Ensemble Filtering at NCAR

The Importance of Four-Dimensional Localization

Shree P. Khare and Jeffrey L. Anderson

`khare@ucar.edu and jla@ucar.edu`

National Center for Atmospheric Research

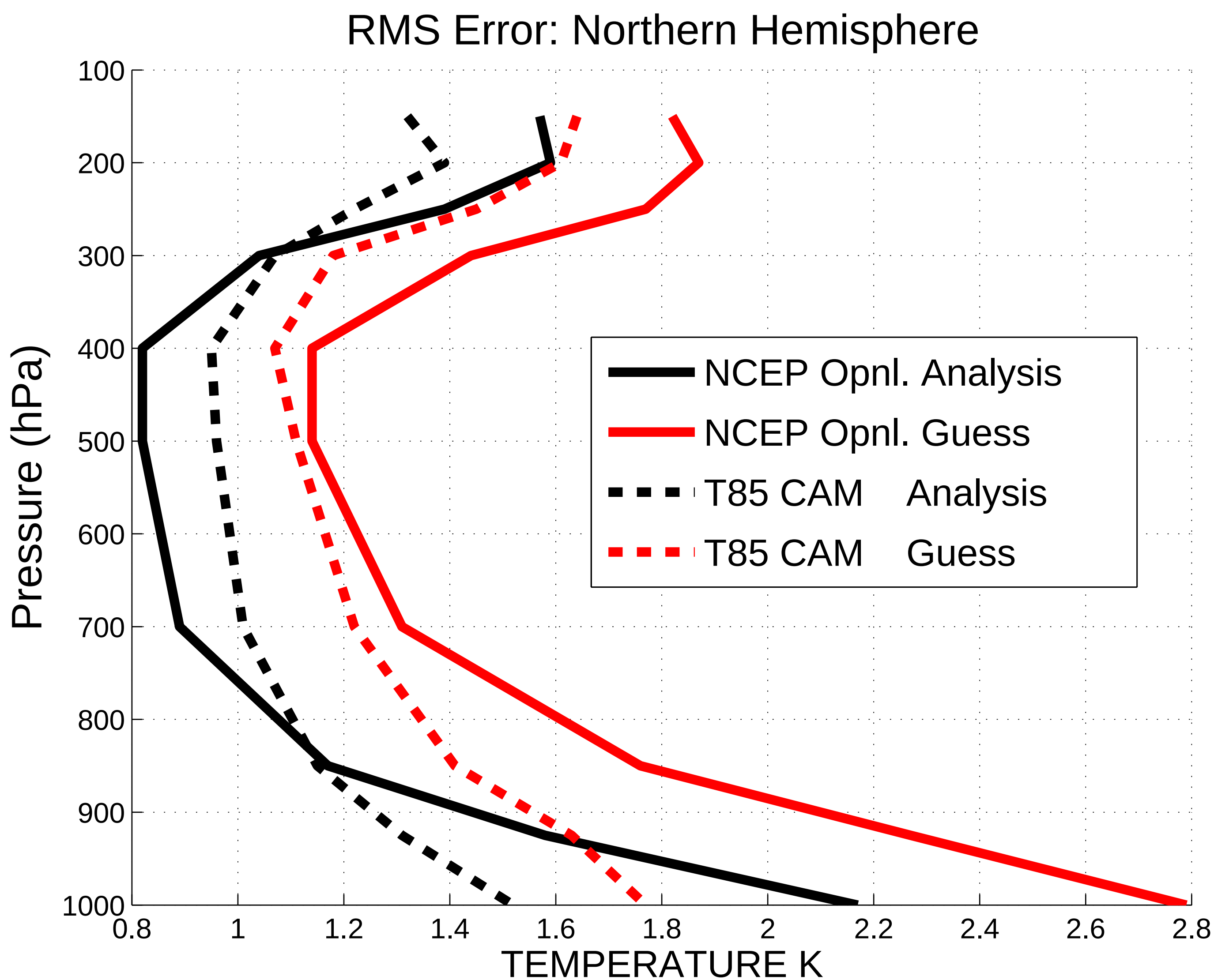
Talk Outline

- 1. Ensemble Filtering at NCAR - Comparisons to the NCEP Operational System of Jan 2003
- 2. Four-Dimensional Localization
- 3. Routine Network Design
- 4. Ensemble Smoothers

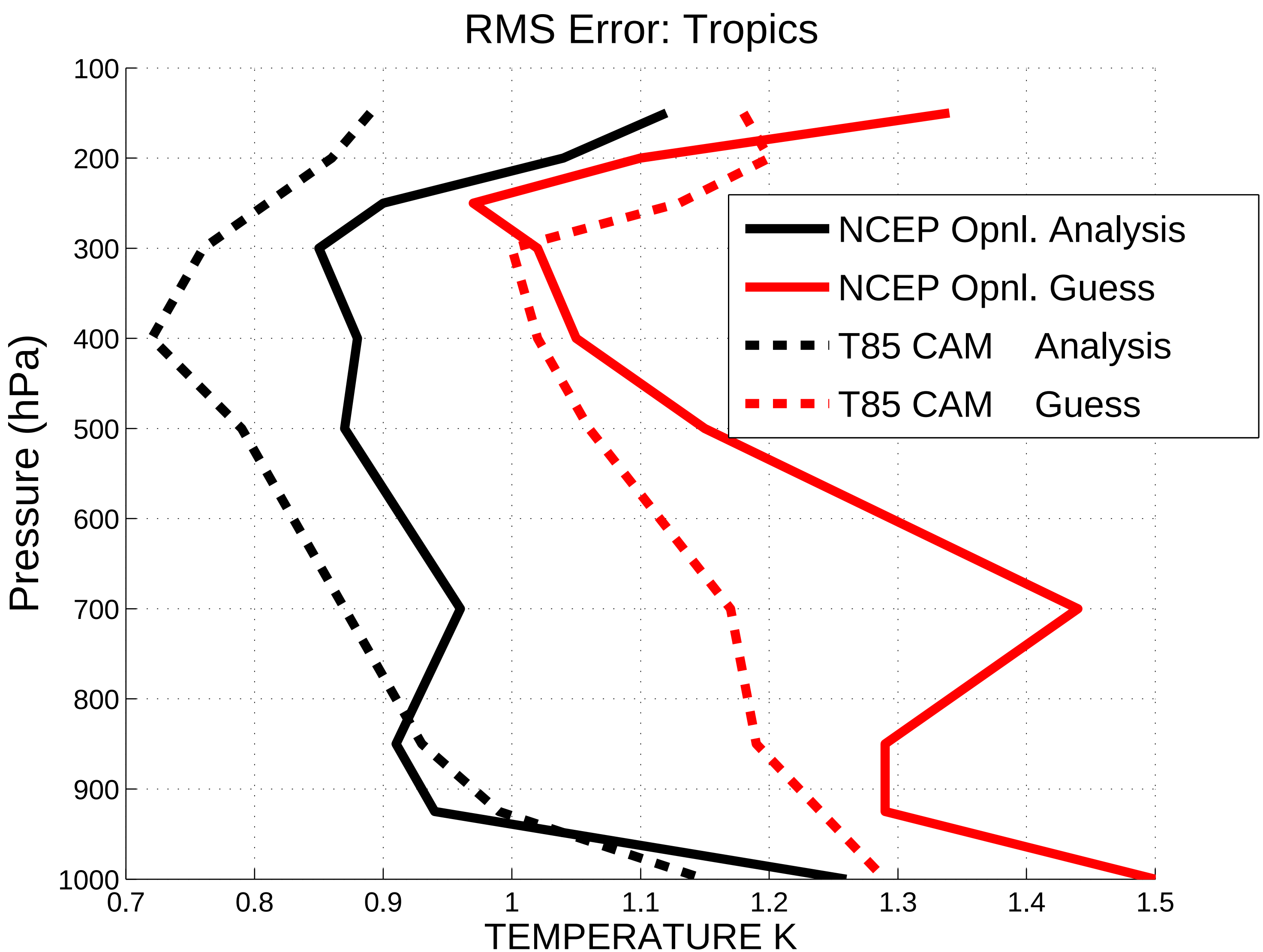
CAM Assimilation - Jan 2003

- **CAM 3.1 T85L26**
  - **U, V, T, Q and PS state variables impacted by observations**
  - **Land model (CLM 2.0) not impacted by observations**
  - **Climatological SSTs**
- **Assimilation/Prediction Experiments**
  - **80 member ensemble divided into 4 equal groups**
  - **Adaptive error correction algorithm**
  - **Initialized from climatological distribution (huge spread)**
  - **Uses most observations used in reanalysis (Radiosondes, ACARS, Sat.Winds ..., no surface obs. or retrievals)**
  - **Assimilated every 6 hours; +/- 1.5 hour windows for obs**
- **Results generated by KEVIN RAEDER**

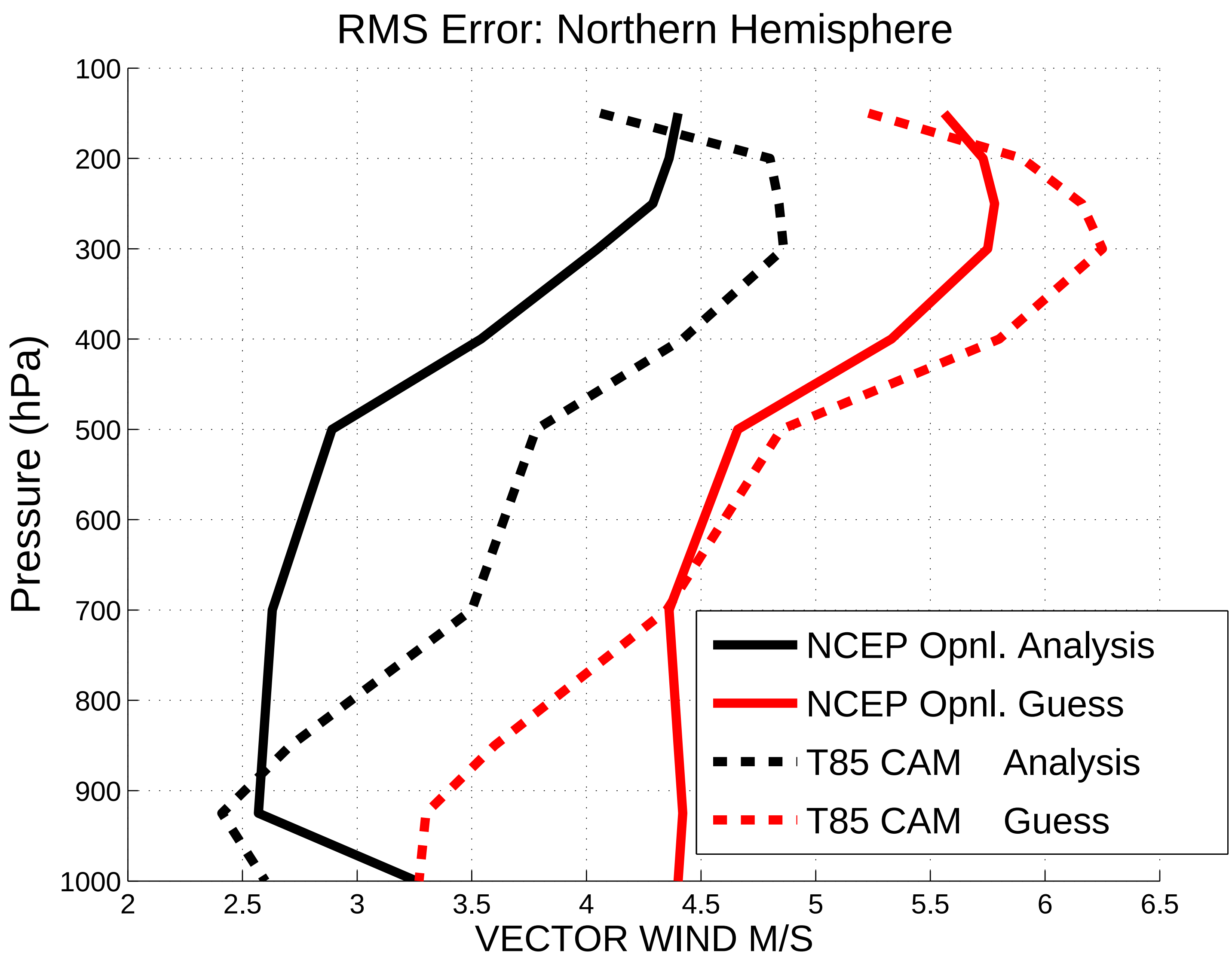
Observation Space Temperature RMS: NH



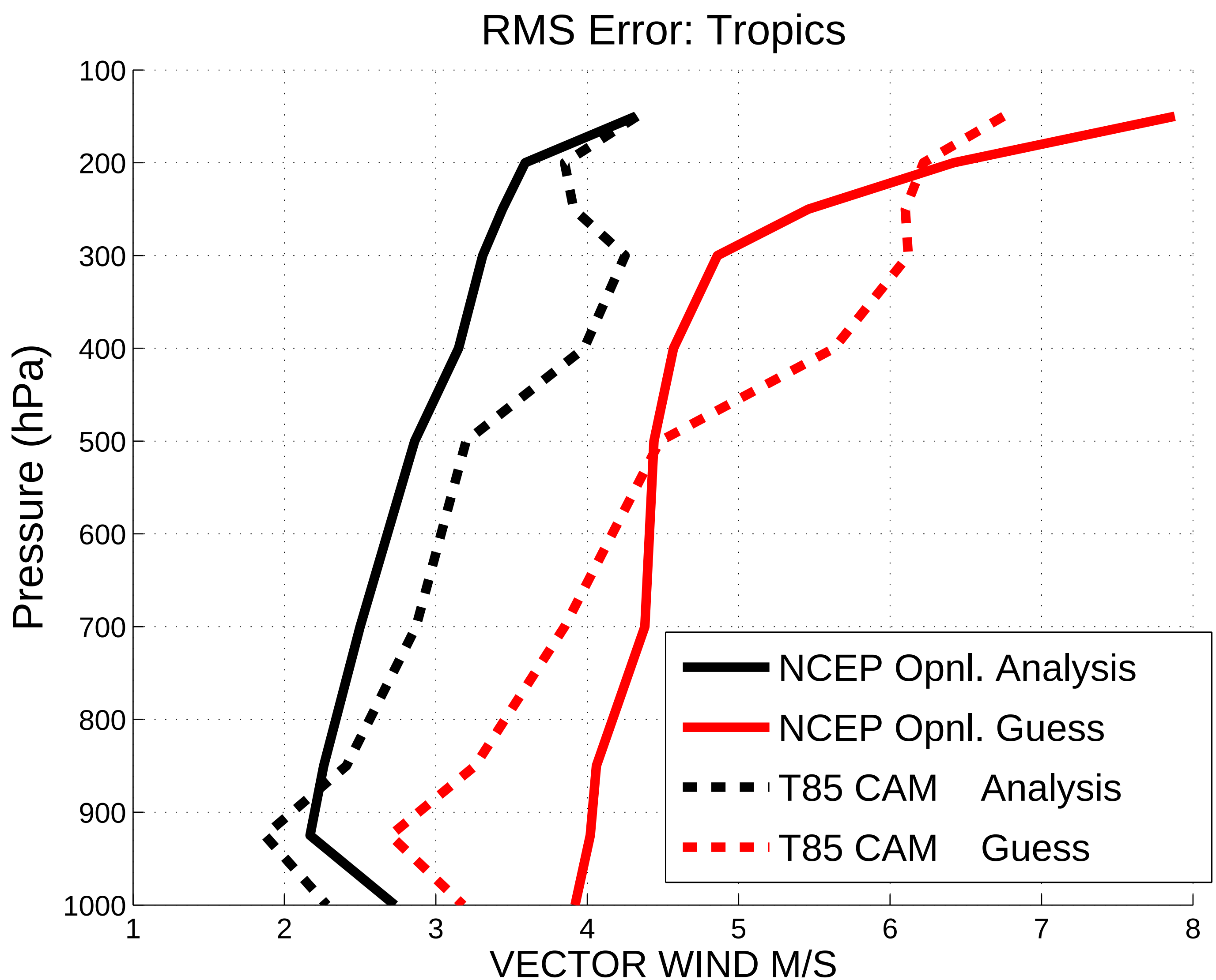
Observation Space Temperature RMS: TROPICS



Observation Space Wind RMS: NH



Observation Space Wind RMS: TROPICS



Next ...

- 1. Ensemble Filtering at NCAR - - Comparisons to the NCEP Operational System of Jan 2003
- 2. Four-Dimensional Localization
- 3. Routine Network Design
- 4. Ensemble Smoothers

Four-Dimensional Localization

- Ensemble Filters - Advantage: Adaptive Observations (ETKF), Routine Network Design (RDA) and Ensemble Smoothers

Four-Dimensional Localization

- Ensemble Filters - Advantage: Adaptive Observations (ETKF), Routine Network Design (RDA) and Ensemble Smoothers
- Restriction: Ensemble sizes are limited

Four-Dimensional Localization

- Ensemble Filters - Advantage: Adaptive Observations (ETKF), Routine Network Design (RDA) and Ensemble Smoothers
- Restriction: Ensemble sizes are limited
- Sequential data assimilation: estimates of $corr(x_1(t), x_2(t))$ - contaminated by effects of sample size

Four-Dimensional Localization

- Ensemble Filters - Advantage: Adaptive Observations (ETKF), Routine Network Design (RDA) and Ensemble Smoothers
- Restriction: Ensemble sizes are limited
- Sequential data assimilation: estimates of $corr(x_1(t), x_2(t))$ - contaminated by effects of sample size
- Adaptive Observations, Routine Network Design, Ensemble Smoothers: Calculations in extended state space

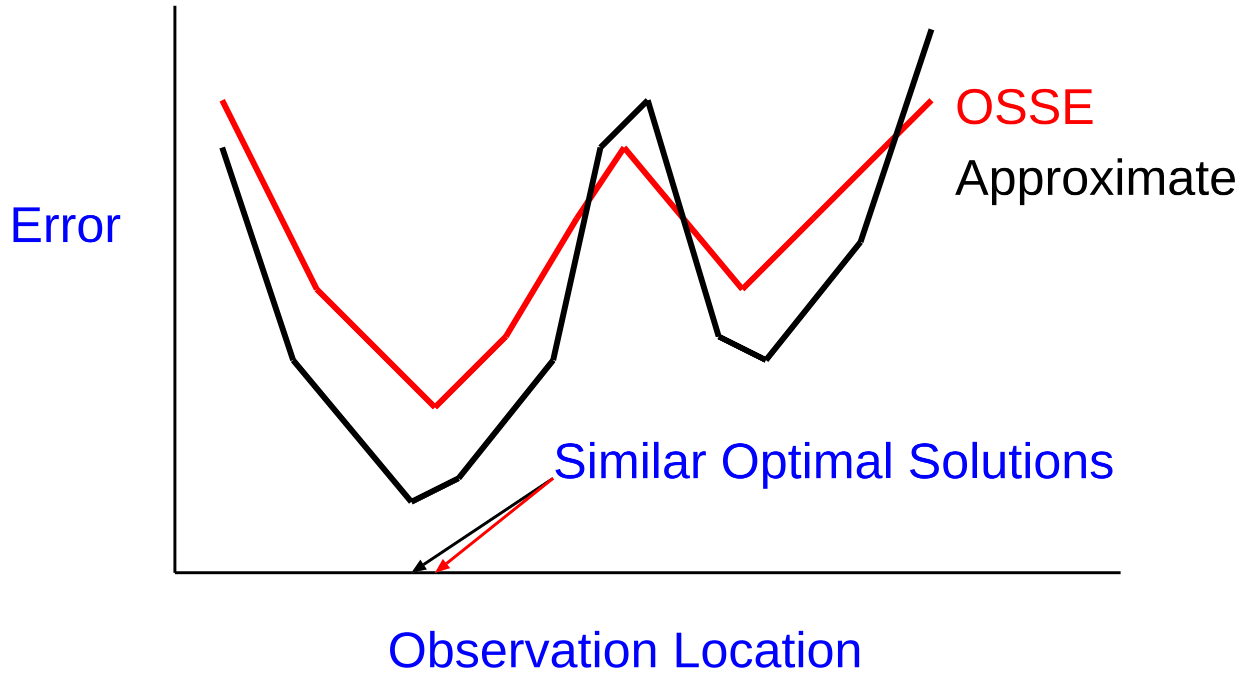
Four-Dimensional Localization

- Ensemble Filters - Advantage: Adaptive Observations (ETKF), Routine Network Design (RDA) and Ensemble Smoothers
- Restriction: Ensemble sizes are limited
- Sequential data assimilation: estimates of $corr(x_1(t), x_2(t))$ - contaminated by effects of sample size
- Adaptive Observations, Routine Network Design, Ensemble Smoothers: Calculations in extended state space
- Consider $corr(x_1(t), x_2(\tilde{t}))$ where $t \neq \tilde{t}$

Next ...

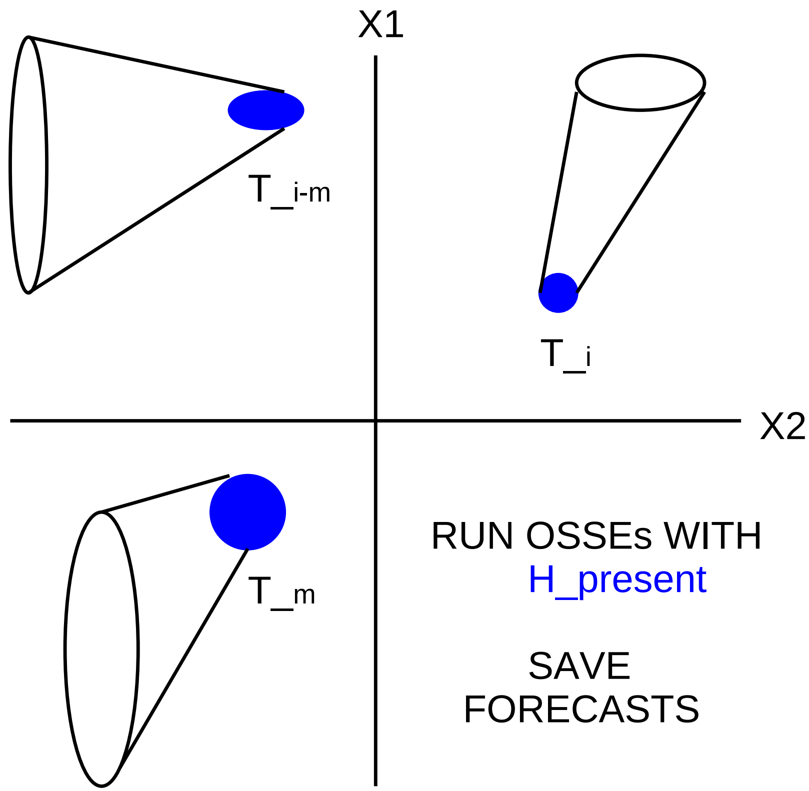
- 1. Ensemble Filtering at NCAR - Comparisons to the NCEP Operational System of Jan 2003
- 2. Four-Dimensional Localization
- 3. Routine Network Design
- 4. Ensemble Smoothers

Approximating Information Derived from OSSEs



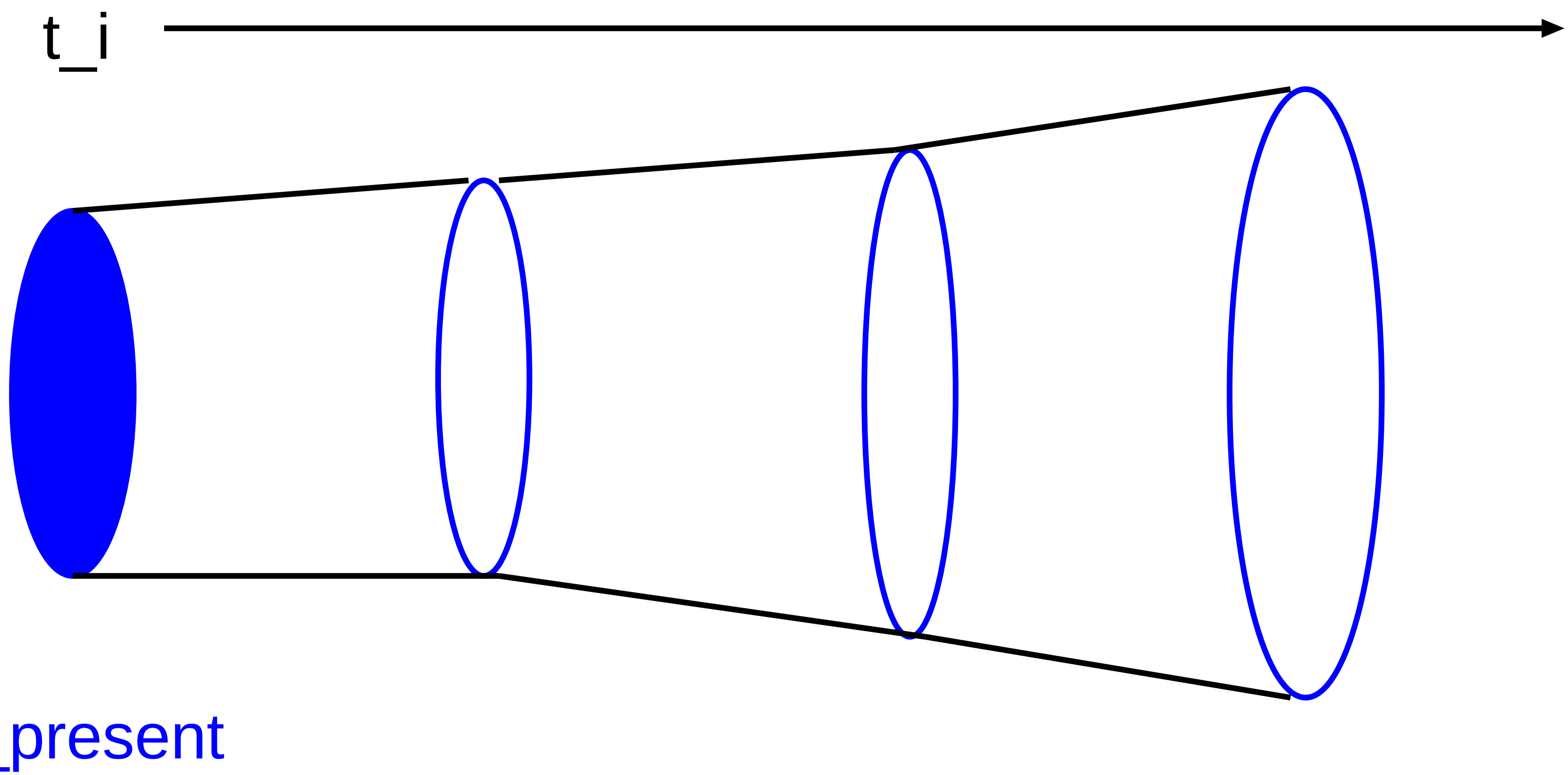
- How can we obtain a statistically and dynamically significant approximation of information derived from OSSEs?

Retrospective Design Algorithm



- Trial network $H_{trial} = [H_{present}; H_{additional}]$

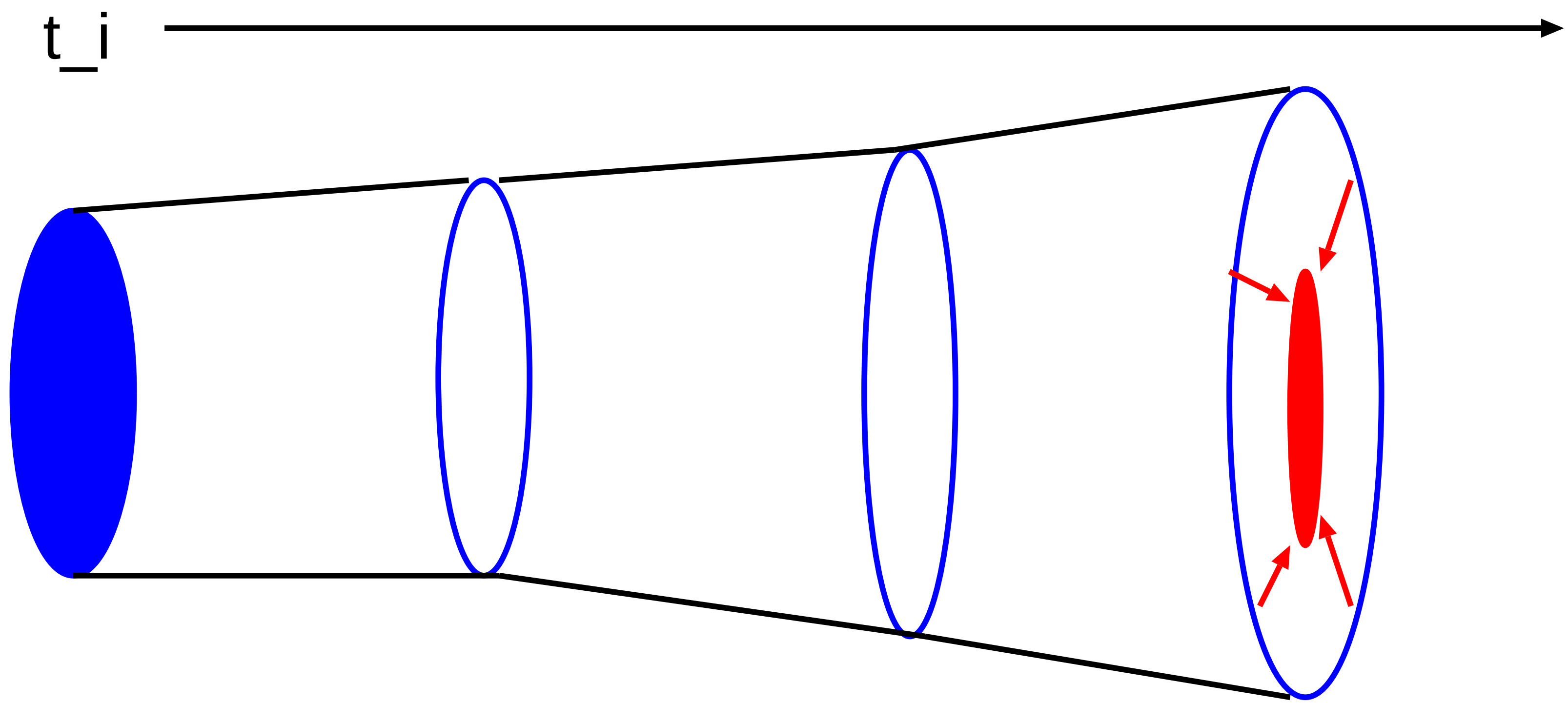
Retrospective Design Algorithm



An ensemble forecast generated at t_i during

the OSSE with H_present

Retrospective Design Algorithm

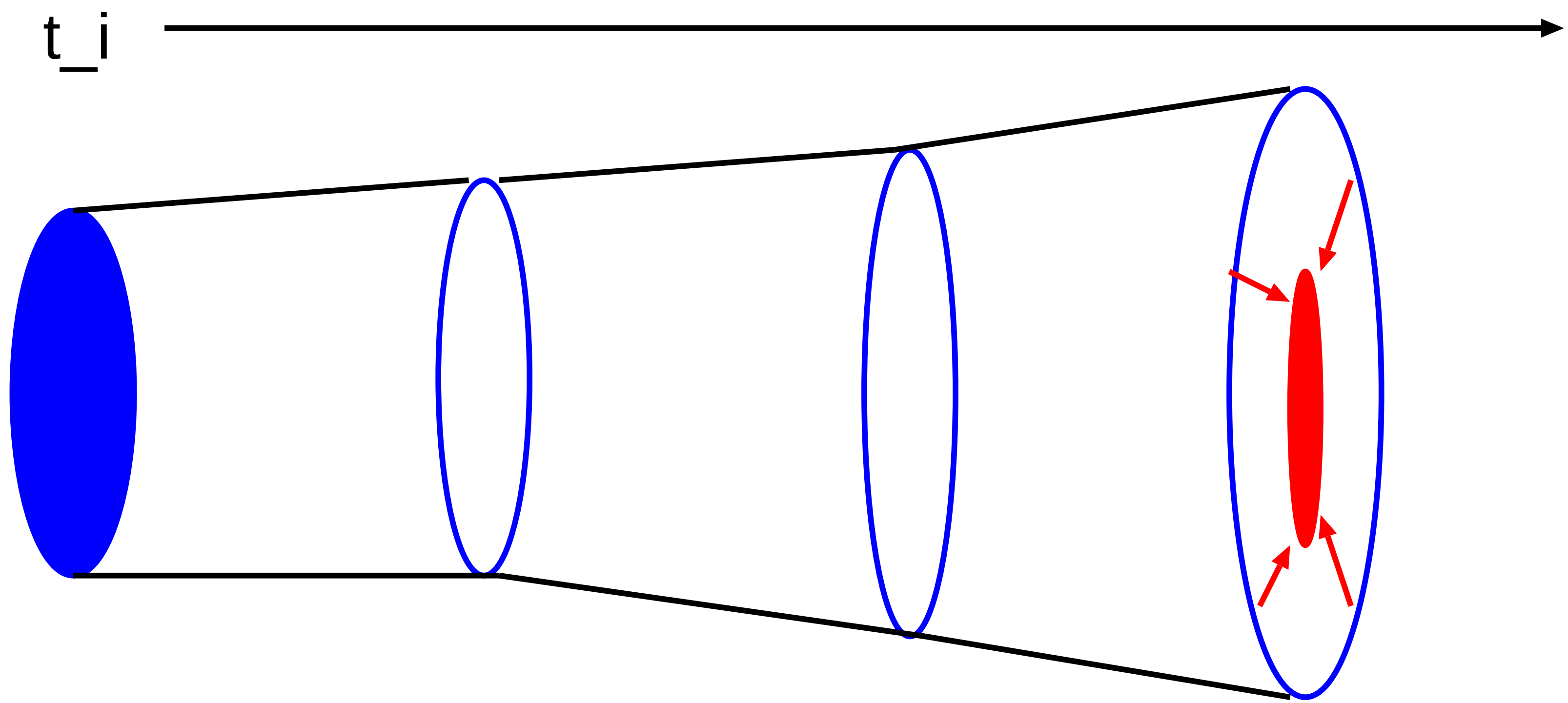


H_present H_trial = H_present + H_additional

From t_i+1 onward, assume the observing network is H_trial - the trial network

Want to compute the covariance of the atmosphere given H_trial for some time t > t_i

Retrospective Design Algorithm

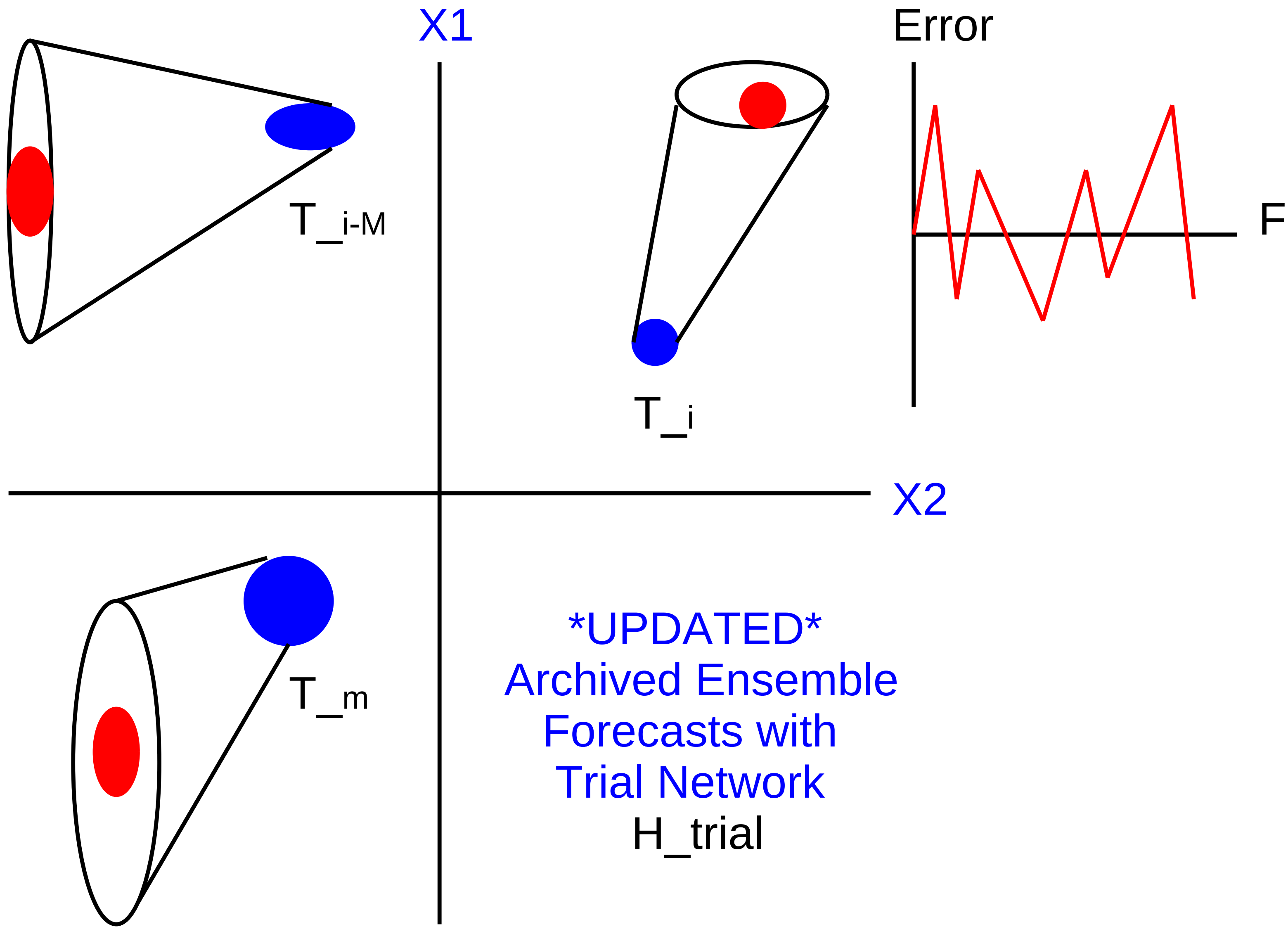


H_present H_trial = H_present + H_additional

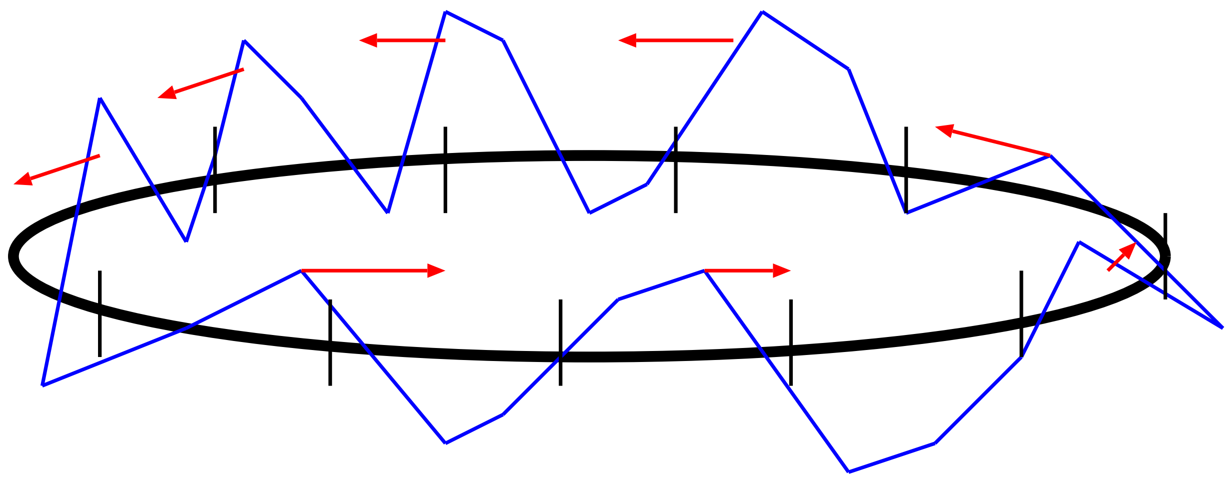
Without re-running the forecast model - an EnKF based algorithm exists for computing the atmosphere's covariance at t > t_i given trial network

H_trial = H_present + H_additional - KEY POINT

Evaluate the Objective Function using the RDA



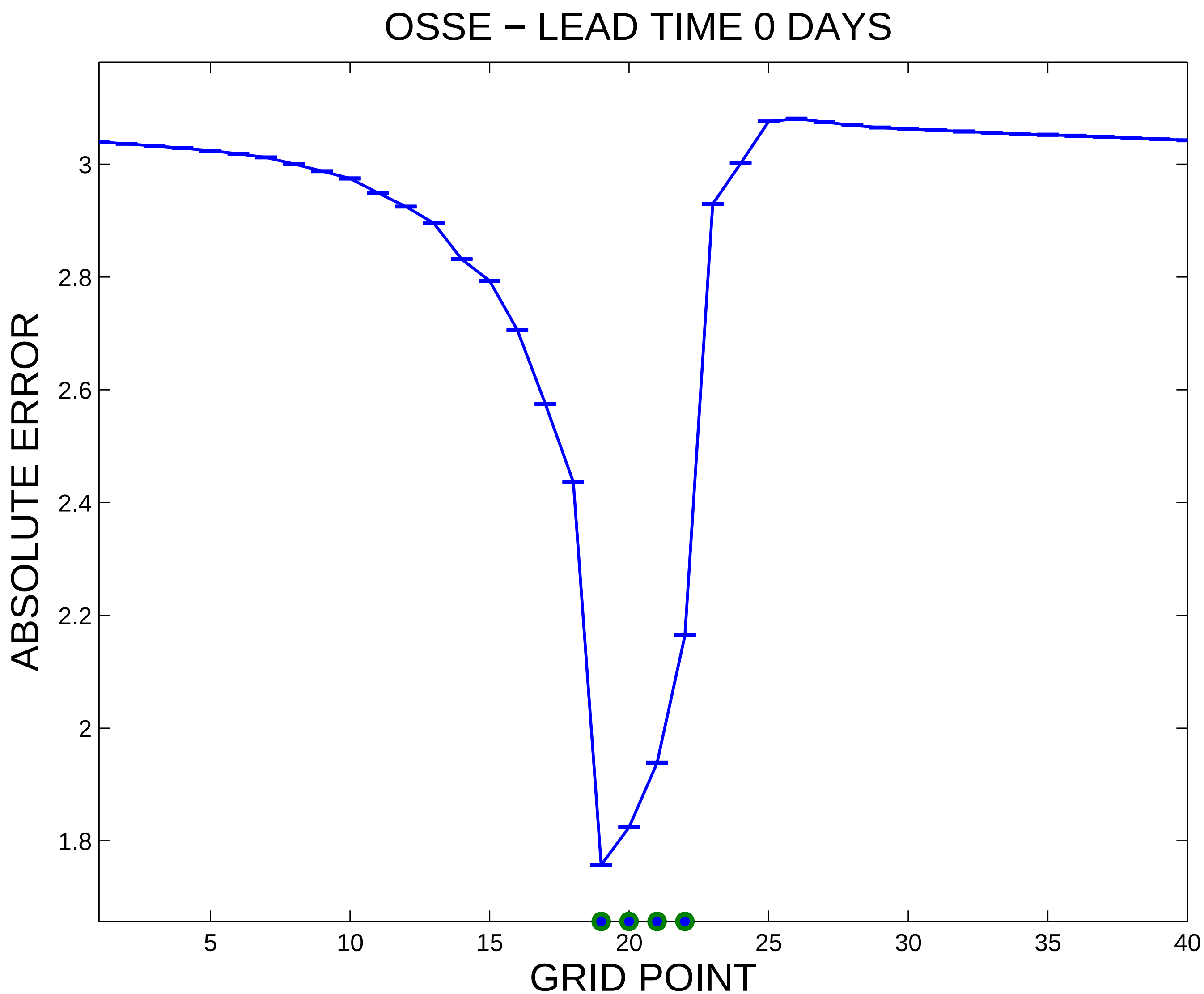
Experiments in an Atmospheric 'Toy Model'



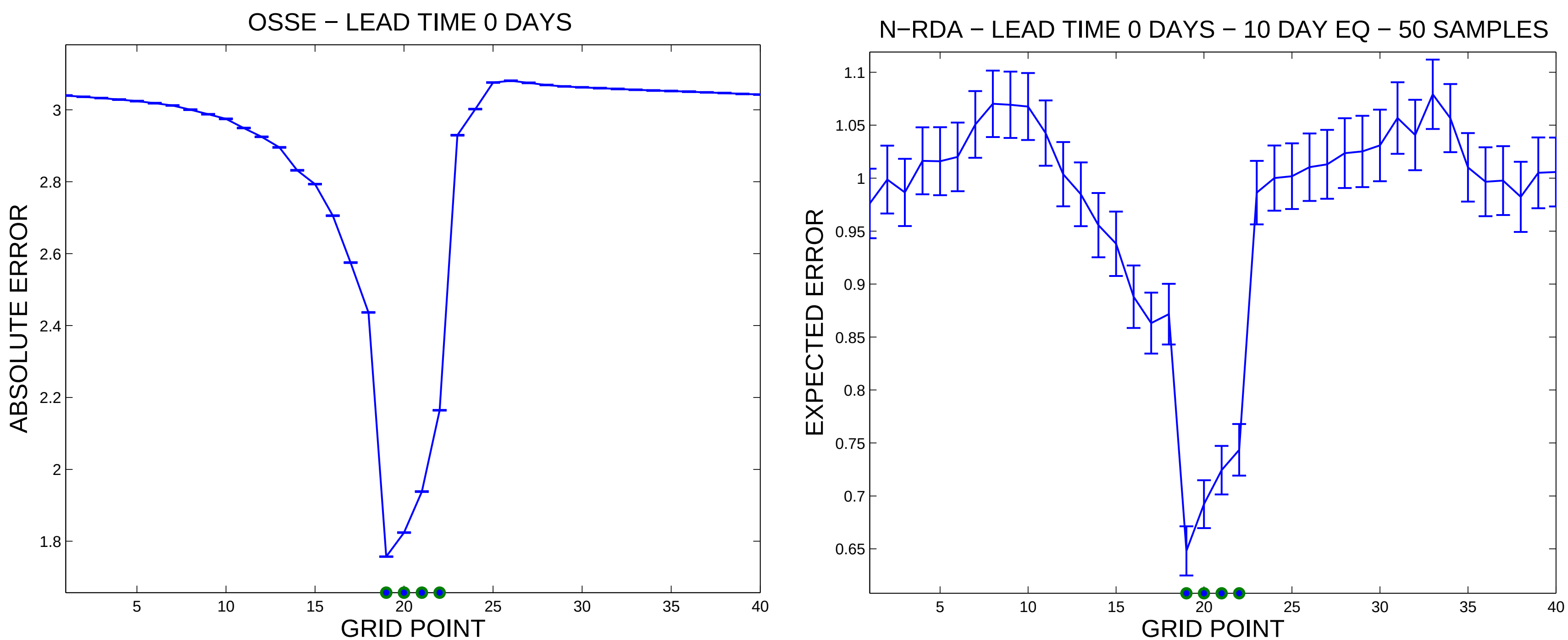
$$\frac{dx_j}{dt} = -x_{j-2}x_{j-1} + x_{j-1}x_{j+1} - x_j + F$$

- Non-linear advection, forcing, linear dissipation, energy conservation $E = (x_1^2 + ... + x_{40}^2)$
- $F = 8$ sensitive dependence on IC's, $j = 1, ..., 40$, cyclic
- Climatology $\sqrt{\sigma_{climate}} = 3.6$ and $\overline{x} = 2.3$

OSSEs in Lorenz 1996

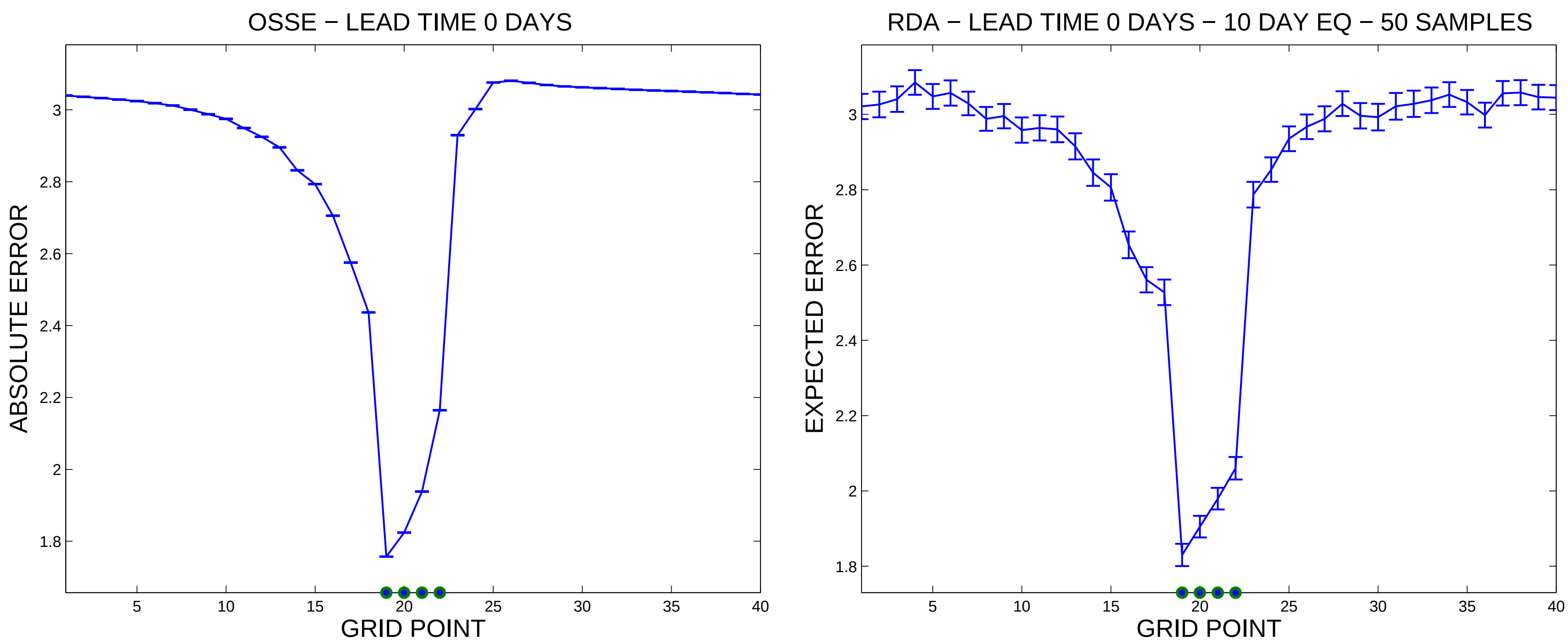


OSSEs vs. RDA (No Localization)



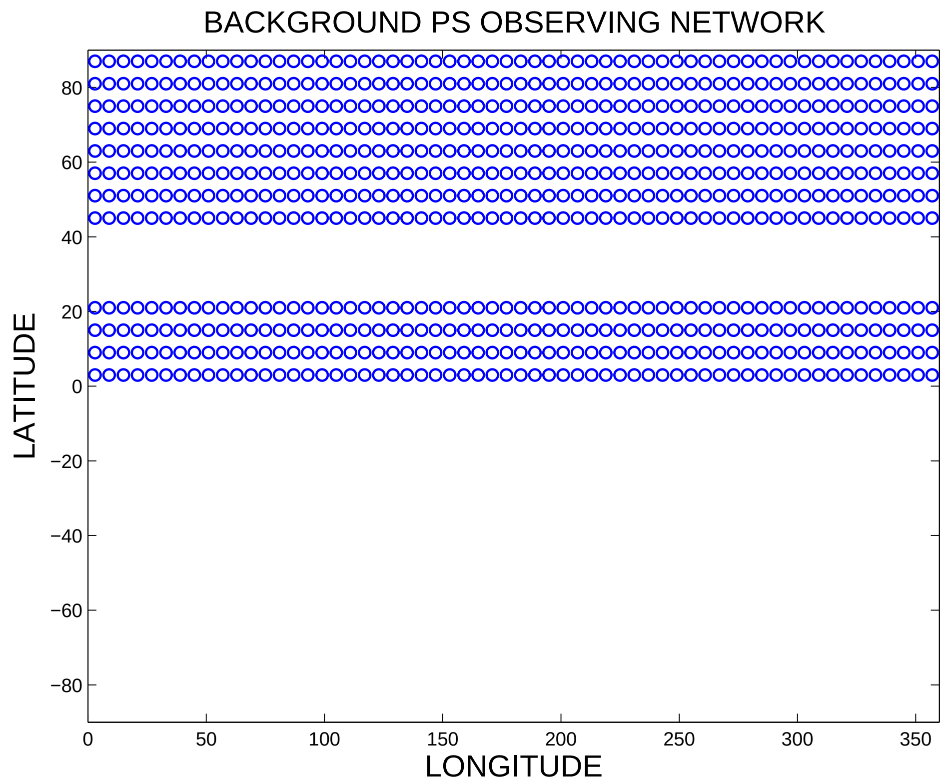
- Verification region - continuous
- Forecast lead time - 0 days
- 50 independent samples - noisy

OSSEs vs. RDA (With Localization)



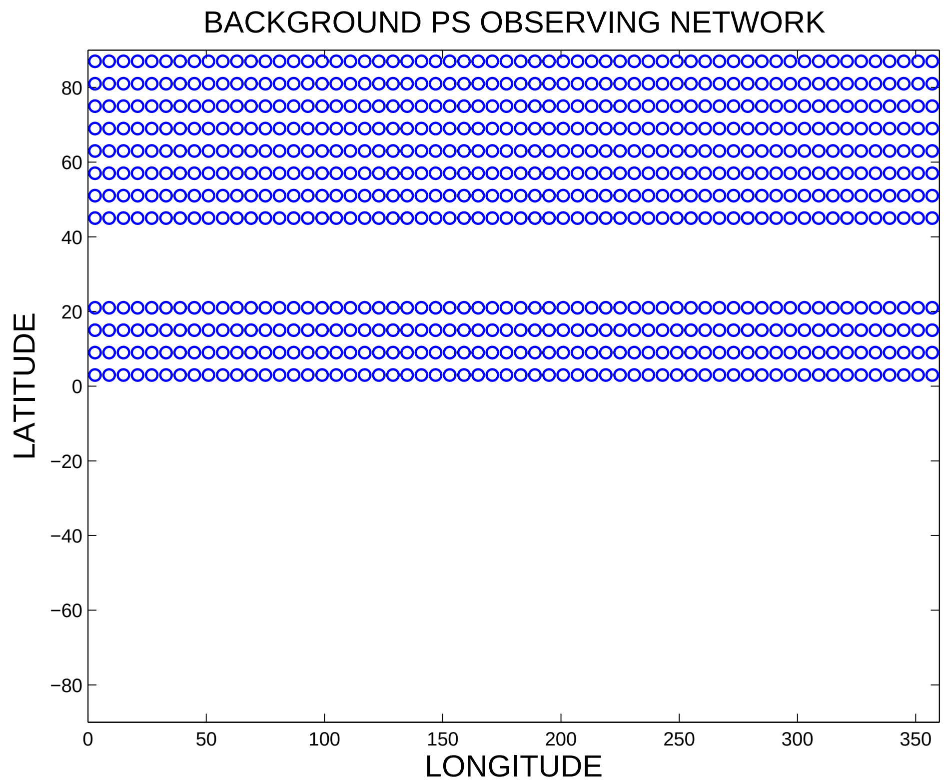
- Verification region - continuous
- Forecast lead time - 0 days
- 50 independent samples - reduced noise

PS Network Design in a GCM



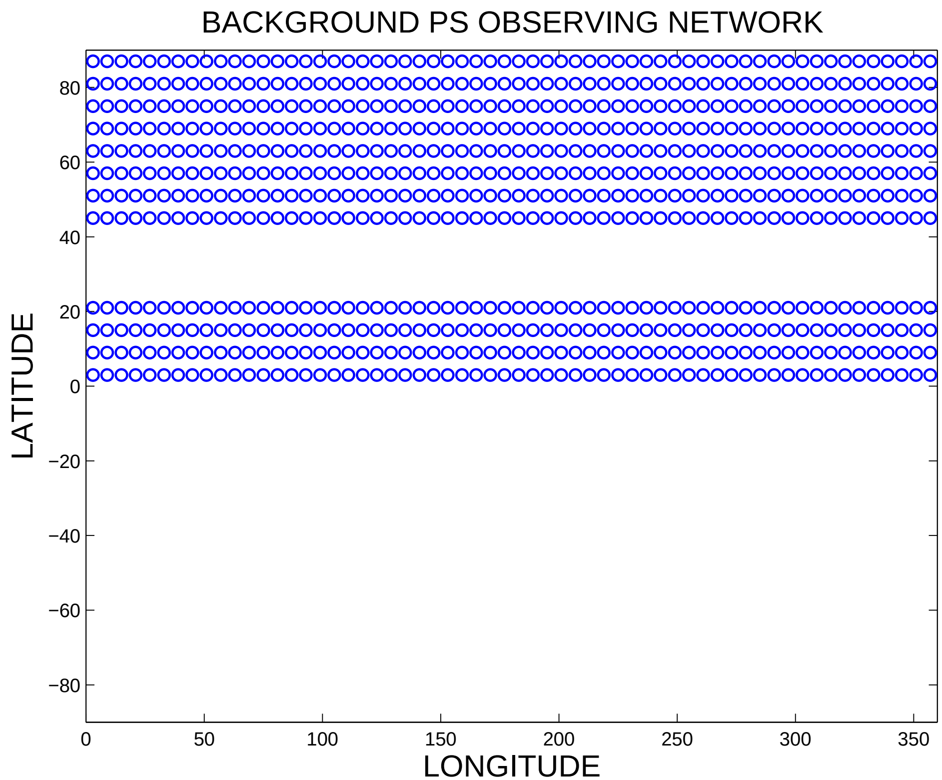
- FIXED network of surface pressure observations - $7\ mb$ observational standard deviation - assimilate every 12 hours

PS Network Design in a GCM



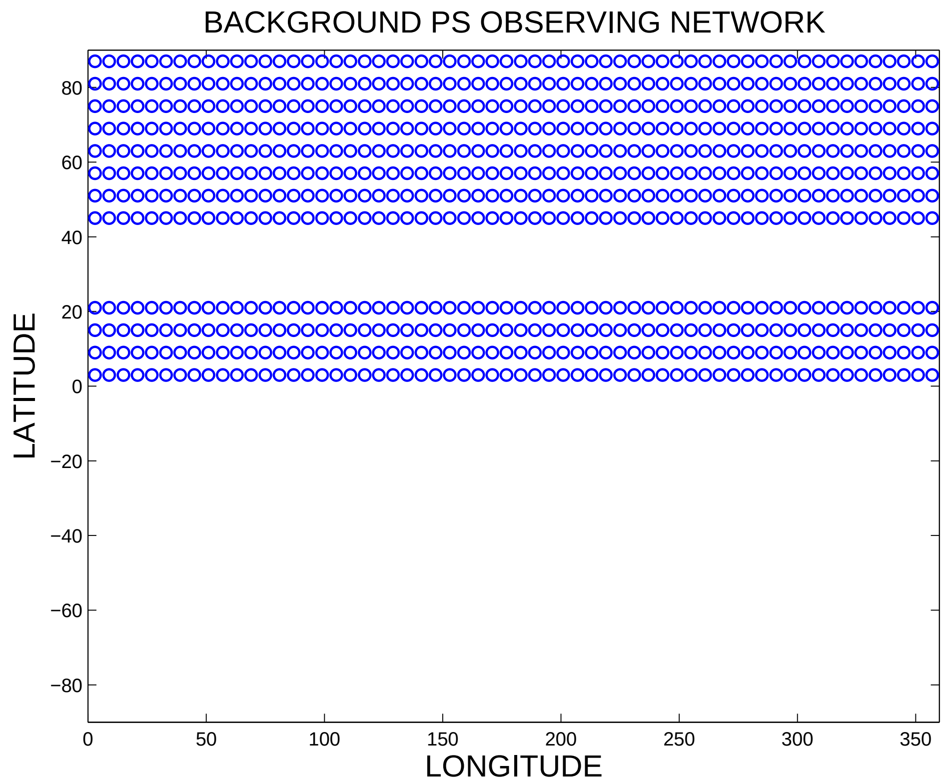
- Run an EAKF with $N = 20$ ensemble members (with localization and no inflation)

PS Network Design in a GCM



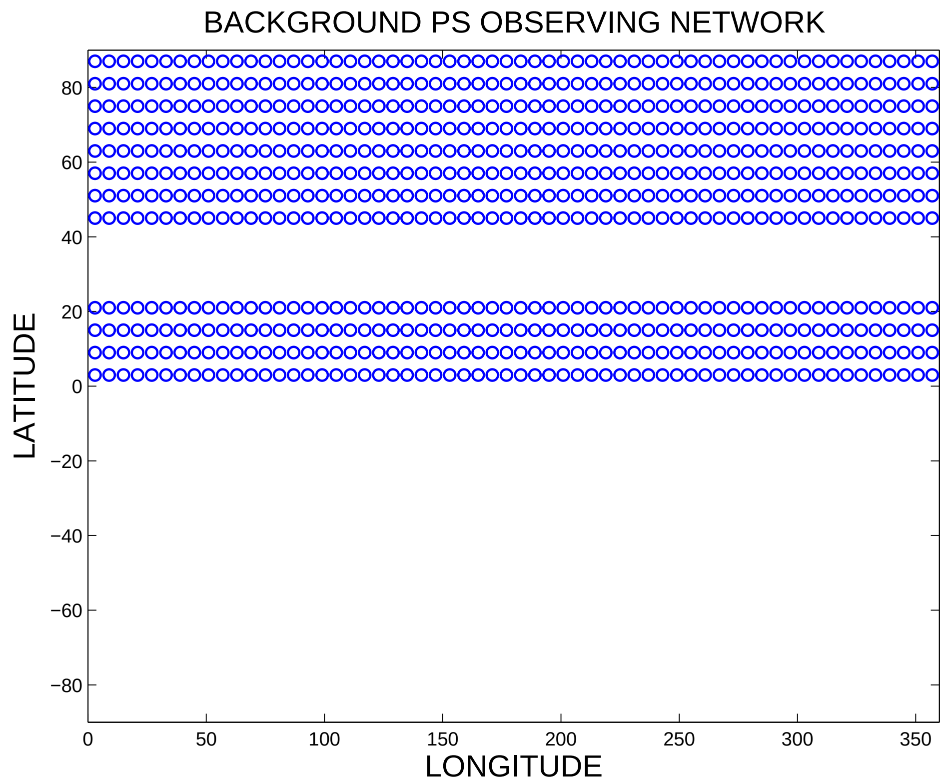
- Zonally symmetric Held-Suarez forcing

PS Network Design in a GCM



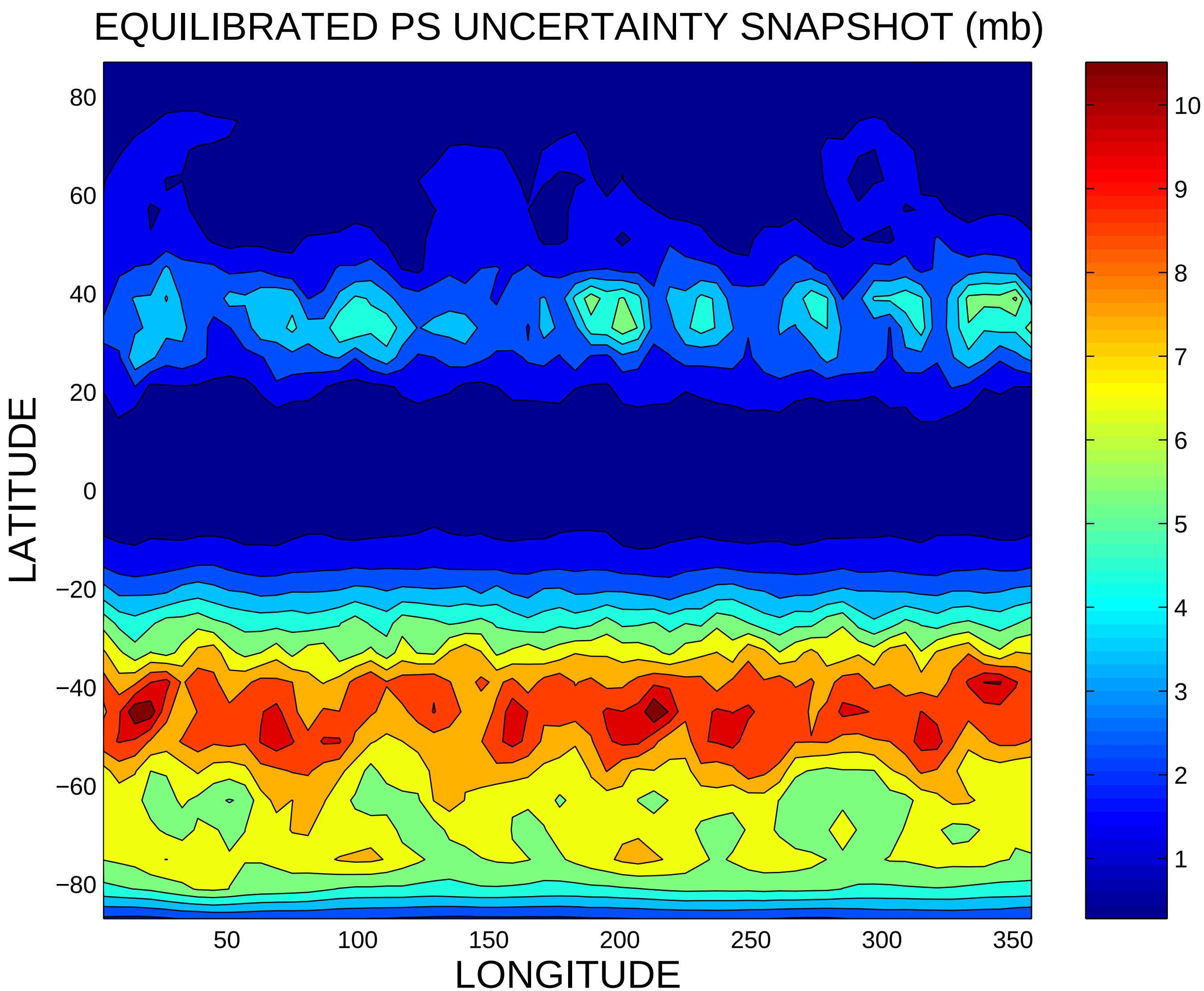
- 6 degrees horizontal resolution (60×30) - 5 vertical levels

PS Network Design in a GCM

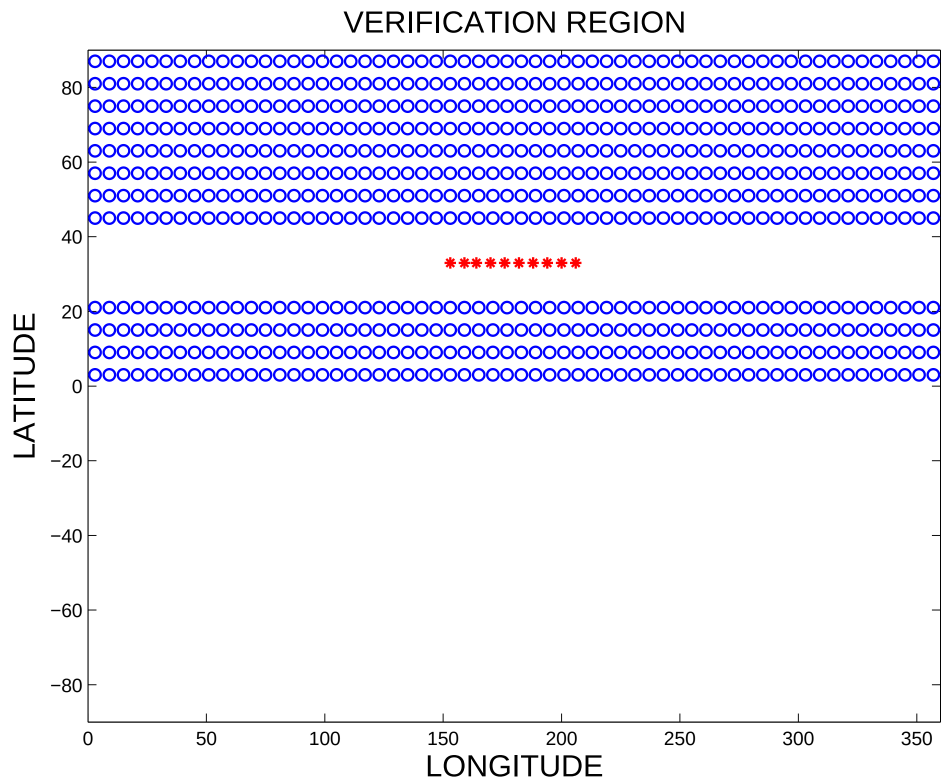


- Temperature gradient drives a baroclinically unstable flow in the mid-latitudes

Assimilation Results: Posterior PS Uncertainty

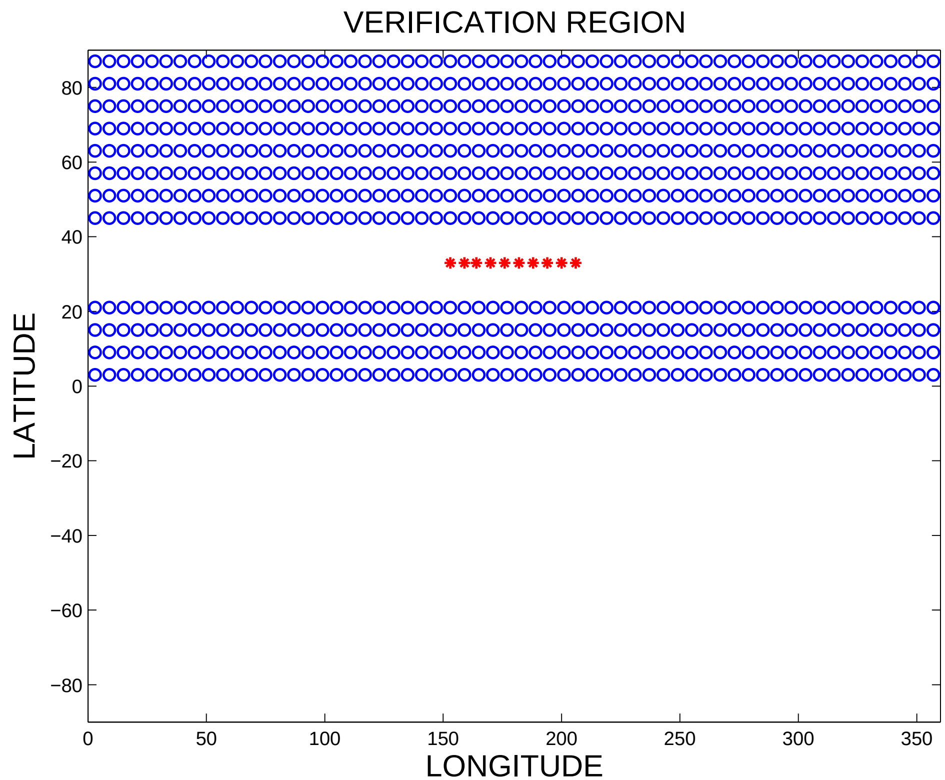


The Experiment



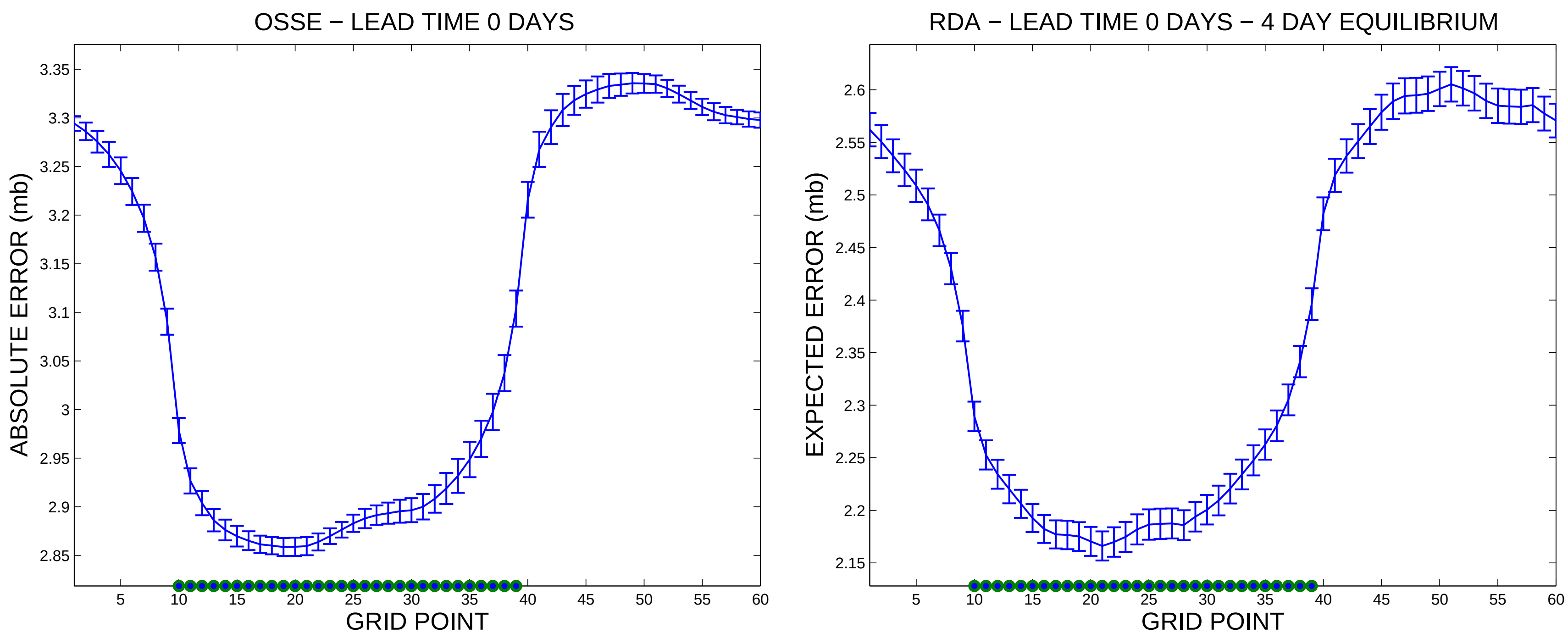
- Find the optimal placement of one *additional* accurate PS observation along the 33 degree latitude band

The Experiment



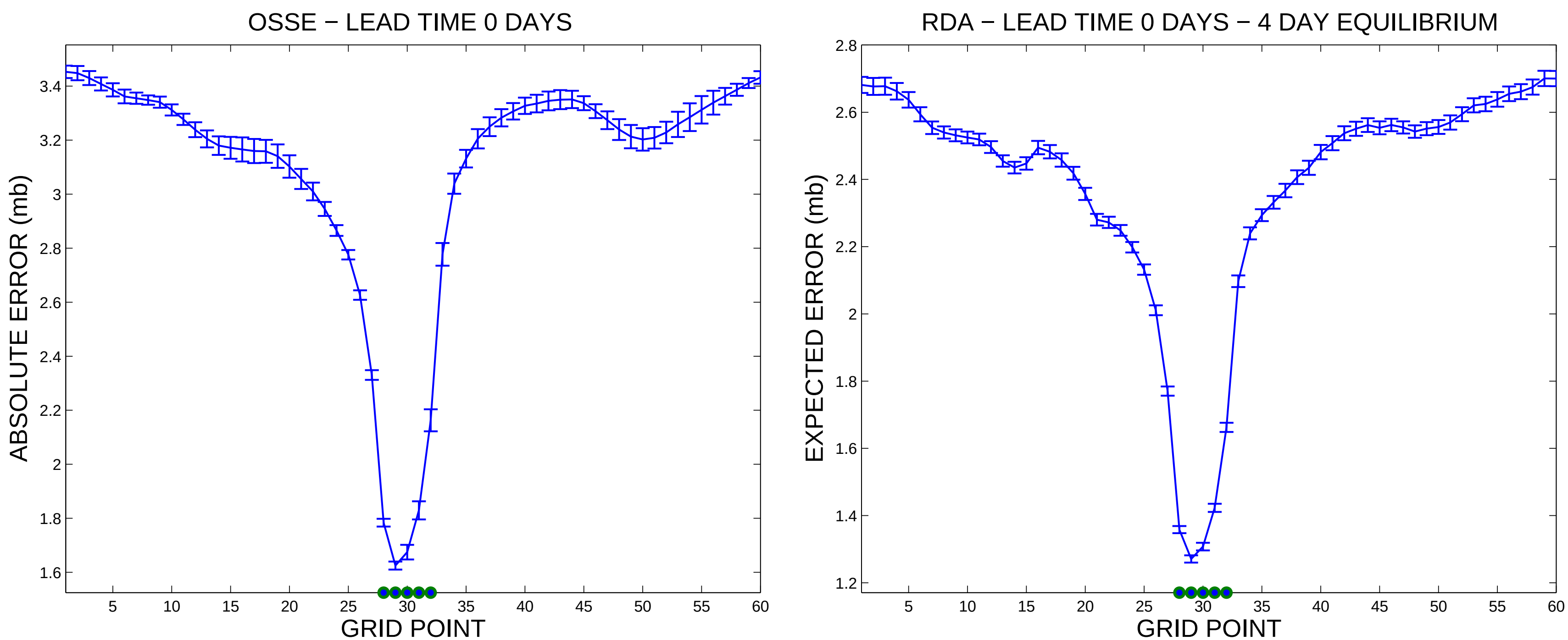
- Strong time mean winds at 33 degrees lat. ensures a strong dynamical signal in this problem

Comparison of Cost Functions I



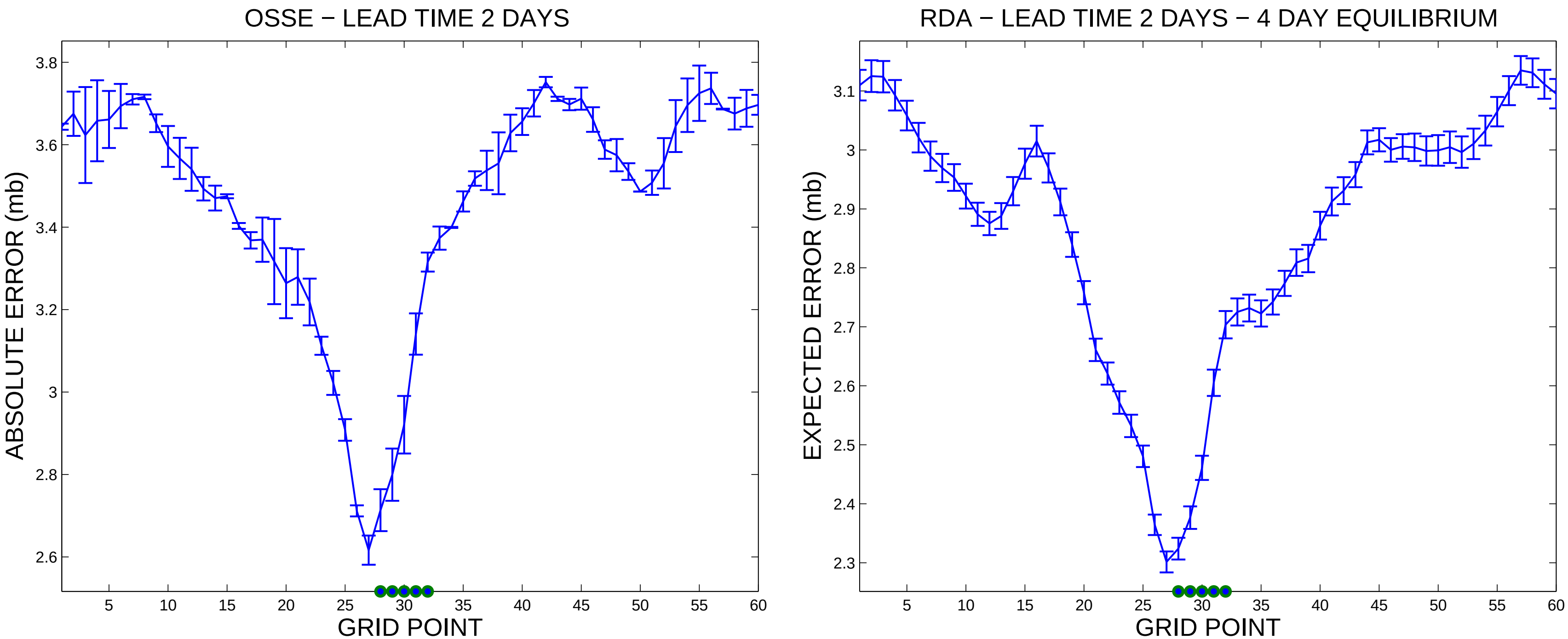
- Verification region - half the latitude band
- Forecast lead time - 0 days
- 1000 independent samples

Comparison of Cost Functions II



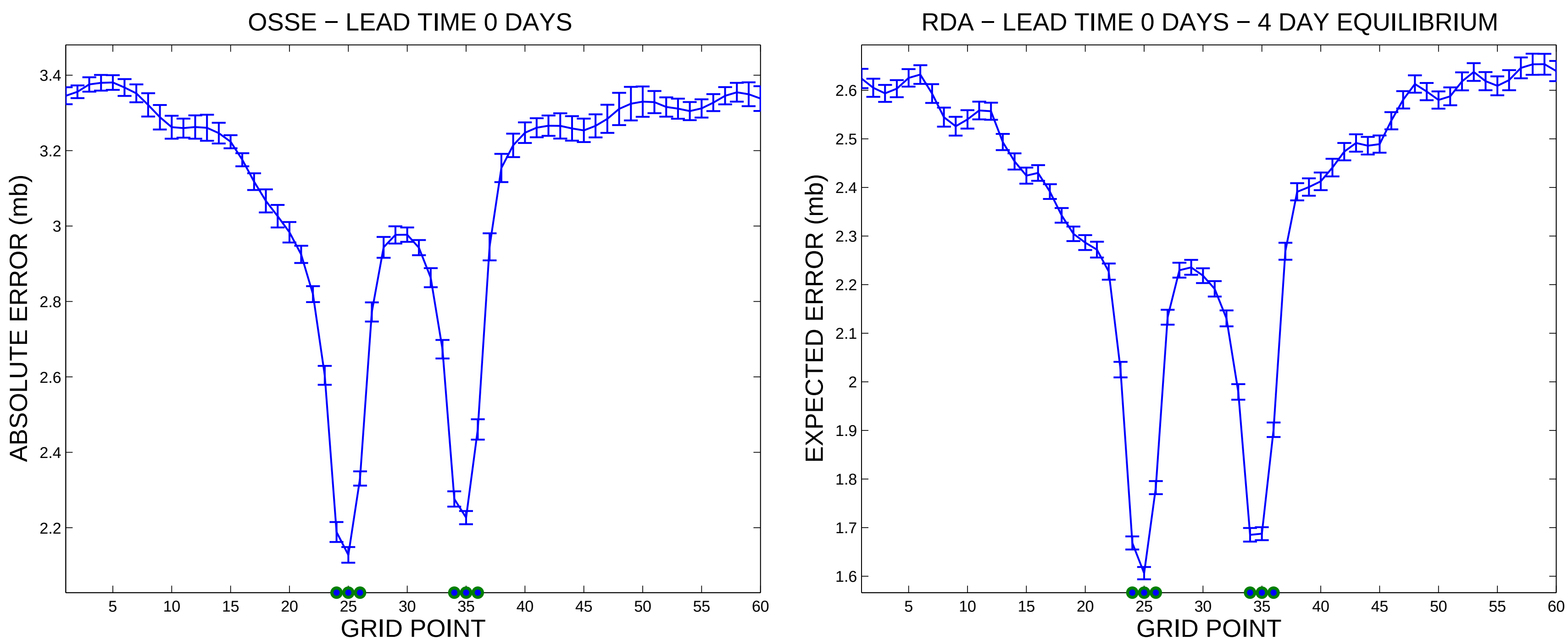
- Verification region - 5 consecutive grid points
- Forecast lead time - 0 days
- 1000 independent samples

Comparison of Cost Functions III



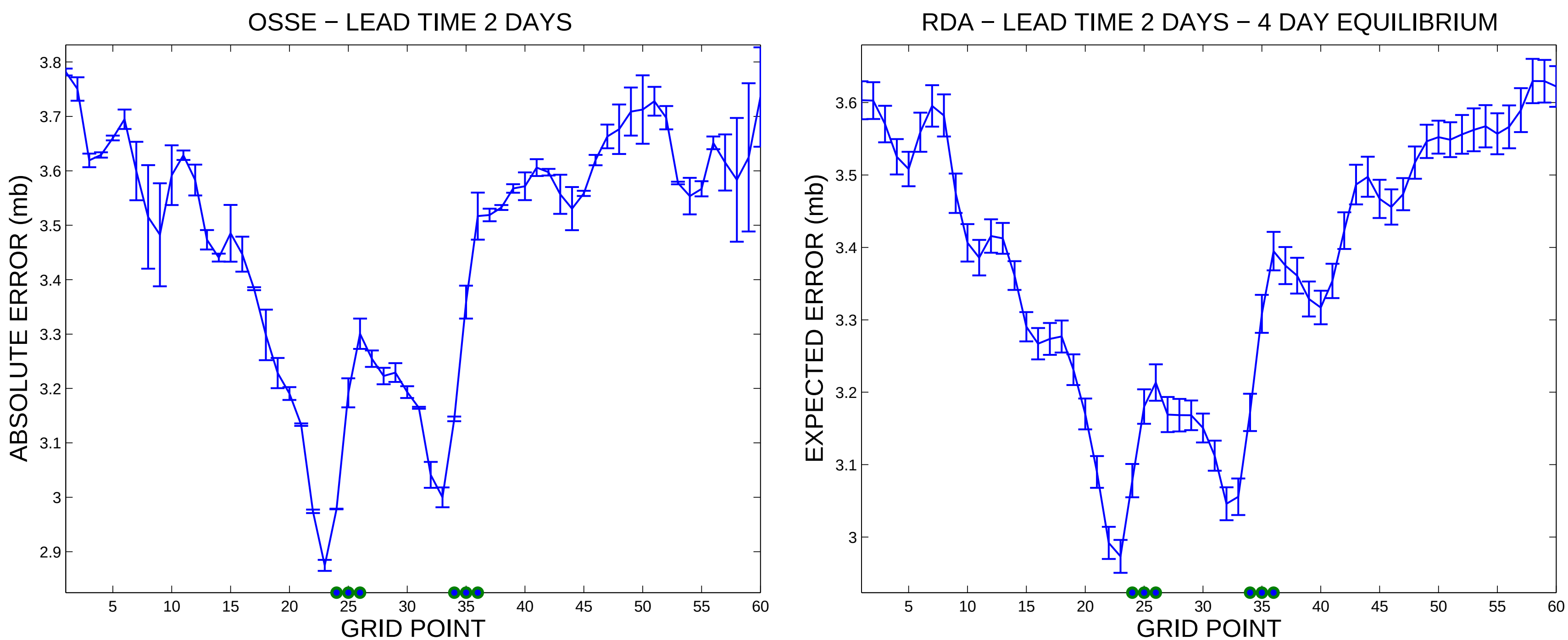
- Verification region - 5 consecutive grid points
- Forecast lead time - 2 days
- 1000 independent samples

Comparison of Cost Functions IV



- Verification region - discontinuous
- Forecast lead time - 0 days
- 1000 independent samples

Comparison of Cost Functions V



- Verification region - discontinuous
- Forecast lead time - 2 days
- 1000 independent samples

Next ...

- 1. Ensemble Filtering at NCAR - Comparisons to the NCEP Operational System of Jan 2003
- 2. Four-Dimensional Localization
- 3. Routine Network Design
- 4. Ensemble Smoothers

Lag k Smoother for State x_o

- $p(x_o|y_o,...,y_k,\mathbf{Y}_o) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)dx_1...dx_k$

Lag k Smoother for State x_o

- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o) dx_1 ... dx_k$
- Start with: $p(x_o|y_o) \stackrel{M}{\rightarrow} p(x_1|y_o, x_o)$

Lag k Smoother for State x_o

- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o) dx_1 ... dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o) = p(x_o|y_o) p(x_1|y_o, x_o)$

Lag k Smoother for State x_o

- $p(x_o|y_o,...,y_k,\mathbf{Y}_o) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o) = p(x_o|y_o)p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o,y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_o,x_1)}{\int(...)dx}$

Lag k Smoother for State x_o

- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o) = p(x_o|y_o)p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o, y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_o,x_1)}{\int(...)dx}$
- $p(x_o, x_1, x_2|y_o, y_1) = p(x_o, x_1|y_o, y_1)p(x_2|y_o, y_1, x_o, x_1)$

Lag k Smoother for State x_o

- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o) = p(x_o|y_o)p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o, y_1) = \frac{p(x_o, x_1|y_o)p(y_1|x_o, x_1)}{\int(...)dx}$
- $p(x_o, x_1, x_2|y_o, y_1) = p(x_o, x_1|y_o, y_1)p(x_2|y_o, y_1, x_o, x_1)$
- $p(x_o, x_1, x_2|y_o, y_1, y_2) = \frac{p(x_o, x_1, x_2|y_o, y_1)p(y_2|x_o, x_1, x_2)}{\int(...)dx}$

Lag k Smoother for State x_o

- $p(x_o|y_o,...,y_k,\mathbf{Y}_o) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o) = p(x_o|y_o)p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o,y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_o,x_1)}{\int(...)dx}$
- $p(x_o,x_1,x_2|y_o,y_1) = p(x_o,x_1|y_o,y_1)p(x_2|y_o,y_1,x_o,x_1)$
- $p(x_o,x_1,x_2|y_o,y_1,y_2) = \frac{p(x_o,x_1,x_2|y_o,y_1)p(y_2|x_o,x_1,x_2)}{\int(...)dx}$
- continue ... $p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)$

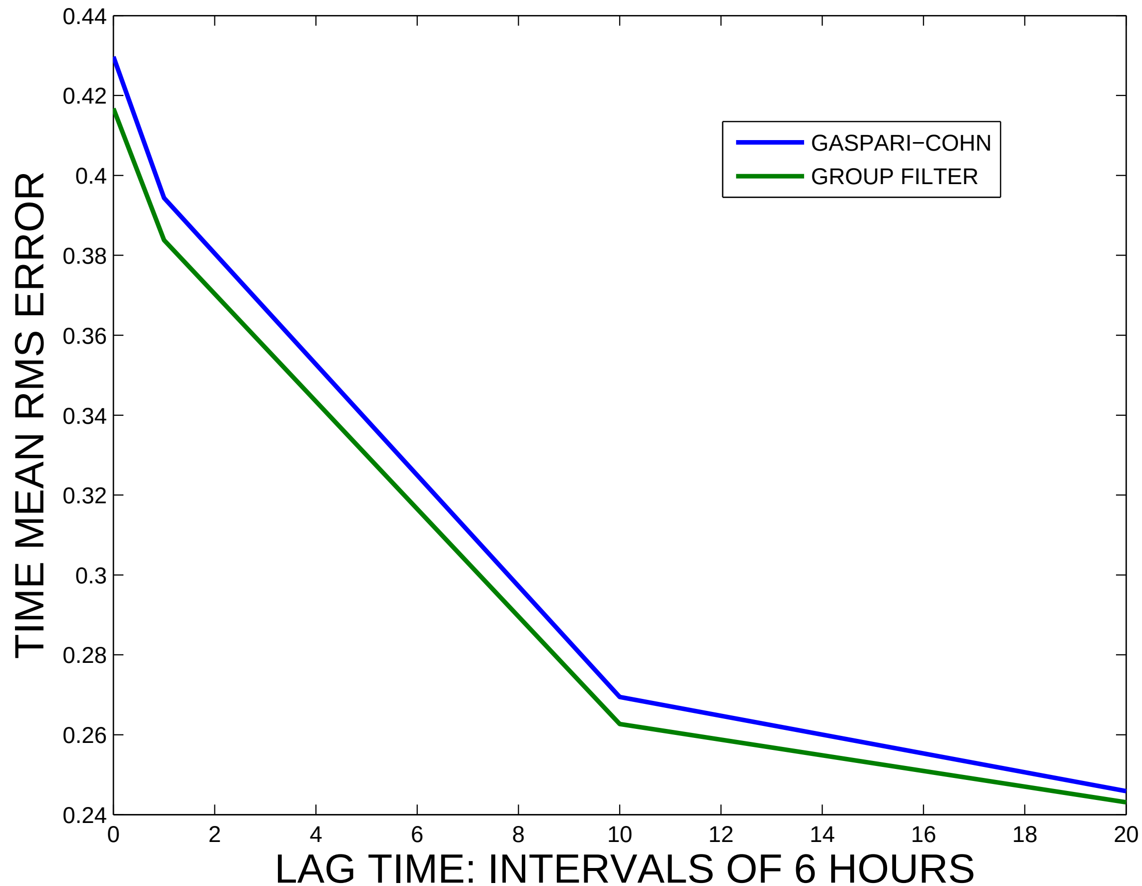
Lag k Smoother for State x_o

- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o) = p(x_o|y_o)p(x_1|y_o, x_o)$
- $p(x_o, x_1|y_o, y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_o,x_1)}{\int(...)dx}$
- $p(x_o, x_1, x_2|y_o, y_1) = p(x_o, x_1|y_o, y_1)p(x_2|y_o, y_1, x_o, x_1)$
- $p(x_o, x_1, x_2|y_o, y_1, y_2) = \frac{p(x_o,x_1,x_2|y_o,y_1)p(y_2|x_o,x_1,x_2)}{\int(...)dx}$
- continue ... $p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o)$
- $p(x_o|y_o, ..., y_k, \mathbf{Y}_o) = \int p(x_o, ..., x_k|y_o, ..., y_k, \mathbf{Y}_o)dx_1...dx_k$

Lag k Smoother for State x_o

- $p(x_o|y_o,...,y_k,\mathbf{Y}_o) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)dx_1...dx_k$
- Start with: $p(x_o|y_o) \xrightarrow{M} p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o) = p(x_o|y_o)p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o,y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_o,x_1)}{\int(...)dx}$
- $p(x_o,x_1,x_2|y_o,y_1) = p(x_o,x_1|y_o,y_1)p(x_2|y_o,y_1,x_o,x_1)$
- $p(x_o,x_1,x_2|y_o,y_1,y_2) = \frac{p(x_o,x_1,x_2|y_o,y_1)p(y_2|x_o,x_1,x_2)}{\int(...)dx}$
- continue ... $p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)$
- $p(x_o|y_o,...,y_k,\mathbf{Y}_o) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}_o)dx_1...dx_k$
- Replace distributions with samples: use EnKF based update method: Extended state space localization

Ensemble Smoother in DART: Lorenz 1996



- $N = 40$, $\gamma = \sqrt{1.04}$, $c = 0.25$, $\mathbf{H} = \mathbf{I}$, $\mathbf{R}_{i,i} = 1.0$