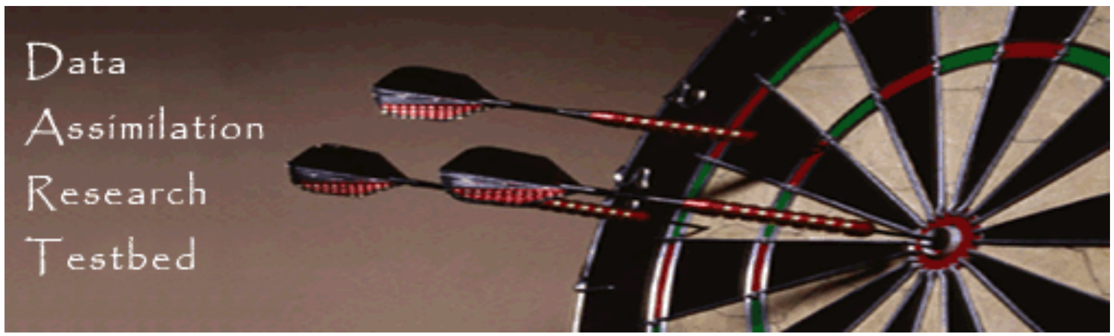
Localization and Correlation in Ensemble Kalman Filters

Jeffrey Anderson



NCAR Data Assimilation Research Section

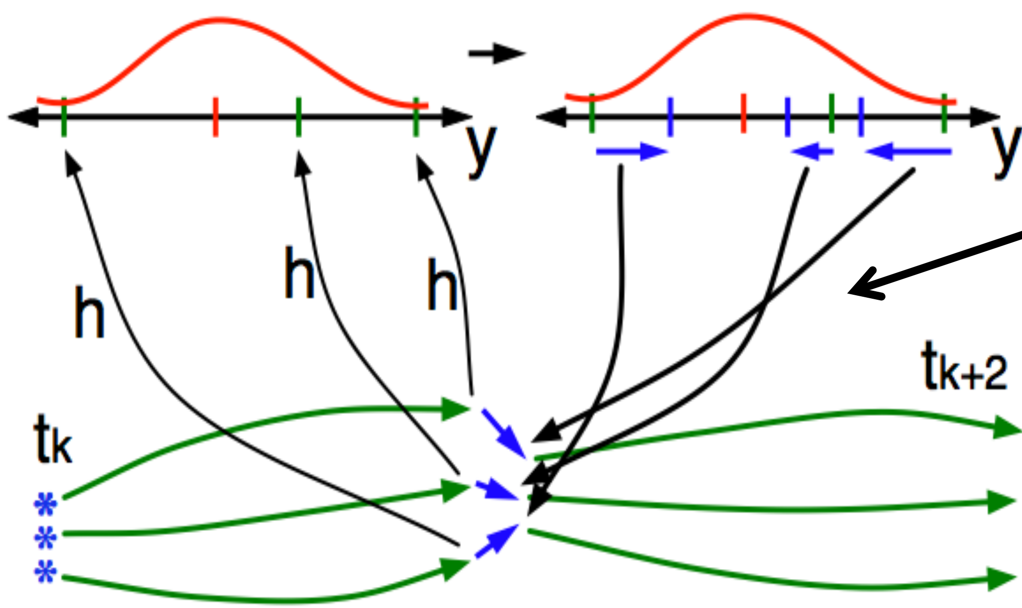
The National Center for Atmospheric Research is sponsored by the National Science Foundation.

NCAR

Sequential ensemble filter assimilation

- Impact of single ob y on scalar state component x_i.

Regress y increments onto each component of x.



$\Delta x_{i,n} = b\Delta y_n,$
$n=1,\ldots N$

N is ensemble size.
b is regression coefficient.

Definition of localization

- For impact of observation y on state component x_i, multiply regression coefficient by factor α

$$\Delta x_{i,n} = \alpha b \Delta y_n , \qquad n=1,\ldots N.$$

- α often function of distance between y and x_i.
- Gaspari Cohn compact pseudo-gaussian function.
- Here, localization used to correct sampling error.

Using additional prior information

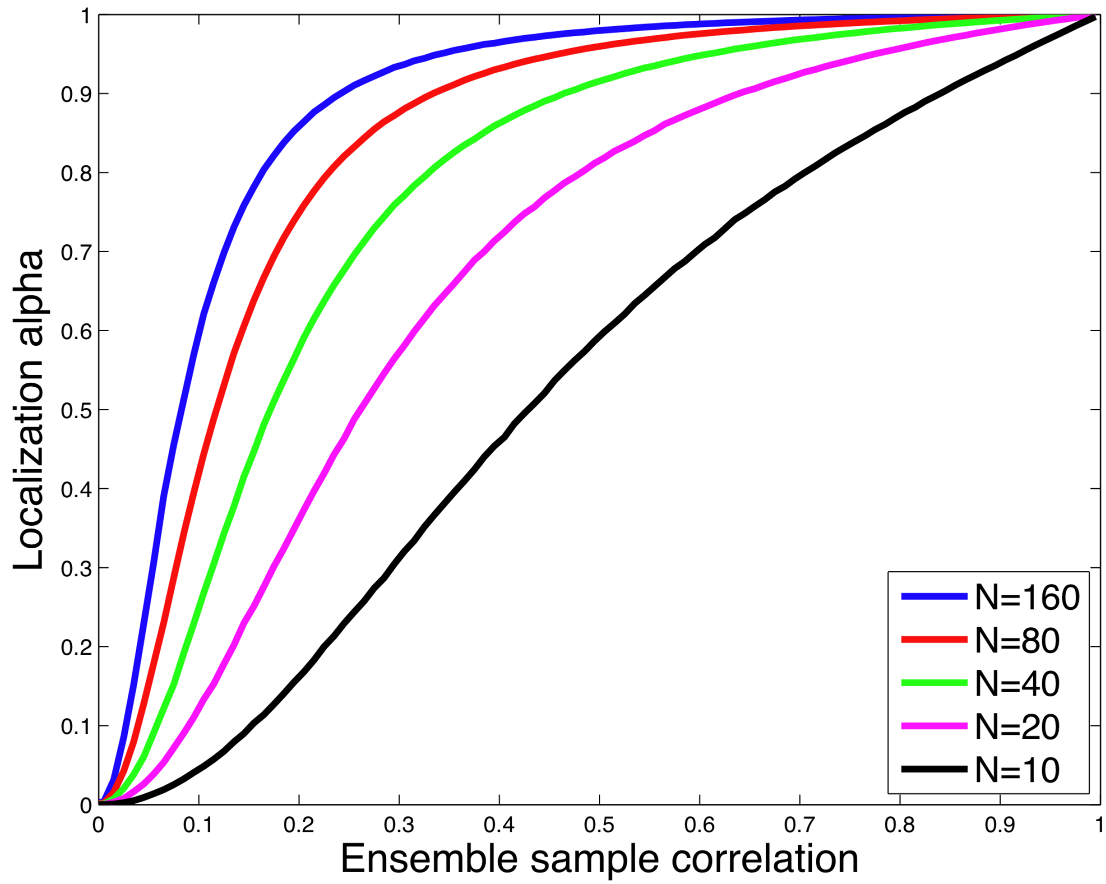
- Need for localization implies that standard ensemble is suboptimal.
- Need additional prior assumptions to improve it.

Assume a prior distribution for correlation r

- Assuming that correlation of y and x_i is U(-1, 1), when nothing else is known, is natural.

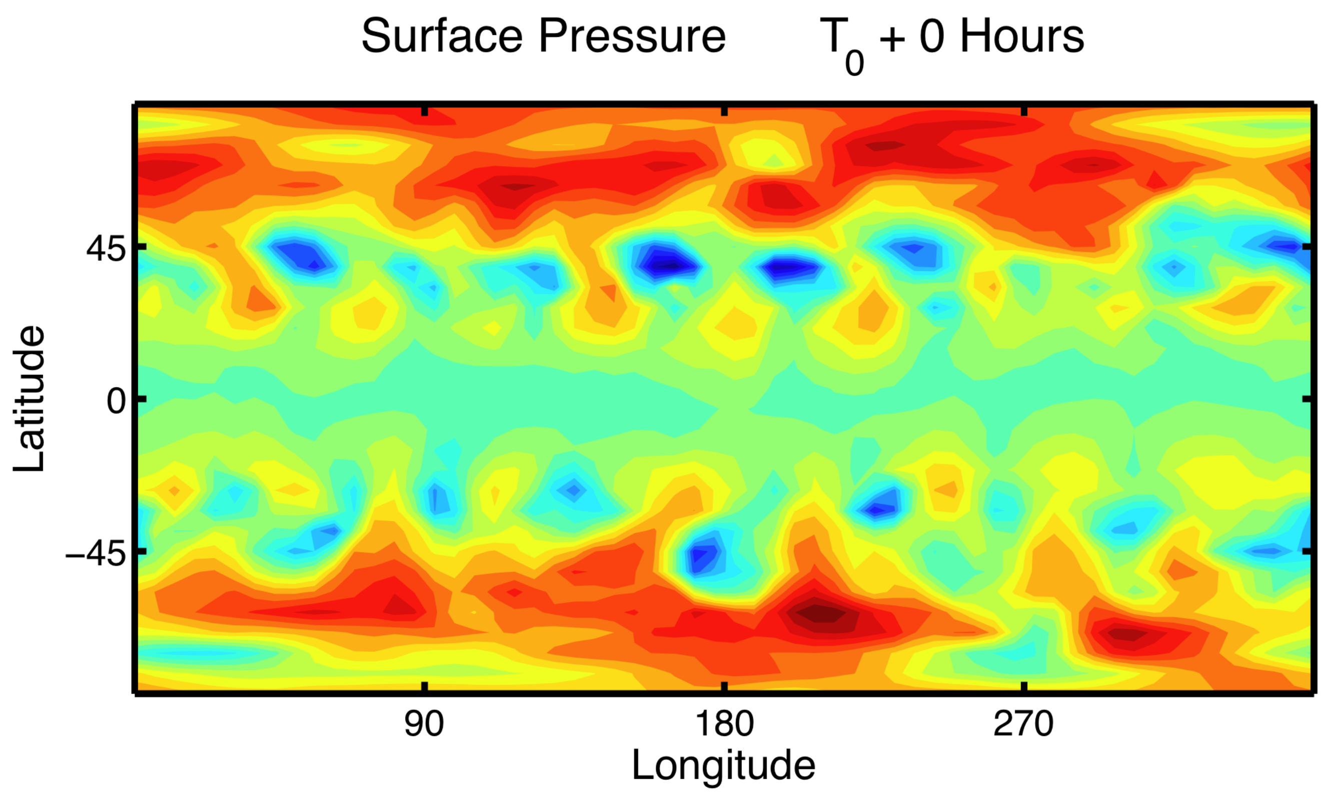
- Given prior U(-1, 1) and ensemble sample correlation $\hat{r}$ can compute:
  - Most likely value of 'true' correlation r,
  - α that minimizes expected RMSE.

Localization α as function of ensemble size N and sample correlation $\hat{r}$.



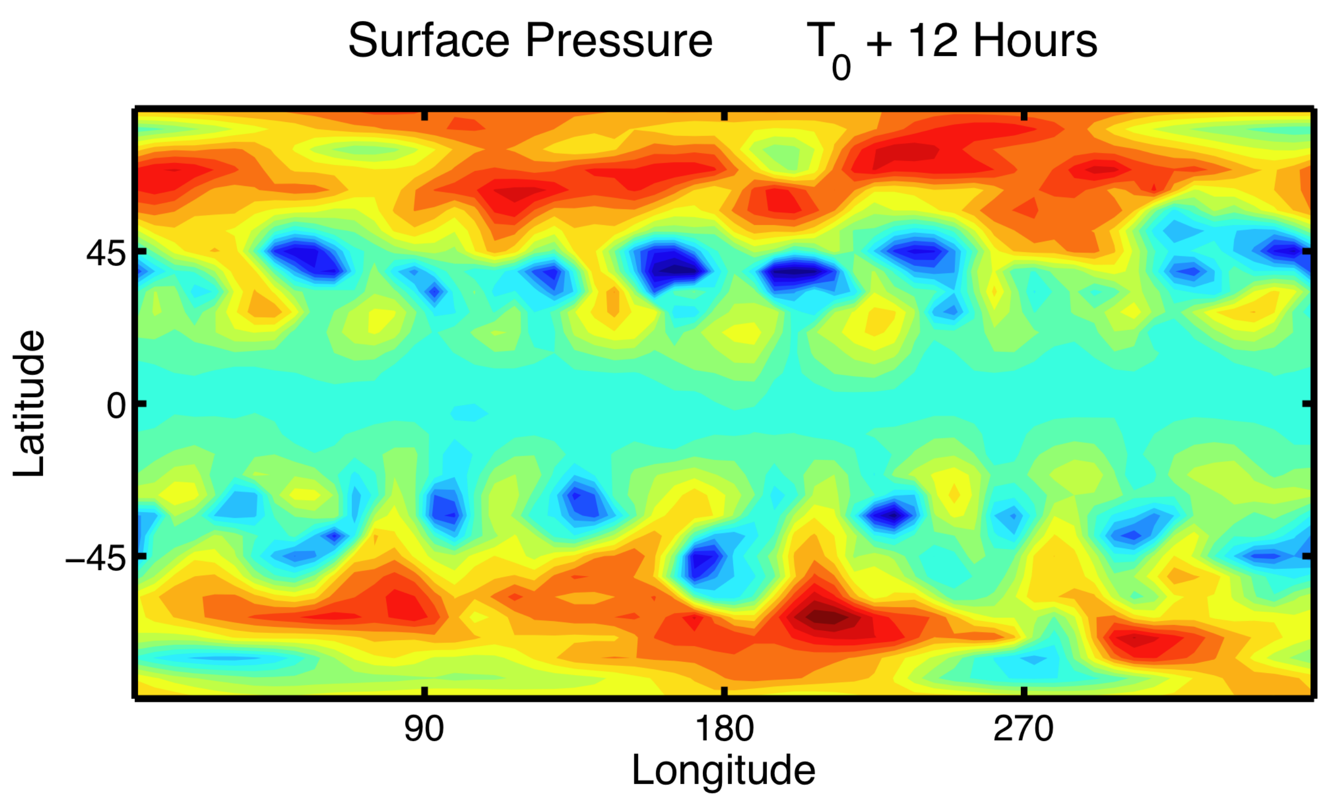
This is precomputed for table look-up.

Test: Low-order dry dynamical core



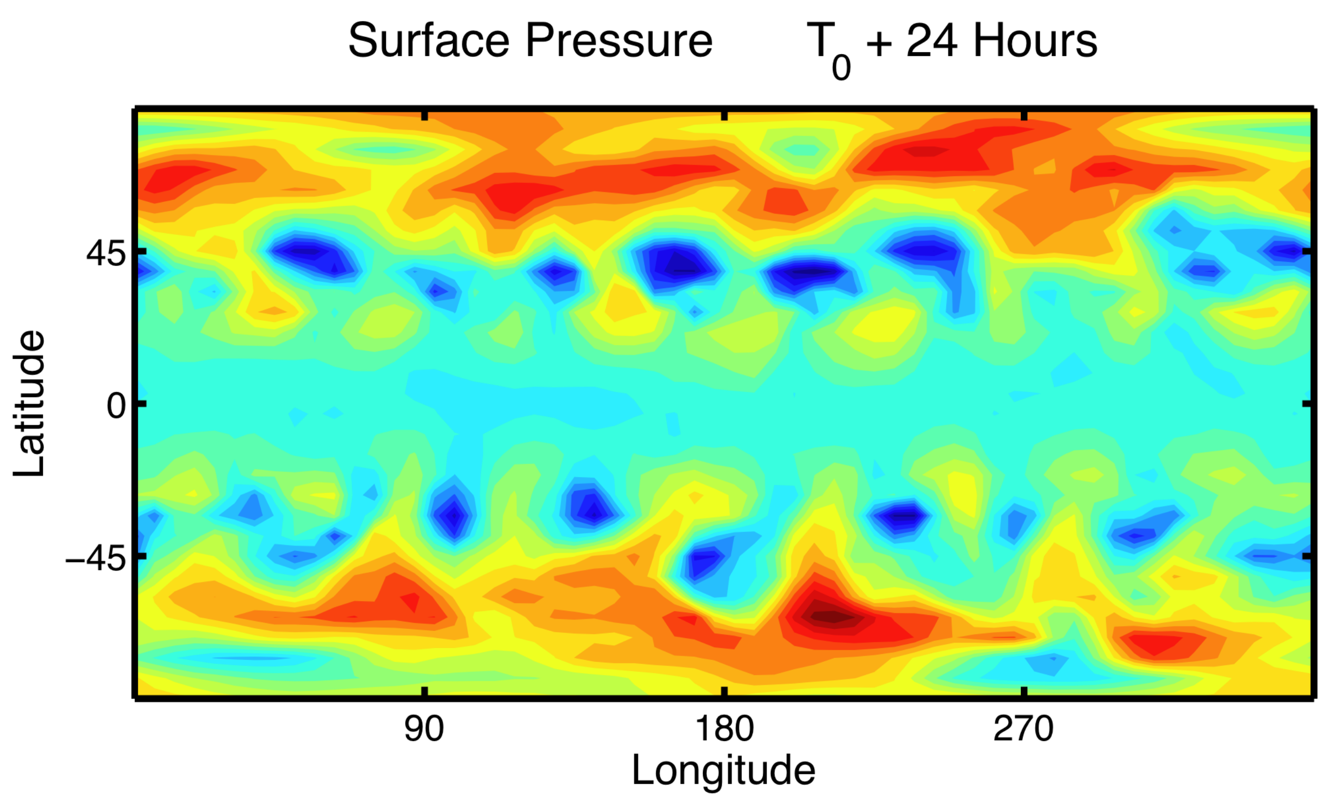
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



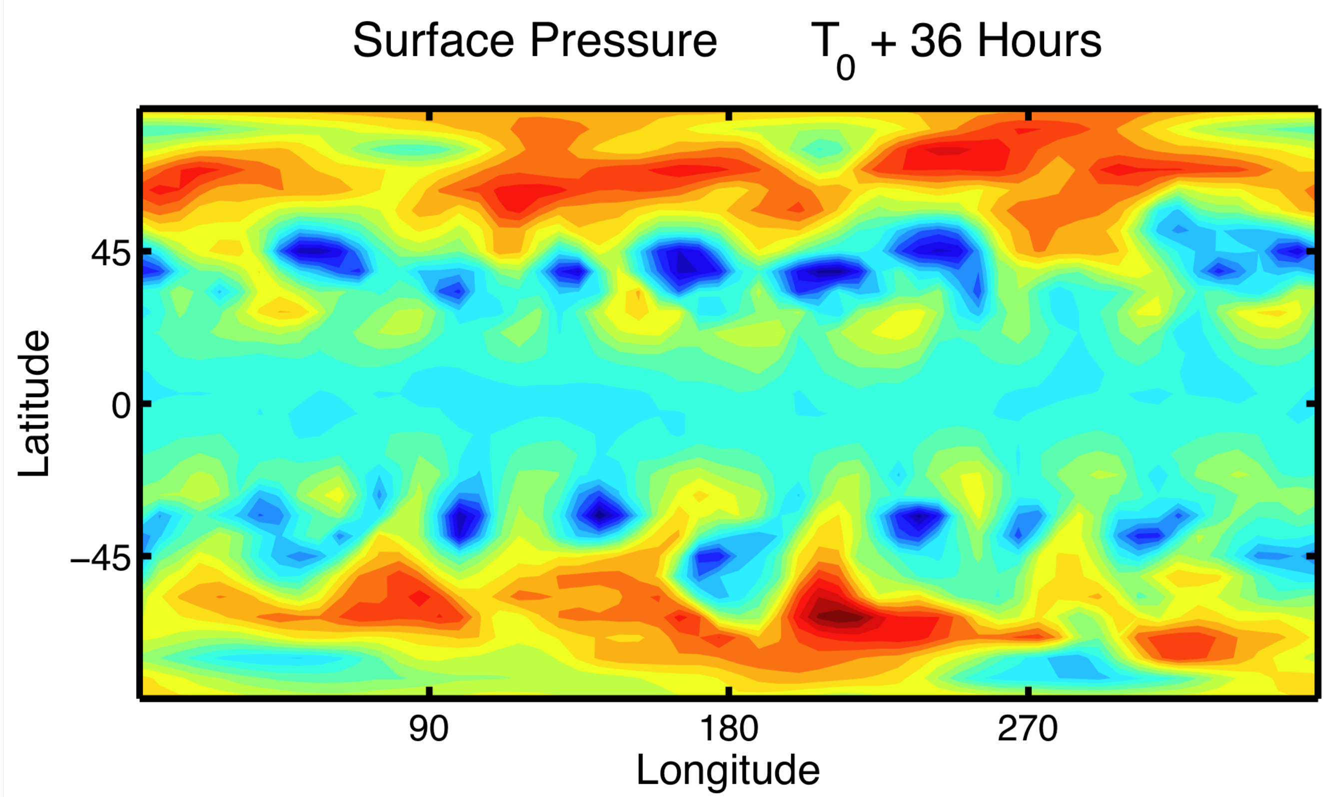
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



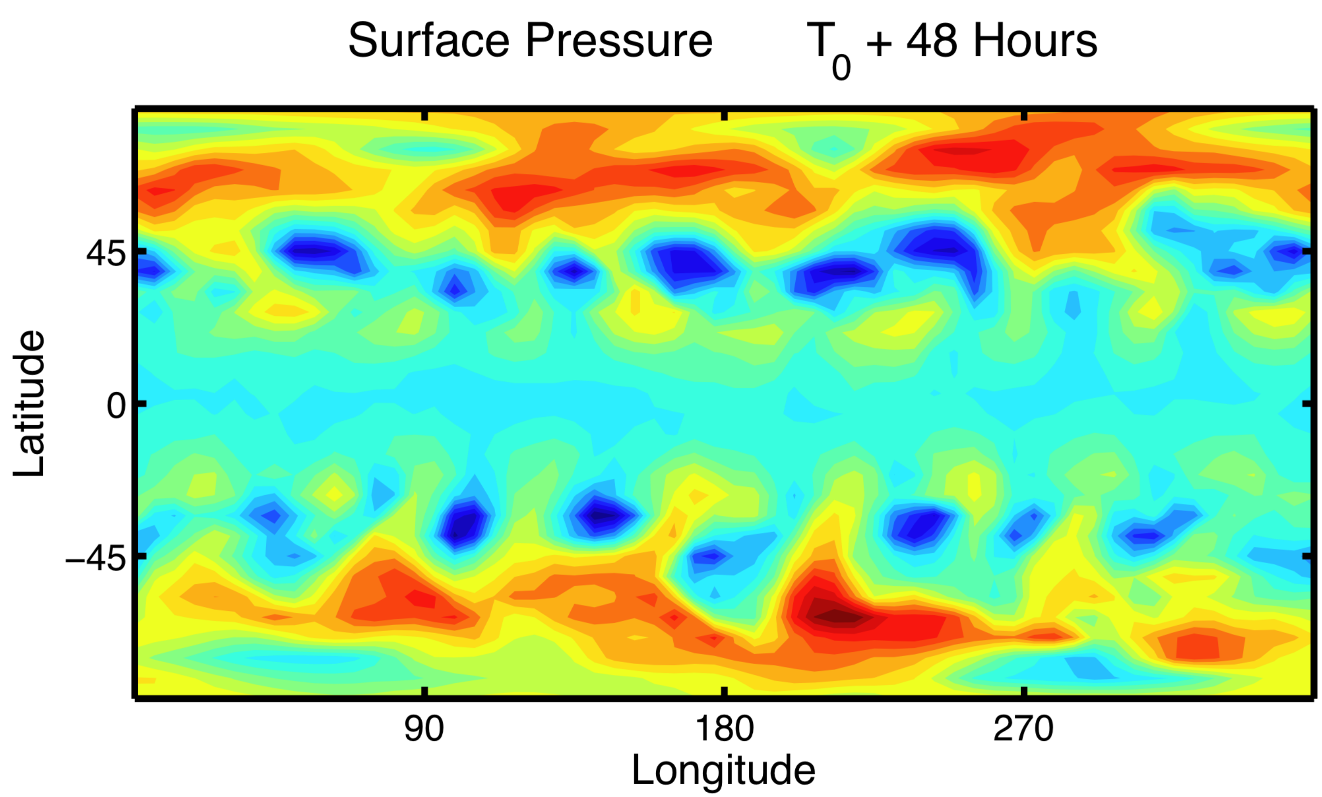
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



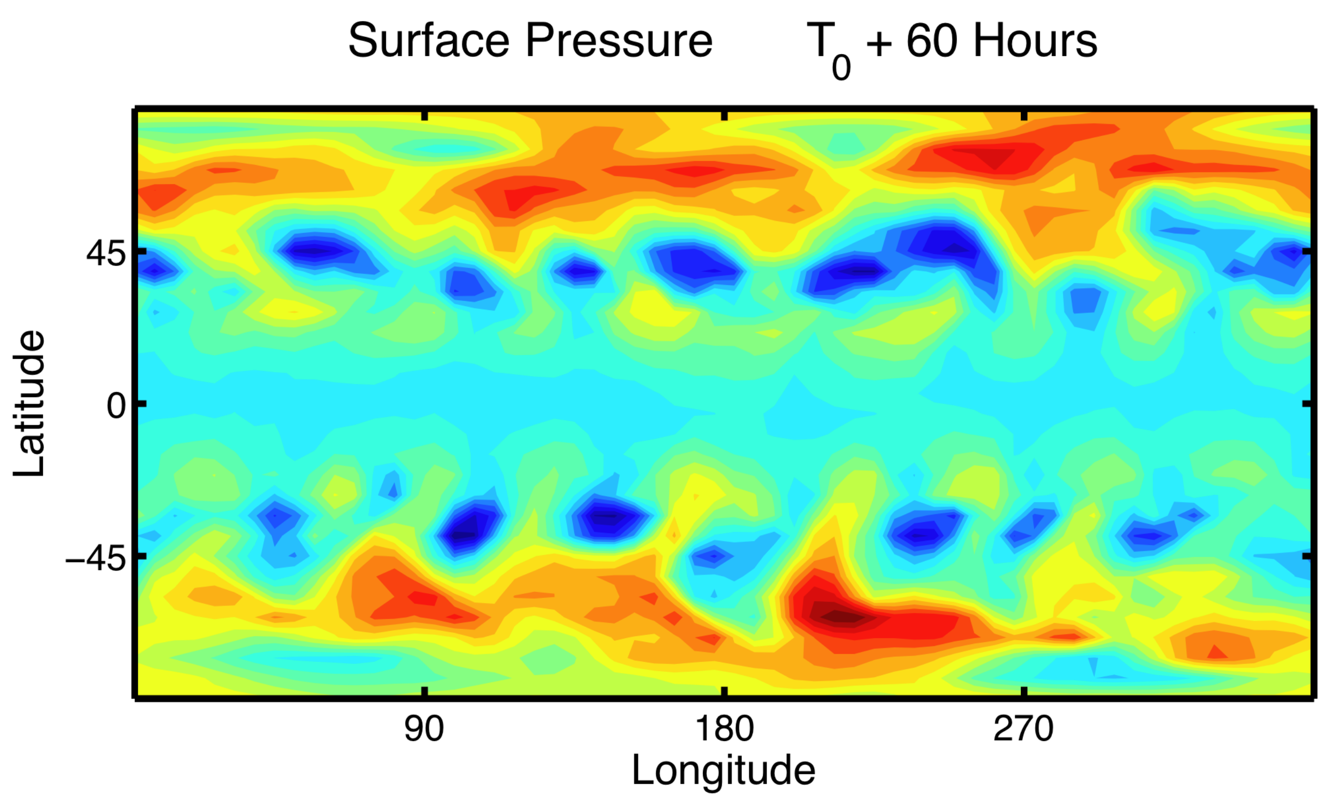
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



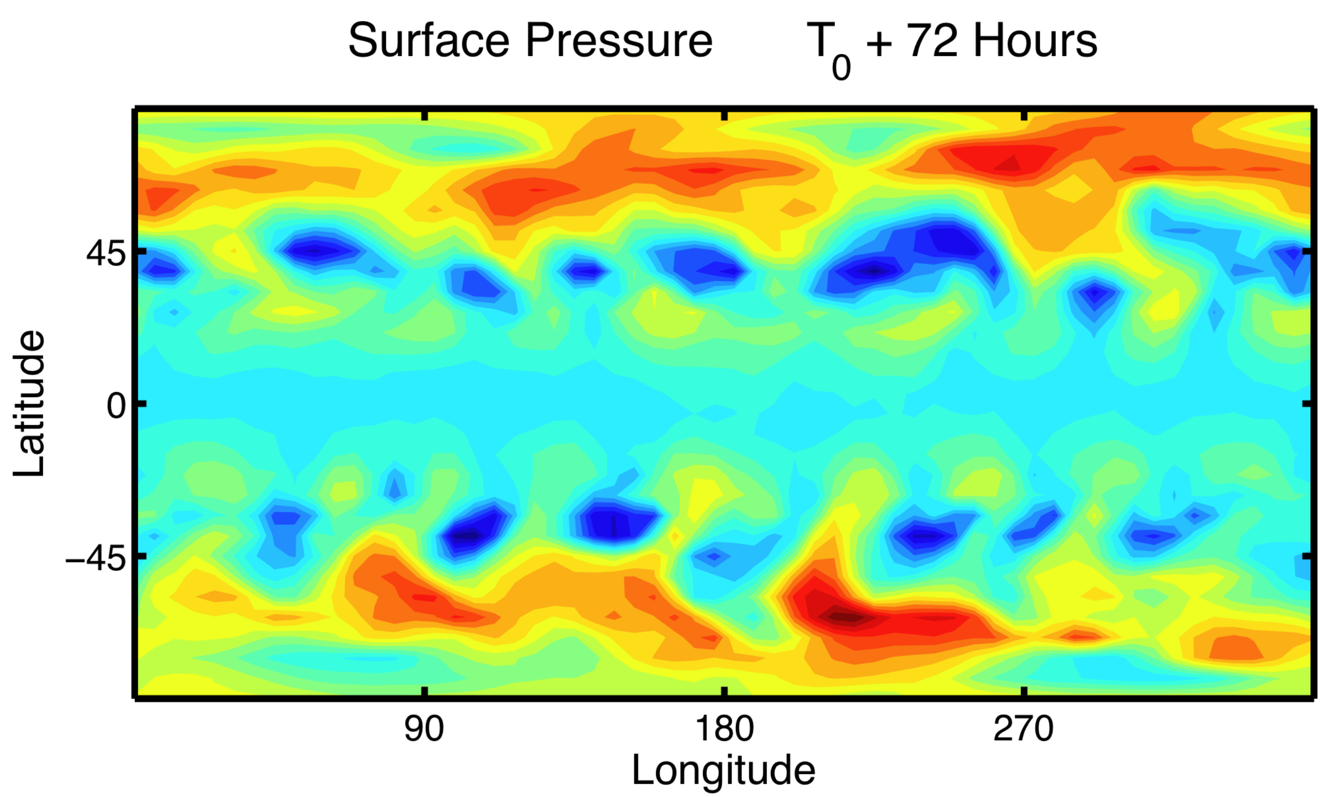
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



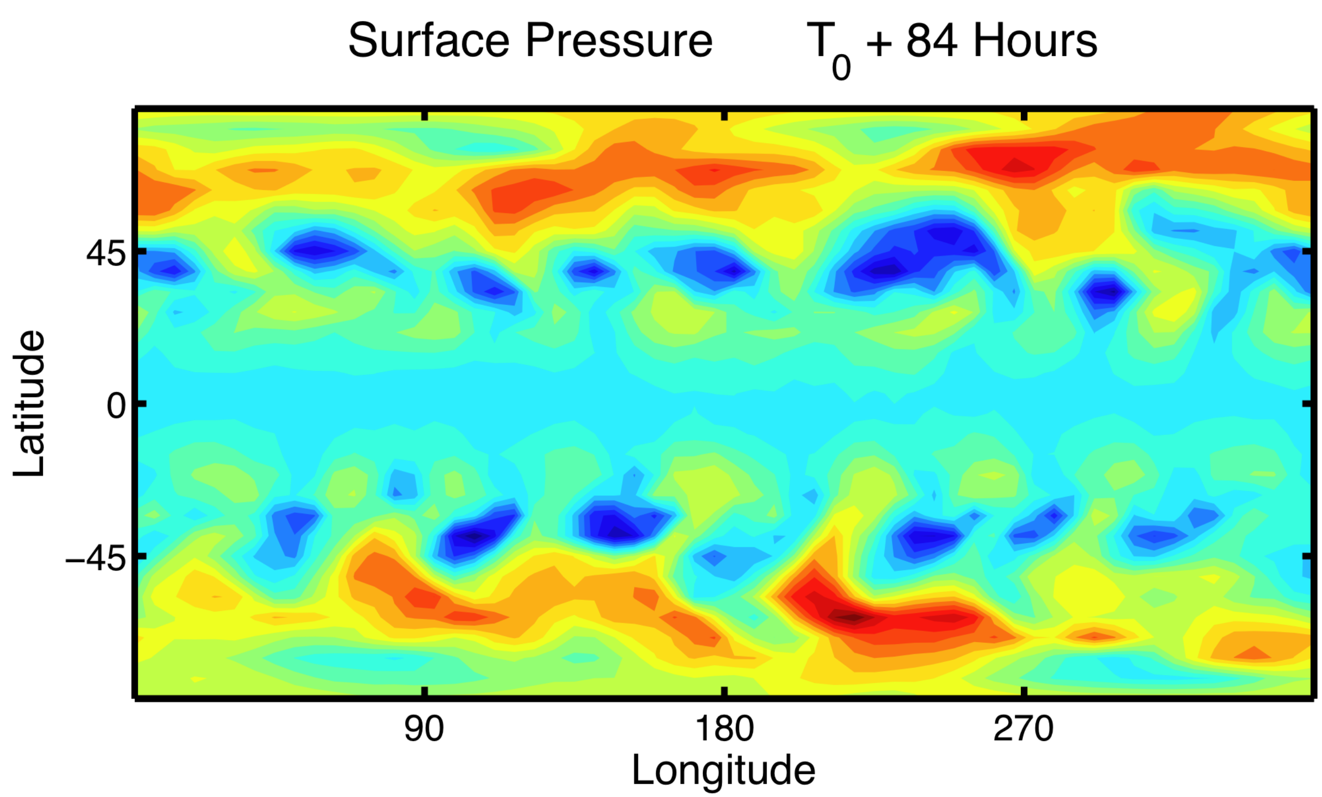
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



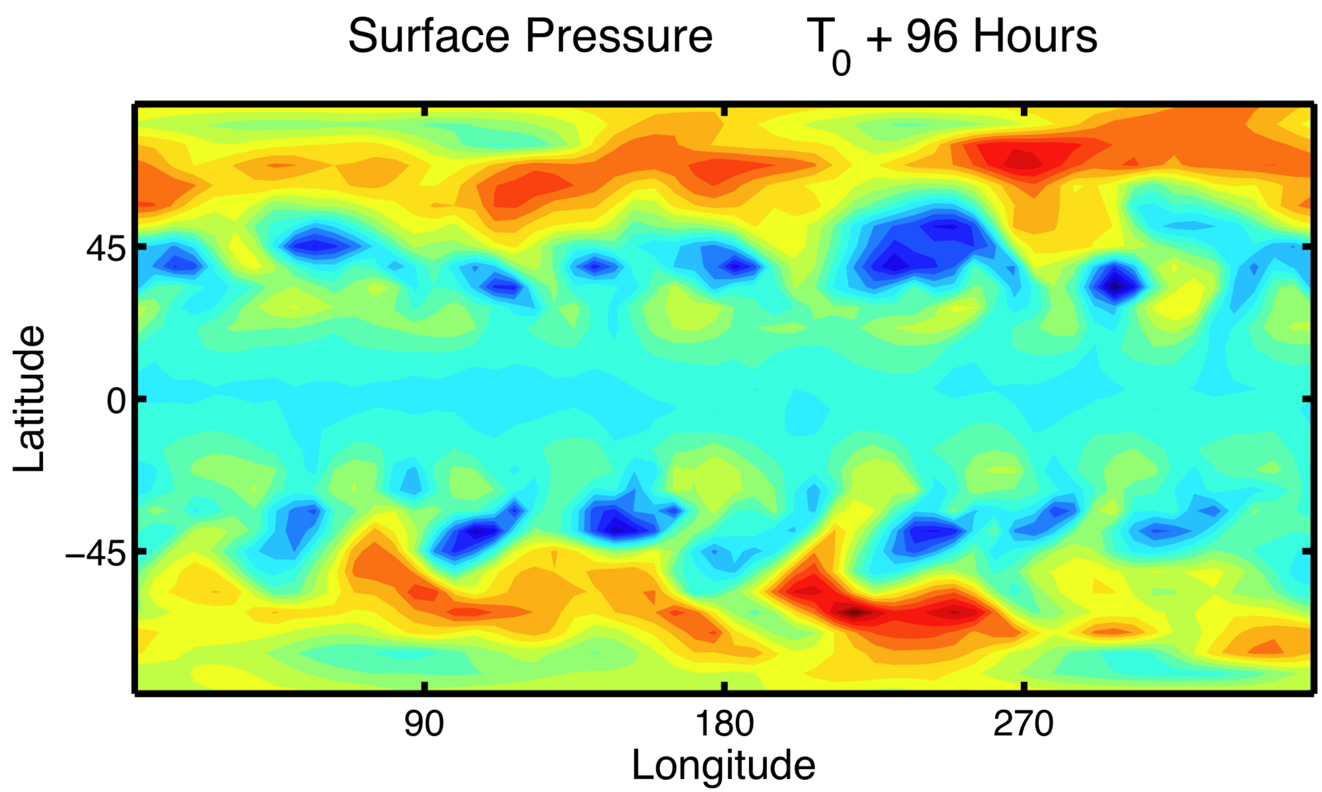
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



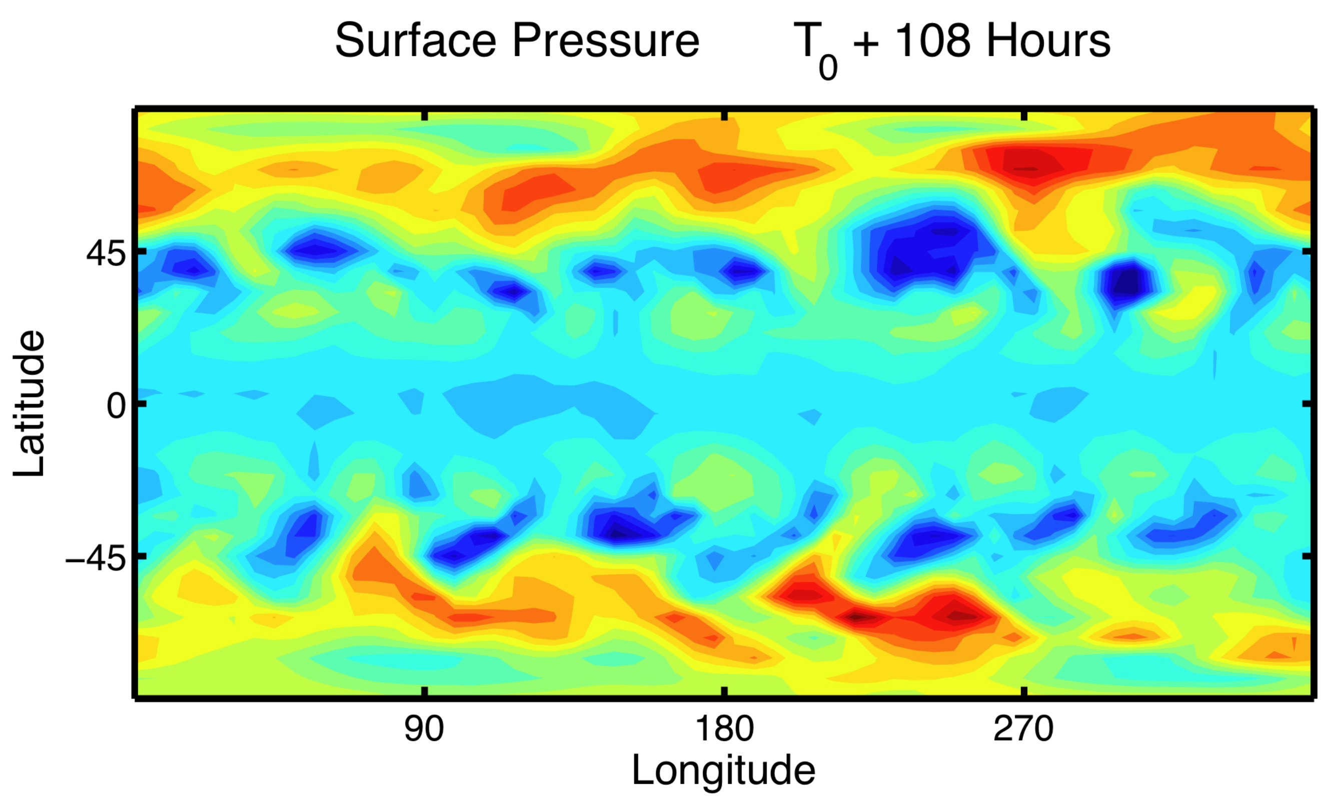
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



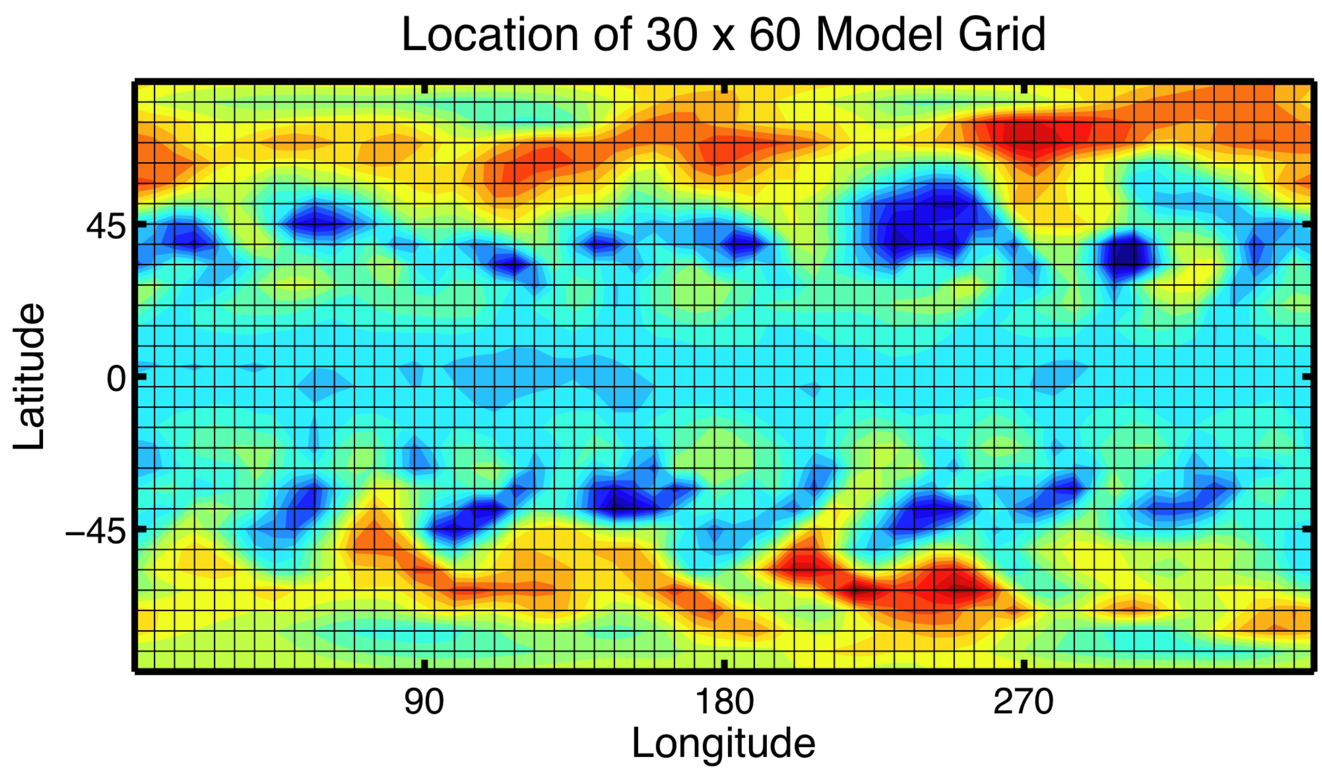
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Test: Low-order dry dynamical core



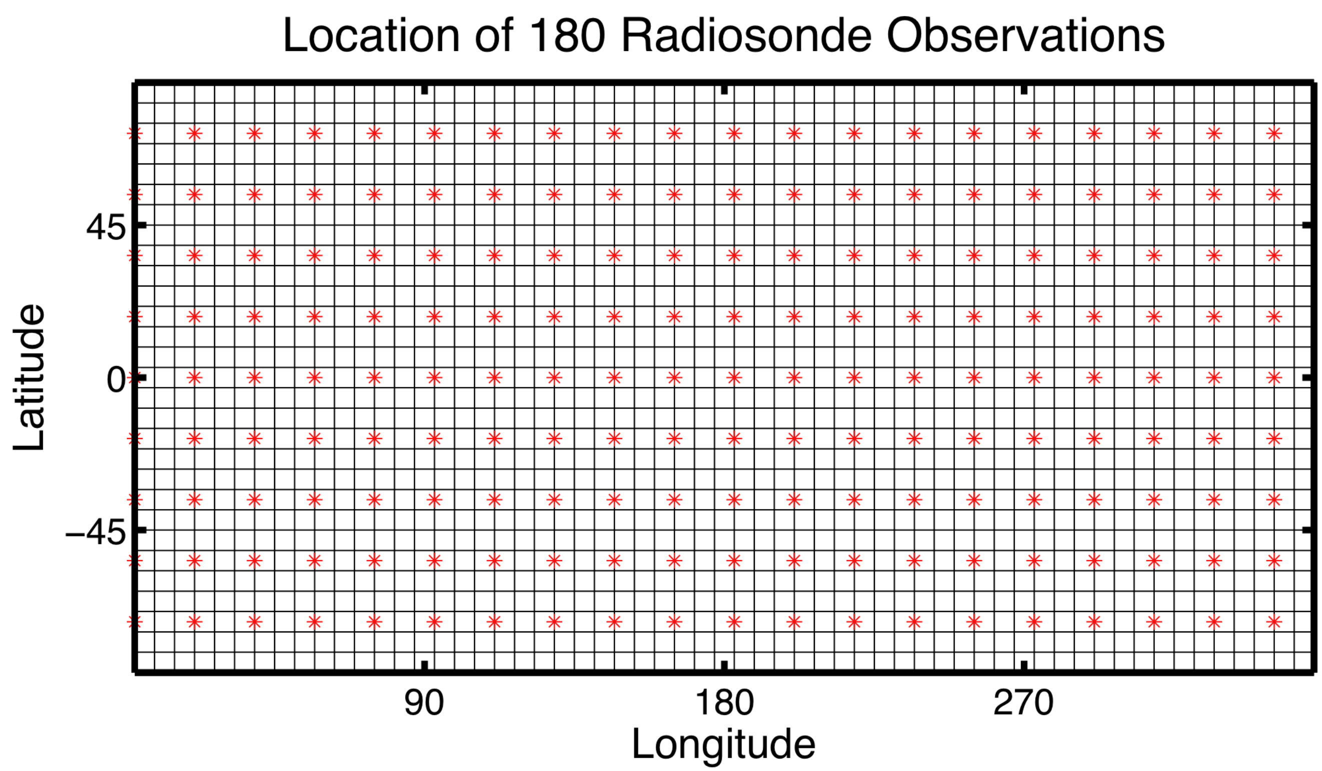
Evolution of surface pressure field every 12 hours.
Has baroclinic instability: storms move east in midlatitudes.

Grid for dynamical core



30x60 horizontal grid, 5 levels.
Surface pressure, temperature, wind components.
28,800 variables.

Simulated radiosonde observing network

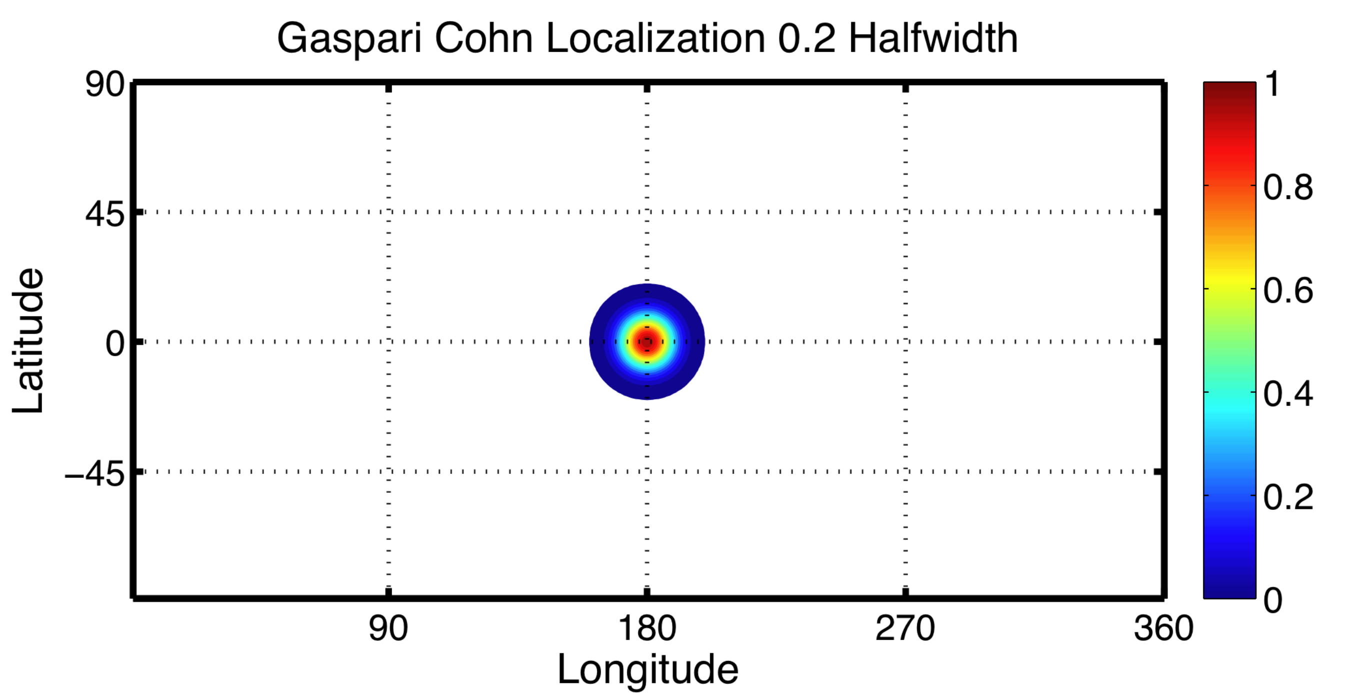


Observe every 12 hours.

Observe all 16 variables in 180 columns shown.

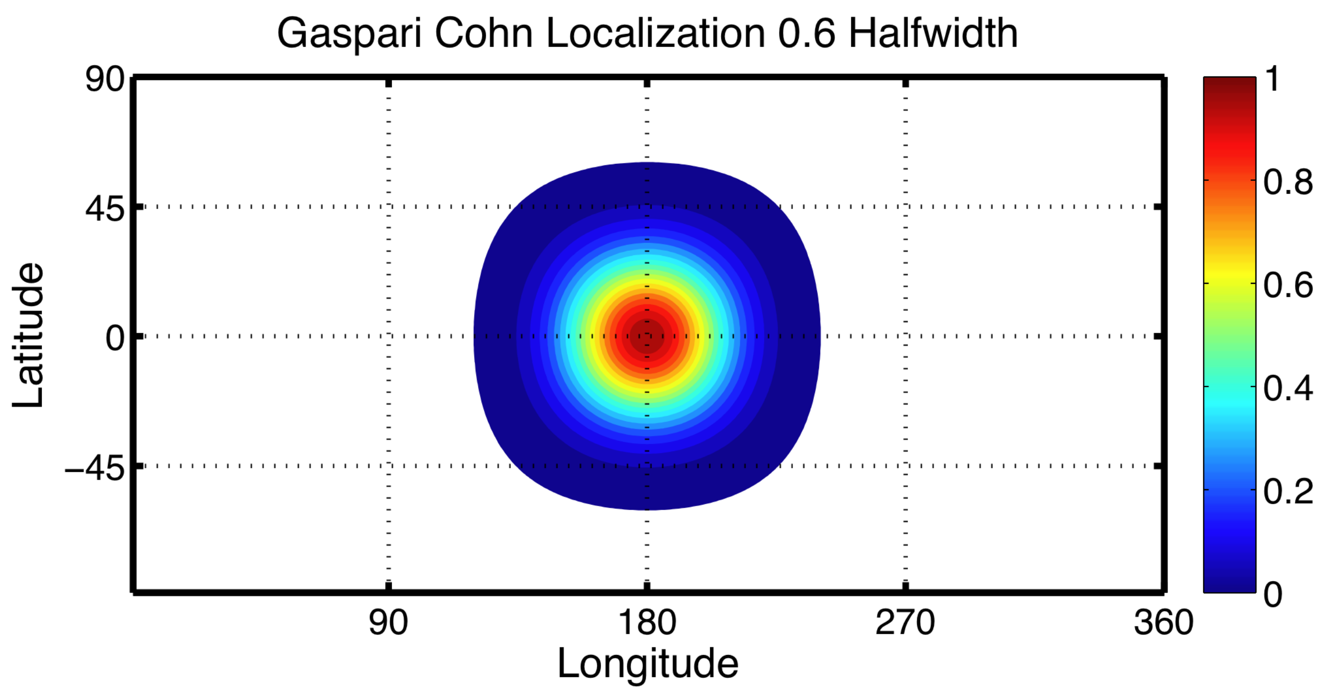
Error SD: Ps=2hPa, T=3K, winds=3m/s.

Background Gaspari Cohn localization: 0.2



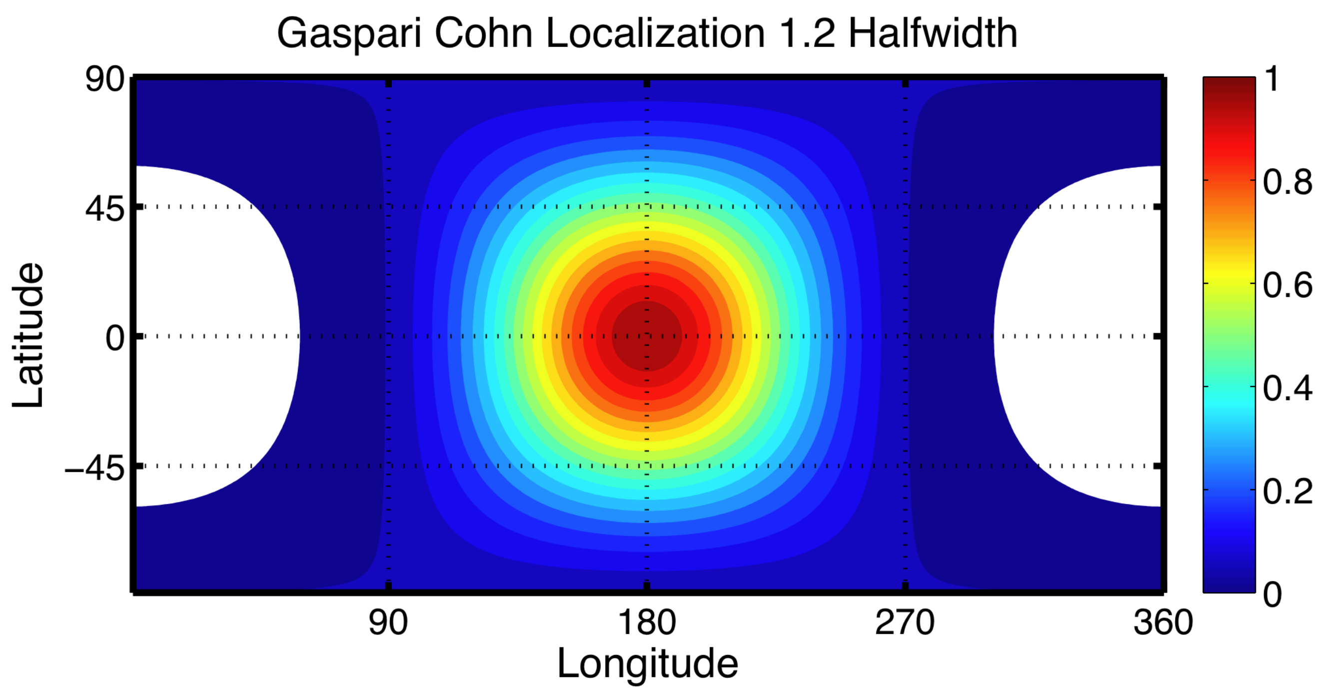
Use a range of Gaspari Cohn background localizations.
Larger ensemble sizes can use broader localizations.

Background Gaspari-Cohn Localization: 0.6



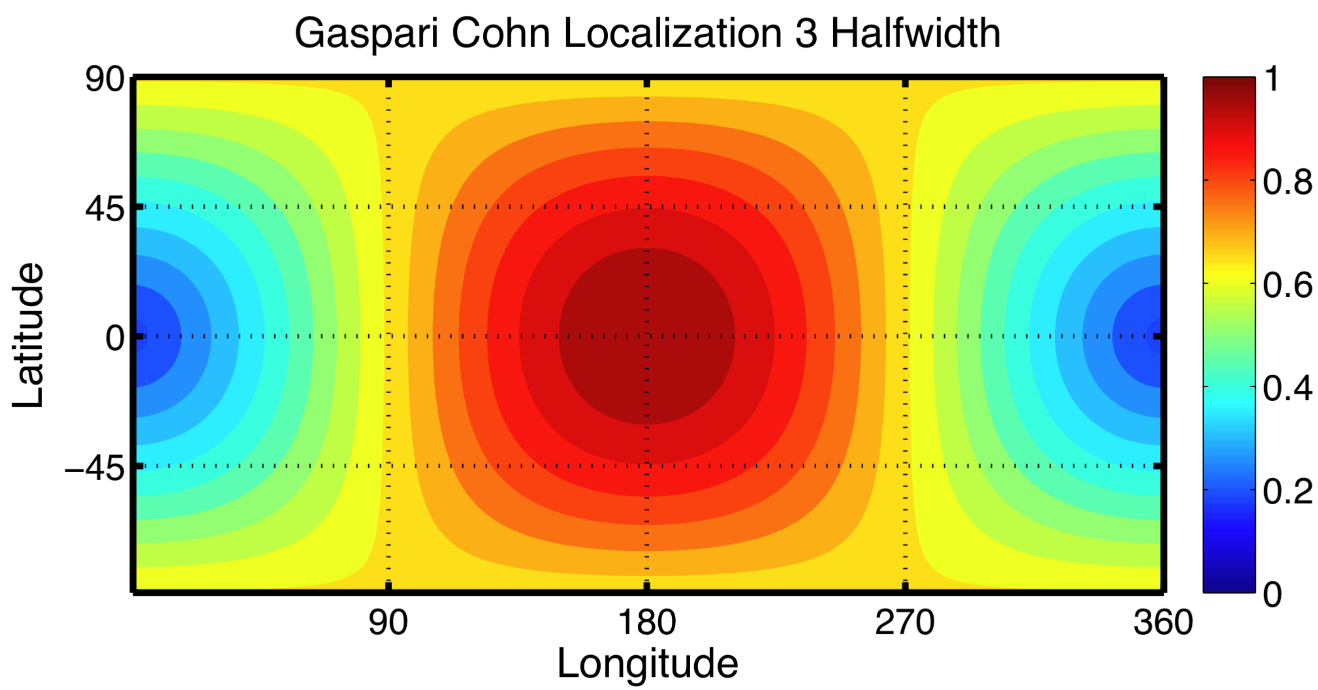
Use a range of Gaspari Cohn background localizations.
Larger ensemble sizes can use broader localizations.

Background Gaspari-Cohn Localization: 1.2



Use a range of Gaspari Cohn background localizations.
Larger ensemble sizes can use broader localizations.

Background Gaspari-Cohn Localization: 3.0



Use a range of Gaspari Cohn background localizations.
Larger ensemble sizes can use broader localizations.

Design of an OSSE

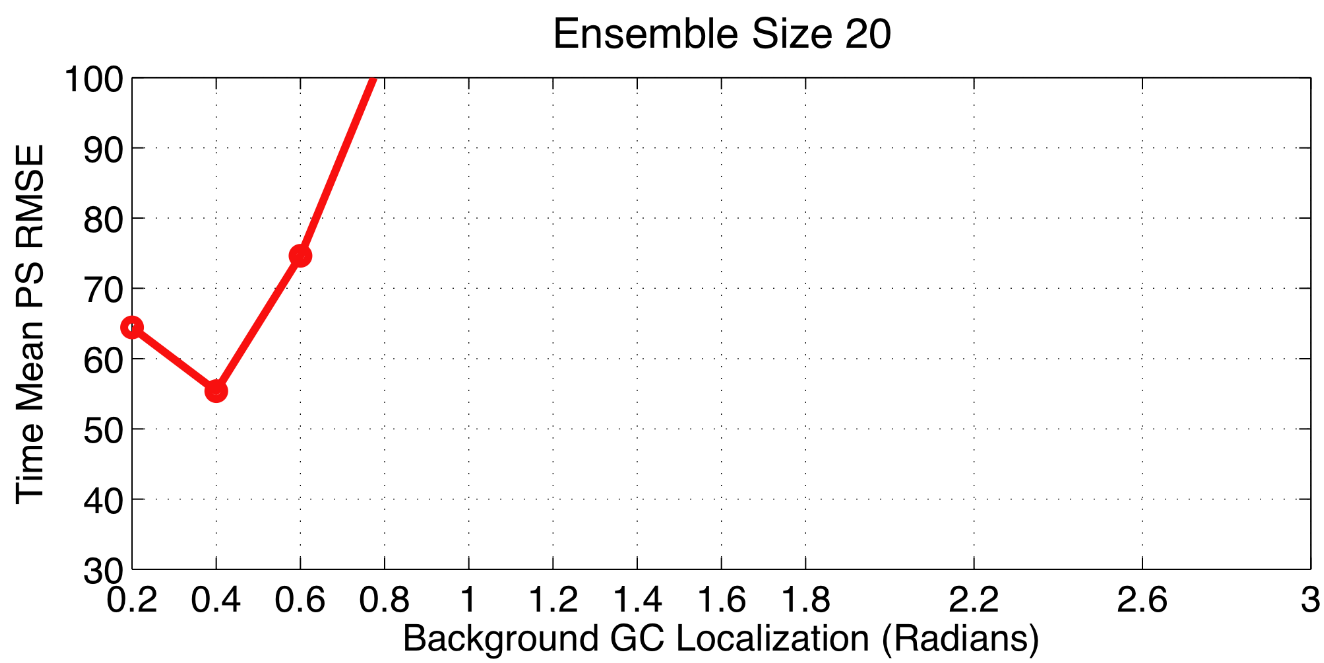
All cases assimilate same observations.
150 day period (300 assimilation cycles).

All use damped adaptive inflation with same settings.

Primary metric is prior RMSE for surface pressure.

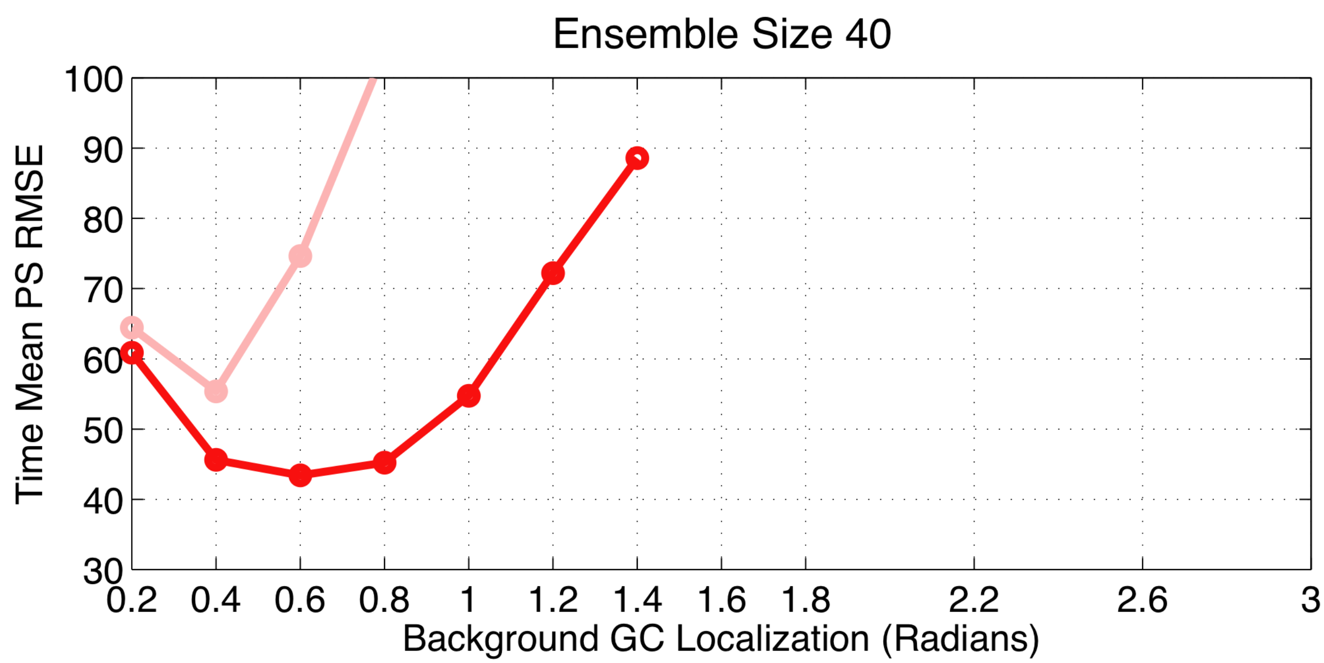
Compare base algorithm to sampling error correction.

RMSE for Base Case



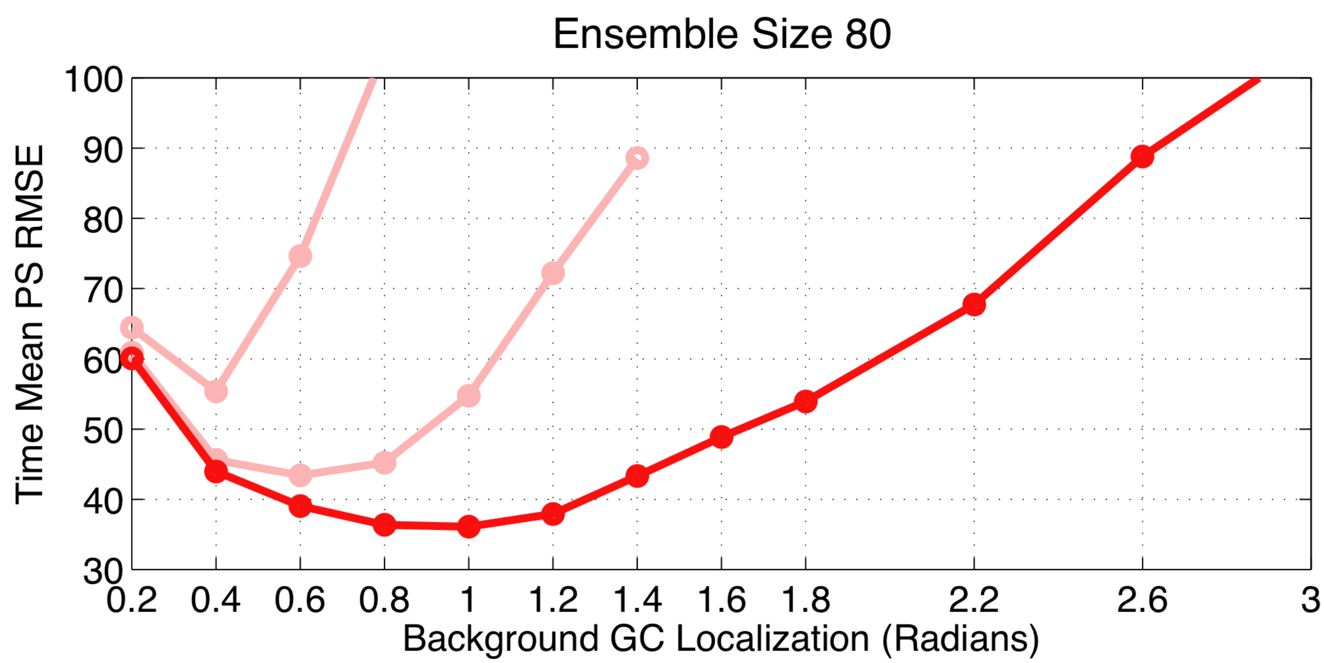
Base GC localization: Ensemble size 20.

RMSE for Base Case



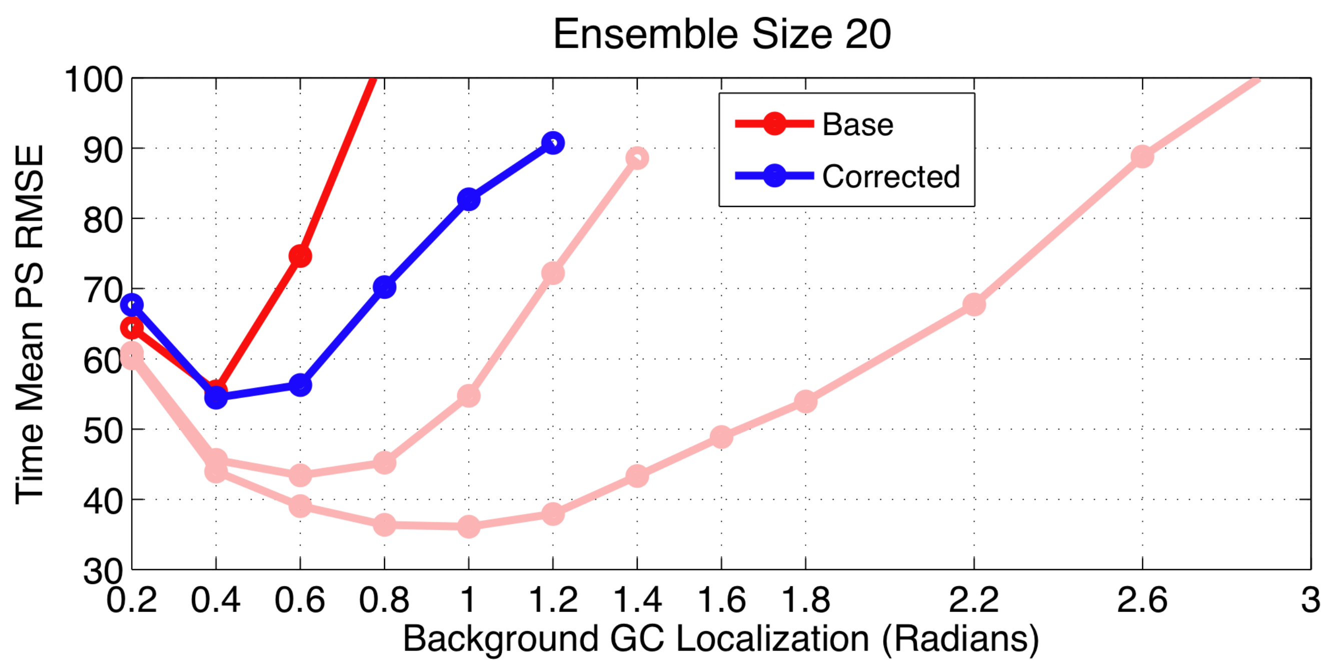
Base GC localization: Ensemble size 40.
RMSE reduced, localization broader for larger ensemble.

RMSE for Base Case



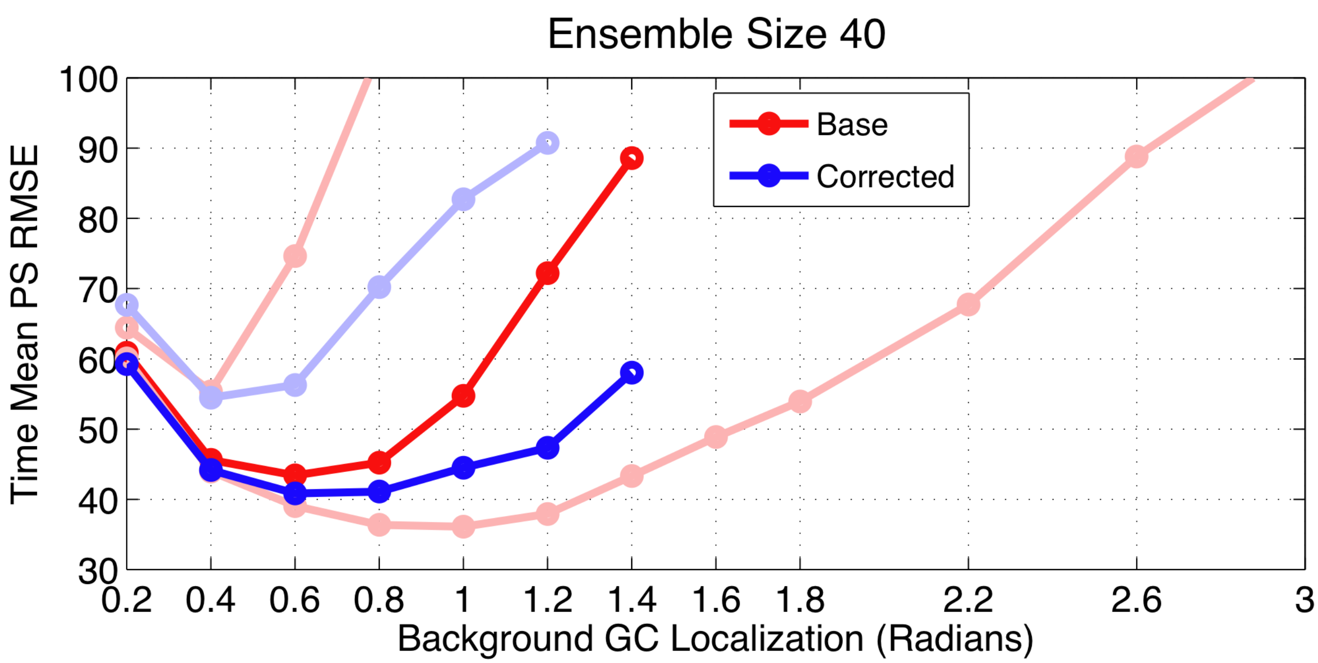
Base GC localization: Ensemble size 80.
RMSE reduced, localization broader for larger ensemble.

Systematic Error Correction Reduces RMSE



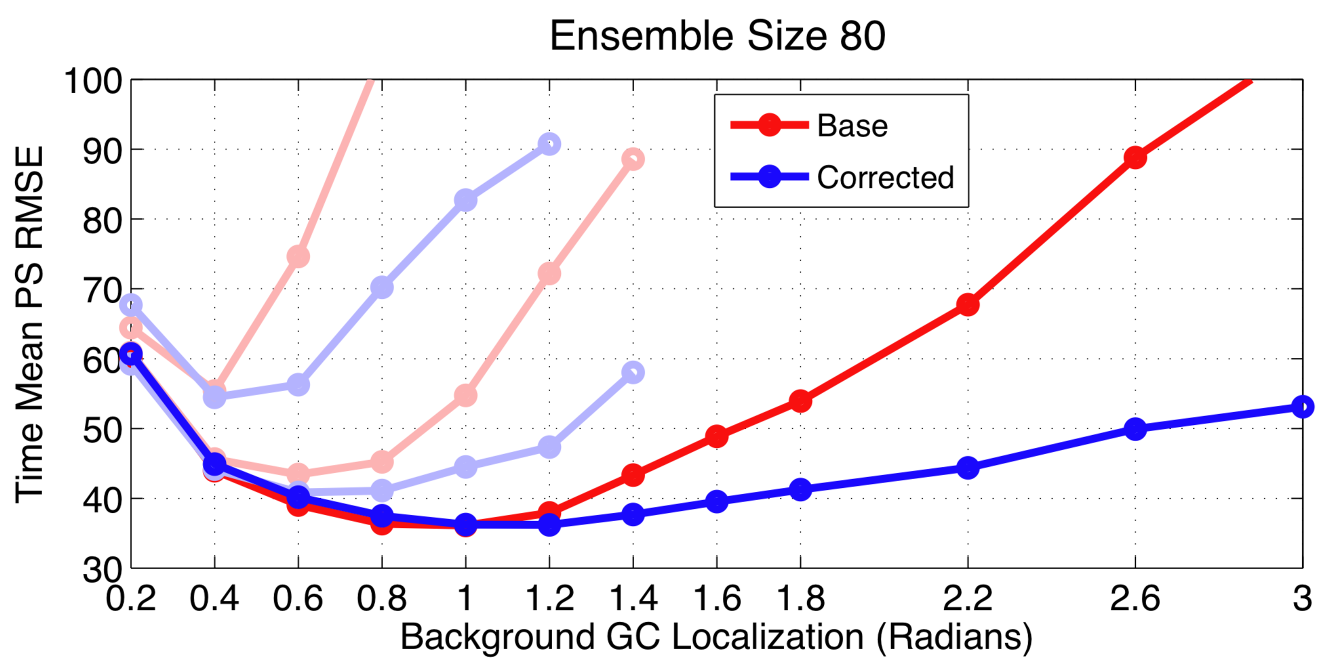
Sampling Error Correction: Ensemble size 20.
Has smaller RMSE for range of background GC localization.

Systematic Error Correction Reduces RMSE



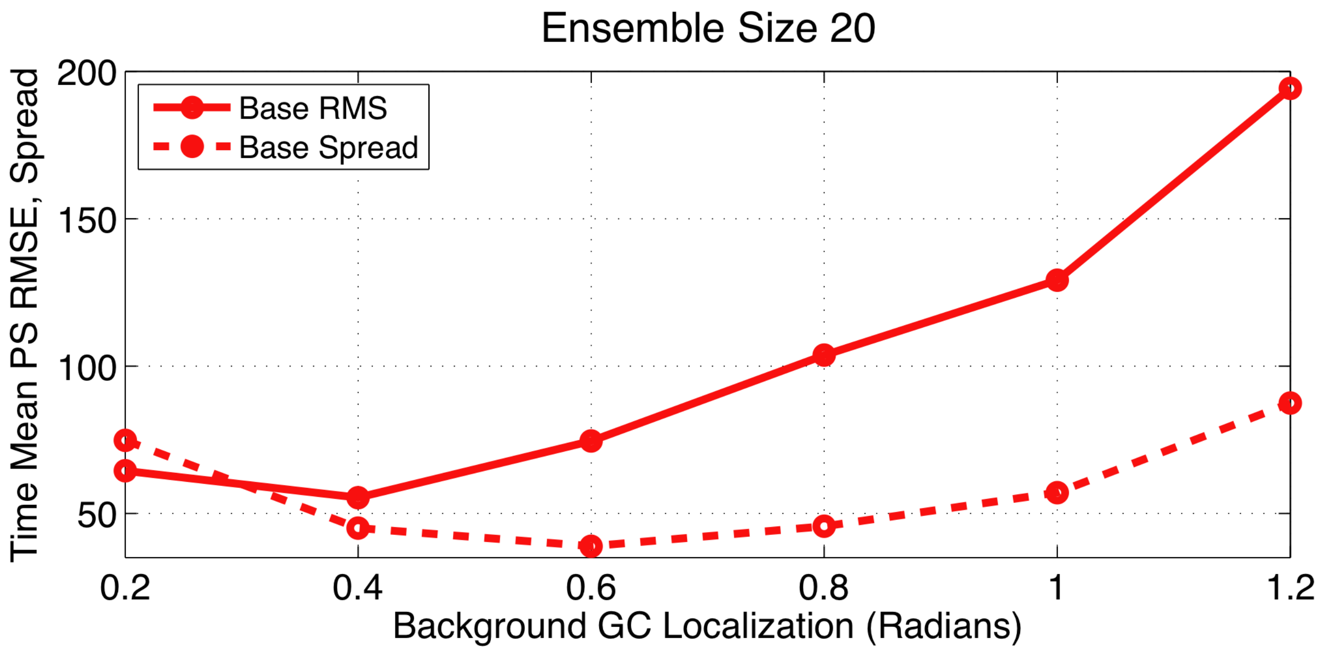
Sampling Error Correction: Ensemble size 40.
Has smaller RMSE for range of background GC localization.

Systematic Error Correction Reduces RMSE



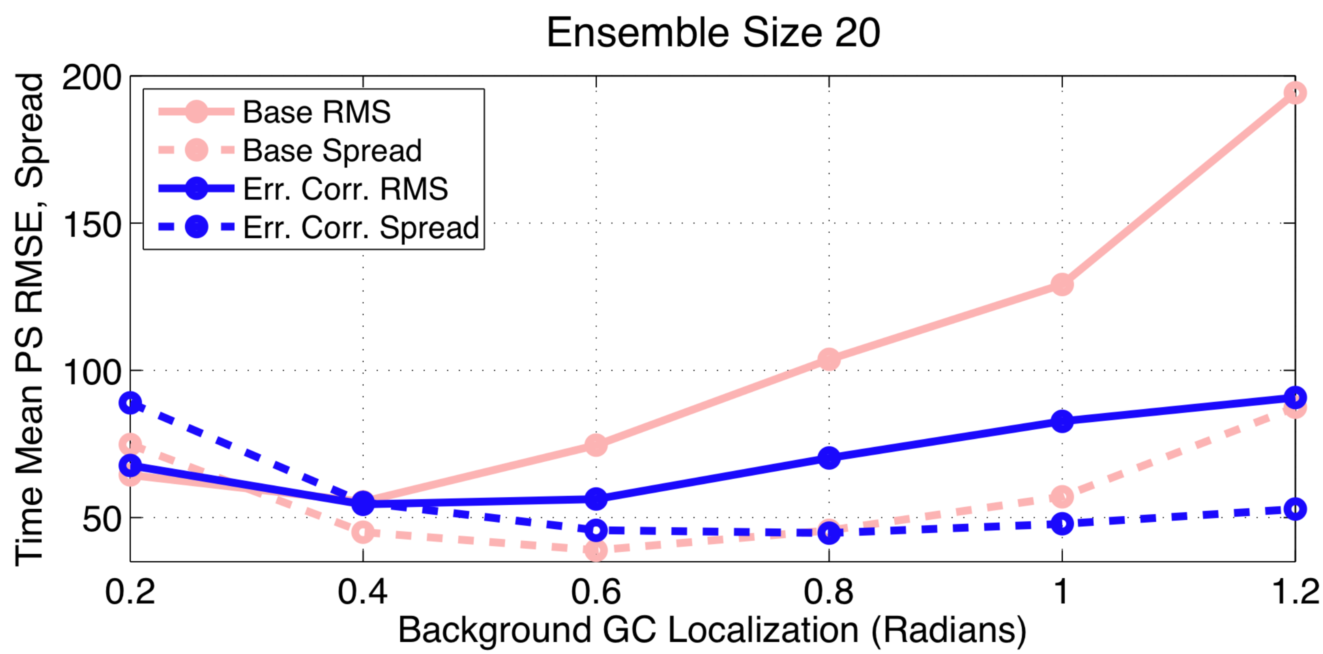
Sampling Error Correction: Ensemble size 80.
Has smaller RMSE for range of background GC localization.

Spread in Base Case is Deficient



Base spread deficient for large GC, even with inflation.
This is expected as more remote observations are used.

Systematic Error Correction Improves Spread



Sample corrected spread is not as deficient.

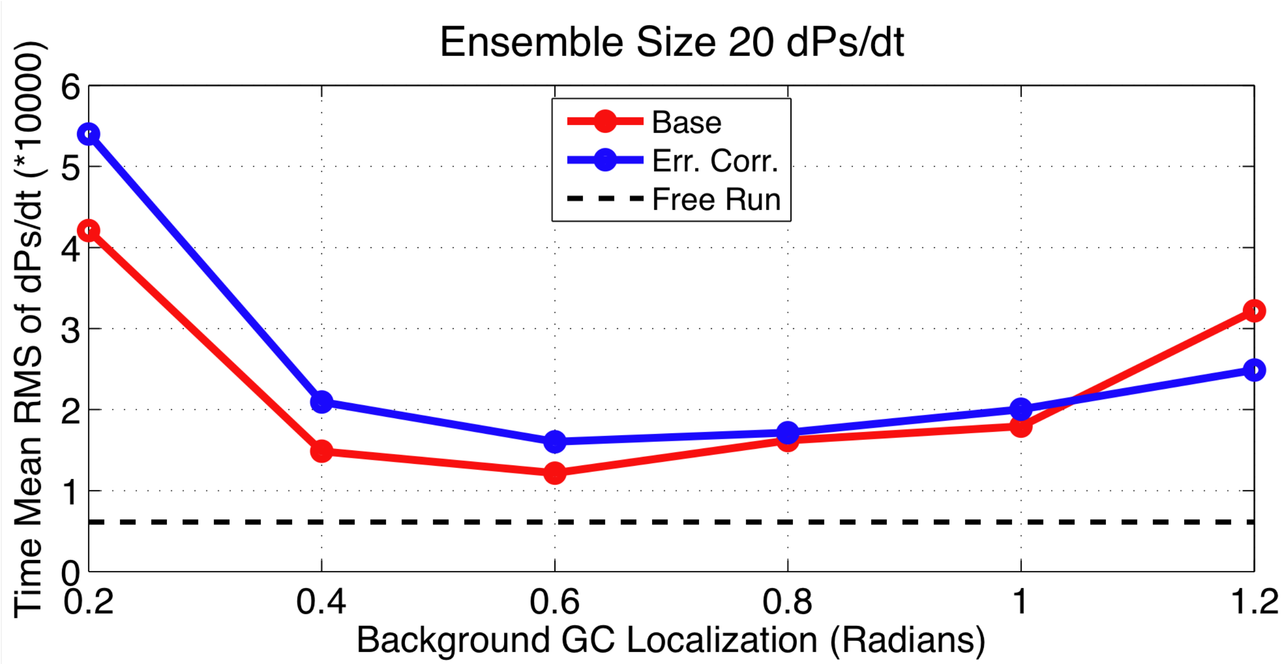
Evaluating Imbalance

Localization associated with increased ‘imbalance’.
Transient off-attractor dynamics after assimilation.

Here, imbalance results in high-frequency gravity waves.
Measured by mean amplitude of surface pressure tendency.
Computed at first timestep after each assimilation.
Averaged globally and for all times.

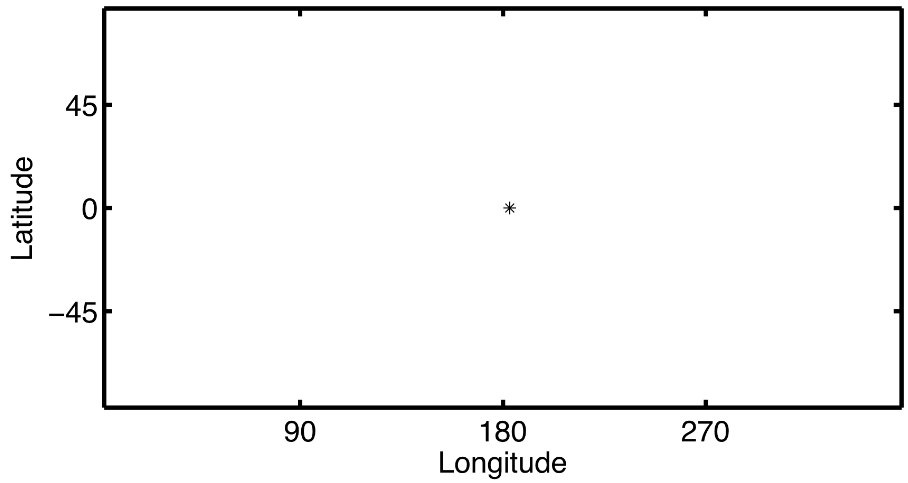
Would like this to be small compared to a free model run.

Systematic Error Correction is Less Balanced



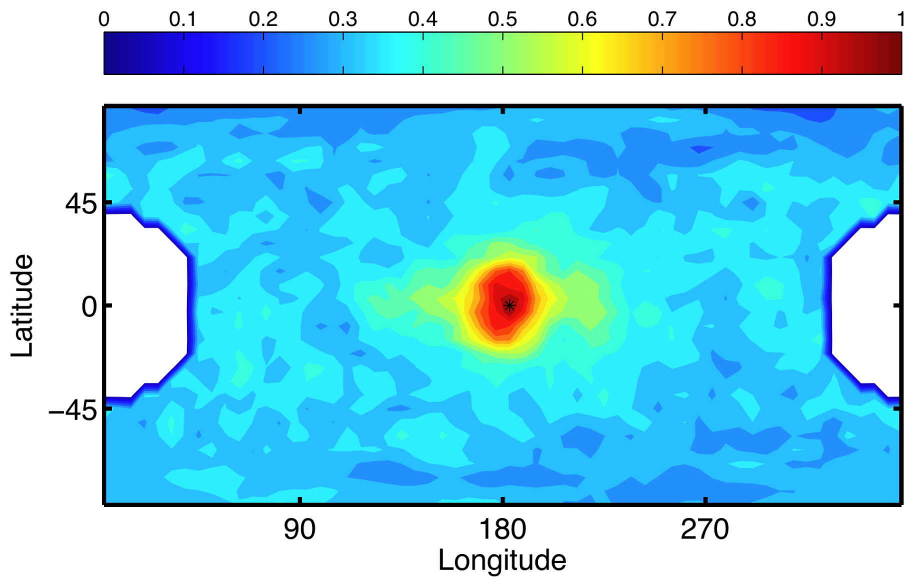
Both cases have enhanced dPs/dt for ensemble size 20.
Sampling error corrected generally worse.
dPs/dt for long free model run shown for comparison.

Time Mean ‘Localization’ for a PS Observation



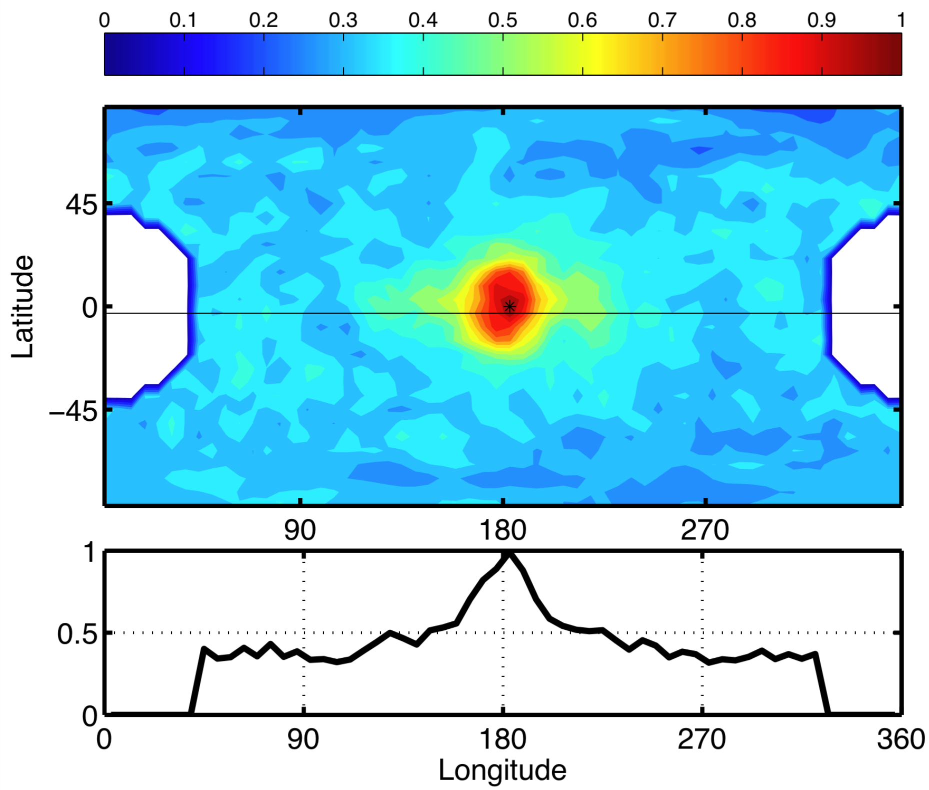
Observation of Ps on the equator.

Time Mean 'Localization' for a PS Observation



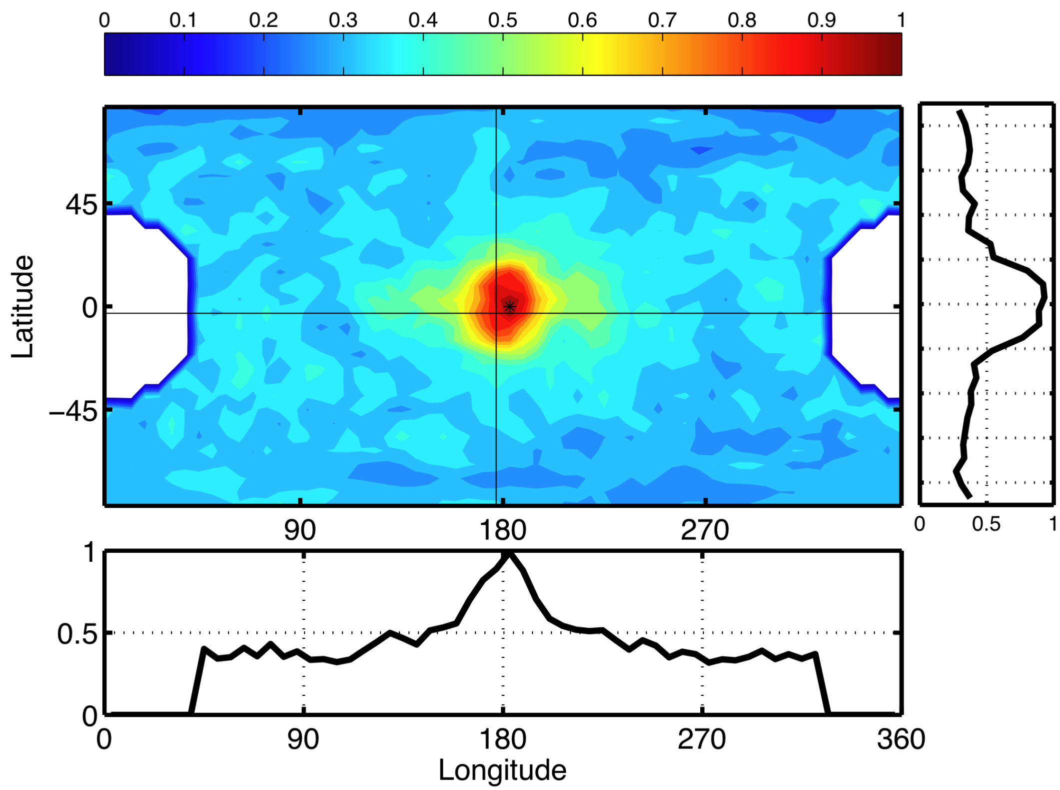
Time mean sampling localization for equatorial Ps ob. on Ps state variables. (Ens. Size 80, GC localization 1.2).

Time Mean 'Localization' for a PS Observation



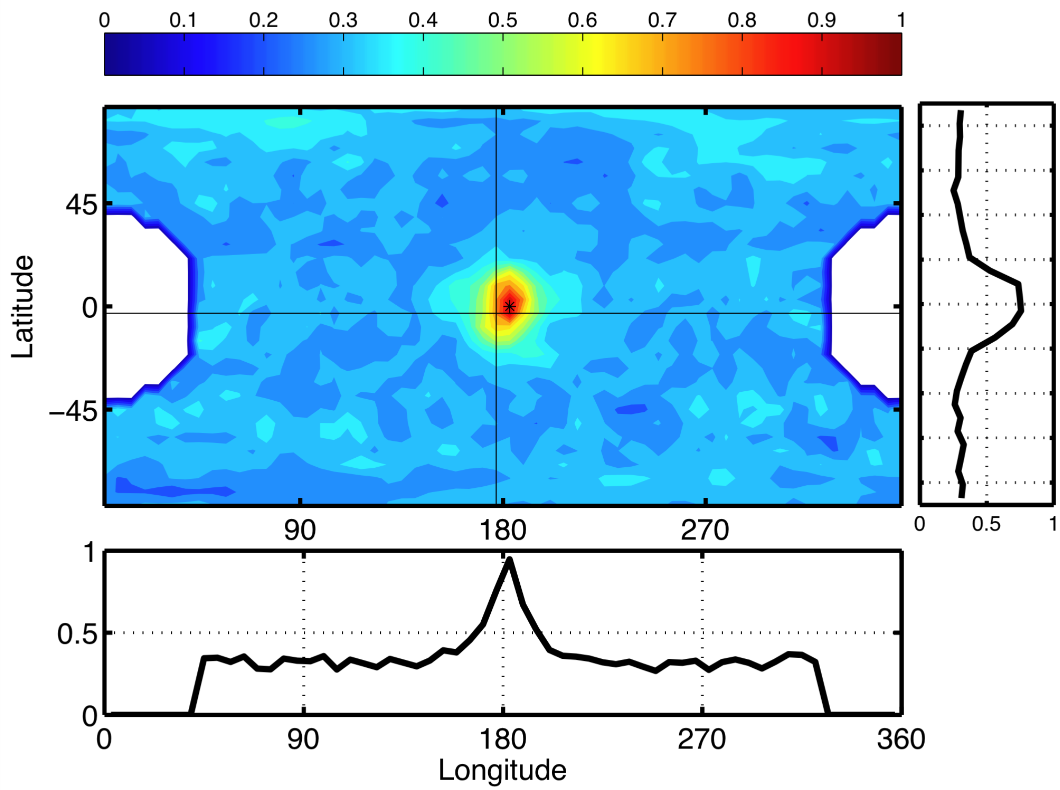
Time mean sampling localization for equatorial Ps ob. on Ps state variables. (Ens. Size 80, GC localization 1.2).

Time Mean ‘Localization’ for a PS Observation



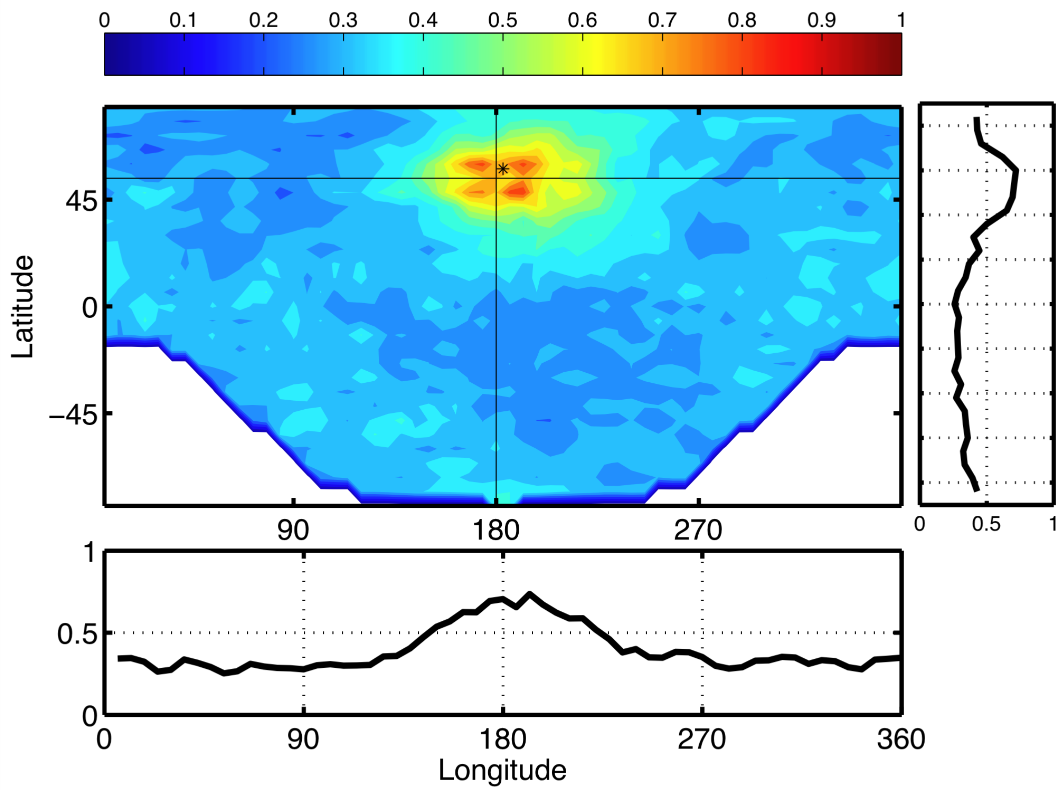
Time mean sampling localization for equatorial Ps ob. on Ps state variables. (Ens. Size 80, GC localization 1.2).
Nearly gaussian locally.

Smaller Ensemble => Tighter Localization



Ensemble size 20 localization of Ps ob. on Ps state.
Localization is tighter for smaller ensemble.

Time Mean 'Localization' for V Observation

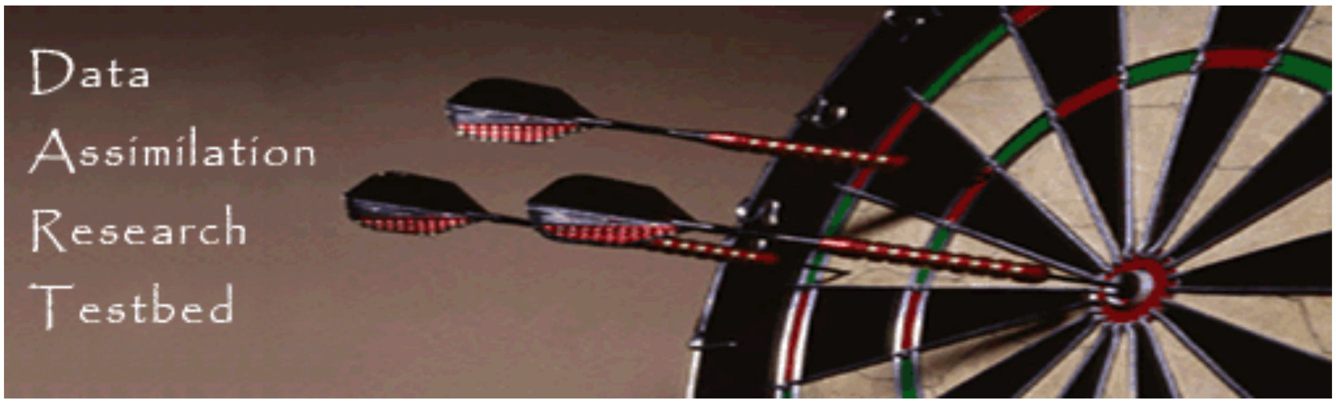


Localization for mid-level V ob. on U state variables (N=80).
Local four maxima.

Conclusions

- Implemented sampling error correction localization.
- Localization is function of sample correlation & ensemble size.
- Helps when localization is not known a priori.
- Could build statistical models of time-mean localization.

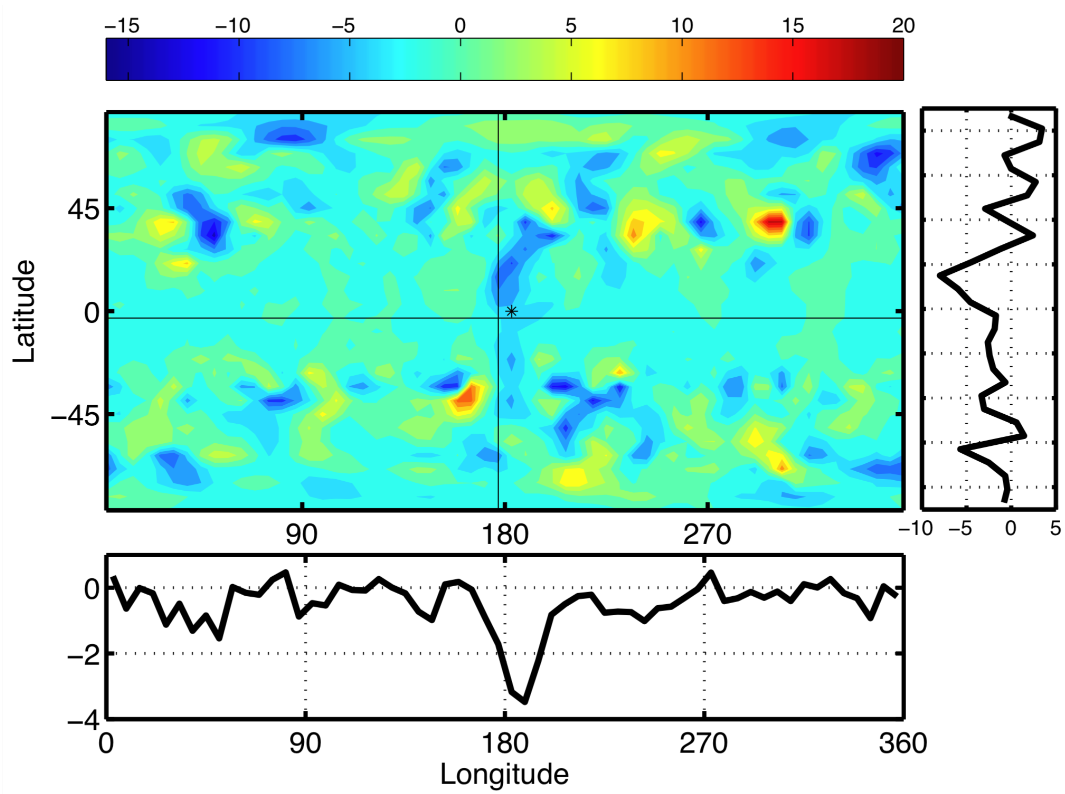
All the tools used to generate this talk are part of DART:



http://www.image.ucar.edu/DAReS/DART

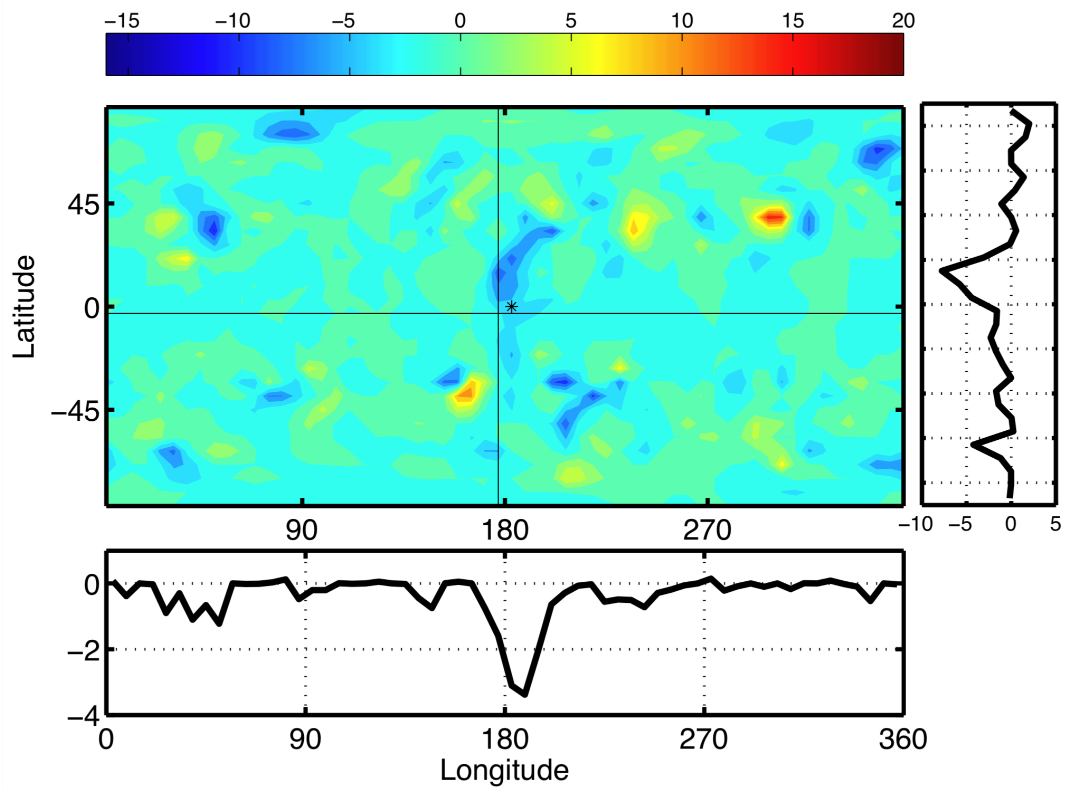
Thanks to Nancy Collins, Tim Hoar,
Kevin Raeder and Glen Romine

Raw Increments from Single PS Obs.



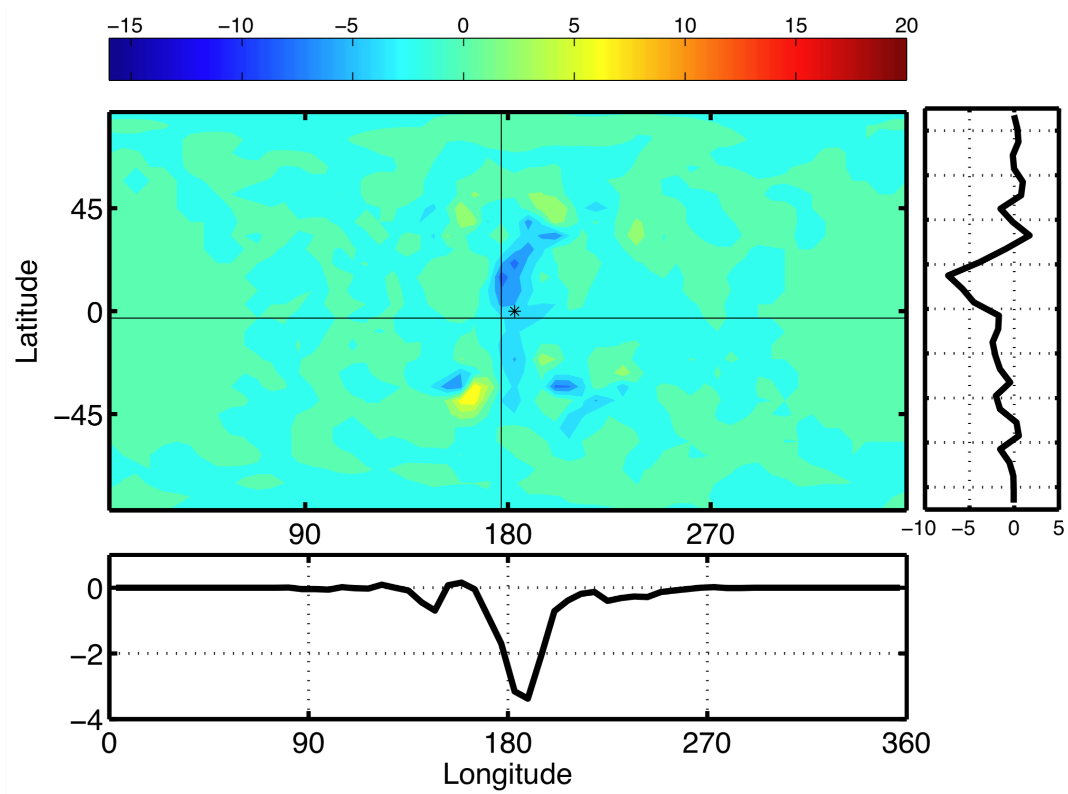
Increments from single unlocalized Ps observation on Ps.
Ensemble size 80, observation at equator, 180.

Increments after Sampling Error Correction



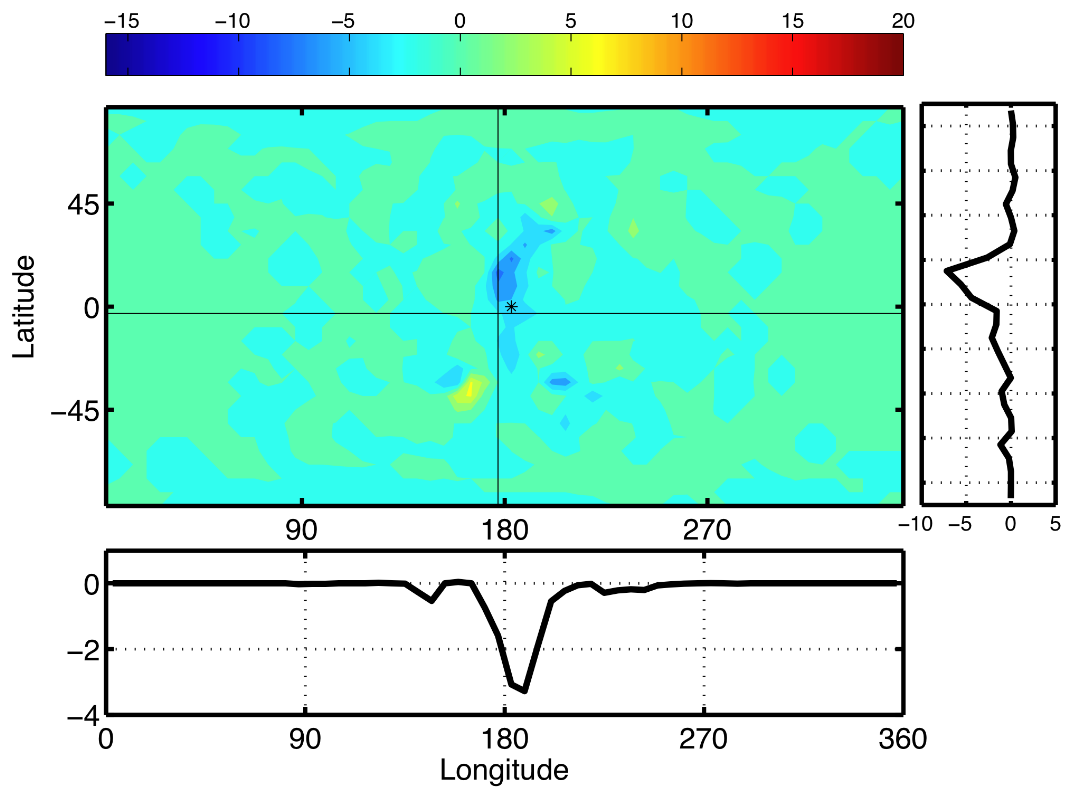
Increments with sampling error localization.
Ensemble size 80, observation at equator, 180.

Increments with Gaspari-Cohn, 1.2 Radians



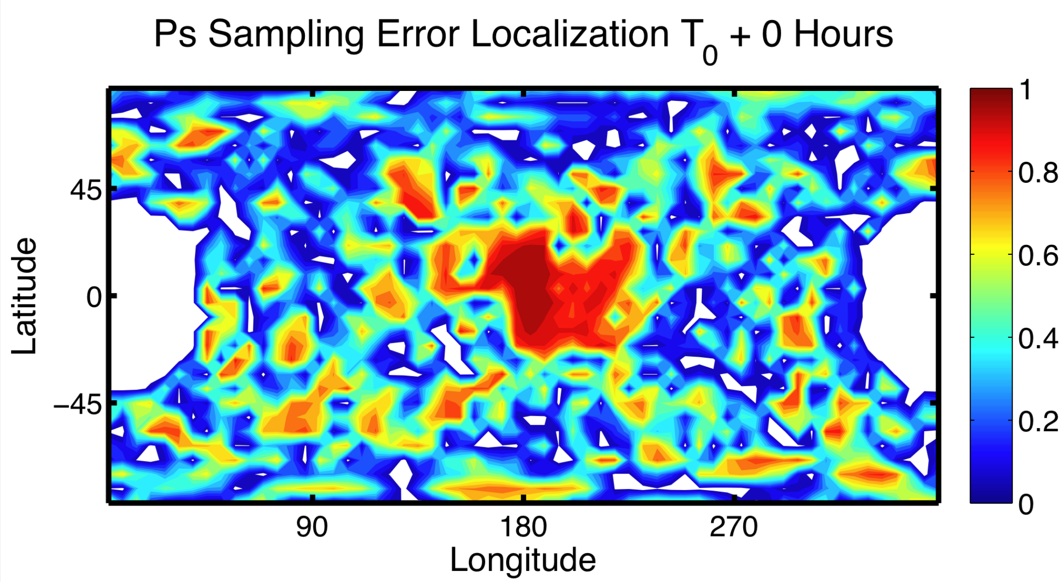
Increments GC=1.2 radians.
Ensemble size 80, observation at equator, 180.

Increments with SEC and GC



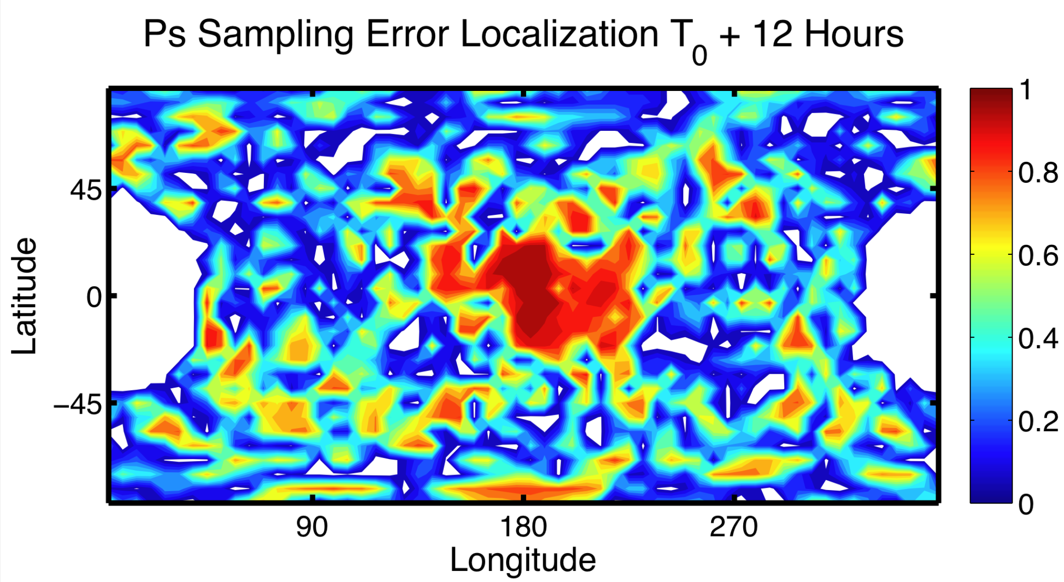
Increments GC=1.2 radians plus sampling error correction.
Ensemble size 80, observation at equator, 180.

Time Variation of Single PS Obs Increments



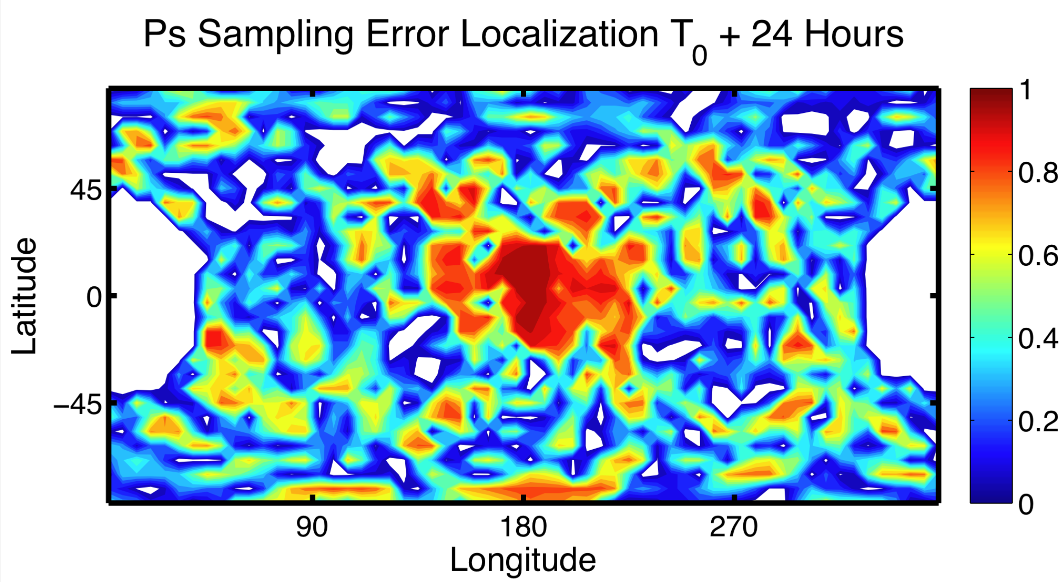
Time variation of sampling error localization on Ps field.
Ensemble size 80, Gaspari Cohn=1.2. (Full obs. set).

Time Variation of Single PS Obs Increments



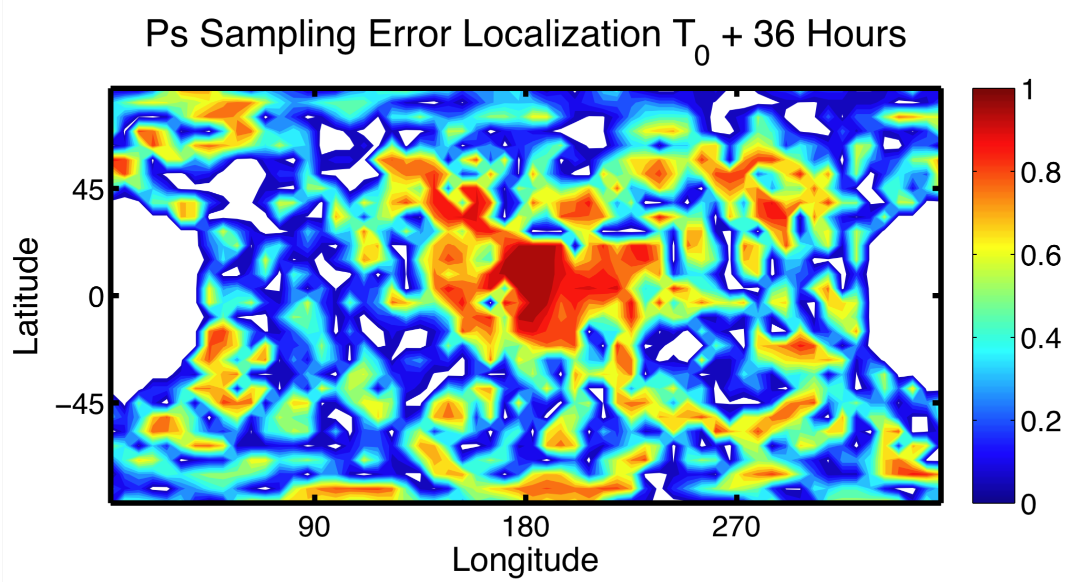
Time variation of sampling error localization on Ps field.
Ensemble size 80, Gaspari Cohn=1.2. (Full obs. set).

Time Variation of Single PS Obs Increments



Time variation of sampling error localization on Ps field. Ensemble size 80, Gaspari Cohn=1.2. (Full obs. set).

Time Variation of Single PS Obs Increments



Time variation of sampling error localization on Ps field.
Ensemble size 80, Gaspari Cohn=1.2. (Full obs. set).