Atmospheric Data Assimilation

Ricardo Todling

NASA/Data Assimilation Office/SAIC

2002 International ACES Meeting

Maui, 5-10 May 2002

Contributions & Discussions: S.E. Cohn, A. da Silva, D. Dee, J. Guo,
L.P. Riishojgaard, B. Zhang & Y. Zhu

http://dao.gsfc.nasa.gov/DAO_people/todling

E-mail: todling@dao.gsfc.nasa.gov

OUTLINE

- Introduction
- Simple Illustrations of Data Assimilation Procedures
  - Chaotic Dynamics
  - Shallow-water Dynamics
- Unique Features of the DAO Data Assimilation System
  - Adaptive Quality Control
  - Model Bias Estimation and Correction
  - Estimation of Analysis Errors
  - Retrospective Data Assimilation
- Lessons Learned So Far

Introduction

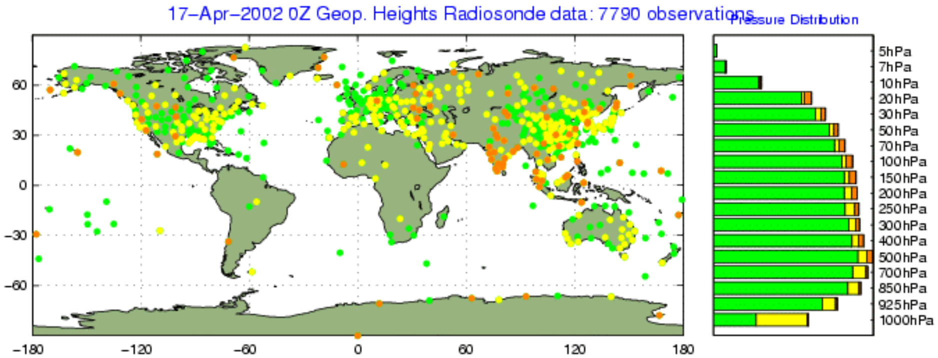
- **Objective: generate best possible consistent global dataset for weather prediction and climate research.**

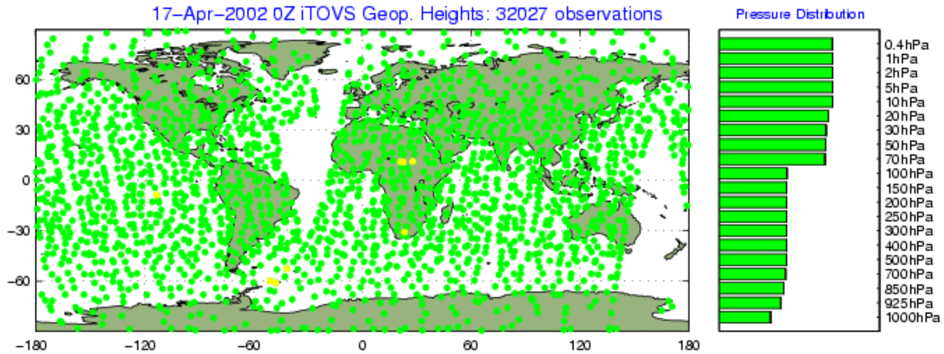
- **How: by combining observations of the atmosphere with general circulation model predictions and obtaining global estimate of the state of the atmosphere at any given time.**

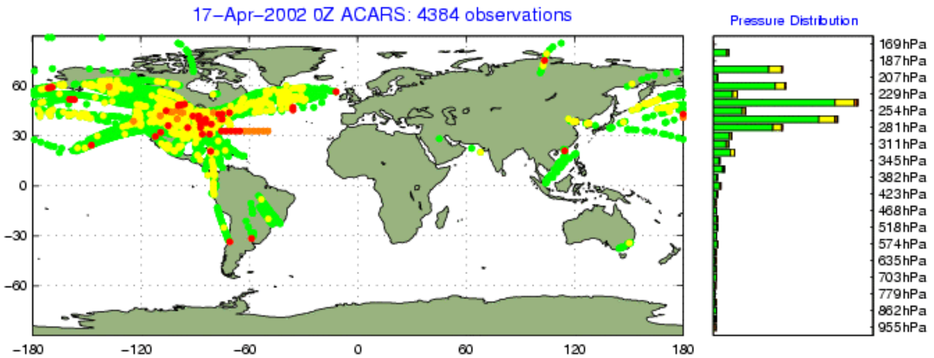
- **Observations:**
  noisy
  sparse in space
  available at a variety to time frequencies
  of order 1,000,000/6h but shrunk down to order 100,000/6h
    (the number per 6 h is rapidly increasing)

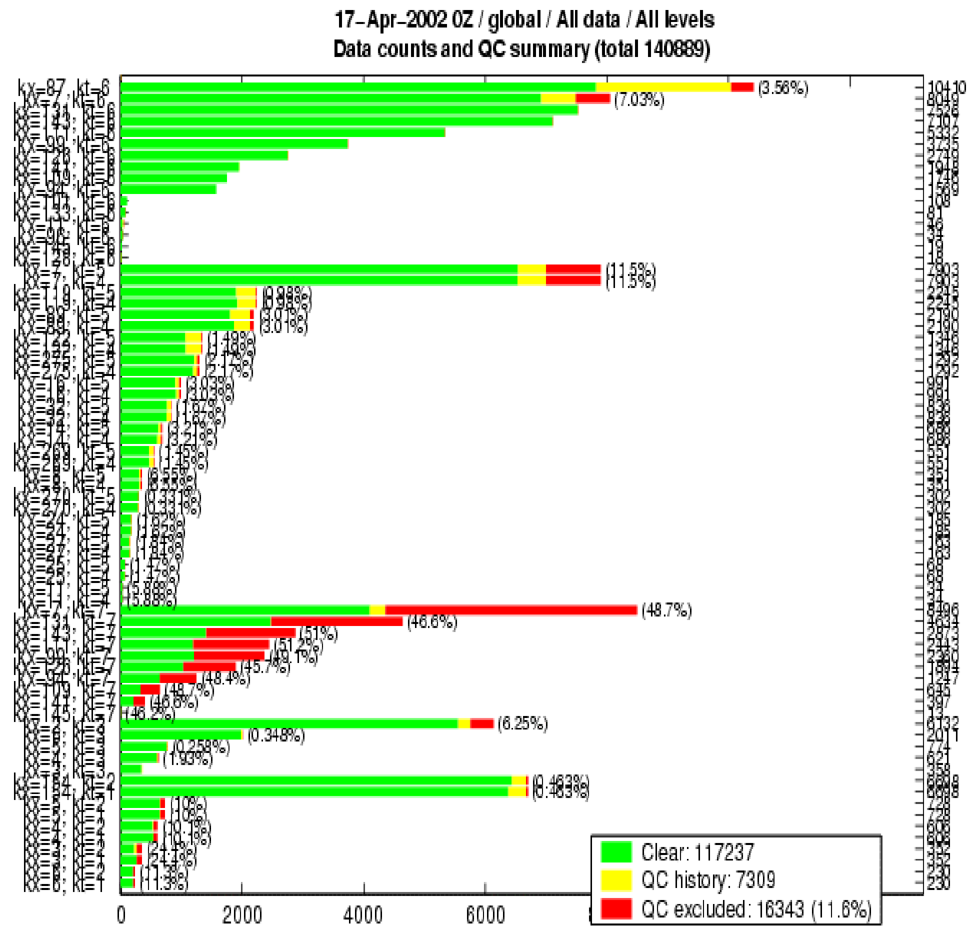
- **Models:**
  imperfect
  of order 10,000,000 degrees of freedom

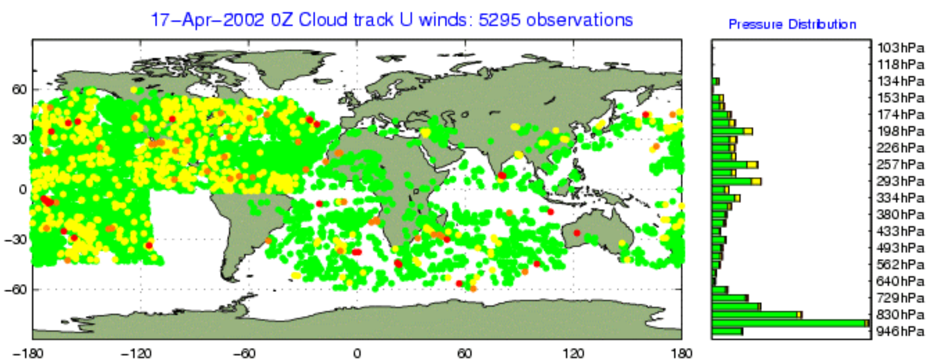
A few observing systems for 00GMT 17 April 2002 and total observation count











Stochastic System Formulation

True n–dimensional atmosphere

$$\mathbf{w}_k^t = \mathbf{f}(\mathbf{w}_{k-1}^t) + \mathbf{b}_k^t$$

Model error statistics:

$$\begin{aligned} \mathcal{E}\{\mathbf{b}_k^t\} &= \mathbf{0} \\ \mathcal{E}\{\mathbf{b}_k^t(\mathbf{b}_{k'}^t)^T\} &= \mathbf{Q}_k\delta_{kk'} \end{aligned}$$

The p-dimensional observation process:

$$\mathbf{w}_k^o = \mathbf{h}(\mathbf{w}_k^t) + \mathbf{b}_k^o$$

Observation error statistics:

$$\begin{aligned} \mathcal{E}\{\mathbf{b}_k^o\} &= \mathbf{0} \\ \mathcal{E}\{\mathbf{b}_k^o(\mathbf{b}_{k'}^o)^T\} &= \mathbf{R}_k\delta_{kk'} \end{aligned}$$

Uncorrelated model and observation errors:

$$\mathcal{E}\{\mathbf{b}_k^t(\mathbf{b}_{k'}^o)^T\} = \mathbf{0}$$

Kalman filtering and smoothing techniques can be invoked to solve the problem.

In practice …

- Errors are in many cases non-Gaussian

- System's dimensionality is too high to allow calculation of error covariance propagation

- Statistics of observational errors are poorly known

- Statistics of model error (system noise) is unknown

- Dynamics and observations nonlinearities can be quite complex

What do we do?

Design reliable approximations and feasible algorithms

Approximations to the Filtering Problem

- Error covariance modeling/parameterization
- Simplified dynamics
- Reduced resolution (and reduced rank)
- Limiting filtering (including representers)
- Ensemble methods (including breeding)

Two Illustrations for Simple Dynamics

1. Data assimilation for chaotic dynamics
2. Different KF approximations for SW dynamics

The Physical-space Statistical Analysis System (PSAS)

At any given time, PSAS calculates analysis as

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{K}(\mathbf{w}^o - \mathbf{H}\mathbf{w}^f)$$

where the gain matrix

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

is never explicit. Furthermore, the forecast error covariance matrix is parameterized and also never exists as a matrix.

PSAS avoids explicit inversion of the innovation cov matrix by using a global conjugate gradient method to determine $\mathbf{x}$ as in

$$\left(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)\mathbf{x} = (\mathbf{w}^o - \mathbf{H}\mathbf{w}^f)$$

So the analysis becomes

$$\mathbf{w}^a = \mathbf{w}^f + \mathbf{P}^f\mathbf{H}^T\mathbf{x}$$

Advances in the DAO's Assimilation System

- **Adaptive quality control**
- **Bias estimation and correction** ←
- **Estimation of analysis errors**
- **Retrospective analysis and assimilation** ←

We are heading toward implementing an <u>adaptive parameterized Kalman filter</u>, and <u>iterated Kalman smoother</u>, for performing atmospheric data assimilation.

Bias Estimation and Correction

In the bias estimation and correction approach bias estimates are obtained through

$$\mathbf{b}^a = \mathbf{b}^f + \mathbf{L}\left[\mathbf{w}^o - \mathbf{H}(\mathbf{w}^f - \mathbf{b}^f)\right]$$

with the weighting matrix $\mathbf{L}$ given by

$$\mathbf{L} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

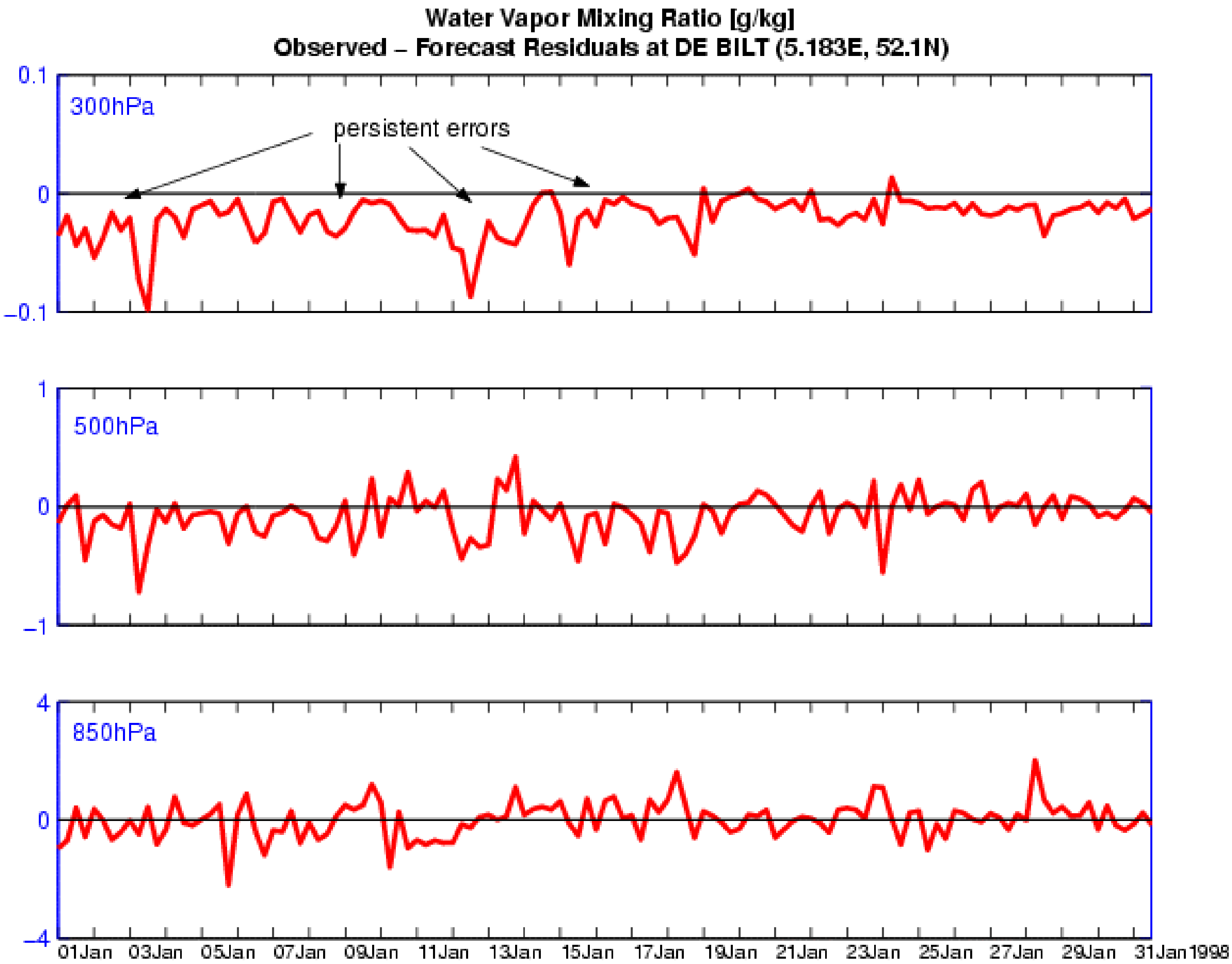
And the bias corrected analysis is obtained by solving

$$\mathbf{w}^a = (\mathbf{w}^f - \mathbf{b}^a) + \tilde{\mathbf{K}}\left[\mathbf{w}^o - \mathbf{H}(\mathbf{w}^f - \mathbf{b}^a)\right]$$

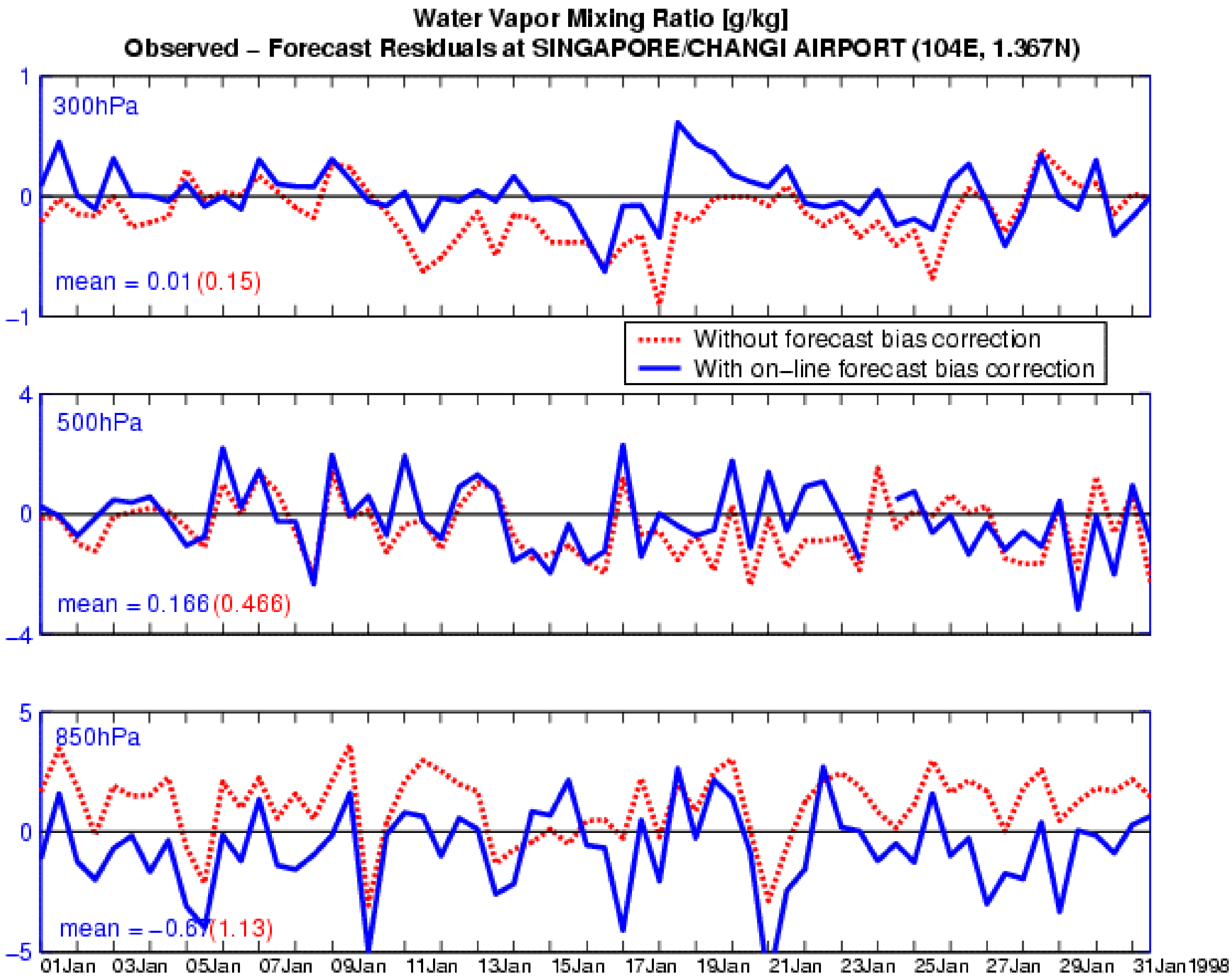
- ▷ remarkably the best choice for $\tilde{\mathbf{K}}$ is $\tilde{\mathbf{K}} = \mathbf{K}$

- ▷ analysis $\mathbf{w}^a$ are unbiased regardless of bias forecast error covariance $\mathbf{P}^b$

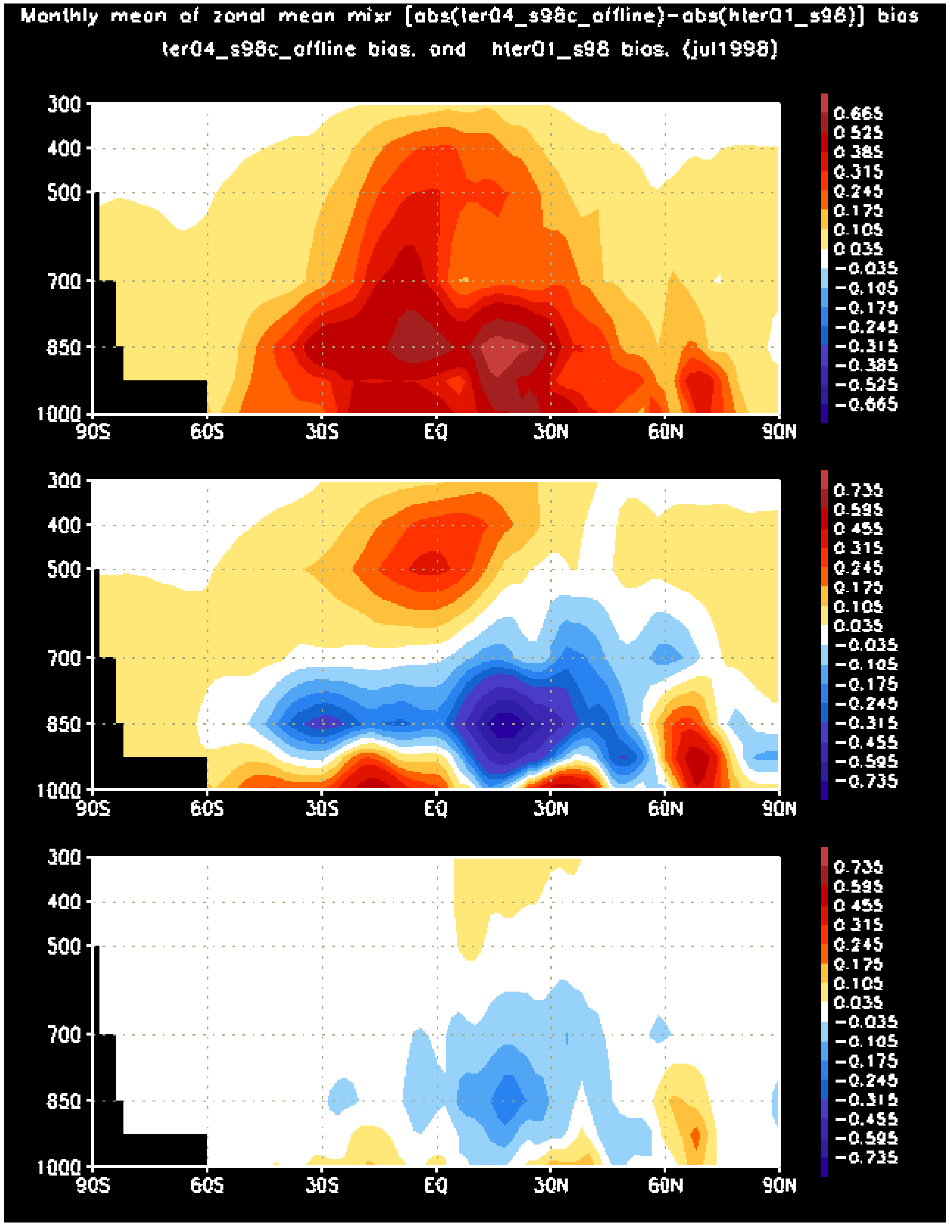
Bias Estimation and Correction (cont.)

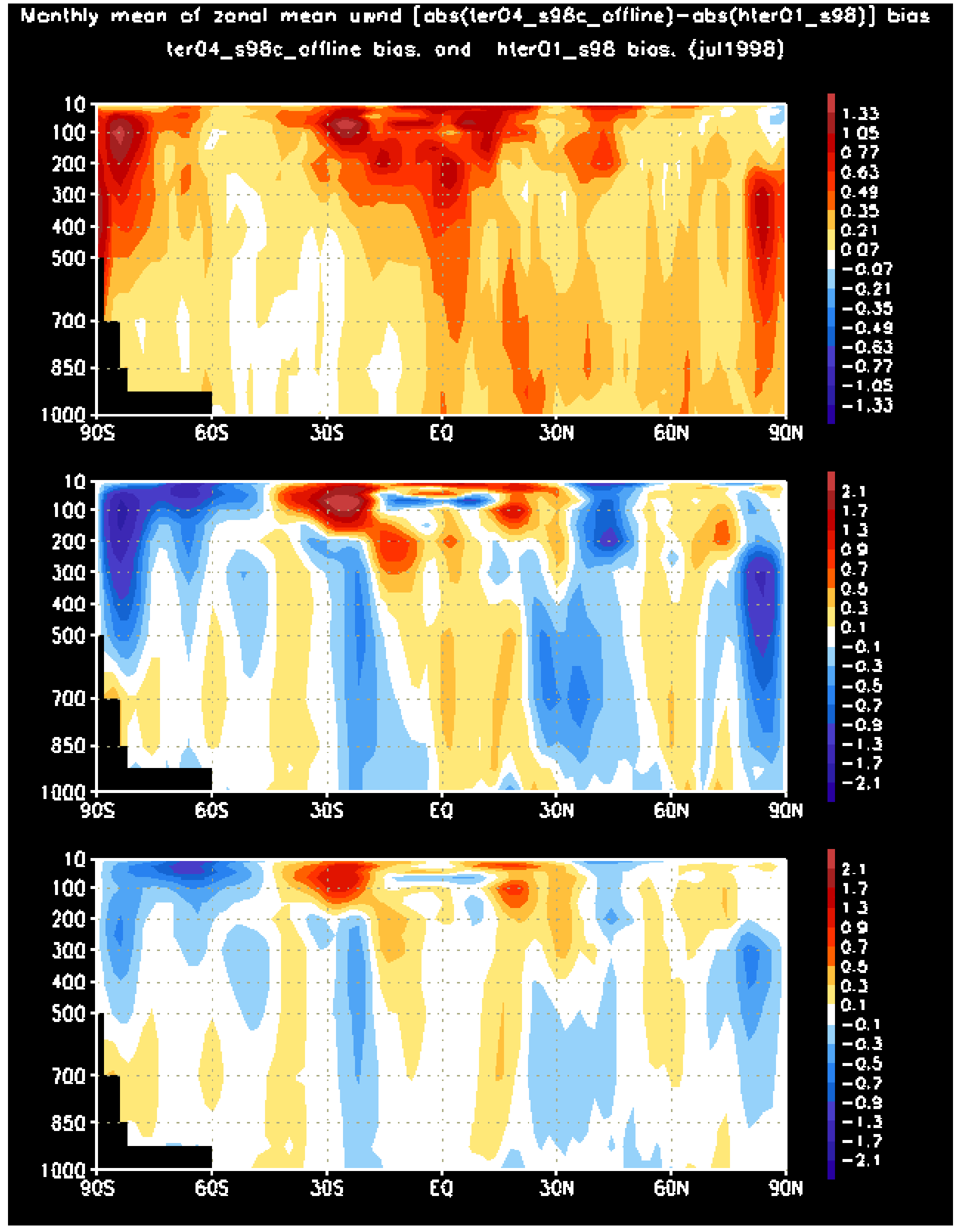


Bias Estimation and Correction (cont.)



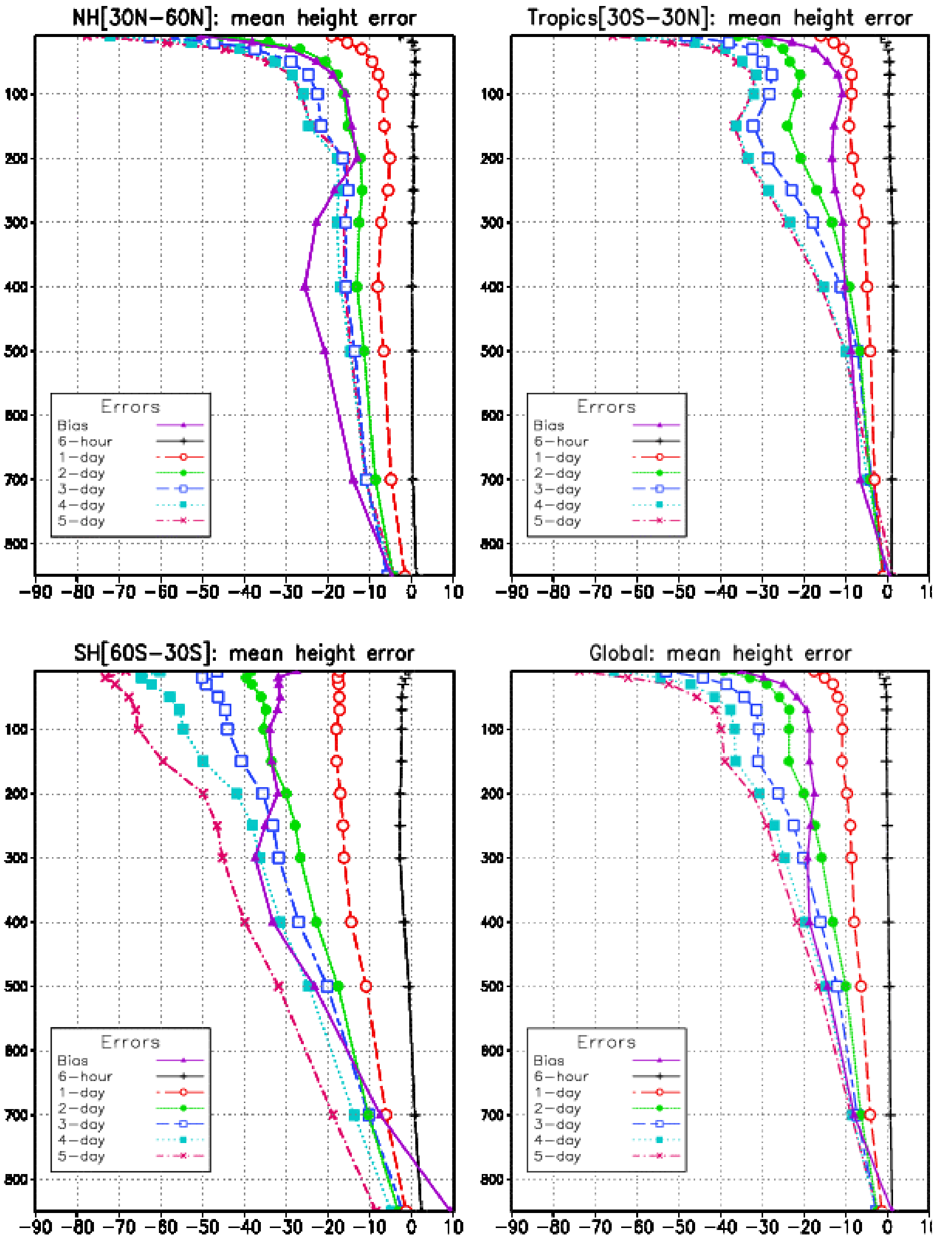
Monthly mean of zonal mean WV mix-ratio **Monthly mean of zonal mean of zonal wind**

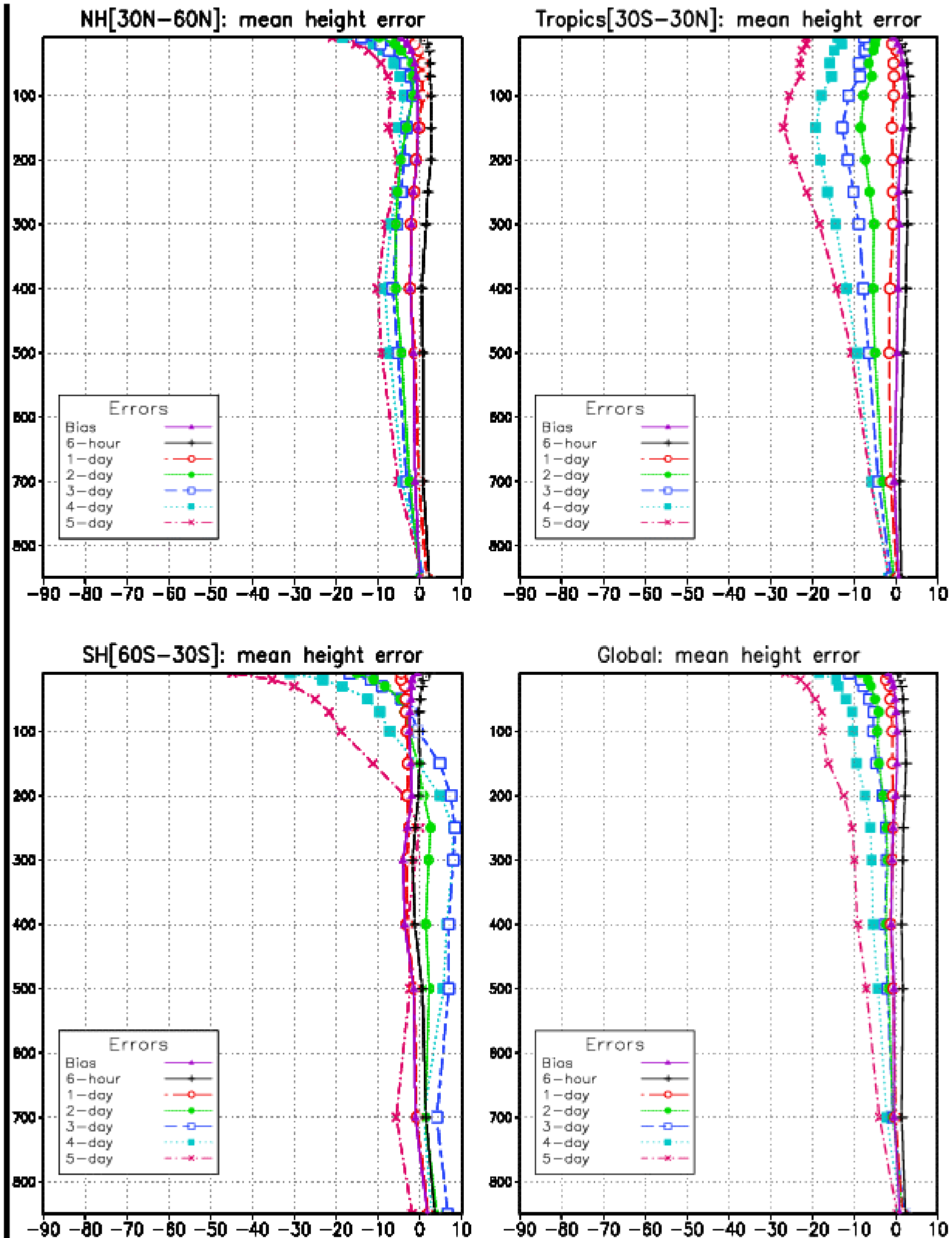




Using Bias Estimates to Improve Forecasts

Errors from regular forecast **geopotential height** *Errors from forced forecast*





Retrospective Data Assimilation

Goal

Use data available over the interval (t,t+T) to improve analysis at each time t

Approach

Fixed-lag Kalman smoother via state augmentation

Remarks

Works well when applied to a SW mode

Requires adjoint of GCM and PSAS solver

Lag-ℓ Retrospective Data Assimilation Algorithm

for $k = 1, 2, \ldots$

- $\mathbf{w}^f_{k|k-1} = f(\mathbf{w}^a_{k-1|k-1})$
- $\mathbf{v}_{k|k-1} = \mathbf{w}^o_k - h(\mathbf{w}^f_{k|k-1})$
- $\mathbf{\Gamma}_k = \mathbf{H}_k \mathbf{P}^f \mathbf{H}_k^T + \mathbf{R}_k$
- Solve PSAS: $\mathbf{\Gamma}_k \mathbf{x} = \mathbf{v}_{k|k-1}$
- $\mathbf{y}^0 = \mathbf{H}_k^T \mathbf{x}$
- Analysis Increment: $\delta \mathbf{w}^a_{k|k} = \mathbf{P}^f \mathbf{y}^0$
- **for** $\ell = 1, 2, \ldots,$ **min**(k, L)
  - Adjoint Integration: $\mathbf{y}^\ell = \mathbf{A}^T_{k-\ell|k-\ell} \mathbf{y}^{\ell-1}$
  - Solve PSAS: $\mathbf{\Gamma}_{k-\ell} \mathbf{x} = \mathbf{H}_{k-\ell} \mathbf{P}^f \mathbf{y}^\ell$
  - $\mathbf{y}^\ell := \mathbf{y}^\ell - \mathbf{H}^T_{k-\ell} \mathbf{x}$
  - Retro–Analysis Increment: $\delta \mathbf{w}^a_{k-\ell|k} = \mathbf{P}^f \mathbf{y}^\ell$
- **endfor**

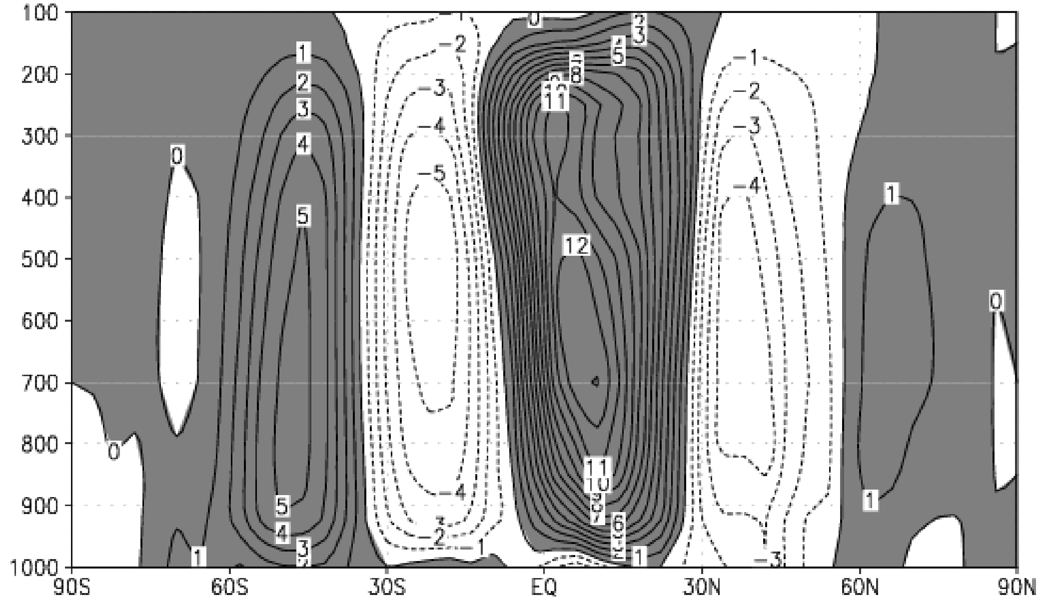
endfor

Retrospective Data Assimilation System

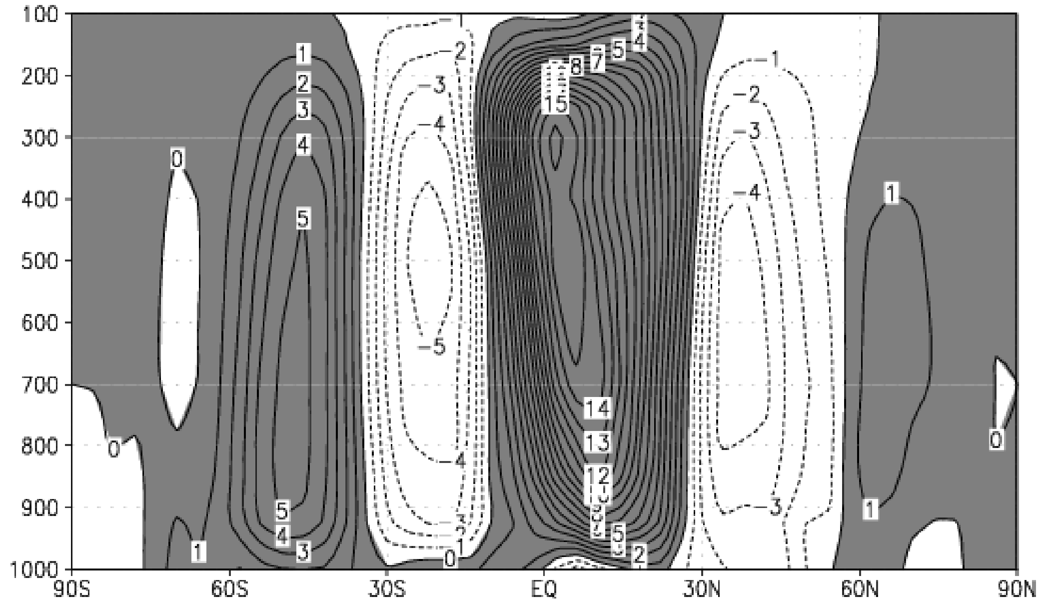
- If the accuracy of retrospective analyses are not sought, the RDAS refers only to the error co-variances normally used in the filter (PSAS).
- In an FLKS-based RDAS future observations are used to correct filter/smoother analyses impaired by lack of observations at earlier analyses times.
- One of the equations solved by the FLKS-based RDAS involves exactly the same operators required to calculate the sensitivity of forecasts to observation changes.
- A simplified version of the FLKS-based RDAS used here can be implemented where the adjoint of the dynamics is replaced by the identity operator. This approximate scheme can be statistically justified in the context of PSAS.

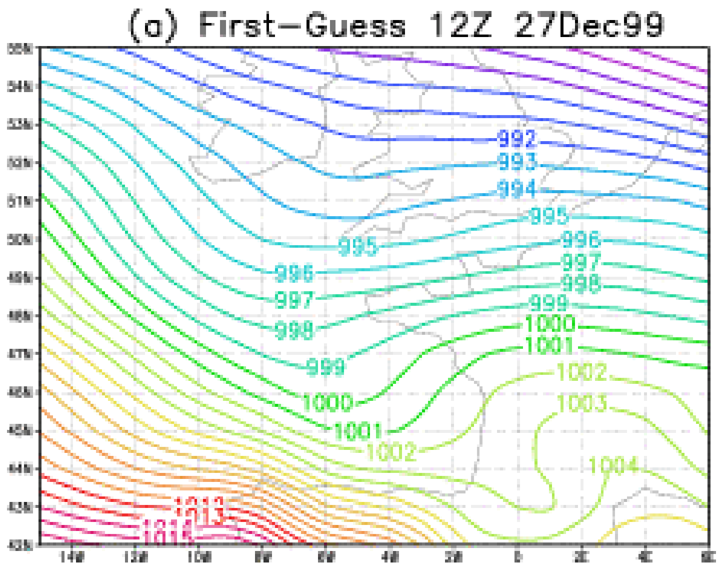
January 1998 Mean Meridional Circulation

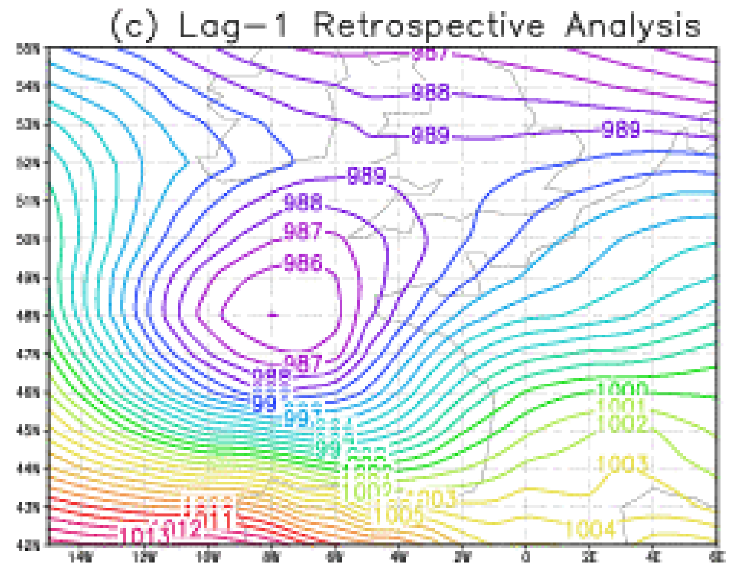
CTL



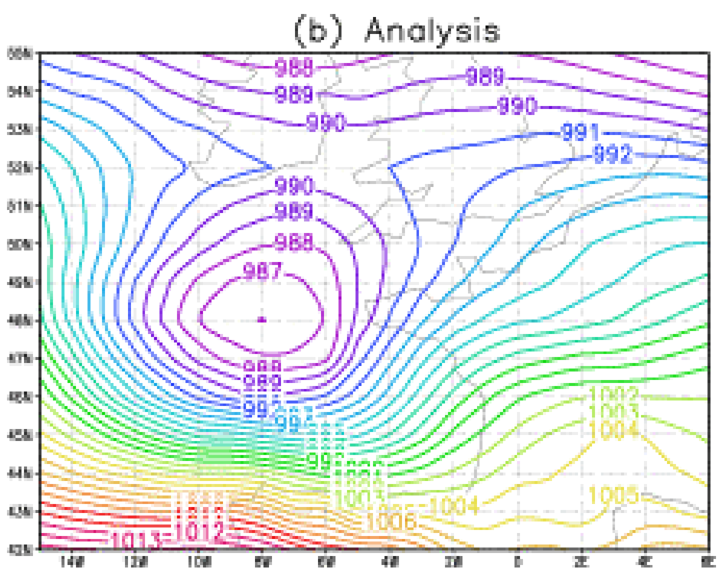
RIA

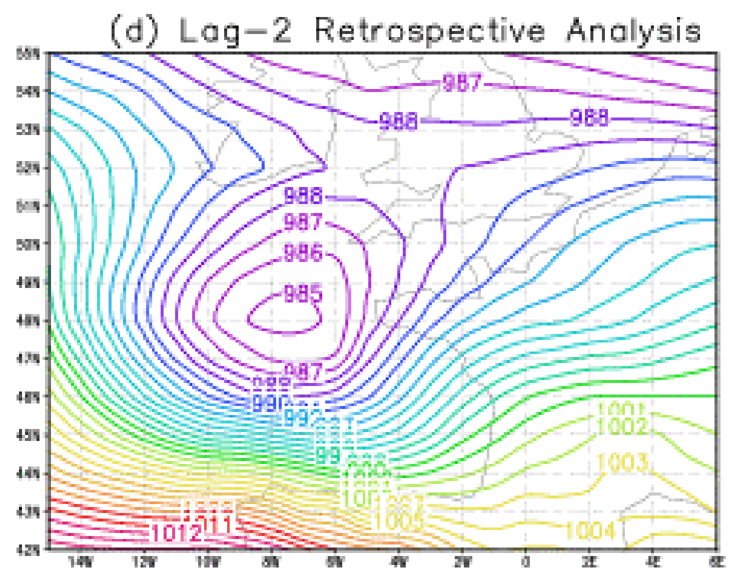






Full Retrospective Analysis: An attempt to improve on synoptic situations





Lessons Learned So Far

(WARNINGS)

Simple things may have a larger beneficial impact on results than implementing a sophisticated DA procedure given all the unknowns:

- assimilating observations at their proper (or near-proper) time

- de-biasing observations and mostly making sure all observing systems used in the DA are telling the same story (what some refer to a homogenizing the observations)

- proper handling of satellite observations (it takes some work to get positive impact)

- estimating and correcting model biases

- proper initialization procedures for preservation of dynamical balances

Ultimately, we would like to be able to:

- Accommodate non-normality in error covariance evolution

- Account for model errors

- Develop suitable and legitimately nonlinear estimation procedures

Adaptive Quality Control

Consists of a gross check and a buddy check:

1) Gross check eliminates largely erroneous data

2) Buddy check eliminates data by testing the null hypothesis: $\mathbf{v} \sim N(0,S)$ where $\mathbf{v}$ stands for the innovation vector

$\mathbf{v}$ is decomposed as $\mathbf{v} = [\, \mathbf{x}^T \;\; \mathbf{y}^T \,]^T$ **where $\mathbf{x}$ is a set of suspect data and $\mathbf{y}$ is a set of buddies**

$\mathbf{x}^{(0)} = \{\, v_i \,\varepsilon\, \mathbf{v}$ such that $|v_i| > \tau_o \; (\mathbf{S}_{ii})^{1/2} \,\}$

for k = 1 , 2, …

$\mathbf{y} = \{\, v_i \,\varepsilon\, \mathbf{v}$ such that v_i not in $\mathbf{x}^{(k-1)} \,\}$

$\mathbf{x}^* = \mathbf{S}_{xy}\, \mathbf{S}_y^{-1}\, \mathbf{y}$

$\mathbf{S}^* = \mathbf{S}_x - \mathbf{S}_{xy}\, \mathbf{S}_y^{-1}\, \mathbf{S}_{xy}^T$

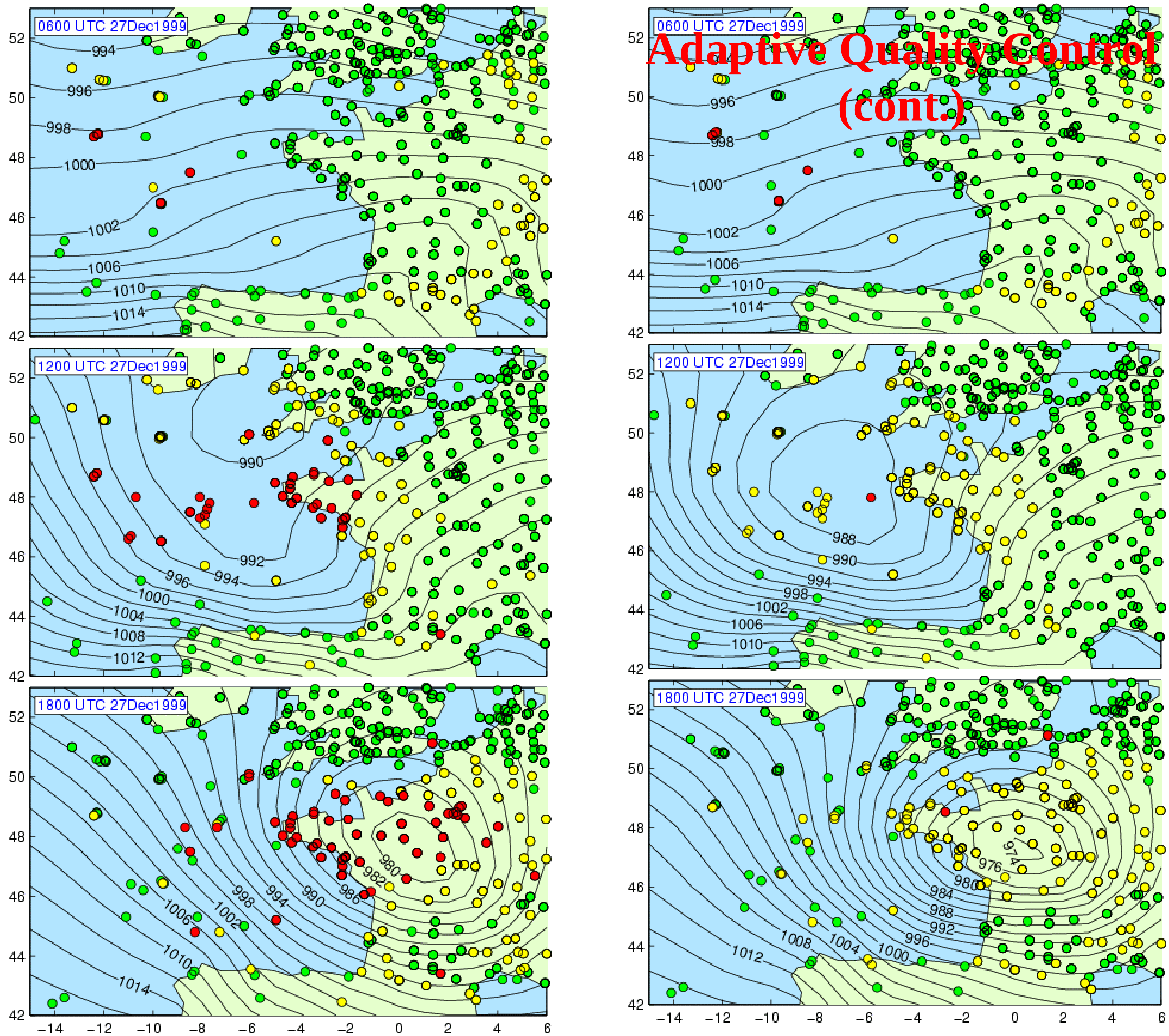
$\alpha = (\, \mathbf{y}^T \mathbf{S}_y^{-1} \mathbf{y} + \dim \mathbf{x}^* \,) \,/\, (\, \dim \mathbf{y} + \dim \mathbf{x}^* \,)$

$\mathbf{x}^{(k)} = \{\, x_i \,\varepsilon\, \mathbf{x}^{(k-1)}$ such that $|x_i - x_i^*| > \alpha\, \tau \; (\mathbf{S}_{ii}^*)^{1/2} \,\}$

if $\dim \mathbf{x}^k = \dim \mathbf{x}^{(k-1)}$ **or** $\dim \mathbf{x}^k = 0$ **then stop**

end

Adaptive Quality Control (cont.)



Linear SW stable dynamics;
nearly-normal dynamics

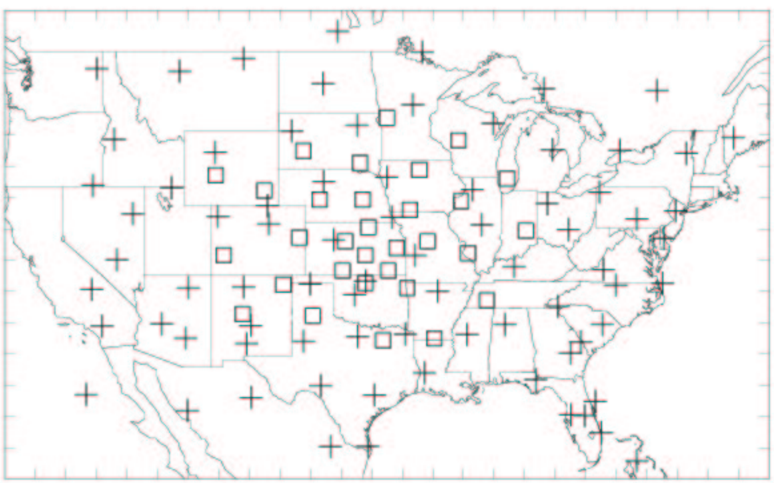
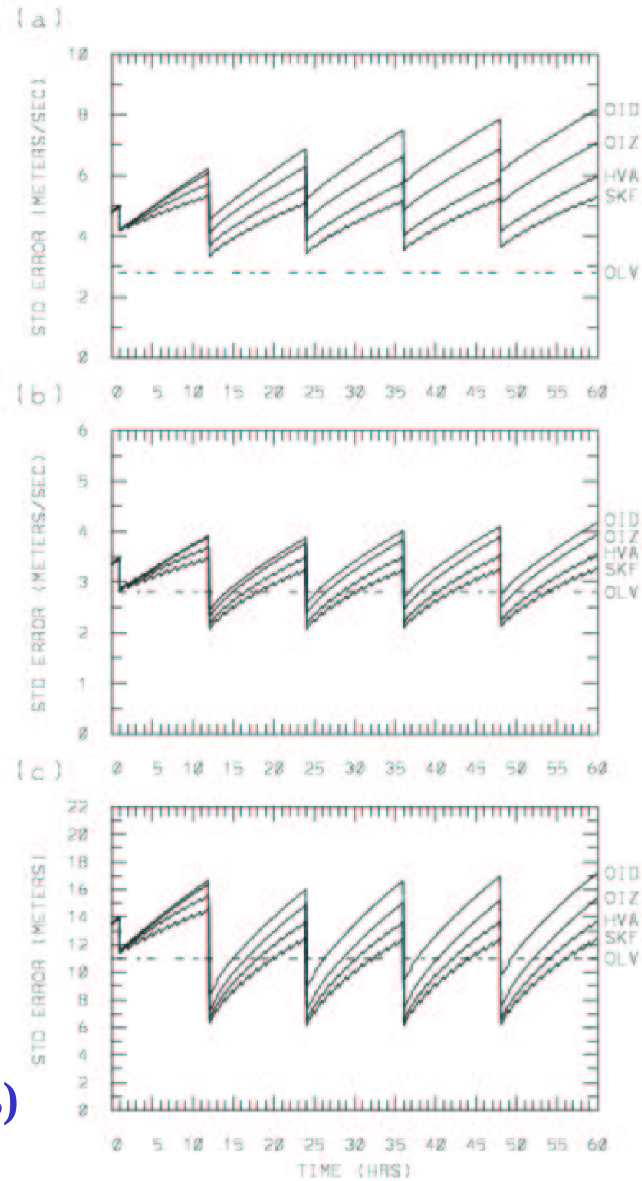


Fig. 1



State Variables: U, V, H

Observations: 6-hourly radiosondes (plus signs)
hourly wind profilers (squares)

Ongoing and Future Developments

We are heading toward implementing an adaptive parameterized Kalman filter, and iterated Kalman smoother, for performing atmospheric data assimilation

- **The adaptive quality control is currently univariate, but can easily be modified to become multivariate**

- **A fully feasible analysis error estimation algorithm, that allows using the complete observing system, is in its final evaluation stage**

- **Validating the improvements of the adjoint-free retrospective scheme still needs to be done for the real observation case**

- **Further testing and evaluation of the complete retrospective scheme, using the adjoint model, is currently being done**

- **We should be able to use the bias estimates to improve forecast skills**