Ensemble Filters for Data Assimilation: Flexible, Powerful, and Ready for Prime-Time?

Jeffrey L. Anderson
NCAR Data Assimilation Initiative
29 June, 2004

Ensemble filters are:

1. Easy to apply to complicated models and observations

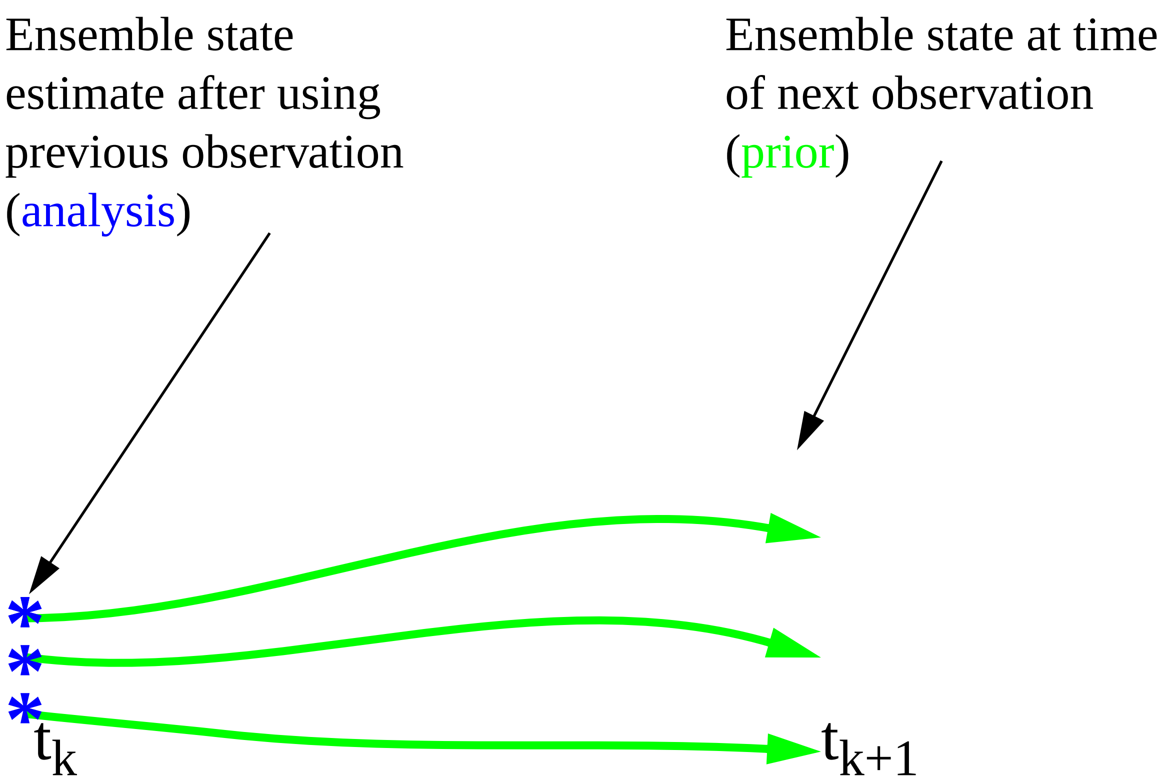
2. Computationally competitive with variational methods

3. Able to extract information about all state variables using multivariate relations

4. Can be augmented to deal with model bias and nasty real observations

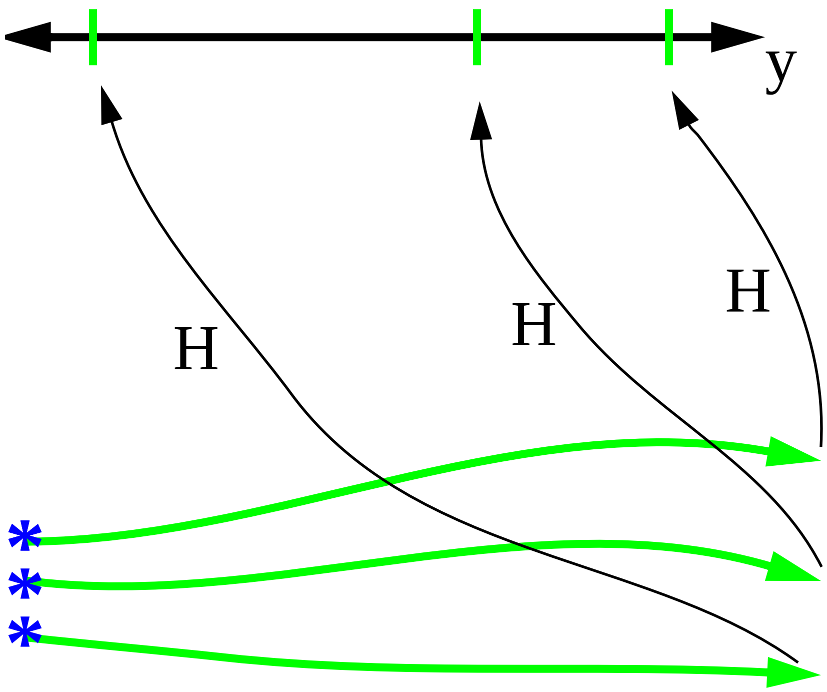
How an Ensemble Filter Works

1. Use model to advance ensemble (3 members here) to time at which next observation becomes available



How an Ensemble Filter Works

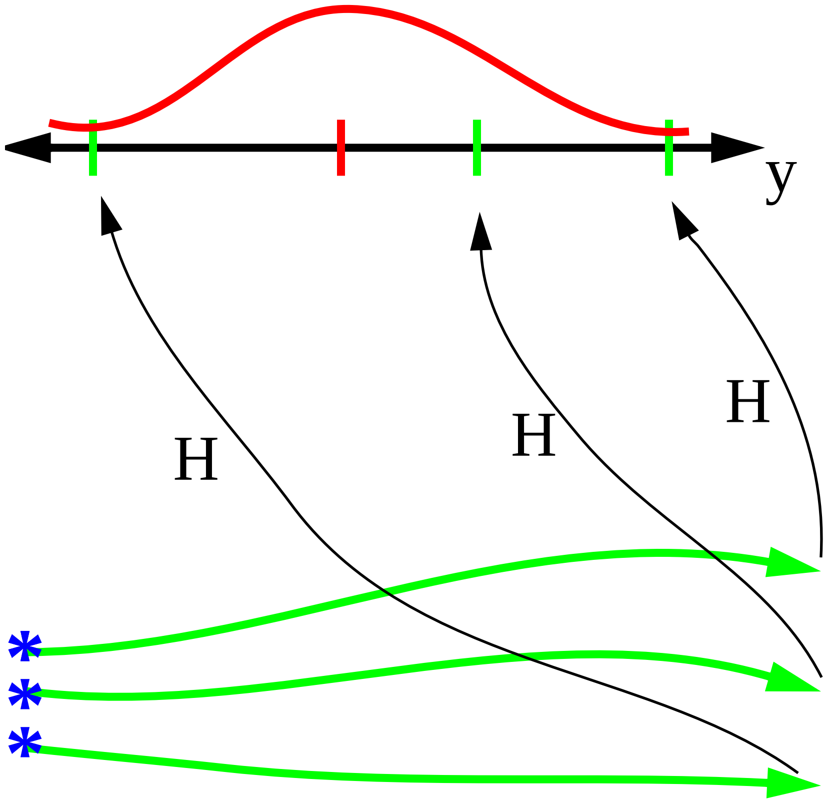
2. Get prior ensemble sample of observation, y=H(x), by applying forward operator H to each ensemble member



Theory: observations from instruments with uncorrelated errors can be done sequentially.

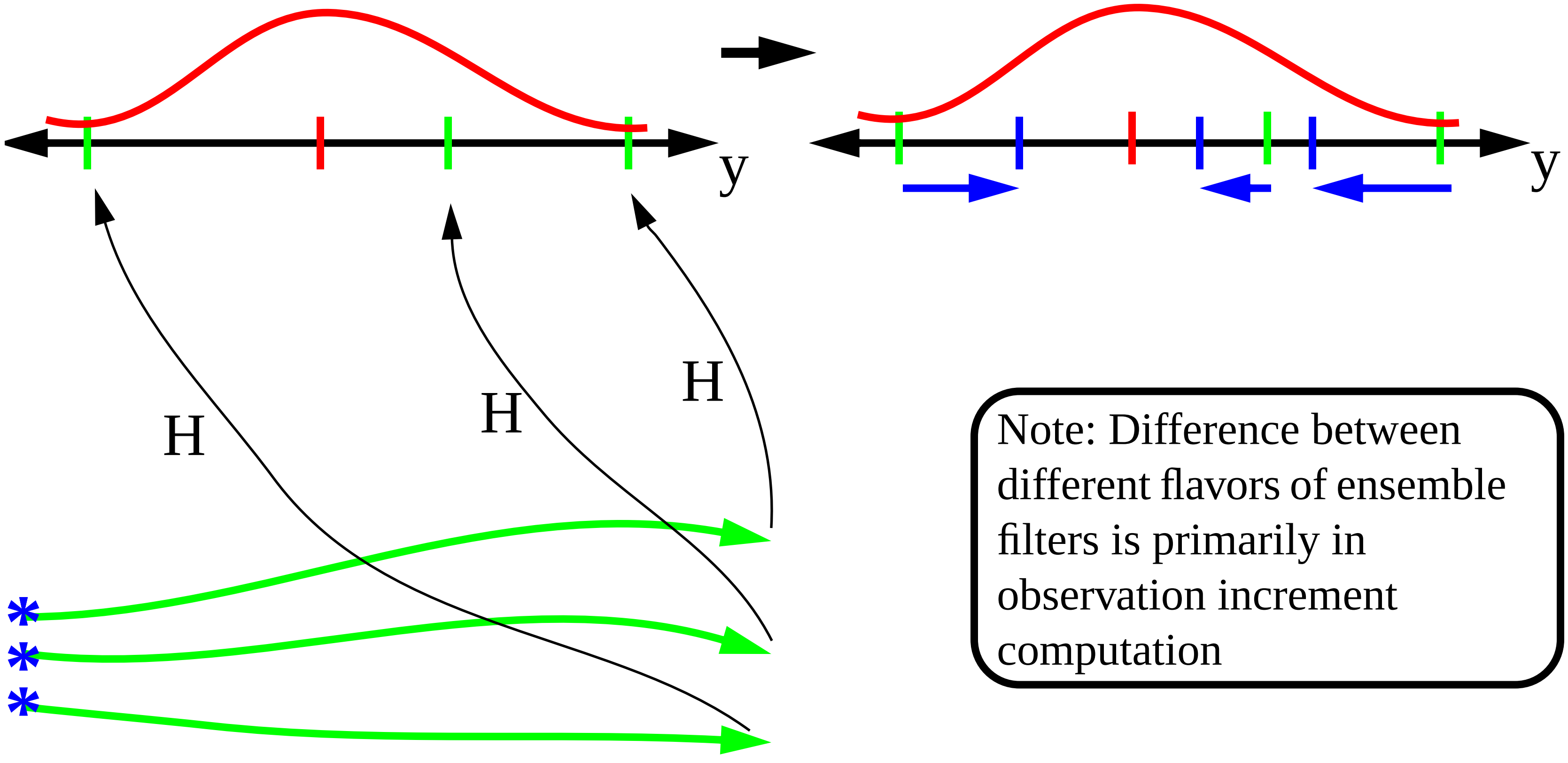
How an Ensemble Filter Works

3. Get observed value and observational error distribution from observing system



How an Ensemble Filter Works

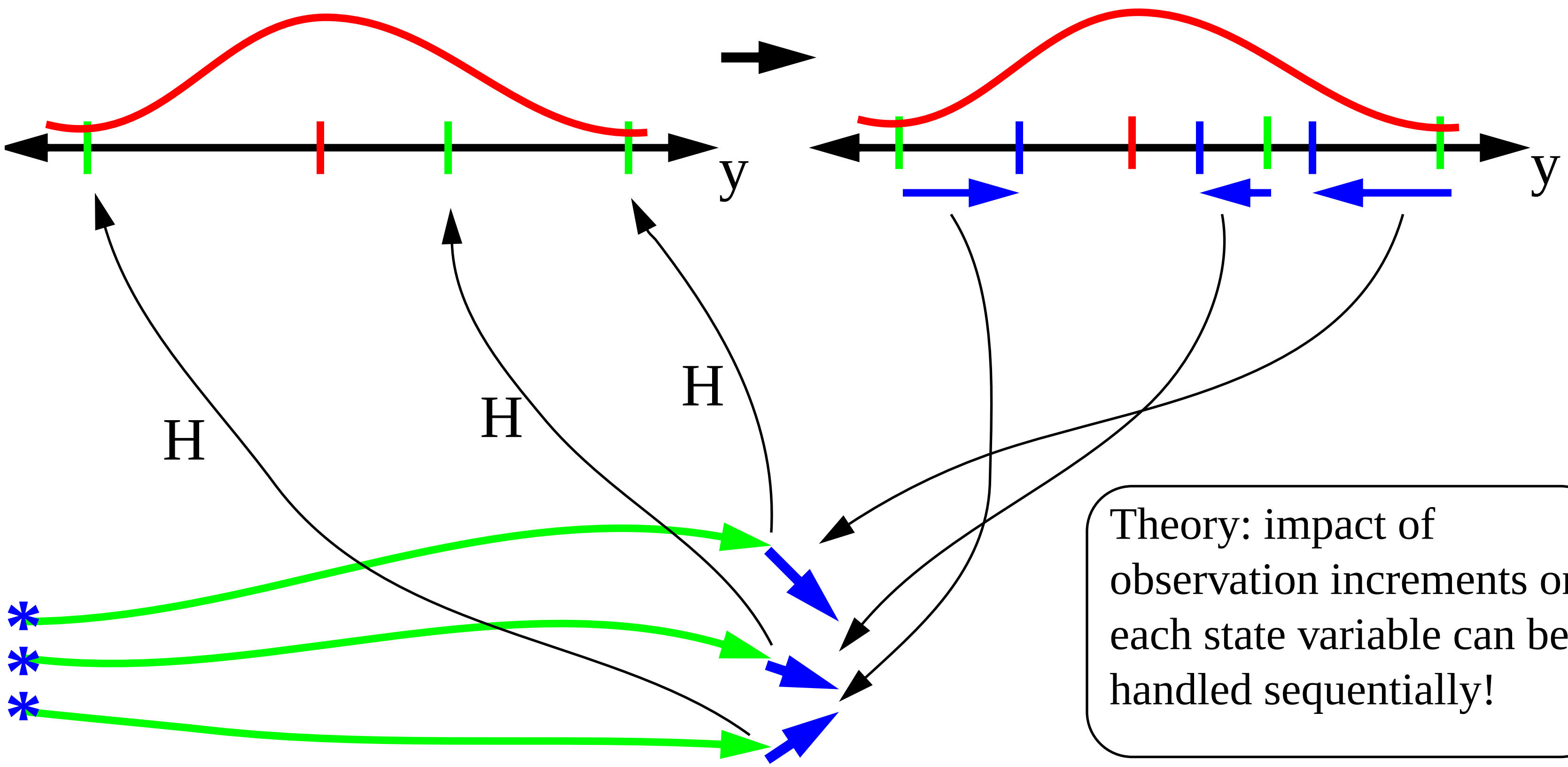
4. Find increment for each prior observation ensemble
(this is a scalar problem for uncorrelated observation errors)



Note: Difference between different flavors of ensemble filters is primarily in observation increment computation

How an Ensemble Filter Works

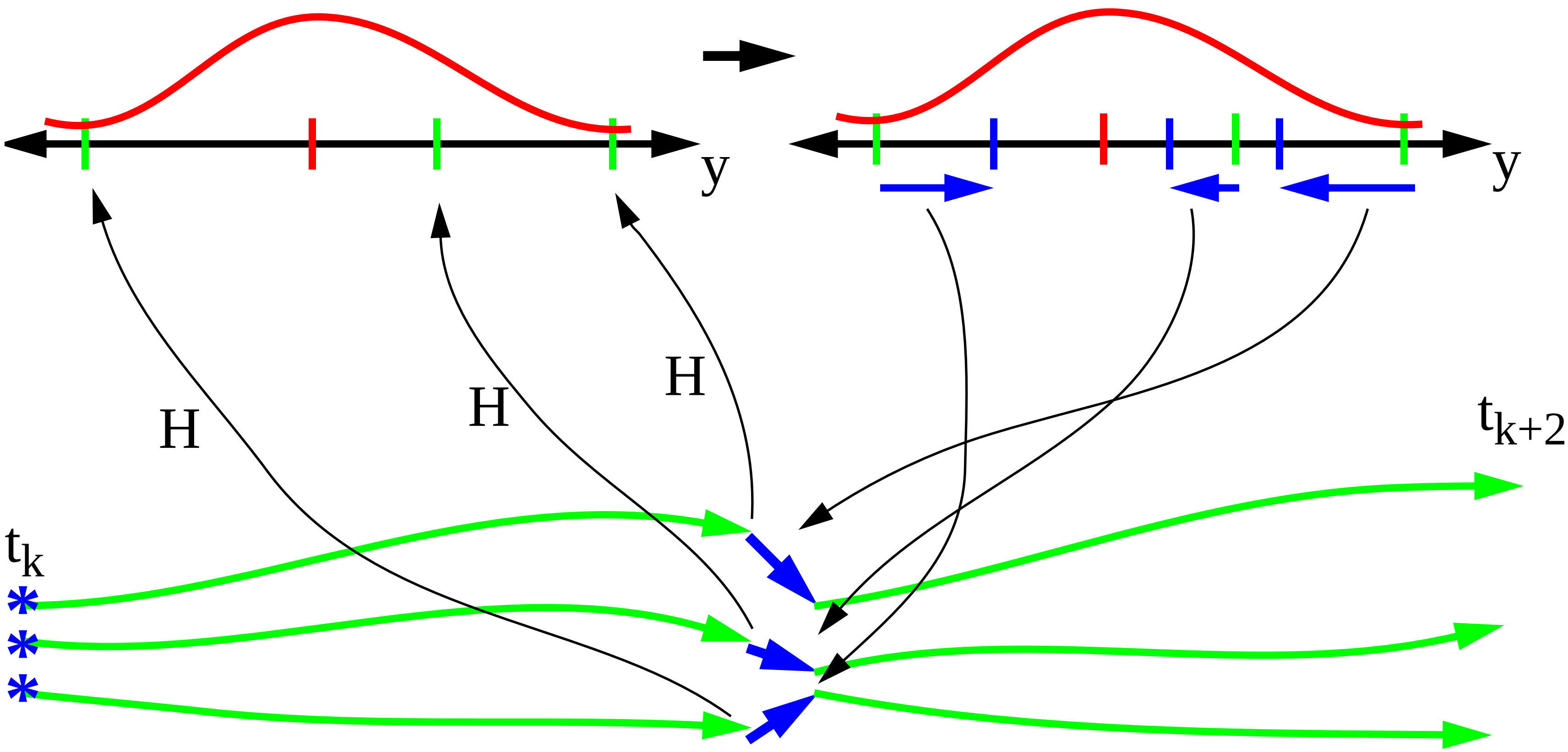
5. Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments



Theory: impact of observation increments on each state variable can be handled sequentially!

How an Ensemble Filter Works

6. When all ensemble members for each state variable are updated, have a new analysis. Integrate to time of next observation...



1. Filters are Easy to Apply

All one needs is:

A. Ability to integrate a model forward in time

B. Ability to compute forward observation operators

Collaborators have put GCMs in our framework in < 1 person month

2. Filters are computationally competitive

A. Model integrations required

Filter requires N forward integrations of model; O(10) sufficient?

4D-var requires K*L forward and backward integrations

K - number of observation intervals over which optimization is performed

L - average number of iterations of minimization solver

K*L at least O(10) for any envisioned application

B. Assimilation algorithm cost

Filter: O(αNnm): N is ensemble size, n is model size, m is number of obs

α related to what fraction of state variables are impacted by given ob.

In certain scenarios this may reduce order of cost

4D-var: O(nm) in best of all possible cases

Relation of constant factors not clear, depends on ensemble size

Data Assimilation Research Testbed (DART)

Basic framework implemented
Primarily implementing ensemble (Kalman) filters
Variational for low-order models only
Plans MAY include a variational (4D-Var) capability

DART compliant models (largest collection ever with assim system)

CGD’s CAM 2.0
GFDL FMS B-grid GCM
Many low-order models available
MMM’s WRF model
NCEP MRF (GFS)
GFDL MOM ocean model partially incorporated in earlier version

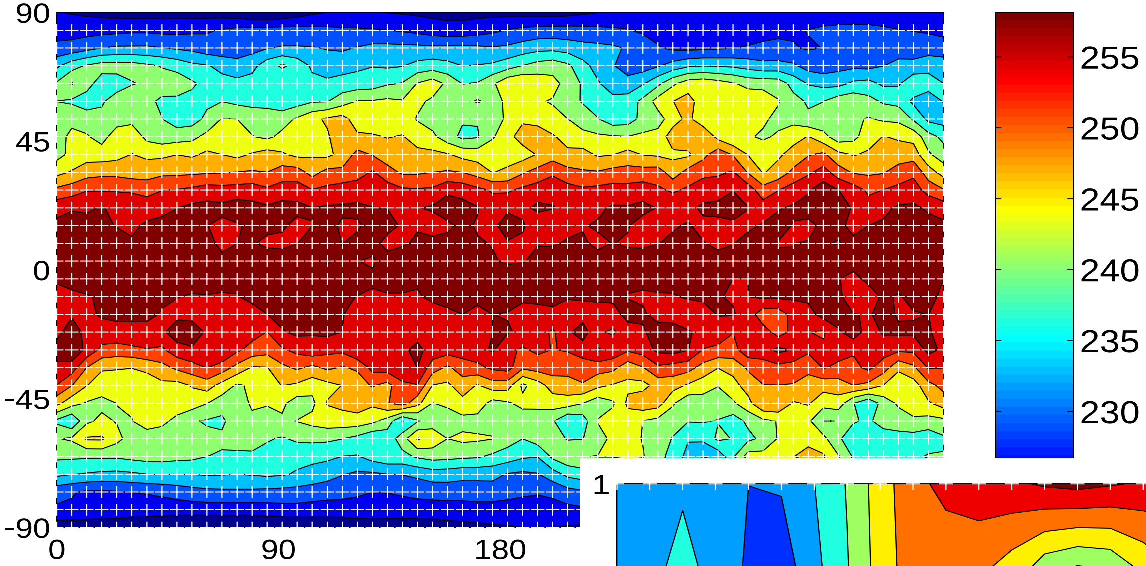
3. Extract amazing amounts of information in perfect model cases

WARNING: View Perfect Model Results with Skepticism

Predictability in an Idealized AGCM: GFDL FMS B-Grid Dynamical Core (Havana)

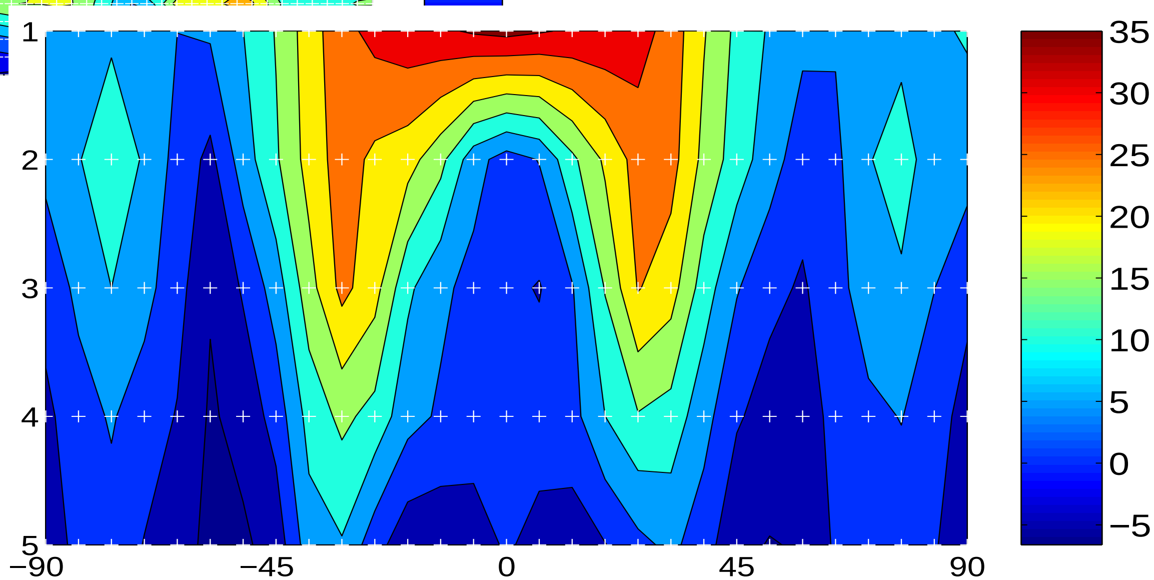
Held-Suarez Configuration (no zonal variation, fixed forcing)

Low-Resolution (60 lons, 30 lats, 5 levels); Timestep 1 hour (less for frequent observations)



T at Level 3
(Middle of Atmosphere)

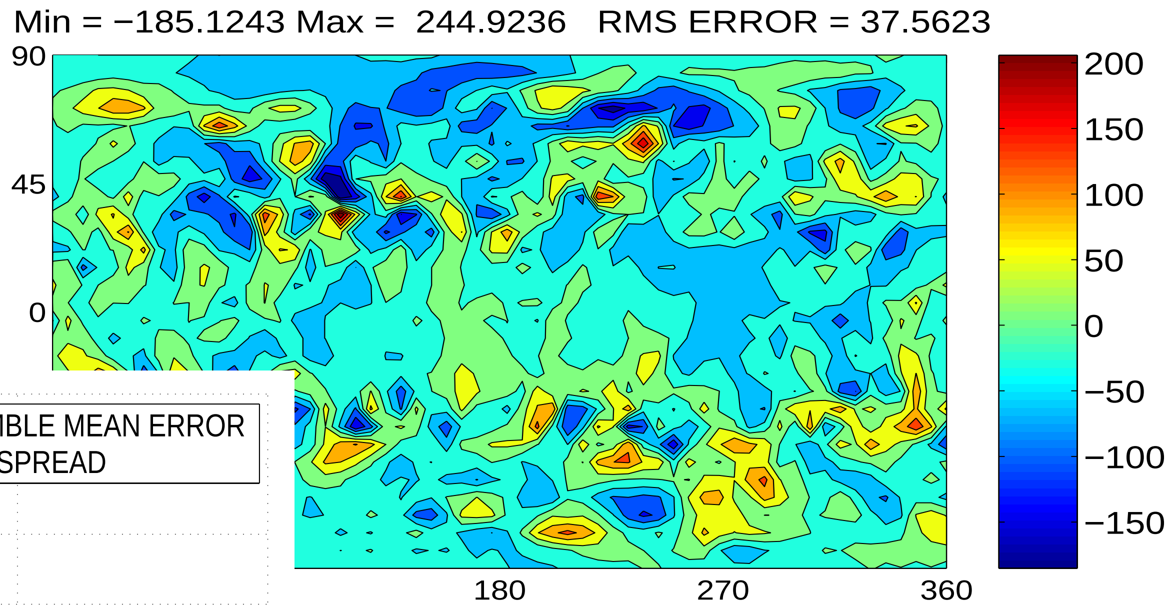
Cross Section of
Zonal Mean U

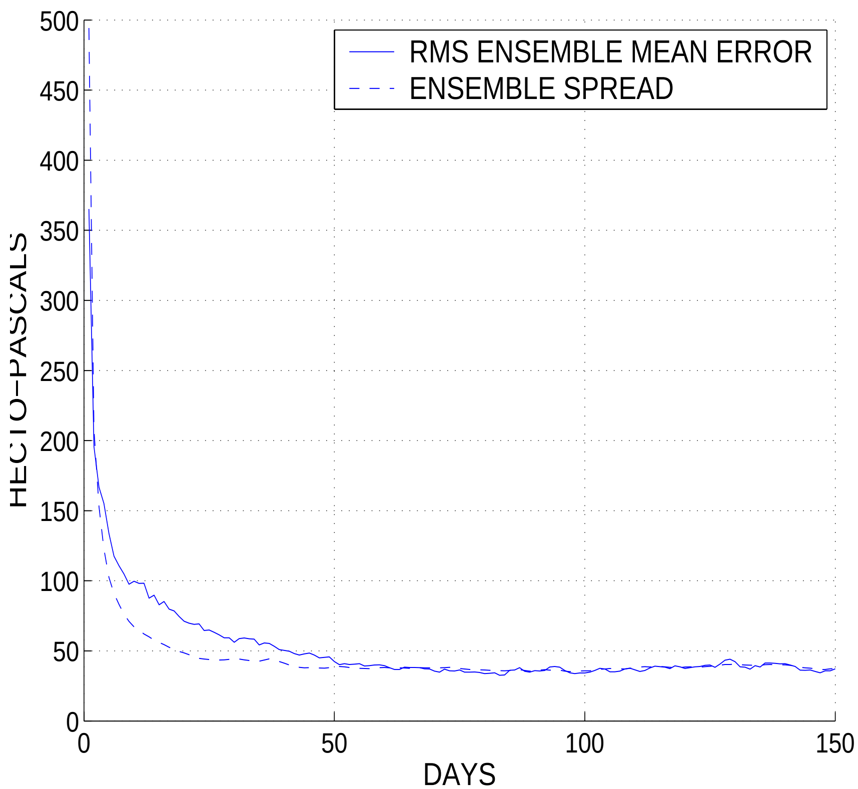


Has Baroclinic Instability

Baseline Case: 1800 PS Obs every 24 hours

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





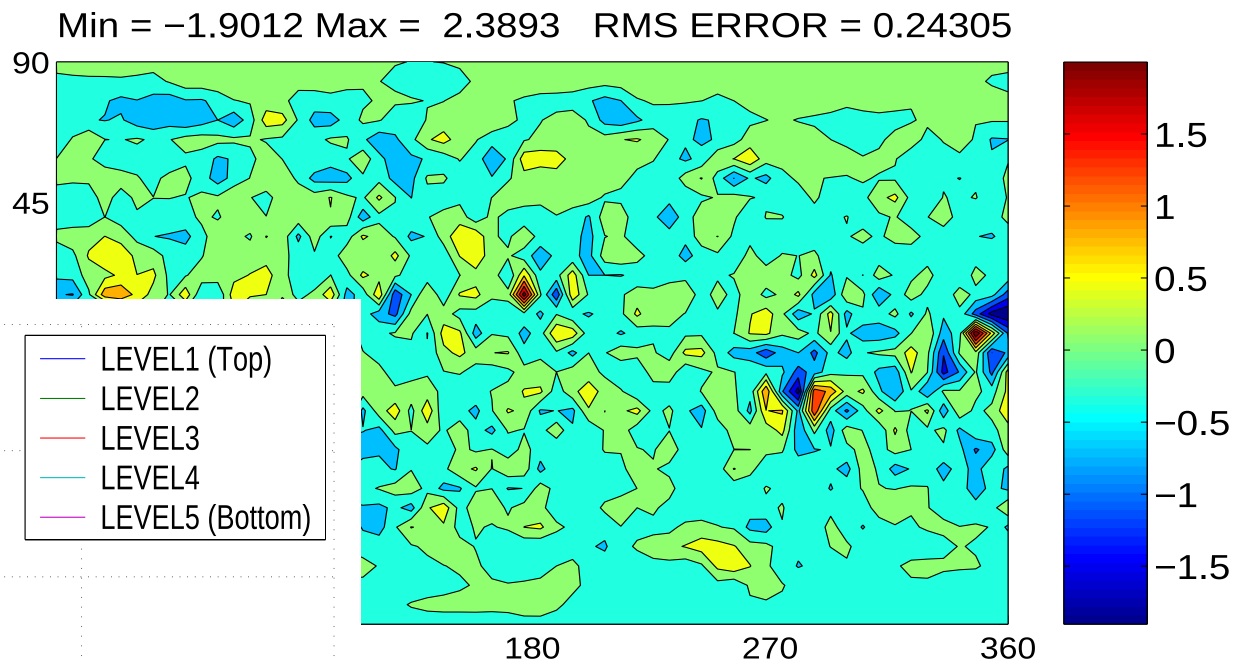
PS Error Reduces by about factor of 10

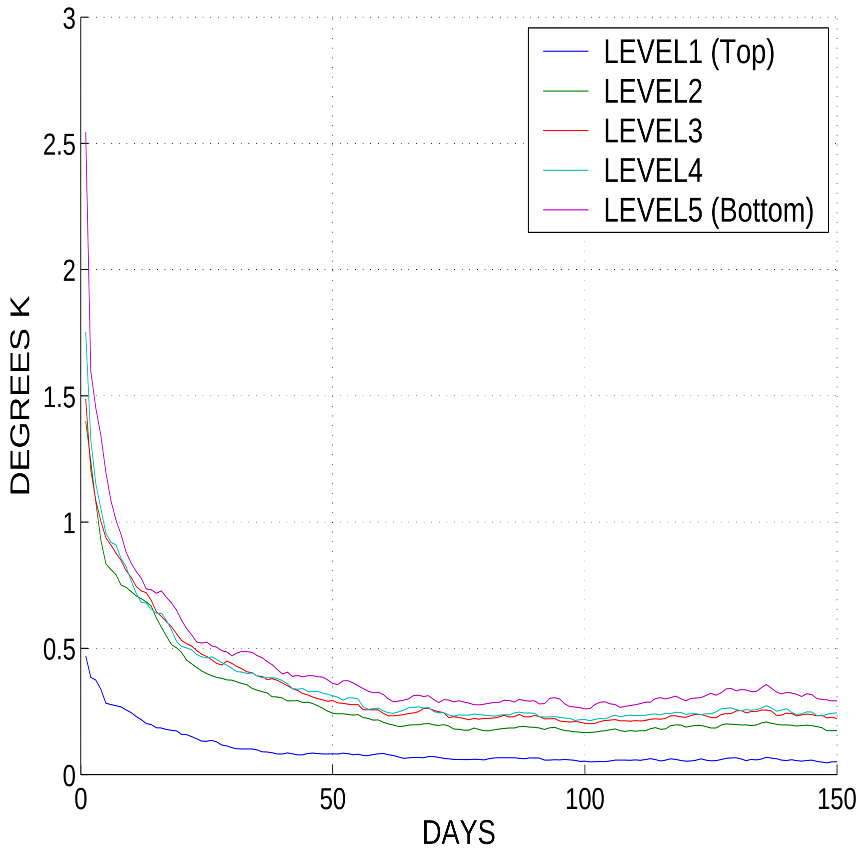
Asymptotes after about 50 days

Ensemble spread is approximately correlated with RMS error

Baseline Case: 1800 PS Obs every 24 hours

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





T Error also reduced by about factor of 10

Asymptotes about day 70

Final error about 0.25 K for interior levels

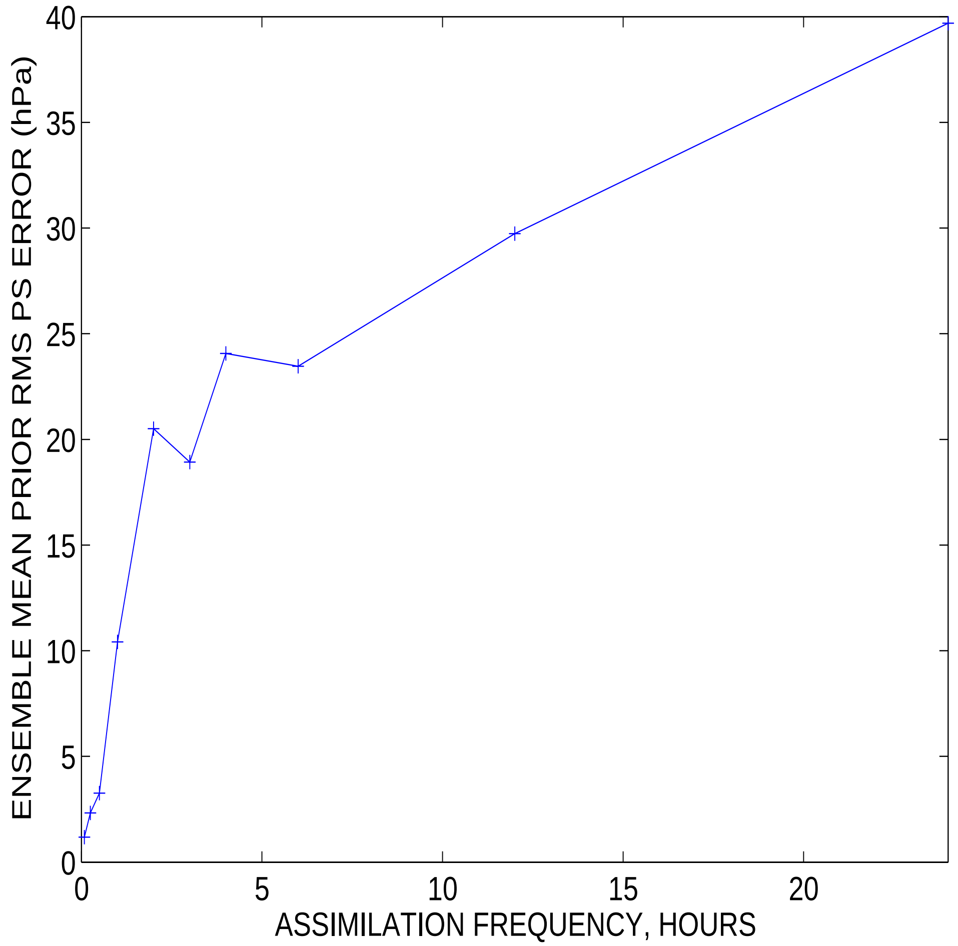
Impacts of frequency of PS observations on PS assimilation error

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

Steady RMS decrease as frequency increases

Much smaller RMS than for high density low frequency obs

RMS < 0.02mb for 5 minutes



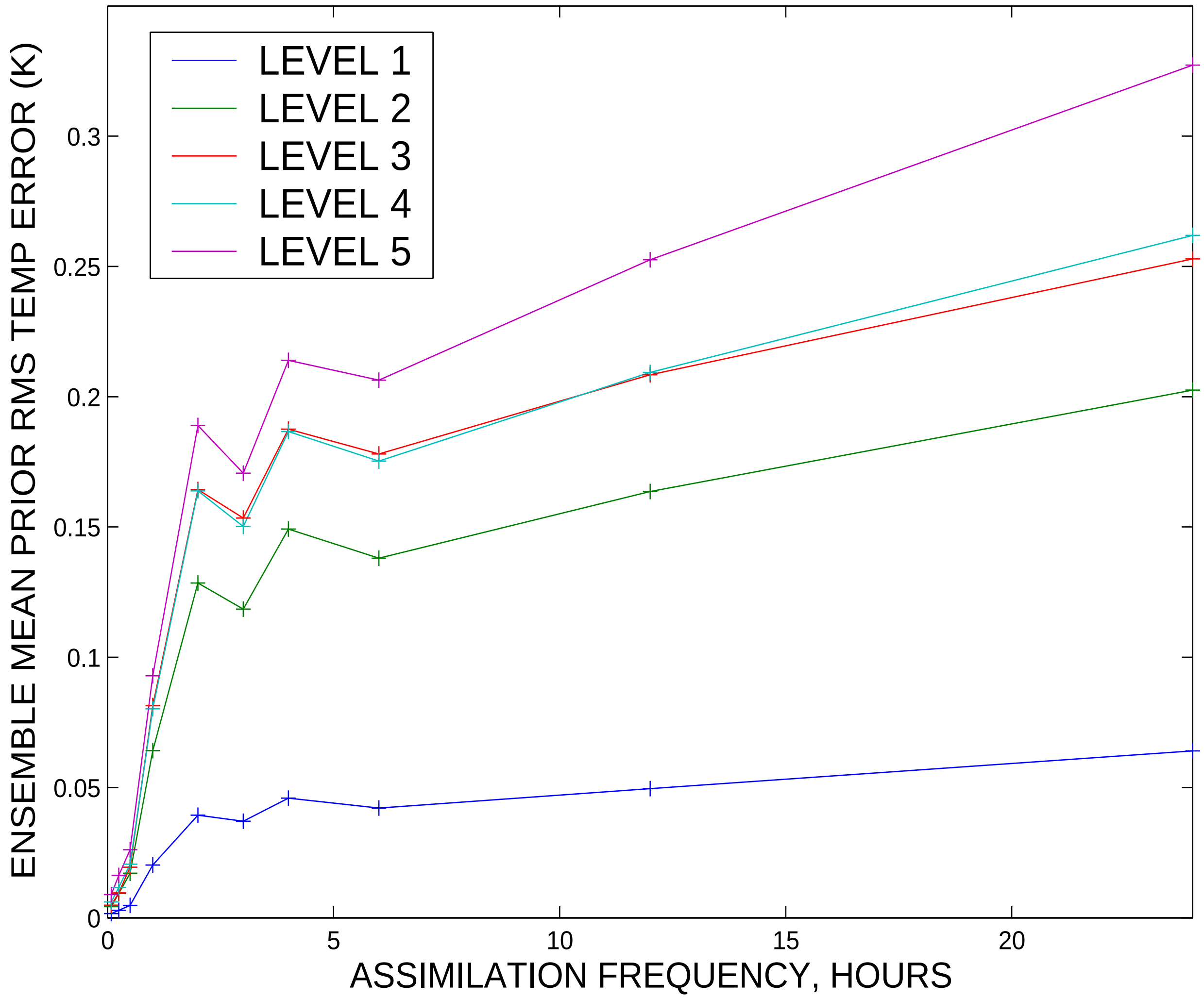
Impacts of frequency of PS observations on T assimilation error

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?



4. Can handle real (biased) model and real (ugly) observations

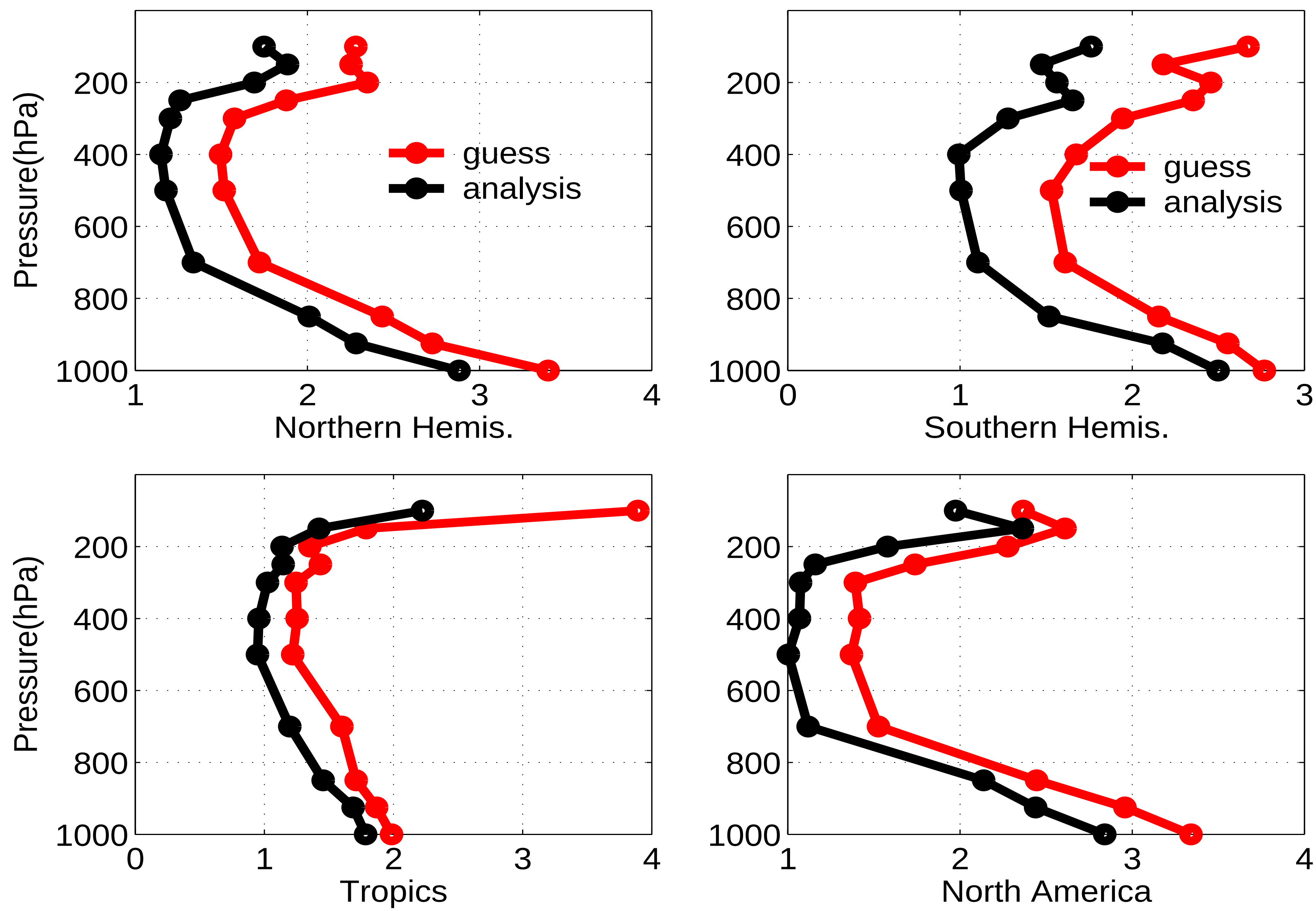
Model:

- CAM 2.0 T42L26
- U,V, T, Q and PS state variables impacted by observations
- Land model (CLM 2.0) not impacted by observations
- Observed SSTs

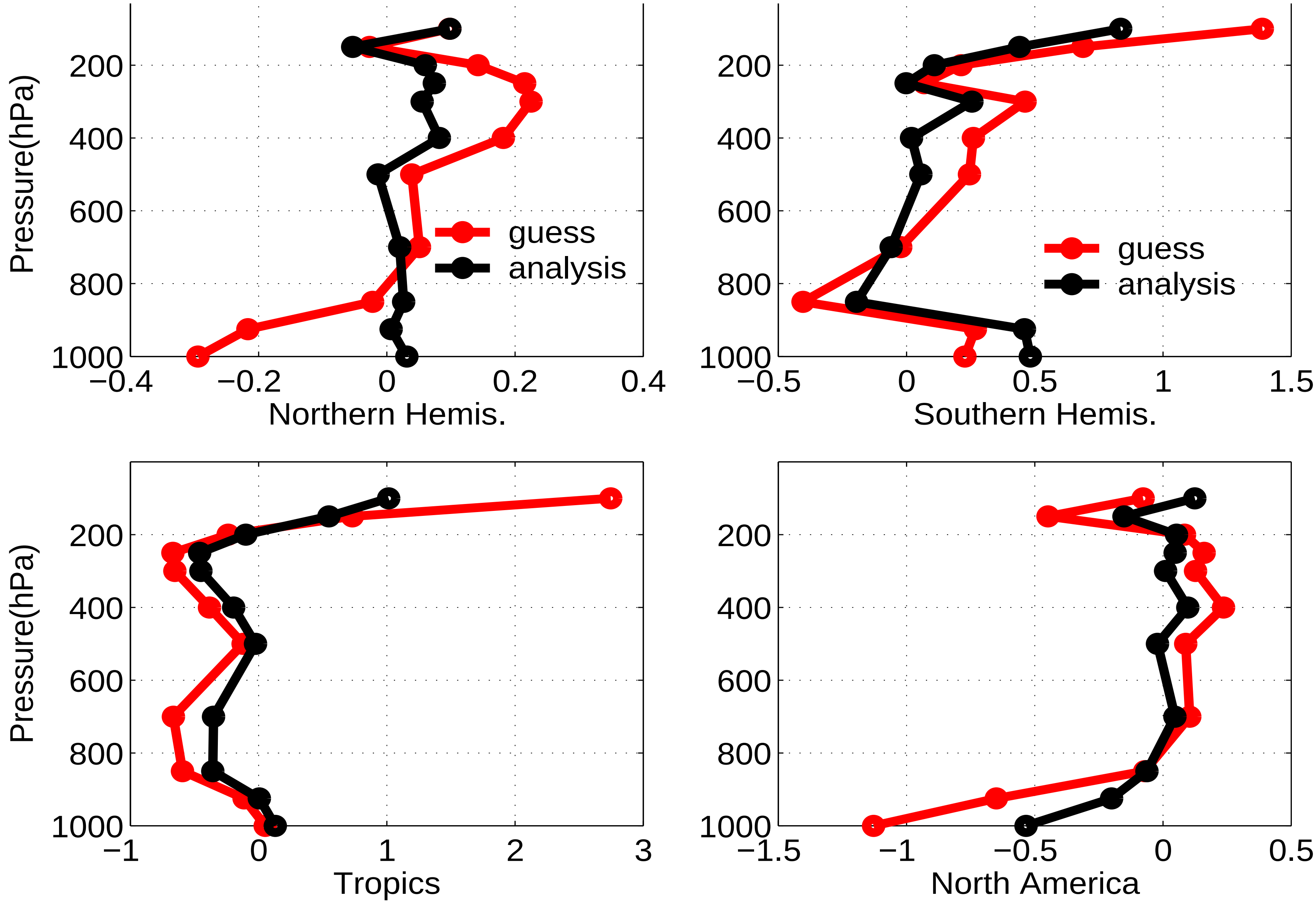
Assimilation / Prediction Experiments:

- Uses observations used in reanalysis
  - (Radiosondes, ACARS, Satellite Winds...)
- Initial tests for first week of January, 2003
- Assimilated every 6 hours
- Run on CGD linux cluster Anchorage

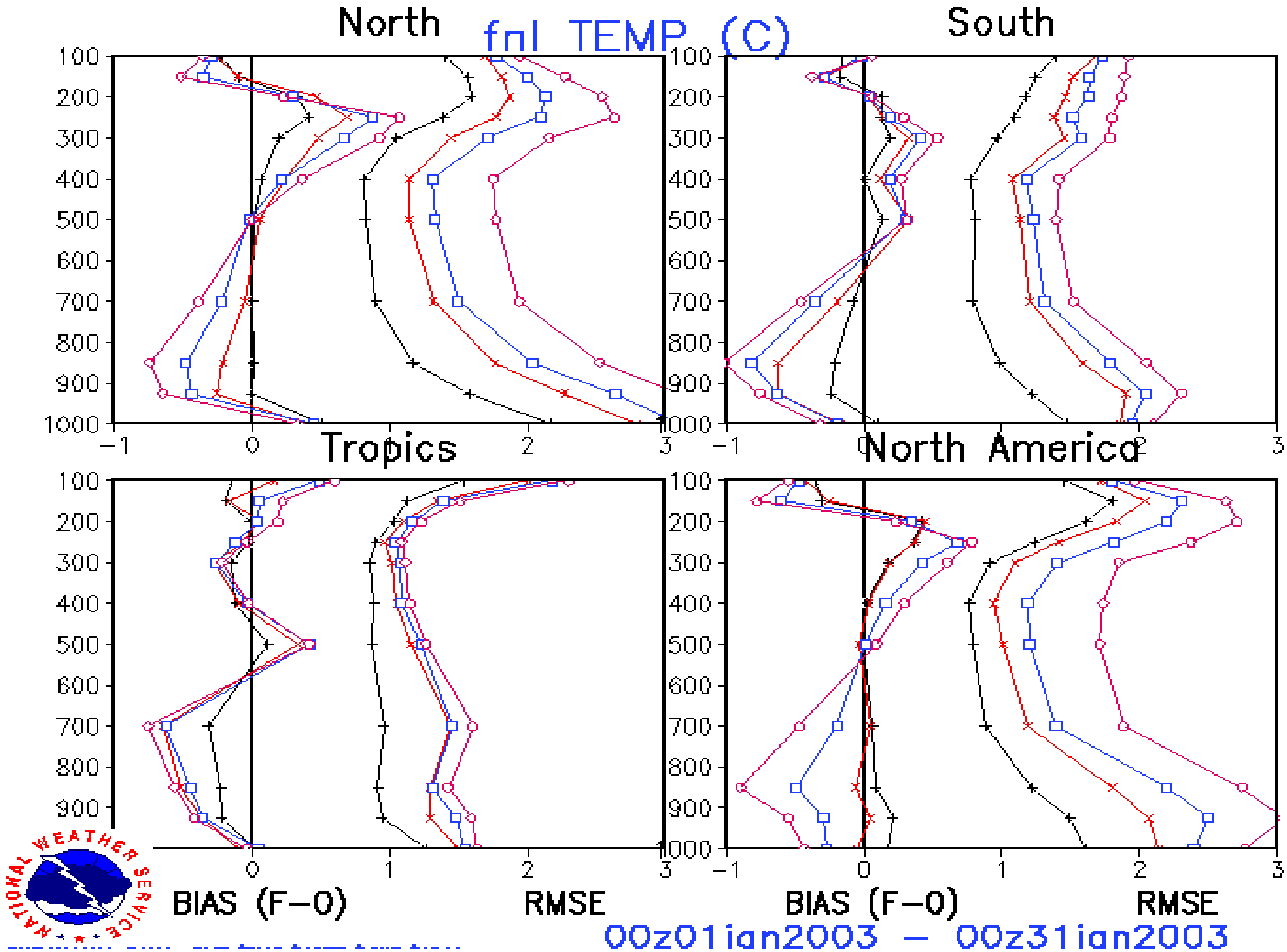
CAM RESULTS: ENSEMBLE MEAN RMS TEMP. ERROR



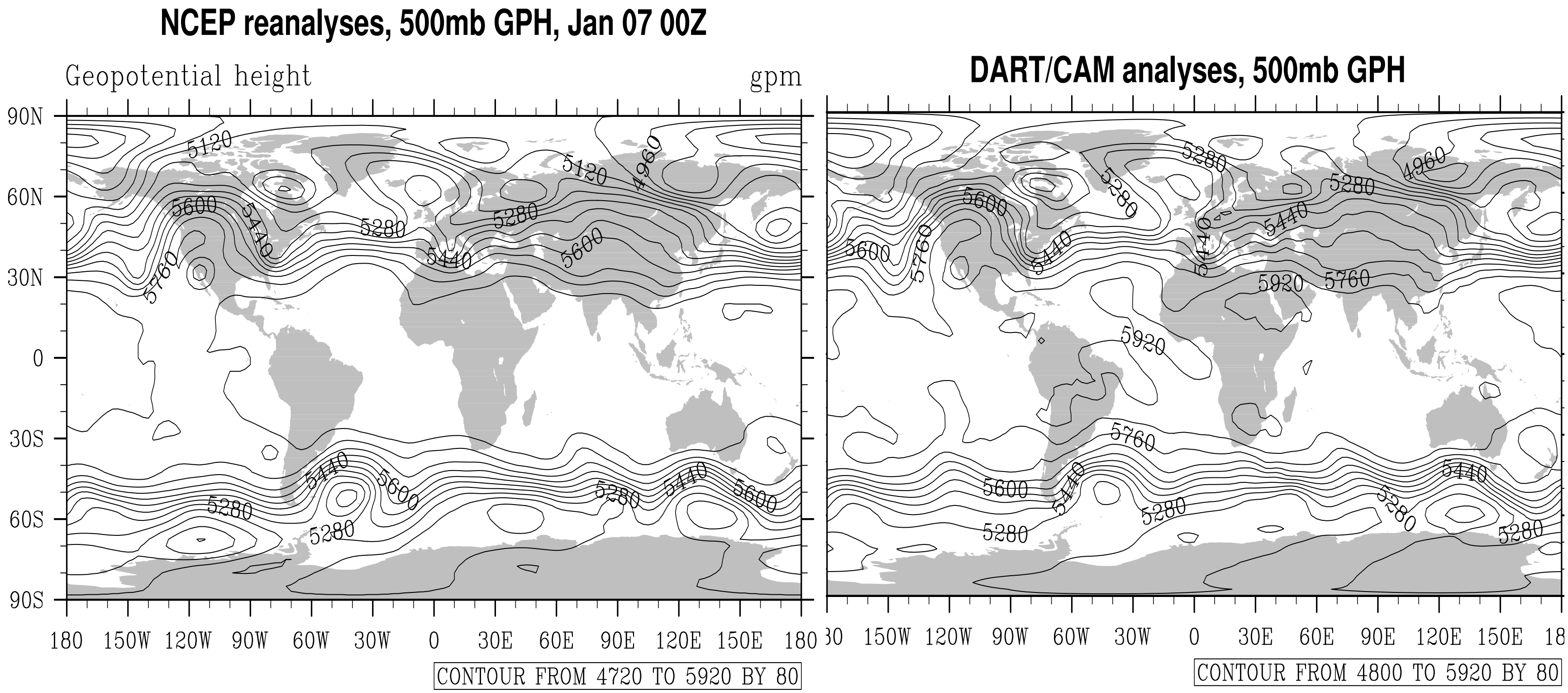
CAM RESULTS: Ensemble Mean Time Mean Temperature BIAS



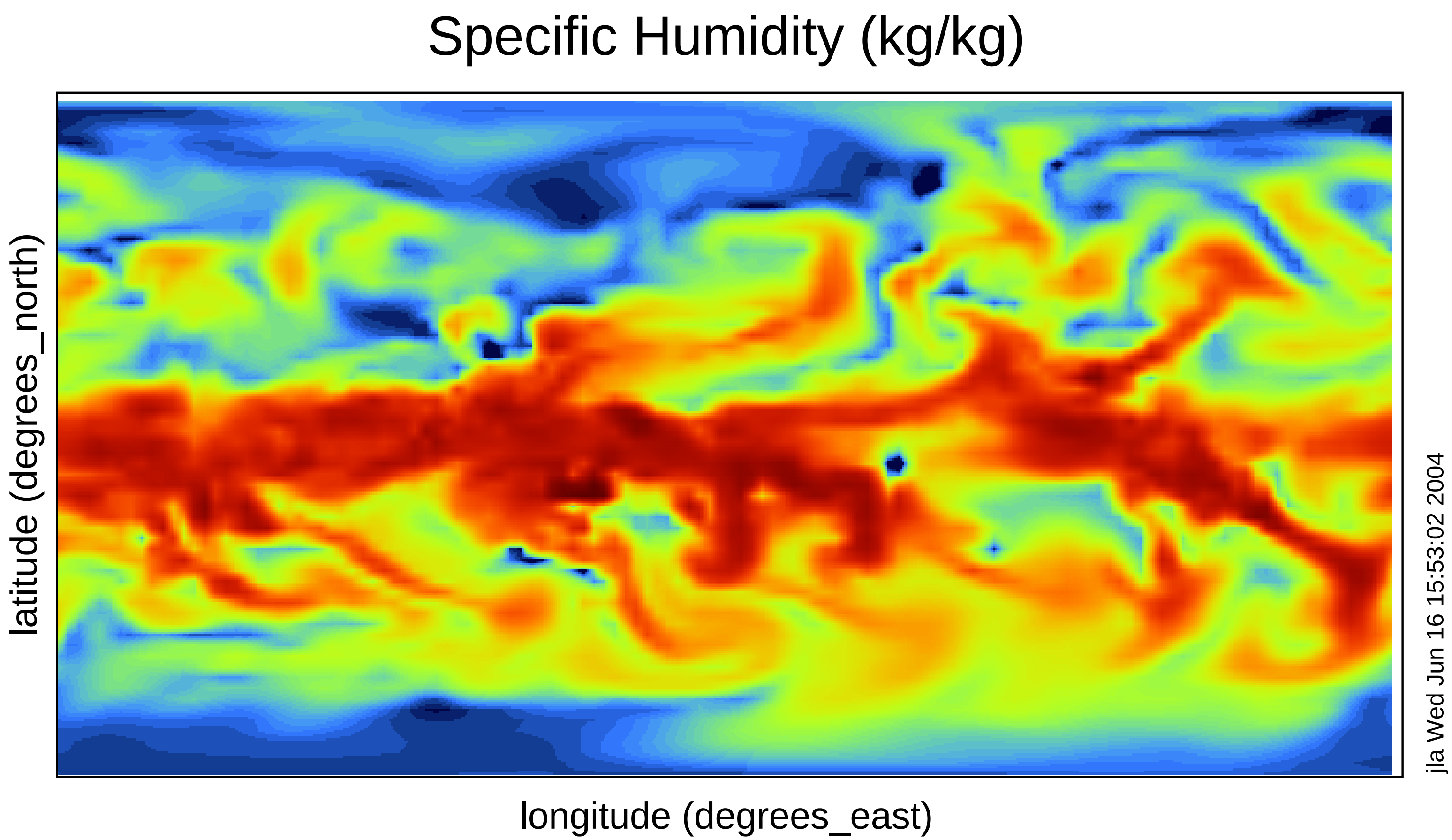
NCEP GFS BIAS (Left), RMS (right): Black Analysis, Red Guess



500mb Height Comparison to NCEP CDAS Analysis; Jan. 7, 2003



Captures details of q without q obs; q increments from other obs!



Conclusions

1. Ensemble filters can do complex, real-data assimilation problems

2. Implementing filters is extremely simple
   (compared to most assimilation techniques)

3. Filters are powerful in extracting multi-variate relations

4. Filters can deal with tracers, observed or unobserved

5. Assimilation is relatively cheap, but ensembles are required