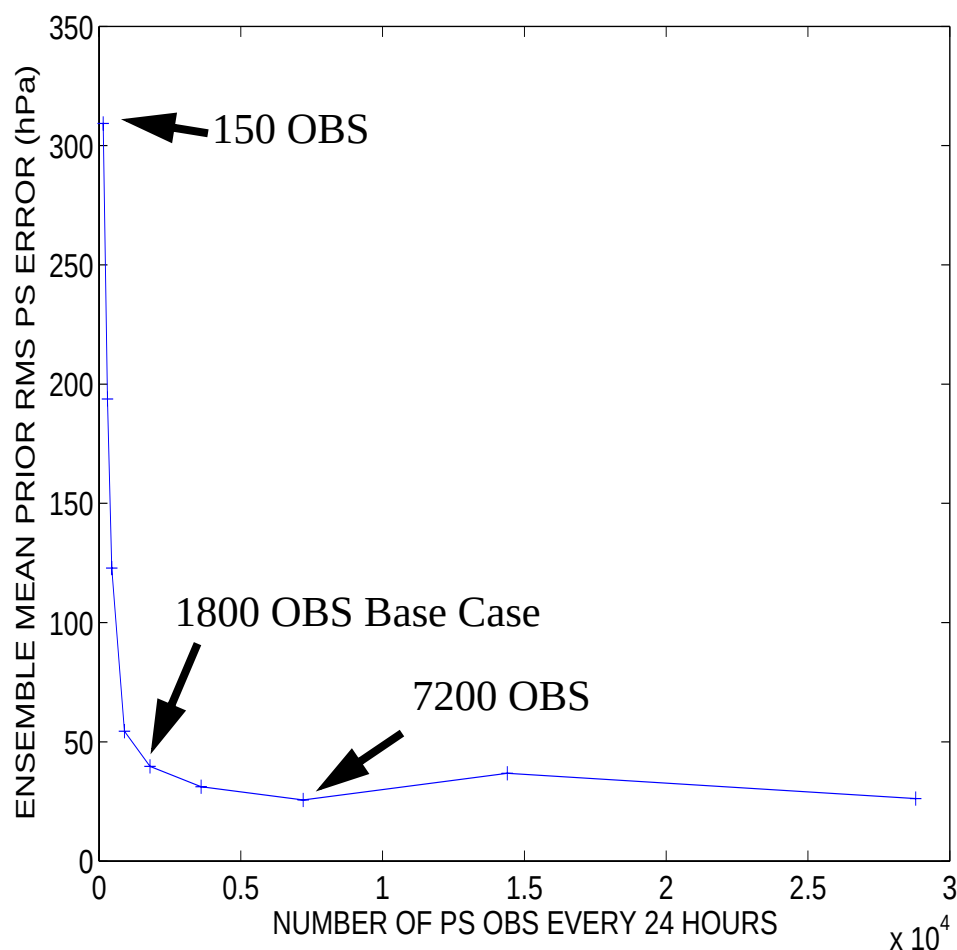
150, 300, 450, 900, 1800, 3600, 7200, 14400 and 28,800 every 24 hours

PS error
reduces to
about 0.3 mb

150 obs
reduces clima-
tological error
by less than
half

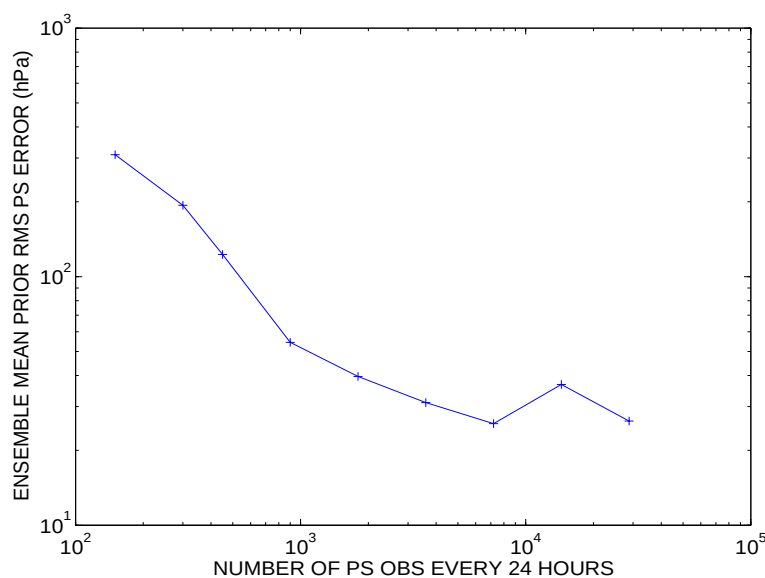
More than 7200
obs appears
superfluous

Why is 14,400
worse? No
clue.

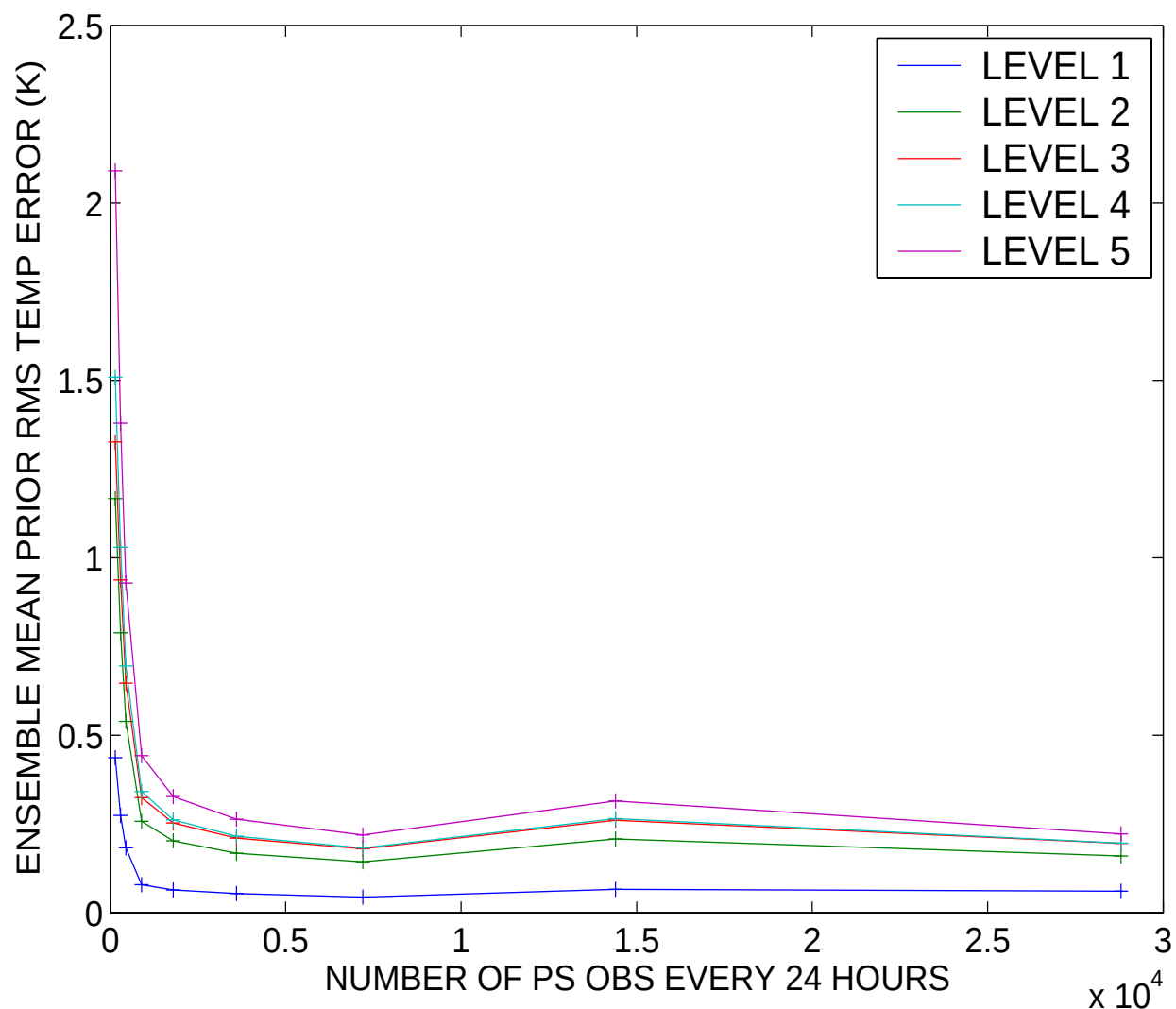


Plotting log /log of RMS
shows approx. linear
decrease from 150 to
7200 obs

Behavior for very large
numbers of obs clearly
different



Impacts of spatial density of PS obs on Temperature RMS
150, 300, 450, 900, 1800, 3600, 7200, 14400 and 28,800 every 24 hours



Behavior for Temperature (and U, V not shown) similar to that for PS
Best results for 7200 PS observations
Interior level mean T RMS of about 0.25 K for best case

Impacts of frequency of PS obs

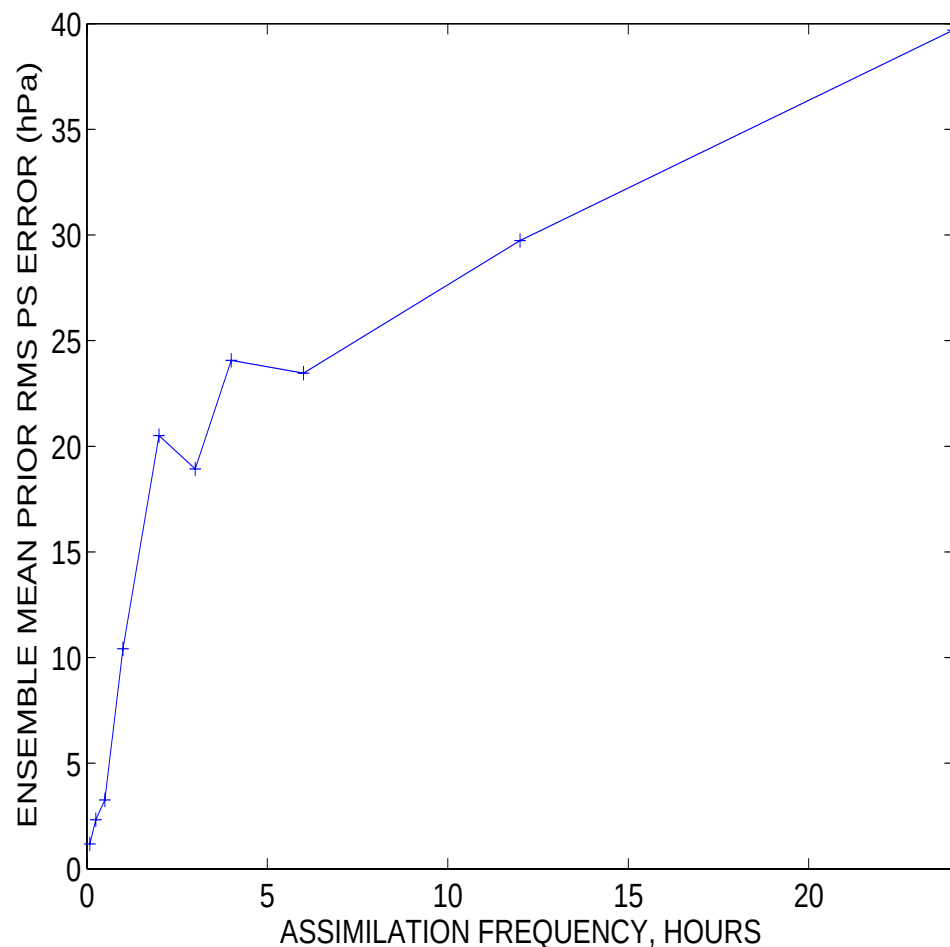
24, 12, 6, 4, 3, 2, 1 hours; 30, 15, 5 minutes; 1800 obs.

Steady RMS decrease as frequency increases

Much smaller RMS than for high density low frequency obs

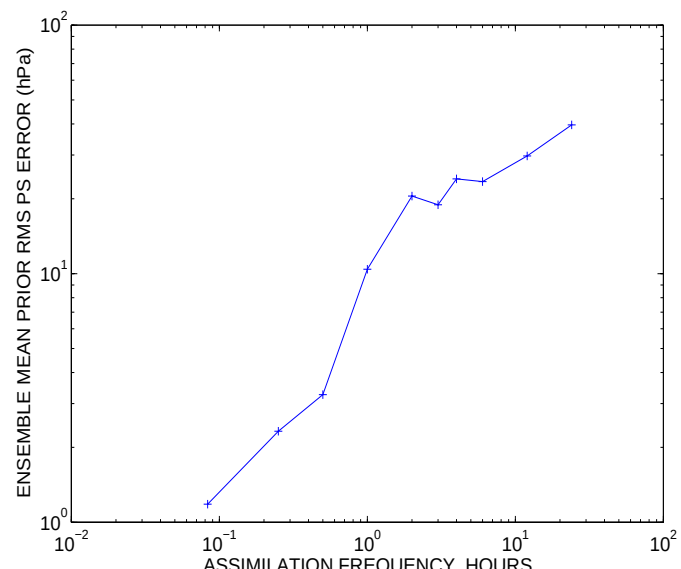
RMS < 0.02mb for 5 minutes

Strange behavior between 1 and 6 hour frequency



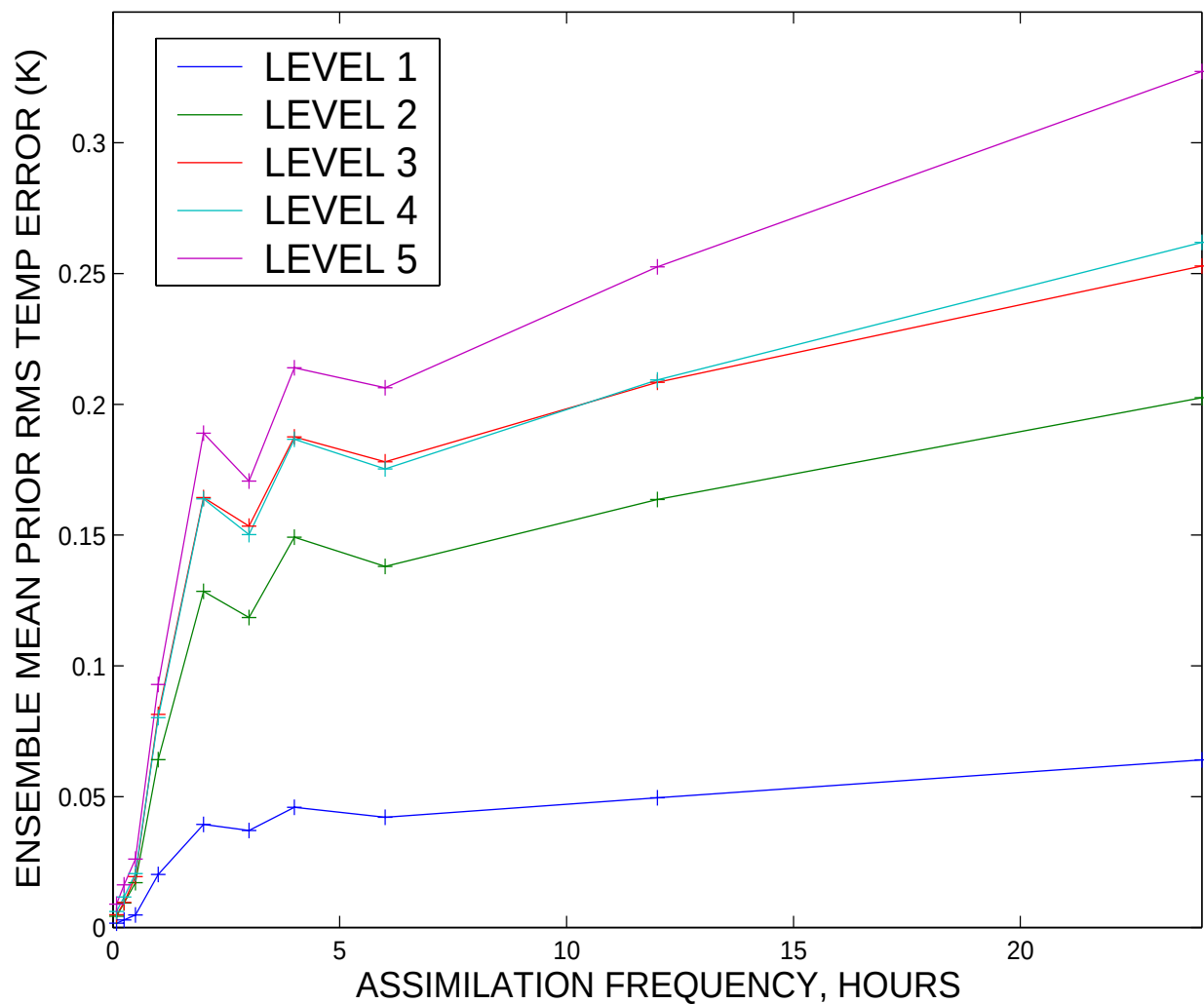
Plotting log /log of RMS shows approx. linear increase with a bump

What's going on in the middle?



Impacts of frequency of PS obs

24, 12, 6, 4, 3, 2, 1 hours; 30, 15, 5 minutes; 1800 obs.



Temperature (and U and V, not shown) similar to PS
Consistent decrease in RMS with increased obs frequency

Errors at 5 minute frequency less than 0.01 K !!!

How low can you go?

What's going on at moderate obs frequencies?

Equilibrated model has very low gravity wave amplitude

When perturbed, 'off-attractor' gravity waves can result

Noise in observations can project off attractor

Ensemble members pulled in same direction; get phased gravity waves

Gravity wave period varies: approximately 4 hours

Gravity waves heavily damped; quickly reduced in amplitude

Low frequency (> 12 hours): gravity waves damped before next obs time

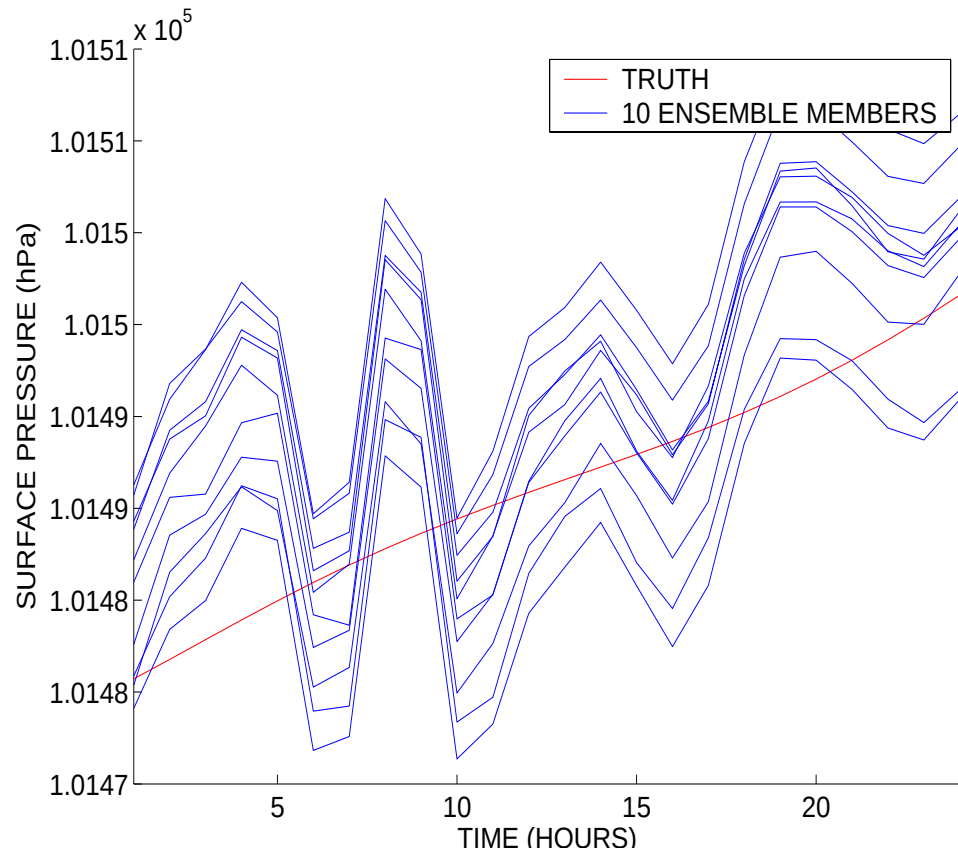
High frequency (< 1 hour): enough obs per period to control amplitude

Moderate frequency (~ 4 hours): get phased gravity waves in ensemble;
large bias; increased assimilation error

Time series of
10 out of 20
ensemble
members at
mid-latitude PS
point.

Forecast initi-
ated from end
of 1800 PS obs
every 4 hours

(Extreme
example)

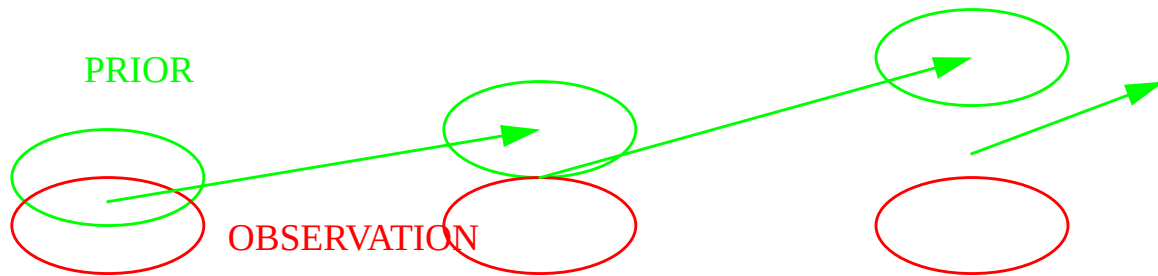


Challenge #1: Model Bias and Atmospheric Balances

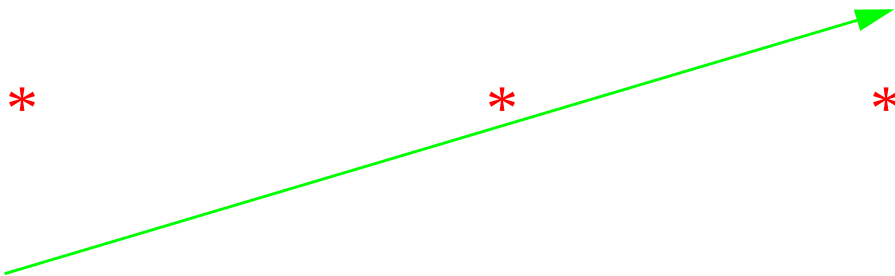
Filter equations assume prior estimate (and observations) are unbiased
Questionable for Observations, ridiculous for Models

Biased prior estimate will cause observations to be given too little weight

Repeated applications lead to progressively less weight, estimate can diverge



Implications are obvious for 4D-Var, too



Dealing with model bias is mostly an open question:

1. Can reduce confidence in model prior estimates by some constant factor
2. Explicitly model the model bias as an extended state vector and assimilate coefficients of this bias model

Model: $dx/dt = F(x)$

Model plus bias model: $dx/dt = F(x) + \epsilon(t)$; $d\epsilon/dt = 0$
where ϵ is a vector of the same length as x

Very tricky: if we knew much about modeling the bias, we could remove

Challenge #2: Balances and Attractors

Many models of interest have balances, both obvious (say geostrophic) and subtle

The structure of the model 'attractors' may be extremely complex

In some cases, off-attractor perturbations may lead to 'large' transient response

Example: High frequency gravity waves in some Primitive Equation models

The behavior of these transients can lead to model bias

In this sense, even perfect model experiments can have large model bias

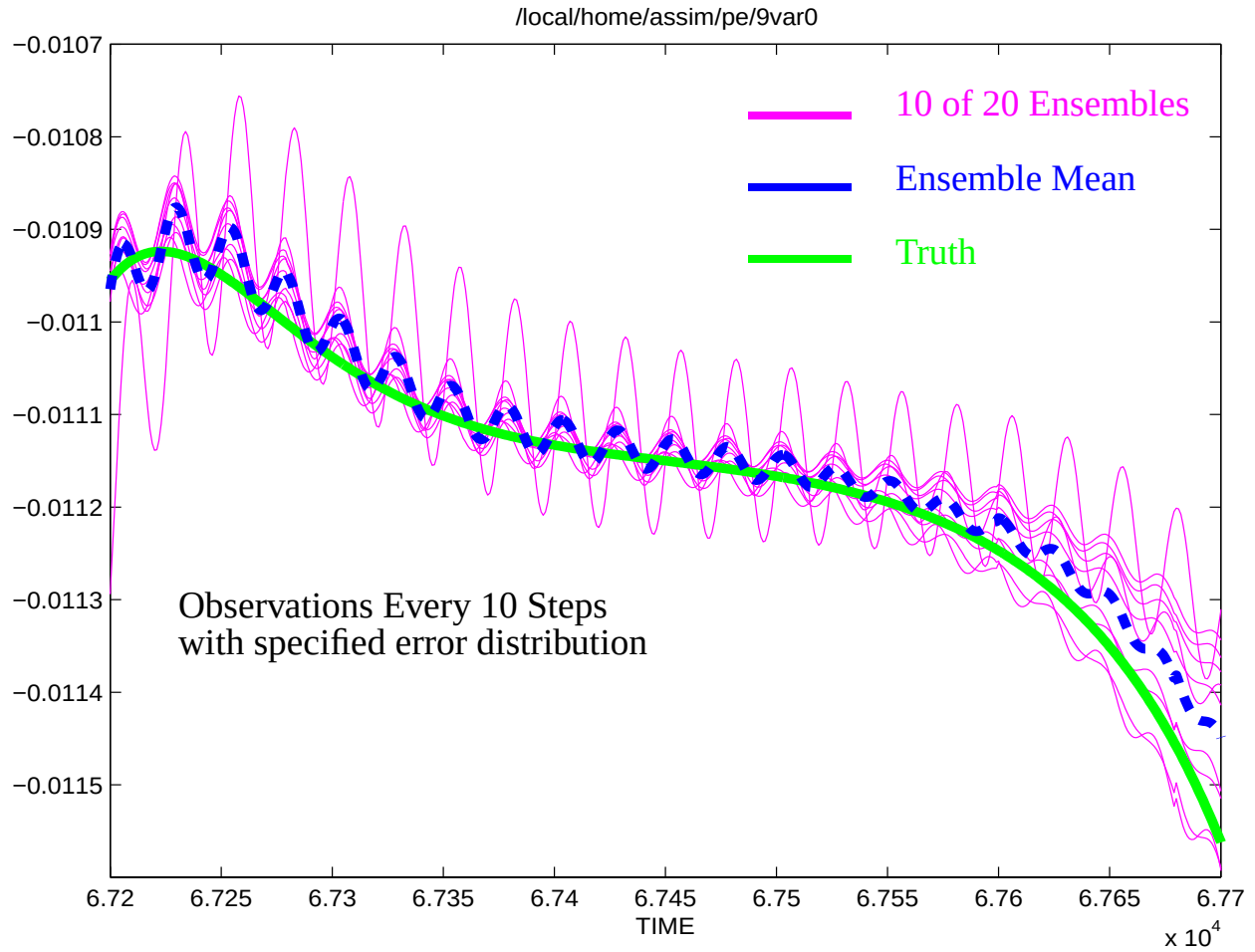
Understanding how to minimize this behavior or limit its impact is a fun problem

The continuous system may also have balances, obvious and subtle

Unclear how differences between model and continuous 'attractors' impacts assimilation

Lorenz 9-Variable Model

Time series of Ensemble Filter Assimilation for variable X1



$$\dot{X}_i = U_j U_k + V_j V_k - v_0 a_i X_i + Y_i + a_i z_i \quad (1)$$

$$\dot{Y}_i = U_j Y_k + Y_j V_k - X_i - v_0 a_i Y_i \quad (2)$$

$$\dot{z}_i = U_j(z_k - h_k) + (z_j - h_j)V_k - g_0 X_i - K_0 a_i z_i + F_i \quad (3)$$

$$U_i = -b_j x_i + c y_i \quad (4)$$

$$V_i = -b_k x_i - c y_i \quad (5)$$

$$X_i = -a_i x_i \quad (6)$$

$$Y_i = -a_i y_i \quad (7)$$

Defined for cyclic permutations of the indices (i, j, k) over the values (1, 2, 3).

X, Y and z variables can be thought of as representing divergence, vorticity and height

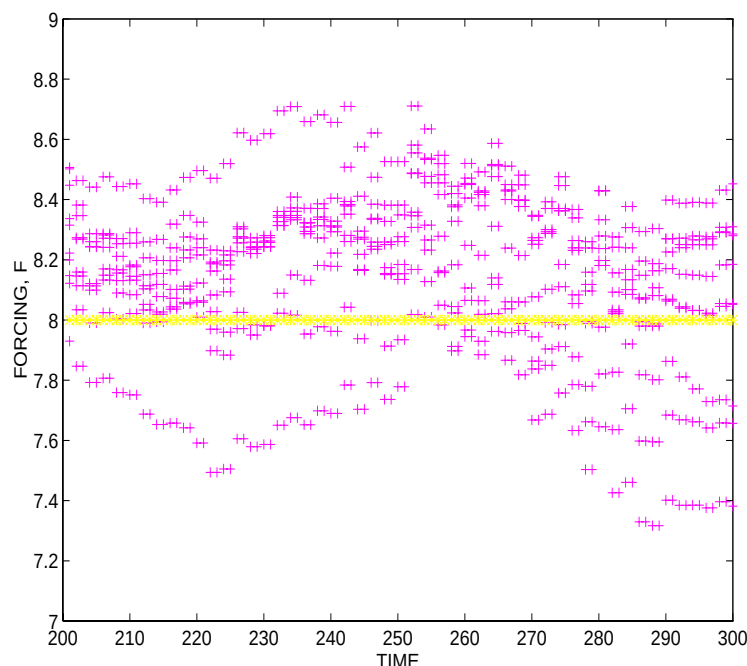
Lorenz-96 Free Forcing Model Filter

20 Member Ensemble (10 Plotted)

Obs Every 2 Steps

Truth (8.0)

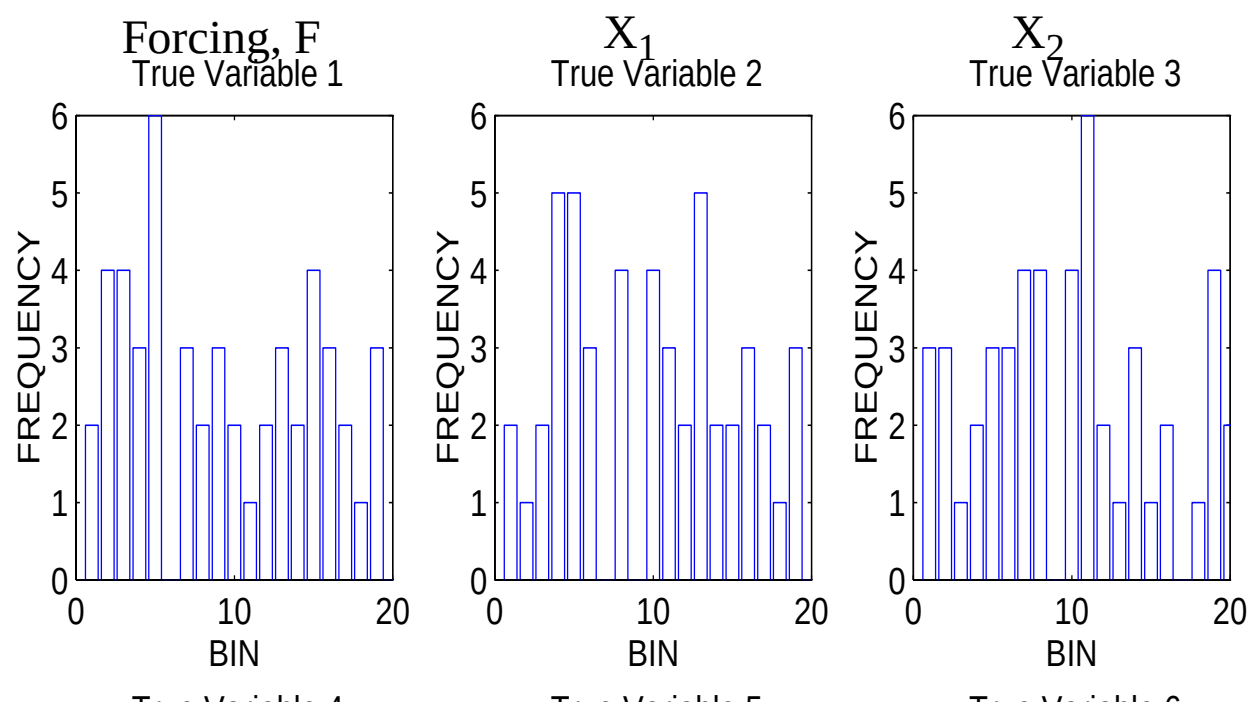
Ensemble



>> Can treat model parameters as free parameters <<

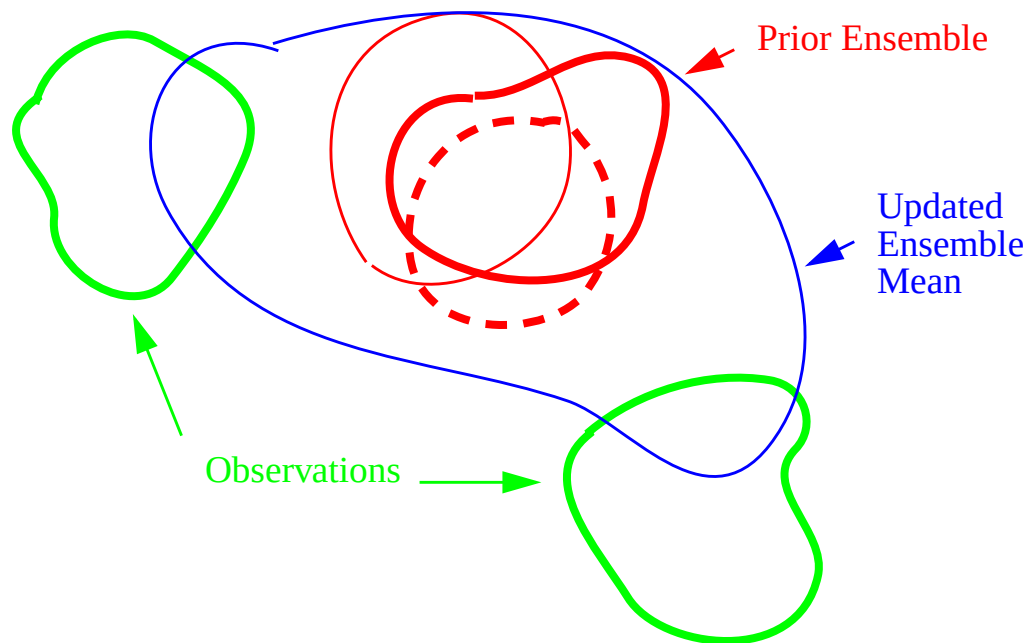
>> Here, the forcing F is assimilated along with the state <<

>> This is potential mechanism for dynamic adjustment of unknown parameters and for dealing with unknown model systematic error <<



Challenge #3: Assimilation of Discrete Distributions

Example: assimilation of convective elements



Prior is 'certain' that there are no convective cells outside the red areas

Observations indicate discrete areas outside the red

This is indicative of highly non-linear problem

Ensemble techniques, at best, tend to smear out prior discrete structures

4D-Var is likely to have non-global local minima

But, we think we know what we want to do

Keep information from prior on larger scale 'background'

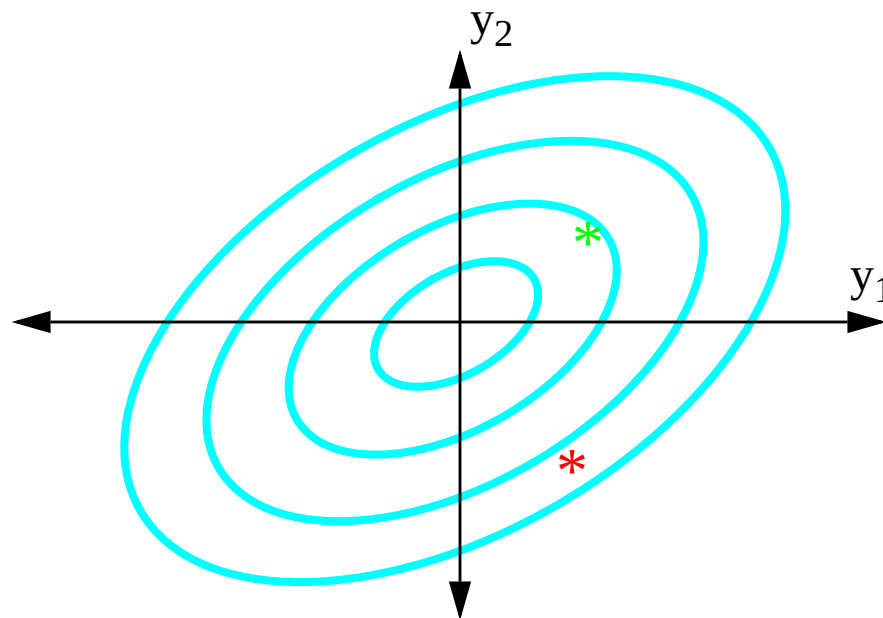
Introduce cells where observed

Requires new norms or ways to deal with model bias as function of scale

Quality Control of Observations

Methods to exclude erroneous observations

1. Discard impossible values (negative R.H.)
2. Discard values greatly outside climatological range
3. Discard values that are more than α prior ensemble sample standard deviations away from prior ensemble mean
4. 'Buddy' checks for pairs of observations: just apply chi-square test using prior ensemble covariance and label pair as inconsistent if threshold value exceeded



5. Could also apply chi-square to larger groups of obs.