Implementation of the LETKF on the NCEP Global Forecast System

Istvan Szunyogh

University of Maryland

ECMWF Workshop on Flow-dependent aspects of data assimilation, Reading, UK, 11-13 June, 2007

Outline

- Brief history of the development of the Local Ensemble Transform Kalman Filter (LETKF)
- Comparison of the LETKF and SSI at a reduced, T62L28, model resolution for a reduced data set (no radiances)
- Some comments on the flow dependent representation of errors and predictability
- Preliminary results with radiances

Ensemble-based Kalman Filter

data assimilation schemes

- The background and the background error covariance matrix are calculated based on a *k*-member ensemble of backgrounds:
  - $\boldsymbol{x_b}$ is calculated by taking ensemble mean
  - ***B*** is calculated by calculating the sample covariance matrix based on the ensemble
- The data assimilation provides an ensemble of initial conditions with the correct statistics
  - The ensemble mean is equal to $\boldsymbol{x_a}$
  - The sample covariance matrix for the ensemble is ***A***

UMCP Weather & Chaos Project
http://weatherchaos.umd.edu

- Started in 2000 by J. Yorke and E. Kalnay with the aim
- To develop an ensemble based data assimilation system for spatio-temporally chaotic systems
- To study predictability in high-dimensional spatiotemporally chaotic systems,
  - the atmosphere is the most intensely studied example
  - numerical weather prediction models provide the most sophisticated and accurate solution of the partial differential equations that govern the evolution of the atmosphere

Practical Motivations for the design of the LETKF approach were the general concerns:

- An estimate of the background error covariance matrix based on a reasonably small ensemble would be hopelessly rank-deficient
- An ensemble-based Kalman filter would be computationally hopelessly expensive
- It seemed to be a good idea to solve the Kalman filter equation
  - locally in physical space,
  - concurrently for the different locations
- (Some scientists also argued that model errors were too large for an indefinitely long cycling of an ensemble base Kalman filter, but we hoped that this was not true)

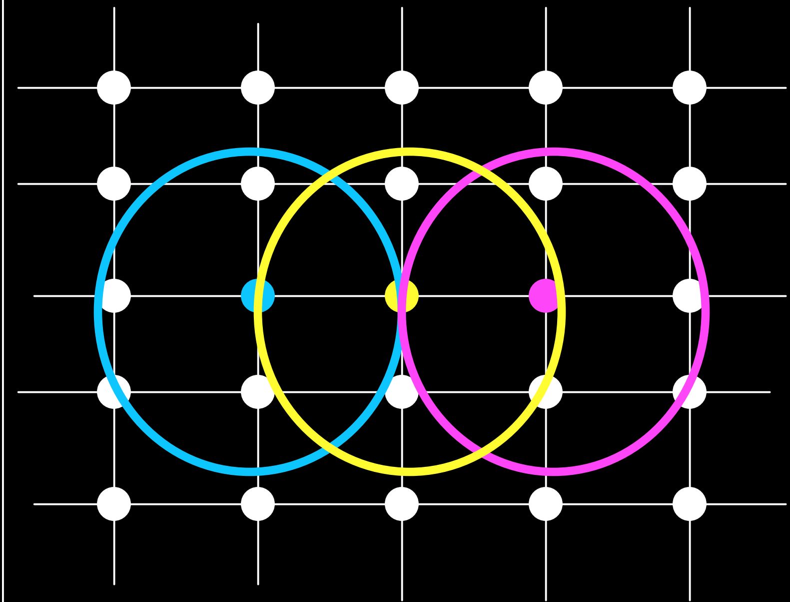
Theoretical Motivation

The unusual aspect of the UMCP approach:

- Dynamical systems theory is usually applied to the global atmosphere by searching for a low-dimensional system behind the complex behavior of the global atmospheric dynamics
- Dynamical systems theory is applied locally in space and time (calculations are done for local neighborhoods of each location at each time), searching for low-dimensional behavior produced by the complex equations of atmospheric dynamics

Illustration of the Local Approach

for a 2D model grid



- A local region is associated with each grid point
- Properties assigned to a grid point are calculated using information from the associated local region
- For instance, the analysis for a given grid point is calculated using $\boldsymbol{x}_a$, $\boldsymbol{x}_b$, $\boldsymbol{K}$, $\boldsymbol{y}$, $\boldsymbol{B}$, and $\boldsymbol{R}$ defined for the local region

LETKF

- Papers (all available at http://weatherchaos.umd.edu)
  - Ott et al. 2004, *Tellus*; Ott et al. 2004, *Phys Let. A:* basic idea, conditions for local approach to provide a smooth global analysis
  - Hunt et al., 2007, *Physica D:* optimized algorithm
- Basic Algorithm
  1. Background ensemble is generated by integration of the model
  2. The observation operator, **H**, is applied to all ensemble members (the forecast and **H** is calculated using the same set of processors for each ensemble member)
  3. Information needed to obtain the analysis at the different grid points is searched for
  4. Grid points and related data distributed between processors
  5. Linear algebra is done
  6. Global ensemble is assembled

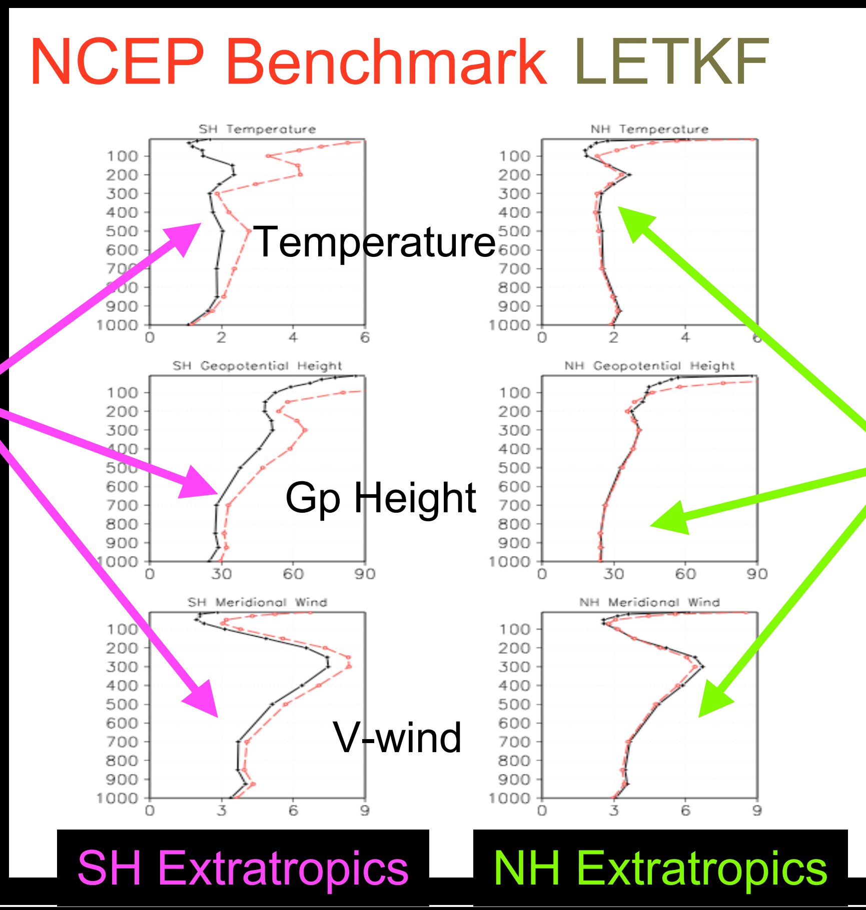
Validation Experiments with the NCEP GFS at resolution T62L28

- Simulated observations at randomly selected grid points (Szunyogh et al. 2005, Tellus)
  - full vertical column soundings of temperature, wind and surface pressure at randomly selected grid points,
  - 10% of grid points selected
- Simulated observations at locations of real observations (except for radiances)
- Real observations except for radiances (Szunyogh et al. 2007, Tellus, in press)
  - The LETKF and the Benchmark SSI system use different **H** operators; the one used with the LETKF is less sophisticated (For a clear comparison, where the same **H** is used in both systems see the talk by Jeff Whitaker). This may affect the results near the surface and in areas of high observational density
  - Benchmark SSI data are provided by NCEP (Y. Song and Z. Toth)
  - 60-member ensemble

Comparison of the LETKF and the SSI

48-hour forecasts with real observations (no radiances)

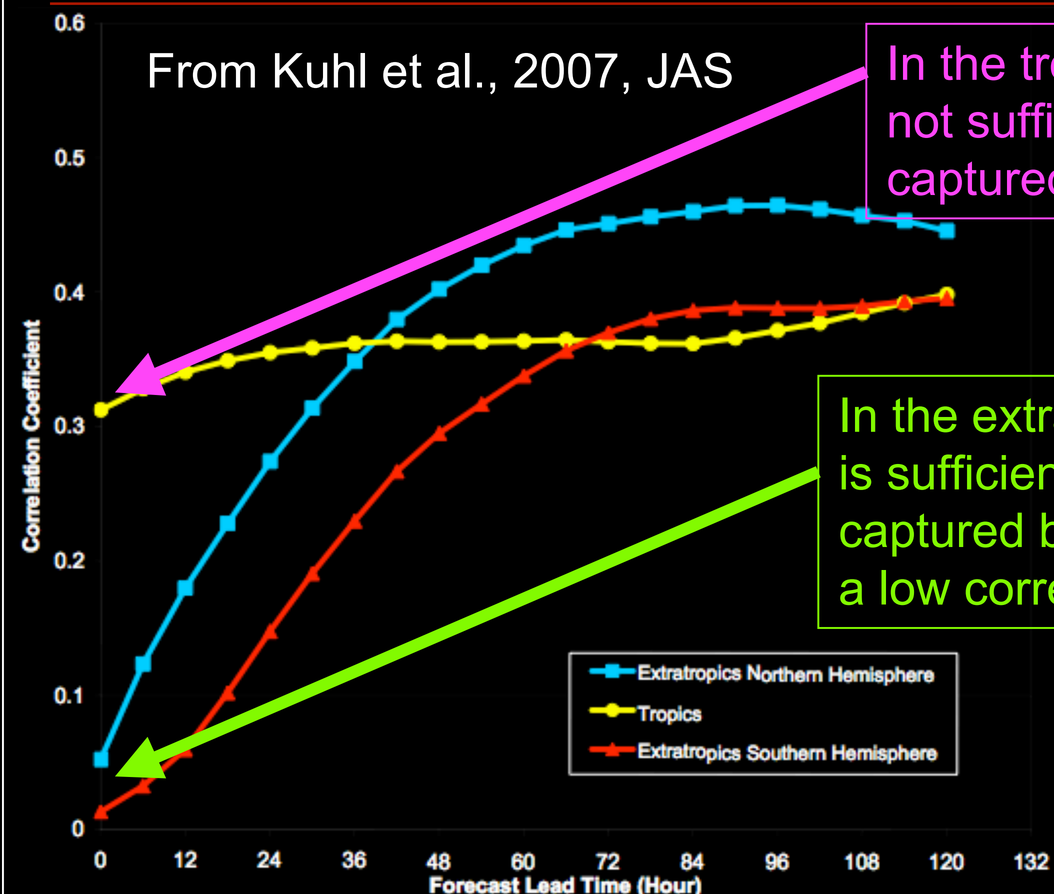
In the SH XT
The LETKF
is more
accurate



In the NH XT
the two systems
are comparable

Skill-Spread Correlation for the randomly distributed simulated vertical soundings

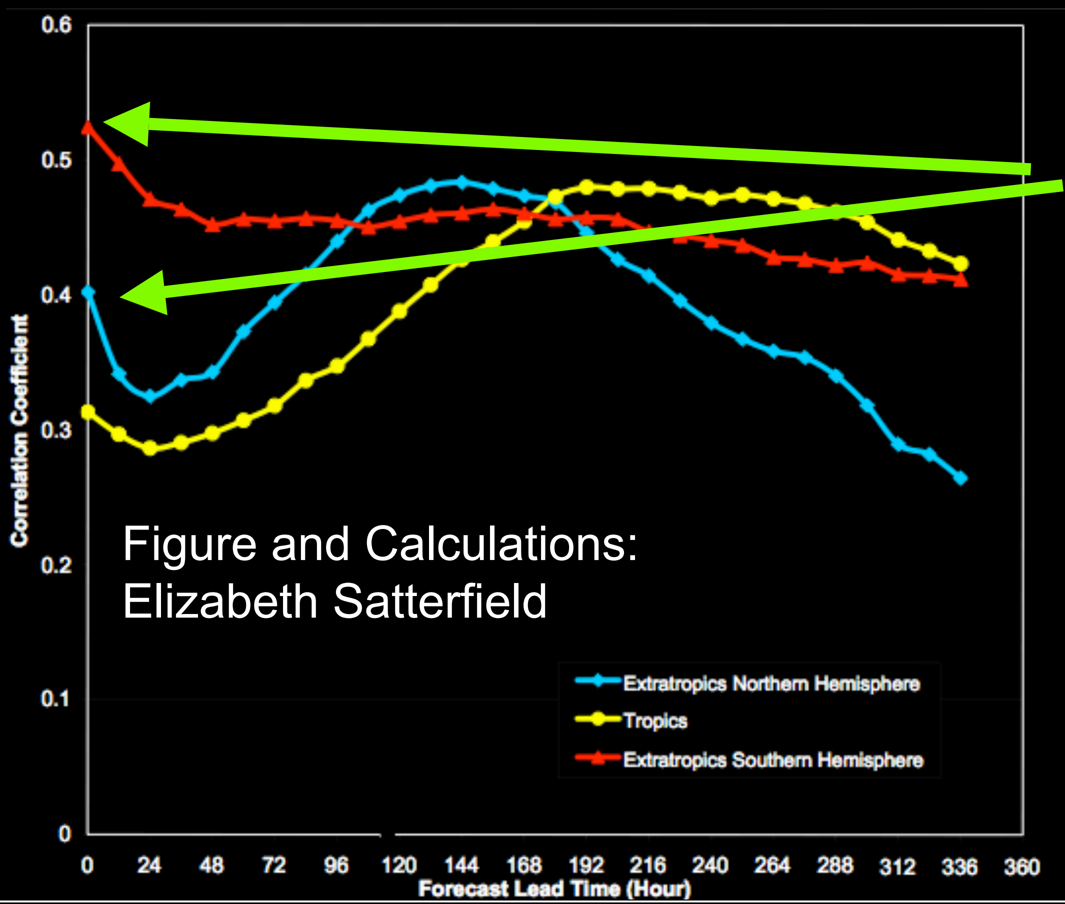
From Kuhl et al., 2007, JAS



In the tropics the data coverage is not sufficient to remove all errors captured by the ensemble

In the extratropics, such a data coverage is sufficient to suppress the errors captured by the ensemble, which leads to a low correlation at analysis time

Skill-Spread Correlation for the simulated observations at the location of real observations

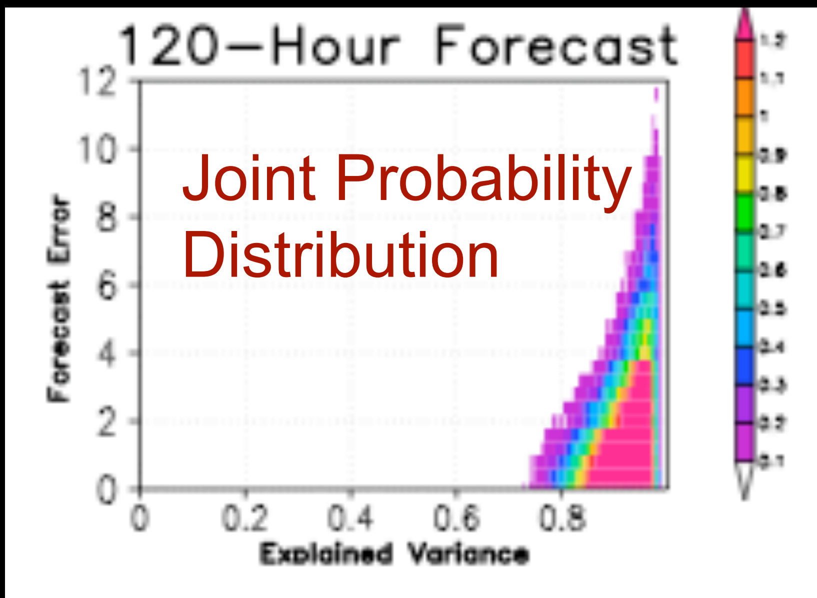


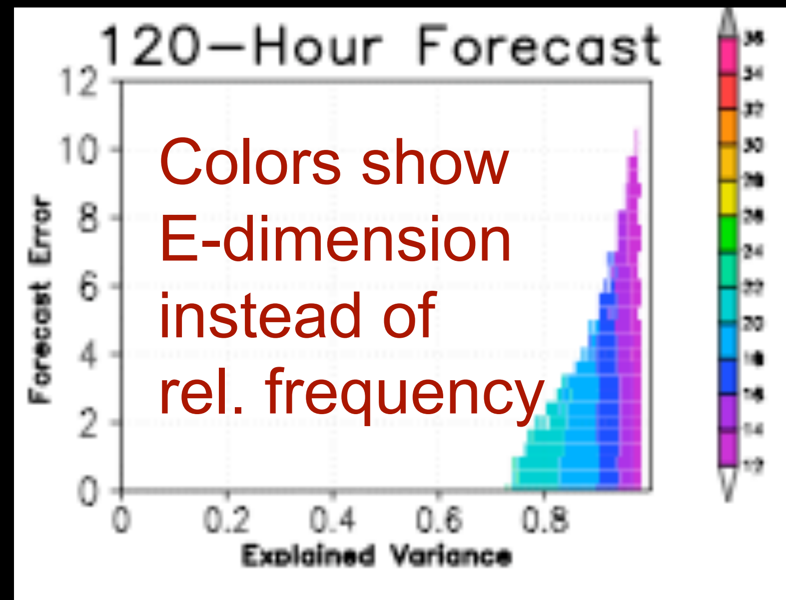
- This time the correlation is much higher in the extratropics: an indication that the data coverage is not sufficient to suppress a large portion of the errors correctly identified by the ensemble.
- This is especially true for the SH extratropics where the data coverage is much sparser
- What would be the correlation for the current operational data sets? (Work in progress)

Error Growth, Ensemble Performance, E-dimension,

- Predictability: Measured by Forecast Error at grid point
- Ensemble Performance (Explained Variance): Portion of the error that projects on the linear space spanned by the ensemble perturbations (difference between ensemble members and ensemble mean); 0 is worst, 1 is best (for our experimental design it is a measure of the predictability of predictability)
- E-dimension: A measure of the steepness of the spectrum of the error covariance matrices; when E-dimension=*k* the spectrum is flat, the smaller the E-dimension the steeper the spectrum (introduced in Patil et al. 2001, *PRL*; discussed in details in Oczkowski et al., 2005, *JAS*)

Predictability of Predictability for the randomly distributed simulated vertical soundings





Rapid Error Growth → Low E-dimension → Good Representation of Uncertainties

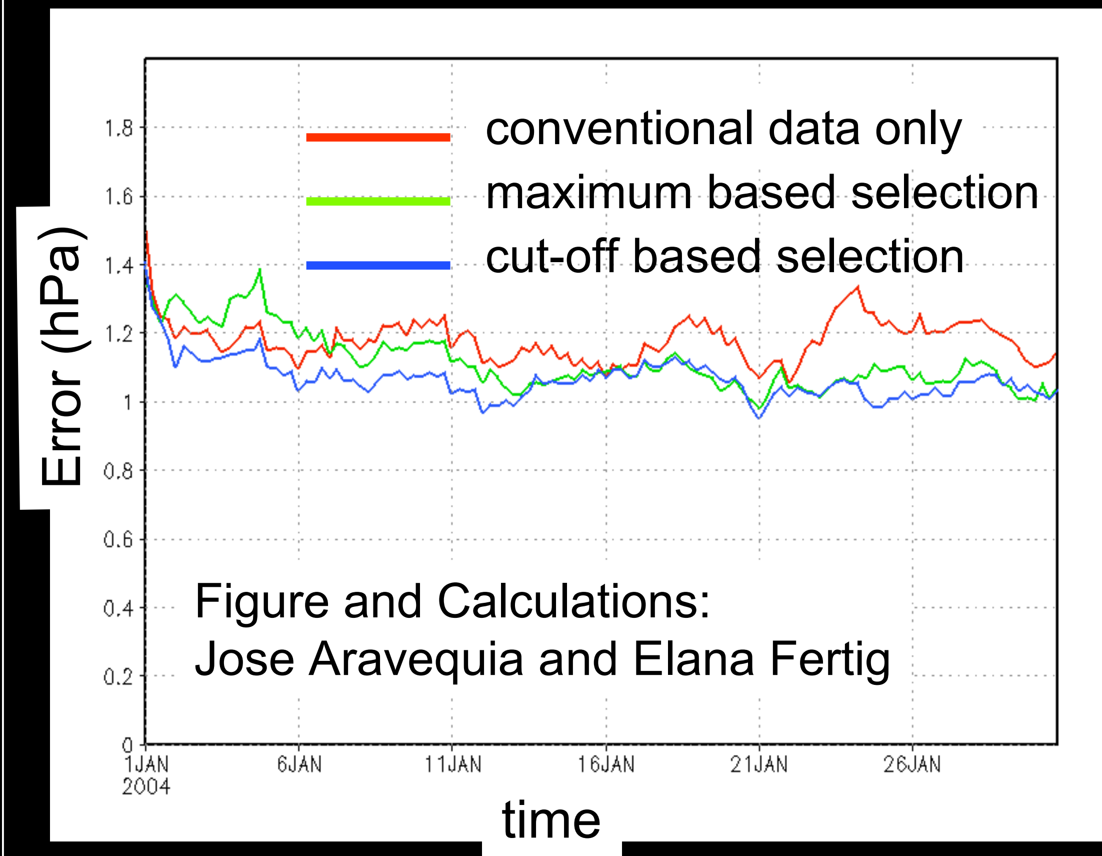
Low predictability → High Predictability of Predictability

From Kuhl et al., 2007, JAS

Assimilation of Satellite Radiances

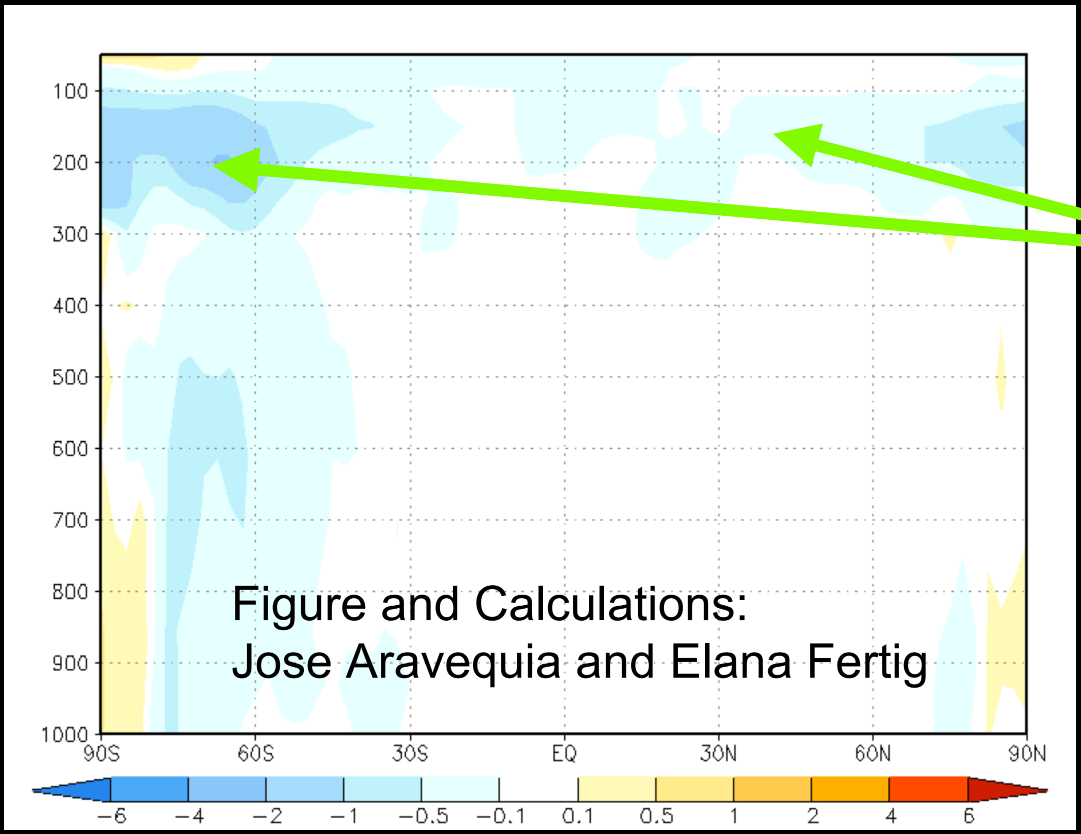
- Cutoff-based data selection strategy a radiance observation is assimilated at a given level, if a significant weight is assigned to any model state vector component within the associated local regions (Fertig et al., 2007, Tellus in press)
- Bias correction is done by augmenting the state with the bias parameters in the LETKF (Fertig et al., 2007, Tellus, under review)
- Example:
  - AMSU-A observations from Aqua, radiances are assimilated from 9 channels (channels with peak response near surface are not used)
  - Cutoff is at 60% of the maximum weight
  - Bias correction: $\mathbf{H}(\mathbf{x})$+bias, bias=aT_s+bs+c; T_s: surface temperature, s: scan angle; estimated are a, b, and c, which are then averaged for each latitude circle
  - $\mathbf{H}(\mathbf{x})$ is calculated with the CRTM from JCSDA
  - 40-member ensemble

Time Evolution of the Surface Pressure RMS Error analyses verified against operational NCEP analysis



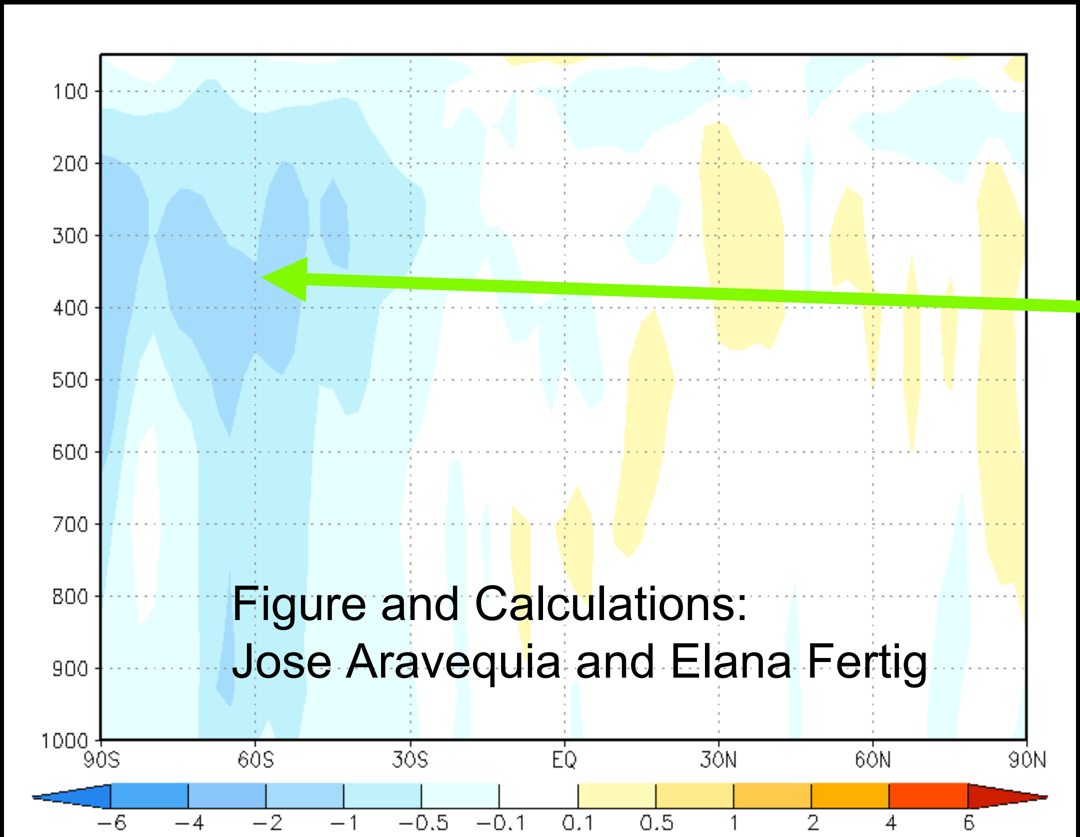
- The error is the largest when only conventional data are assimilated
- The maximum-based data selection leads to a long settling time and occasionally larger errors than those for the cut-off based selection
- The cut-off based selection provides the overall best performance

Effects of the AMSU-A Data on the 48-hour Forecast Errors temperature, 15-day mean



- The AMSU-A data make the most important contribution in the upper troposphere, especially in the SH

Effects of the AMSU-A Data on the 48-hour Forecast Errors meridional wind



- The large improvements in the SH suggests, that there is a lot of useful information in the estimated background error covariance matrix between the temperature (most closely related to the radiances) and the wind

Concluding Remarks

- There seems to be a lot of useful flow-dependent information about the error structures in the data sparse regions
- An issue for discussion: This conclusion is drawn from low-resolution experiments. Is there an advantage to the ensemble based approach at the much higher operational resolutions using full operational data sets?
- Another issue for discussion: Is there any computationally efficient way to transport information about the background error covariance from an ensemble to the currently implemented variational scheme?
- For more information on the UMCP Weather & Chaos Project visit http://weatherchaos.umd.edu