Ensemble Smoothers

Shree P. Khare

khare@ucar.edu

National Center for Atmospheric Research

Background

- BSc, Honours Physics, Univ of British Columbia, May 1999
- Phd, Program in Atmospheric and Oceanic Sciences, Princeton, November 2004
- December 2004 - June 2005, Postdoctoral Fellow at SAMSI, Geophysical Data Assimilation Program
- July 2005 - present, Postdoctoral Fellow in IMAGe, funding from DAReS and GSP

Motivation

- **Data assimilation: Optimal combinations of prediction models and observations**
- **Ensemble data assimilation: Powerful algorithms developed over last 10+ years**
- **Smoothers: Use past and future observations to obtain precise estimates of atmospheric state**
- **Potential application: Re-analysis**
- **Ensemble smoothers literature: Van Leeuwen 2001, Evensen and van Leeuwen 2000**
- **Statistical issues: Sampling errors in realistic model applications**
- **My contribution: Implementation in DART, application to GCM, examination of sampling error impacts**

Talk outline

- 1. Filtering Equations: Notation
- 2. Lag k Smoother Equations
- 3. What's in DART?
- 4. Lorenz 1996 Experiments
- 5. Atmospheric General Circulation Model Experiments

Filtering equations

- $p(x_t|\mathbf{Y}_t)$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$
- $p(x_o|y_o) \overset{M}{\rightarrow} p(x_1|y_o)$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$
- $p(x_o|y_o) \stackrel{M}{\rightarrow} p(x_1|y_o)$
- $p(x_1|y_o, y_1) = \frac{p(x_1|y_o)p(y_1|x_1)}{p(y_1|y_o)}$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$
- $p(x_o|y_o) \stackrel{M}{\rightarrow} p(x_1|y_o)$
- $p(x_1|y_o, y_1) = \frac{p(x_1|y_o)p(y_1|x_1)}{p(y_1|y_o)}$
- continue ...

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$
- $p(x_o|y_o) \stackrel{M}{\rightarrow} p(x_1|y_o)$
- $p(x_1|y_o, y_1) = \frac{p(x_1|y_o)p(y_1|x_1)}{p(y_1|y_o)}$
- continue ...
- $p(x_t|\mathbf{Y}_t)$

Filtering equations

- $p(x_t|\mathbf{Y}_t)$
- Start with: $p(x_o)$
- $p(x_o|y_o) = \frac{p(x_o|y_o)p(y_o|x_o)}{p(y_o)}$
- $p(x_o|y_o) \overset{M}{\rightarrow} p(x_1|y_o)$
- $p(x_1|y_o, y_1) = \frac{p(x_1|y_o)p(y_1|x_1)}{p(y_1|y_o)}$
- continue ...
- $p(x_t|\mathbf{Y}_t)$
- Replace distributions with samples: use EnKF based update method: State space localization

Next ...

- 1. Filtering Equations: Notation
- 2. Lag k Smoother Equations
- 3. What's in DART?
- 4. Lorenz 1996 Experiments
- 5. Atmospheric General Circulation Model Experiments

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

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

Lag k smoother for state x_o

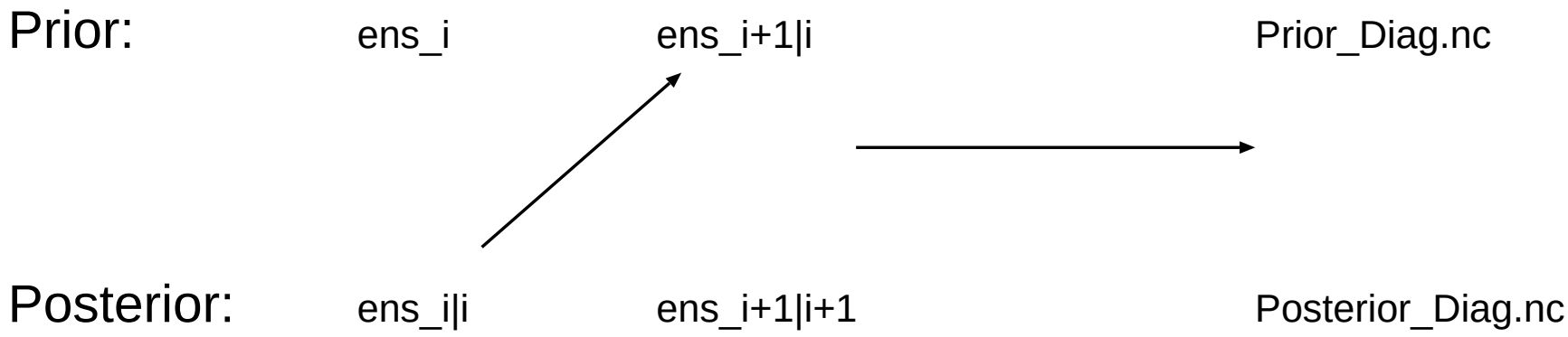
- $p(x_o|y_o,...,y_k,\mathbf{Y}^-) = \int p(x_o,...,x_k|y_o,...,y_k,\mathbf{Y}^-)dx_1...dx_k$
- Start with: $p(x_o|y_o) \stackrel{M}{\rightarrow} p(x_1|y_o)$
- $p(x_o,x_1|y_o) = p(x_o|y_o)p(x_1|y_o,x_o)$
- $p(x_o,x_1|y_o,y_1) = \frac{p(x_o,x_1|y_o)p(y_1|x_1)}{p(y_1|y_o)}$
- $p(x_o,x_1,x_2|y_o,y_1) =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

Next ...

- 1. Filtering Equations: Notation
- 2. Lag k Smoother Equations
- 3. What's in DART?
- 4. Lorenz 1996 Experiments
- 5. Atmospheric General Circulation Model Experiments

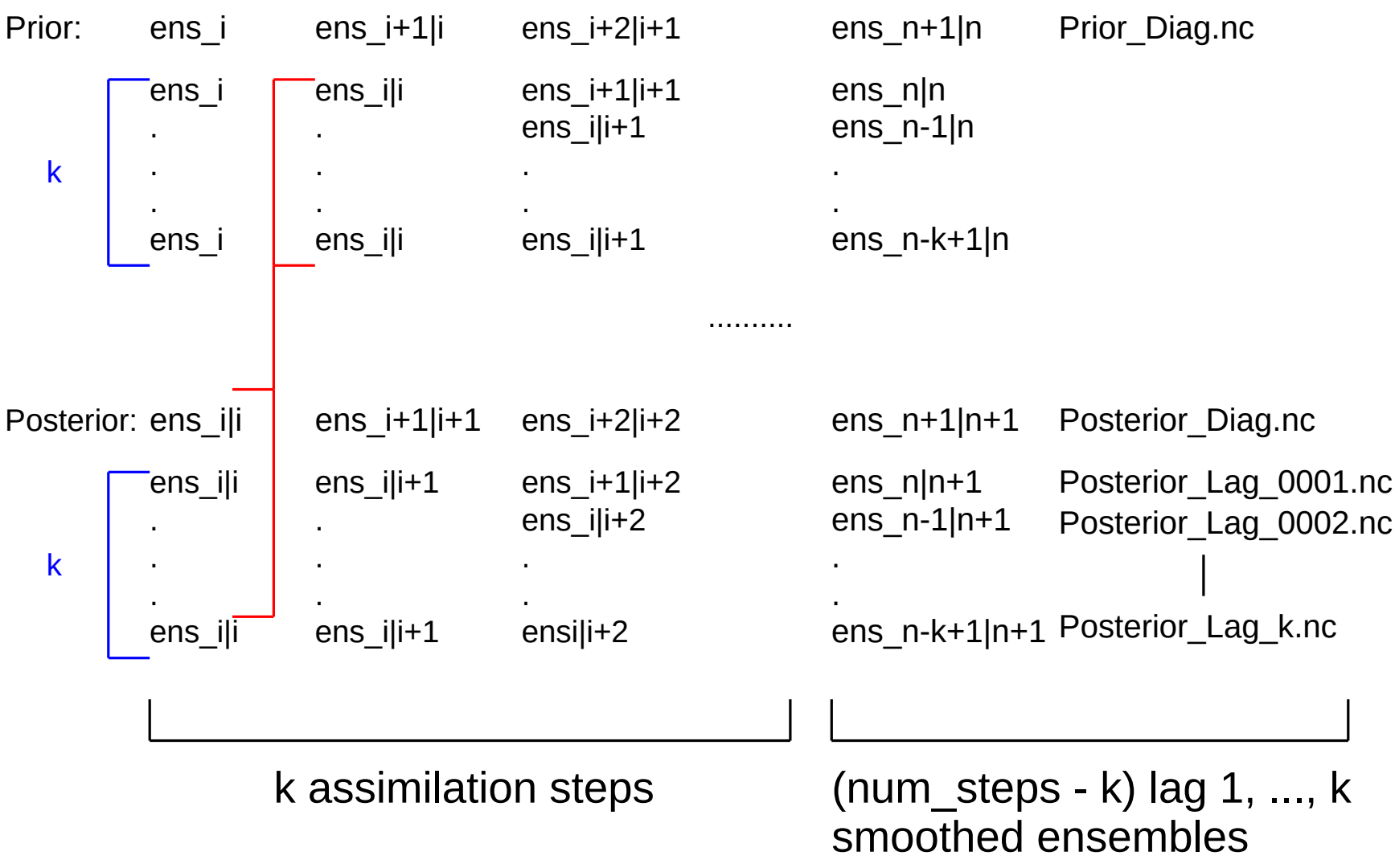
Filtering

Filtering: ens_handle(1)



Smoother

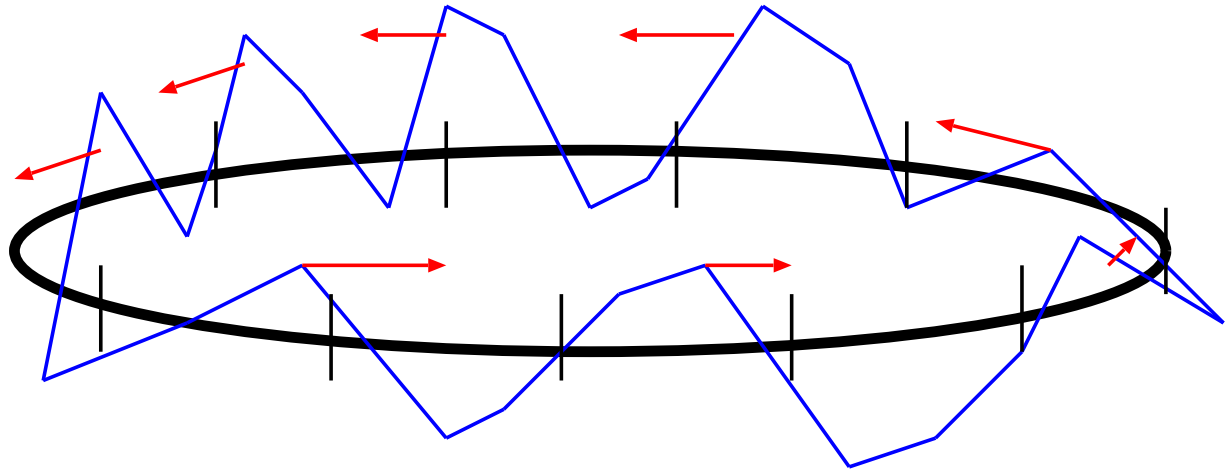
Lag k Smoother: ens_handle(1 + k)



Next ...

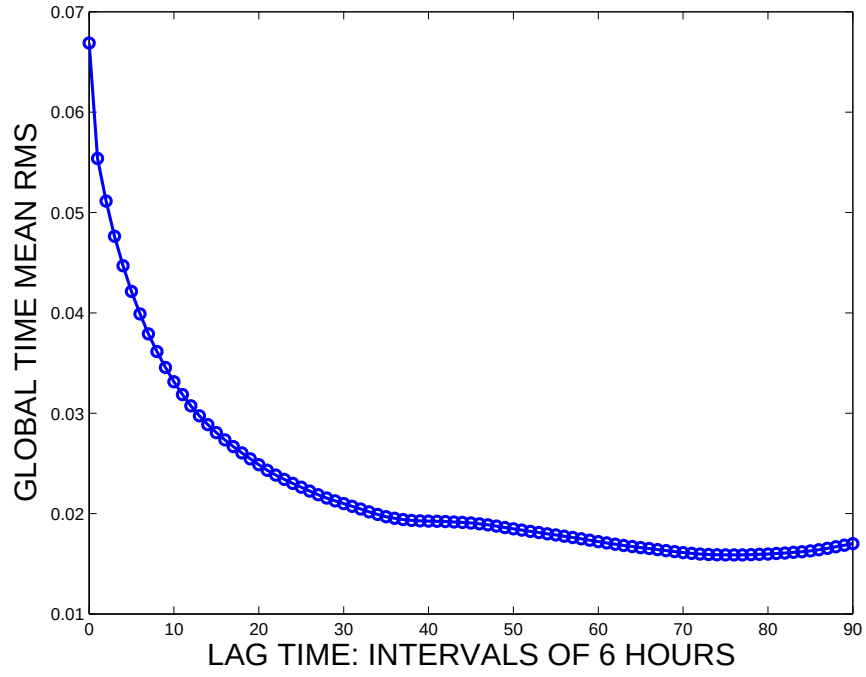
- 1. Filtering Equations: Notation
- 2. Lag k Smoother Equations
- 3. What's in DART?
- 4. Lorenz 1996 Experiments
- 5. Atmospheric General Circulation Model Experiments

Experiments in an atmospheric ‘toy model’



$$\frac{dx_j}{dt} = -x_{j-2}x_{j-1} + x_{j-1}x_{j+1} - x_j + F \; j = 1, ..., 40$$

Ensemble smoother in DART: Lorenz 1996

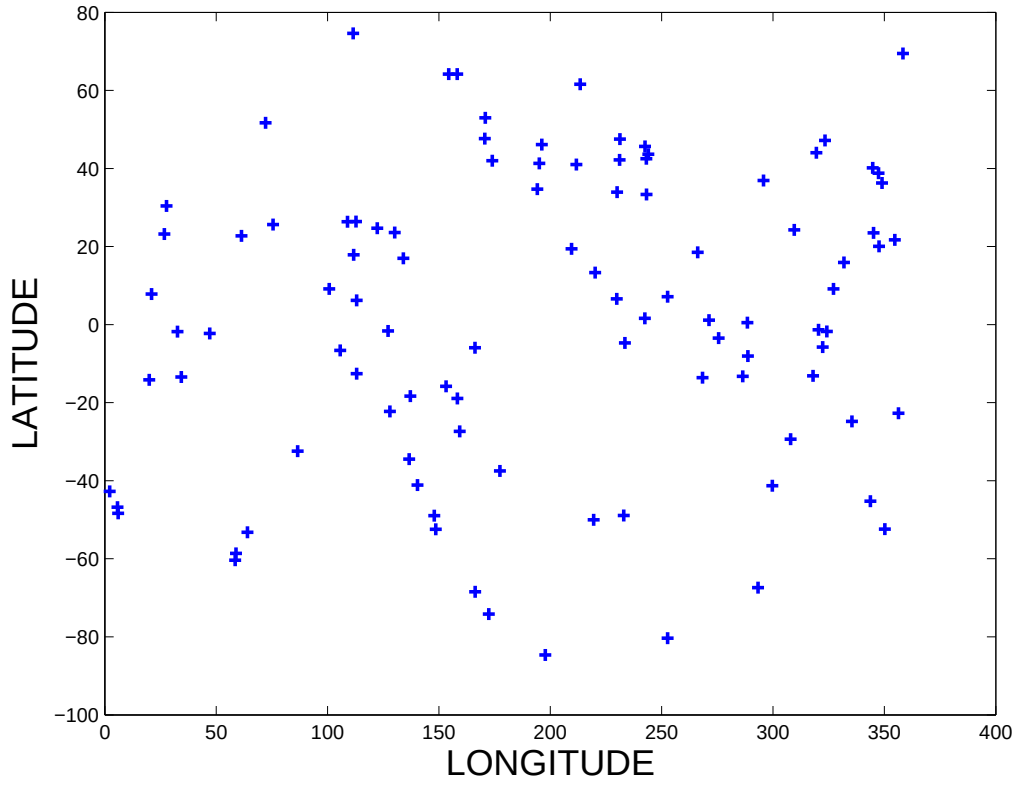


- $N = 200$, No inflation/localization, $\mathbf{H} = \mathbf{I}$, $\mathbf{R}_{i,i} = (0.36)^2$
- Research question: What are the impacts of realistic ensemble sizes?

Next ...

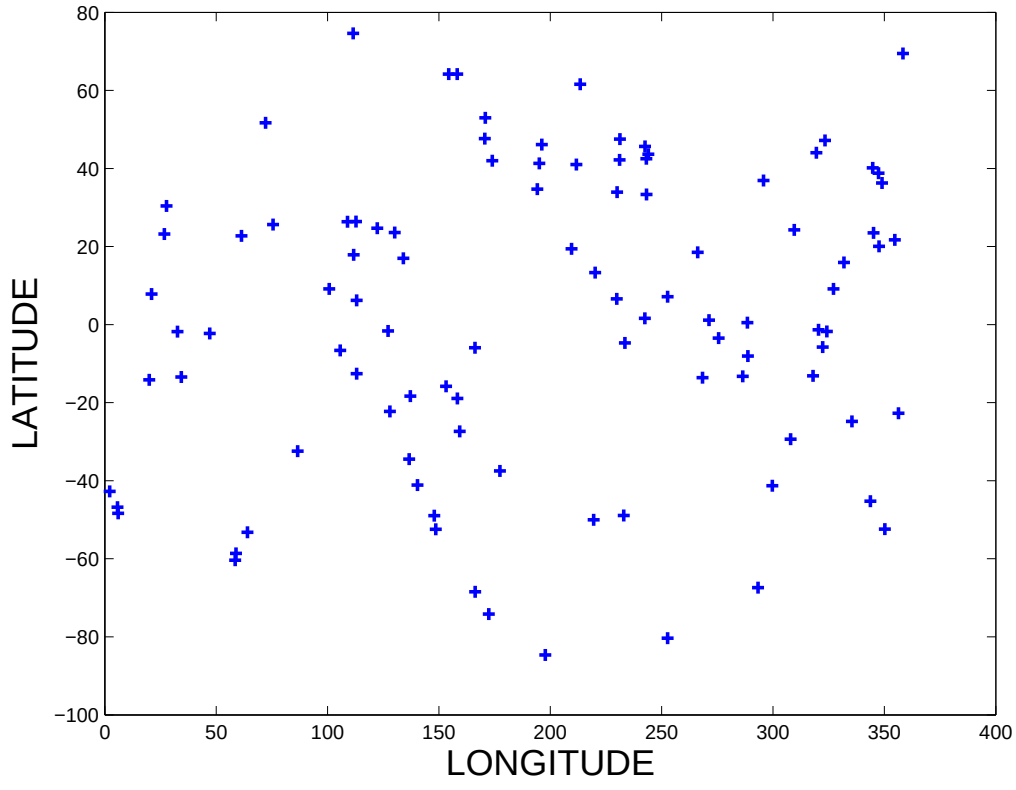
- 1. Filtering Equations: Notation
- 2. Lag k Smoother Equations
- 3. What's in DART?
- 4. Lorenz 1996 Experiments
- 5. Atmospheric General Circulation Model Experiments

Experiments in a simplified GCM



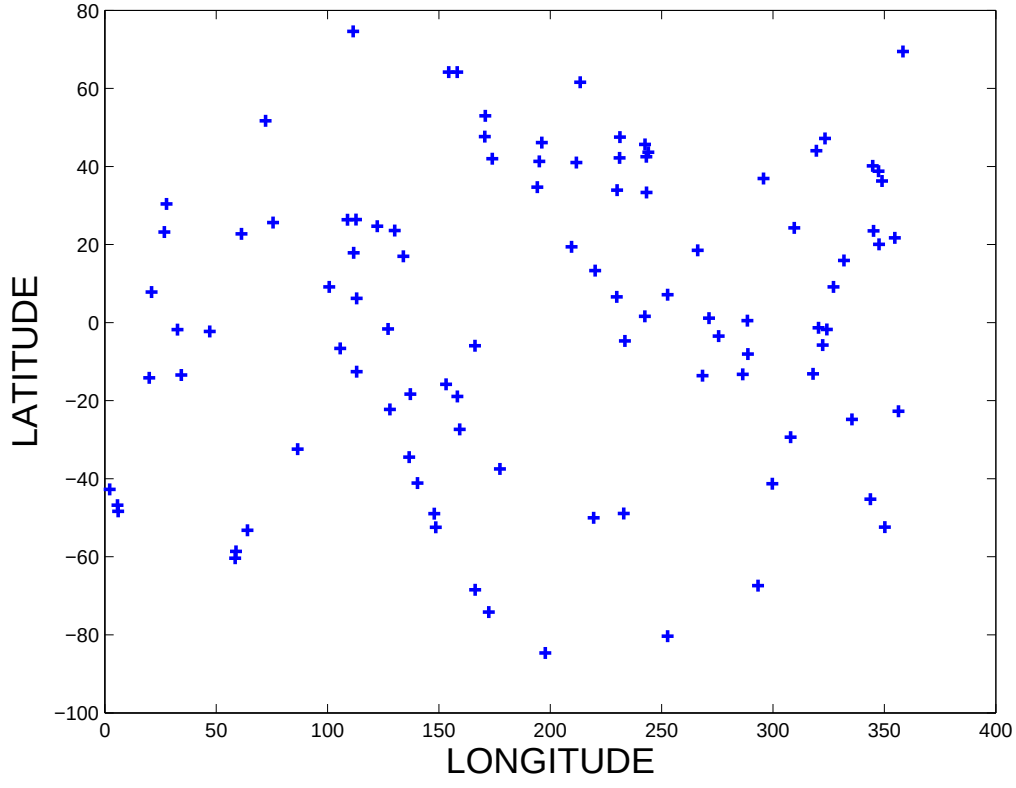
- Simulated column (radiosonde) observations - assimilate every 12 hours - PS, T and winds

Experiments in a simplified GCM



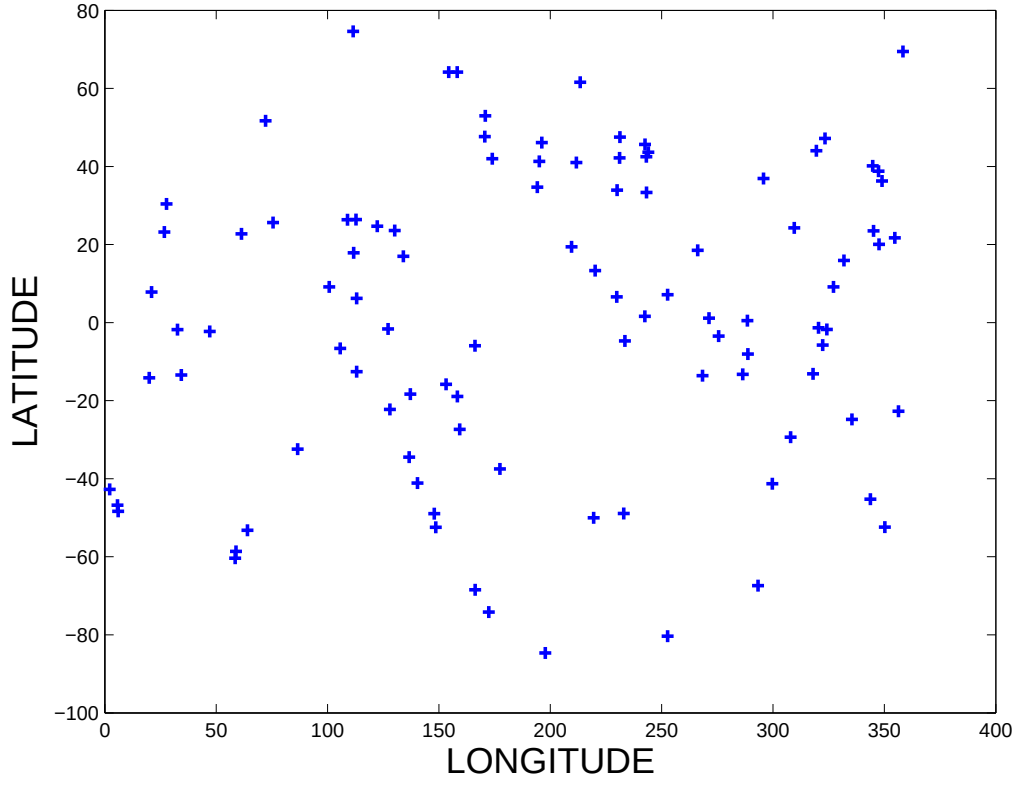
- Run an EAKF with $N = 40$ ensemble members (with localization and no inflation) in a Held-Suarez configuration of an AGCM

Experiments in a simplified GCM



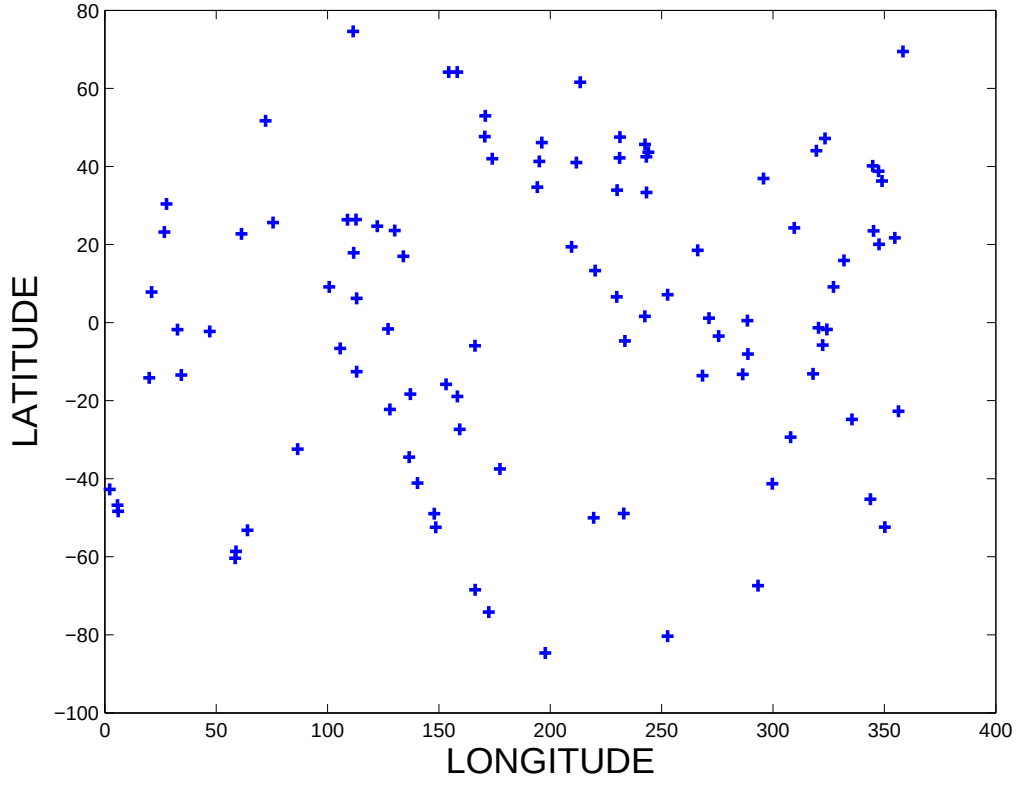
- Forcing - Newtonian cooling, Damping - Rayleigh Friction

Experiments in a simplified GCM



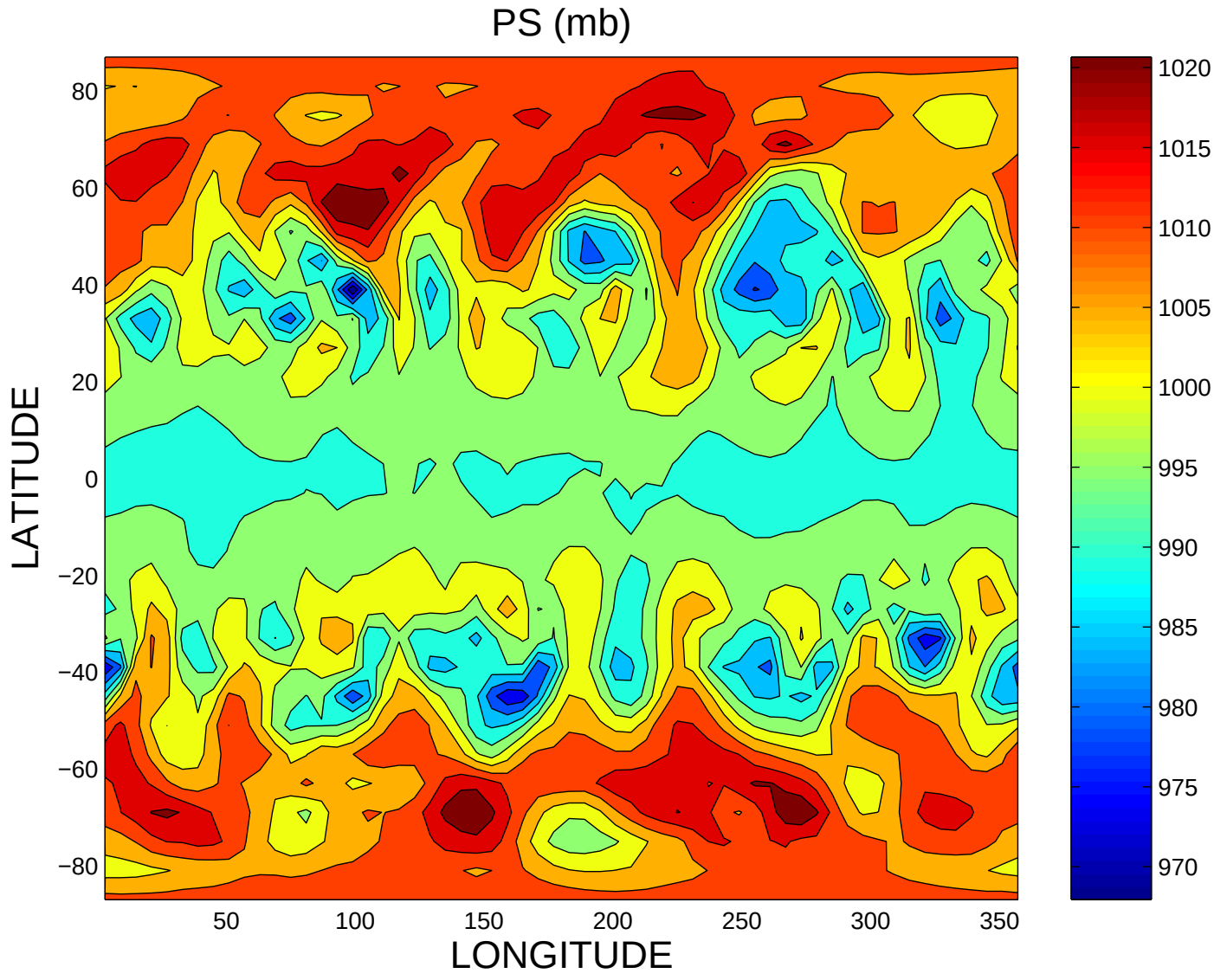
- 6 degrees horizontal resolution (60×30) - 5 vertical levels

Experiments in a simplified GCM

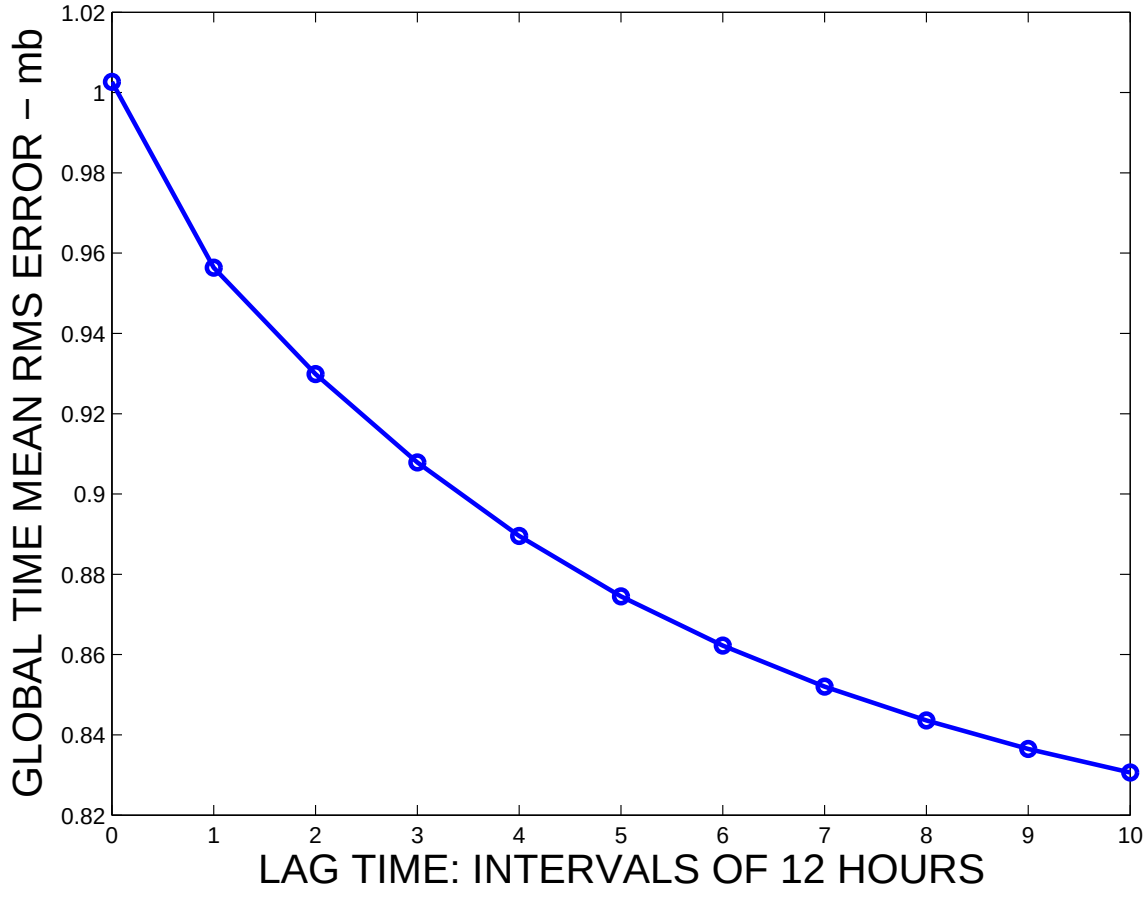


- Temperature gradient drives a baroclinically unstable flow in the mid-latitudes

Atmospheric models: Surface pressure (PS)



Ensemble Smoother in DART: Simplified GCM



- Research questions: Sampling errors? Spatial variation? Optimal time-scales?