# Bias and Data Assimilation

Dick Dee
ECMWF

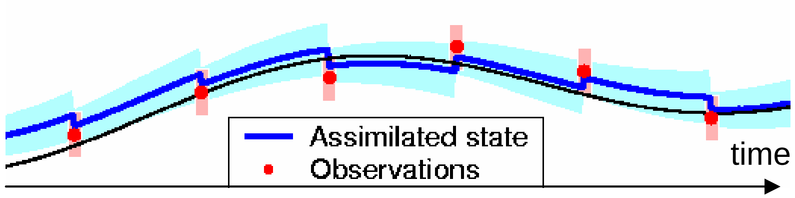
ECMWF/NWP-SAF Workshop on
Bias estimation and correction in data assimilation
8 - 11 November 2005

## Outline

- Bias-blind data assimilation
  - Standard assumptions
  - OSSE example
  - ERA-40 analysis increments
- Time series analysis of station data
  - Long timescales (climate)
  - Short timescales (weather)
- Bias-aware data assimilation
  - Variational correction of observation bias
  - Weak-constraint 4D-Var
  - Sequential model bias correction schemes
- Summary

## Bias-blind assimilation

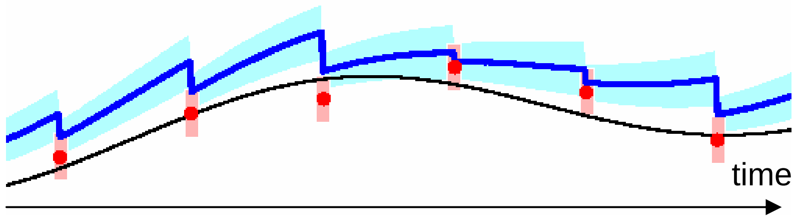
Data assimilation is essentially a sequential procedure for adjusting a model integration to actual observations:

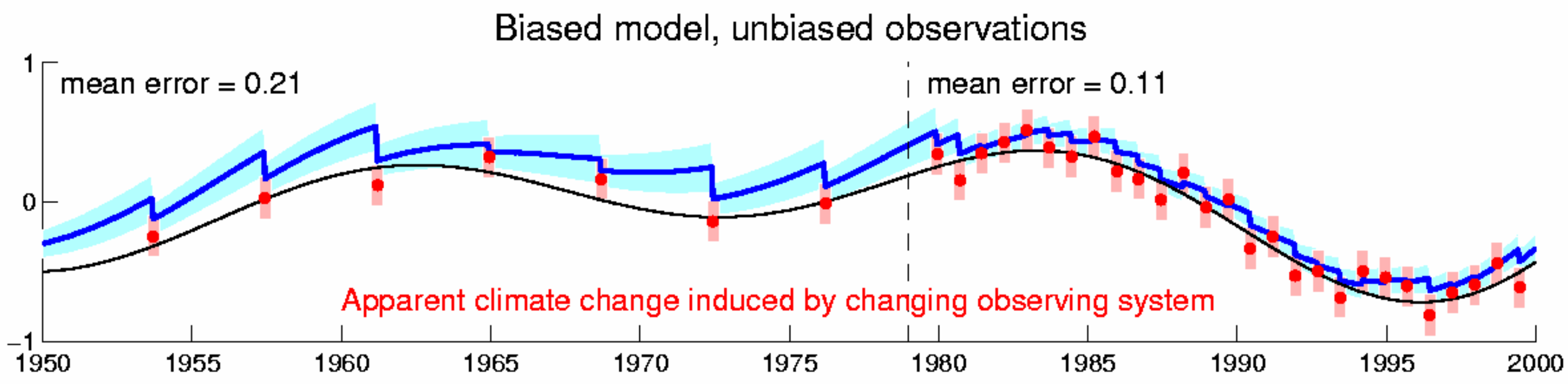



Most data assimilation systems are **bias-blind:** they were **designed to correct random errors** only.

Systematic errors in models and observations cause many **problems in assimilation** systems:

- Suboptimal use of observations
- Biases in the assimilated fields
- Non-physical structures in the analysis
- Extrapolation of biases due to multivariate background constraints
- Spurious trends due to changes in the observing system



# Bias-blind data assimilation, in theory

**Minimize** $J(x) = (x_b - x)^T B^{-1}(x_b - x) + (y - h(x))^T R^{-1}(y - h(x))$

**Output:** $dx = x - x_b$ (analysis increments)

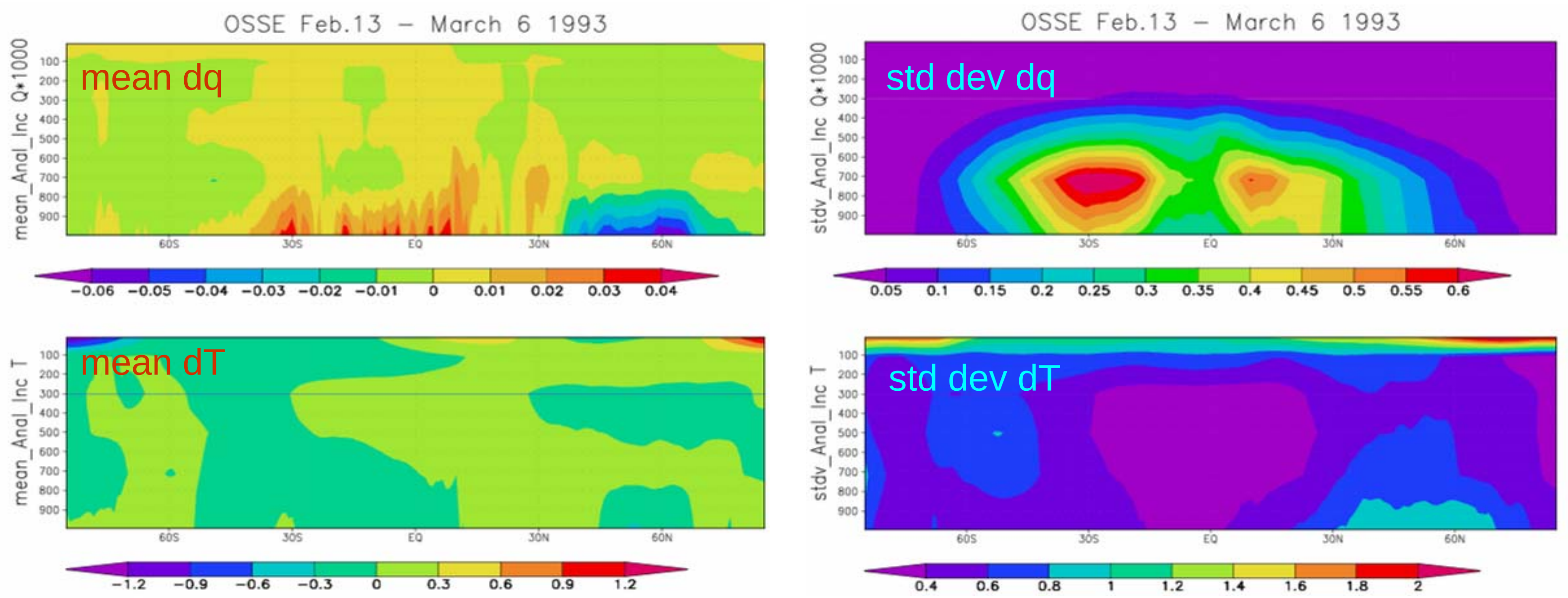
**Input:** $dy = y - h(x_b)$ (background departures)

**In the absence of biases:**

$$\langle dy \rangle \approx 0$$

$$\langle dx \rangle \approx 0$$

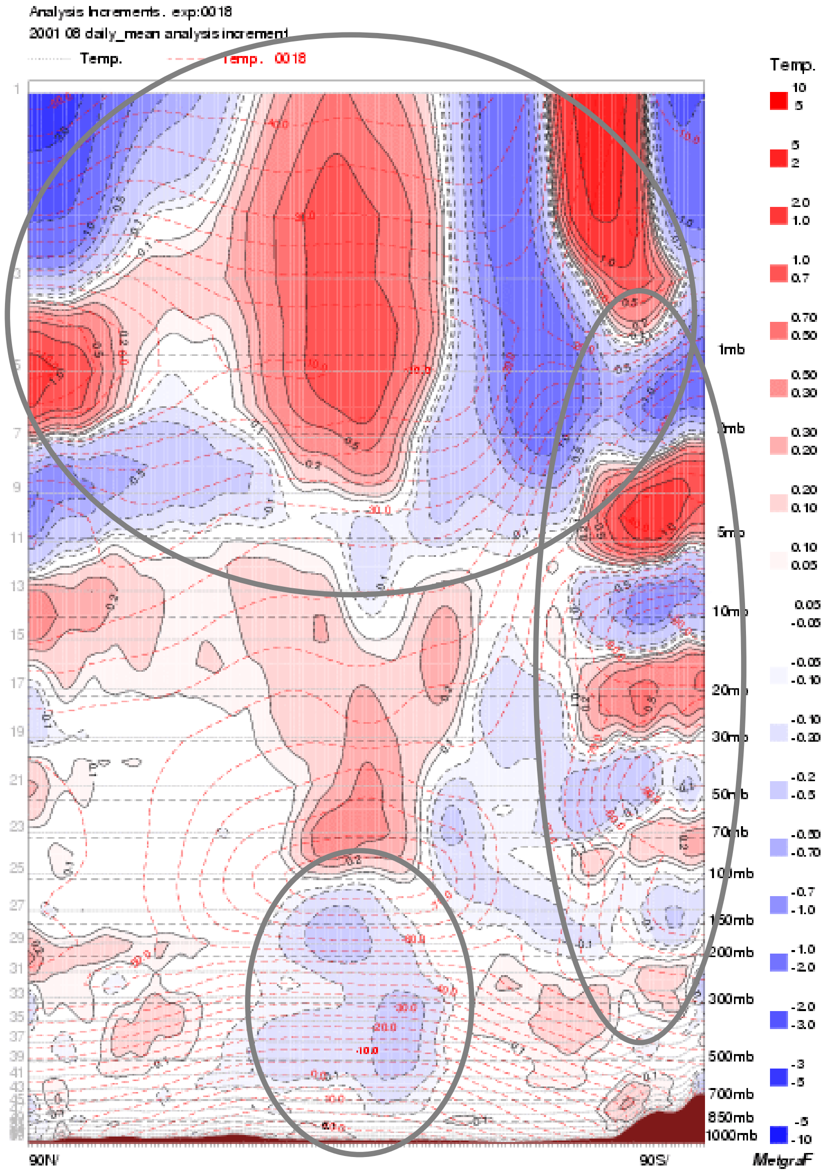
**Zonal mean analysis increments in an Observing System Simulation Experiment (OSSE)**




Courtesy Michiko Masutani (NCEP), Ron Errico, Runhua Yang (GMAO)

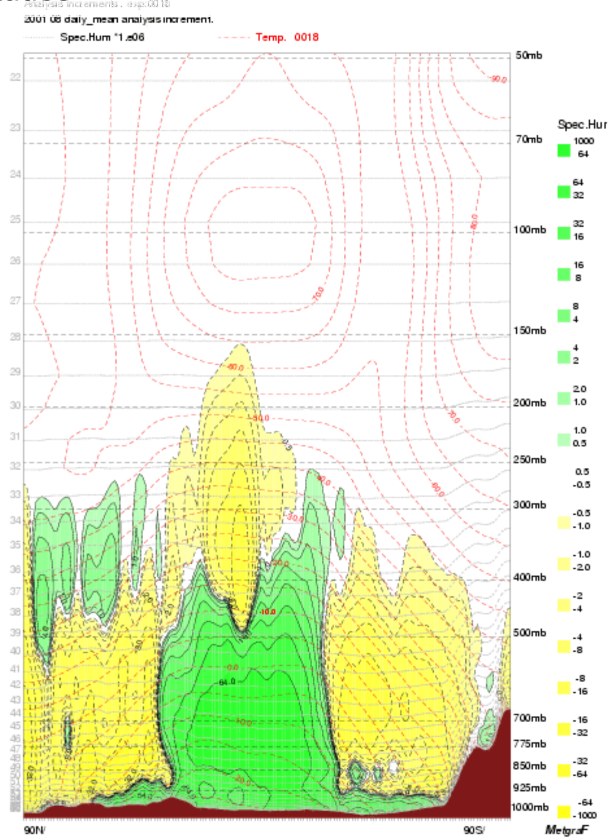
# ERA-40 Monthly Mean Analysis Increments

## August 2001 zonal mean T,q



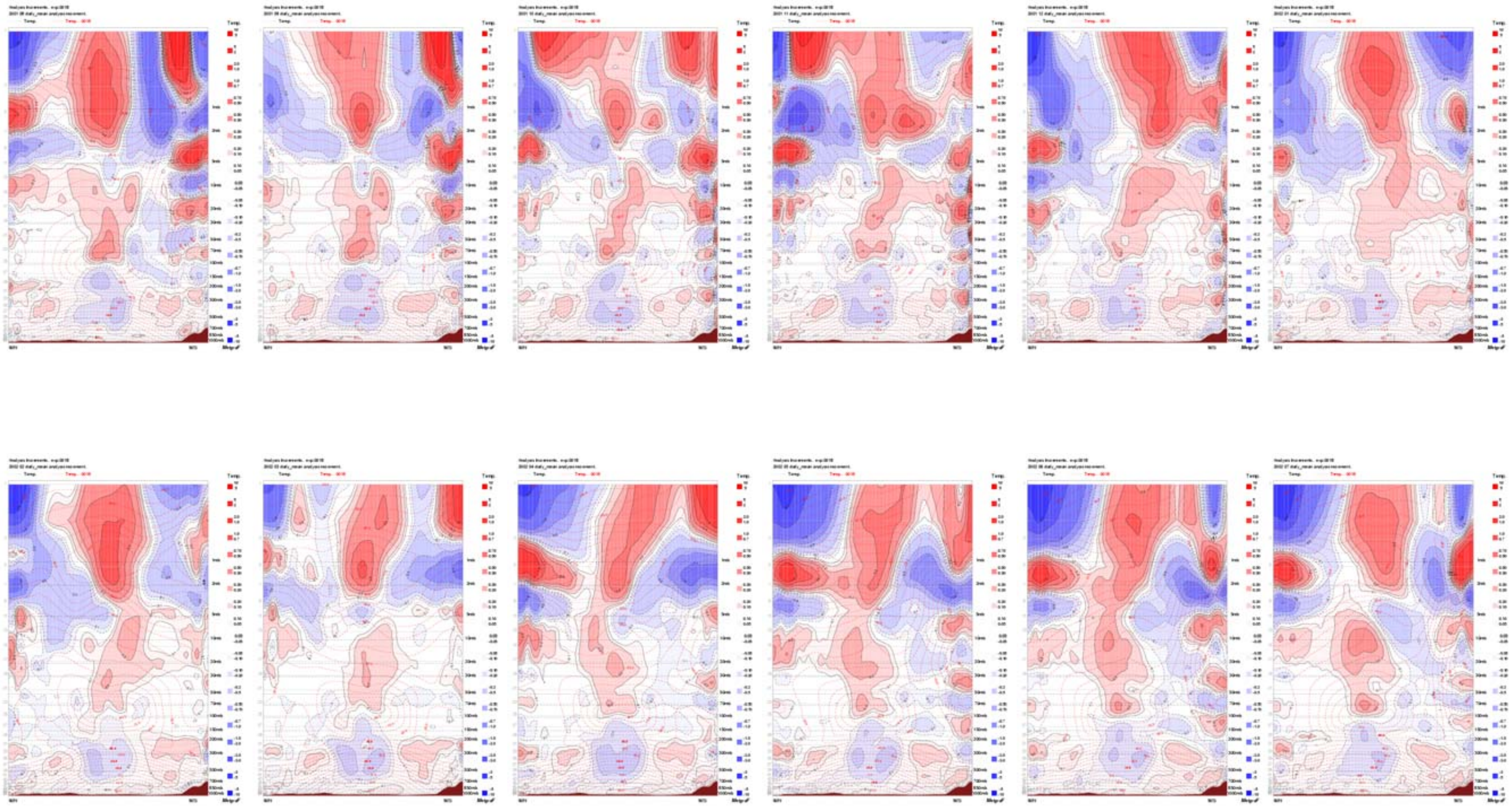

- Large upper stratospheric temperature bias
- Vertical structure of the increments reflect background constraints (B)
- Tropospheric mean temperature increments are large as well, especially in tropics

- Systematic dry bias in tropics, wet bias in high latitudes



# ERA-40 Monthly Mean Analysis Increments

## August 2001 - July 2002 zonal mean T



## Summary so far

- Bias-blind analysis schemes are suboptimal, propagate biases and generate spurious signals and trends in the assimilation

- Persistent patterns in the analysis increments are indicative of systematic errors

- These are present in any data assimilation system, unless it uses synthetic data

- What about the possible sources of these problems?

  - **Model errors**
  - **Data errors**
  - **Observation operators**
  - **Assimilation methodology**

- A great deal of work is being done on identification and correction of biases – much of it outside the assimilation framework

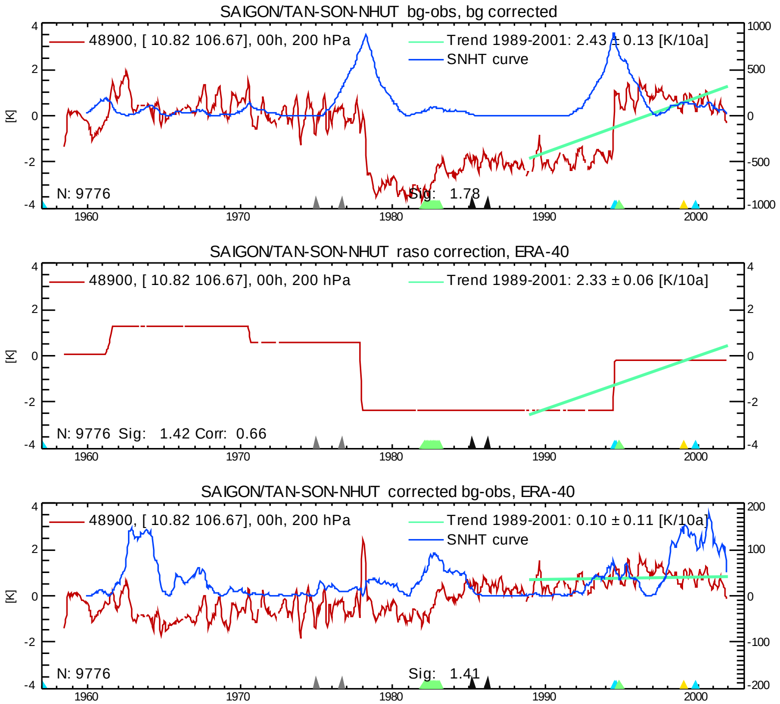
## Outline

- Bias-blind data assimilation
  - Standard assumptions
  - OSSE example
  - ERA-40 analysis increments
- Time series analysis of station data
  - Long timescales (climate)
  - Short timescales (weather)
- Bias-aware data assimilation
  - Variational correction of observation bias
  - Weak-constraint 4D-Var
  - Sequential model bias correction schemes
- Summary

# Break detection in radiosonde time series (climate time scales)

**Leopold Haimberger (Univ. of Vienna and ECMWF)**

Break detection based on stationarity tests of bg-obs differences, combined with available information about changes in radiosonde and/or ground equipment, radiation correction, etc.




Uncorrected bg-obs temperature at 50hPa

Test statistic (SNHT) used for break detection

Corrections applied at this station after break identification

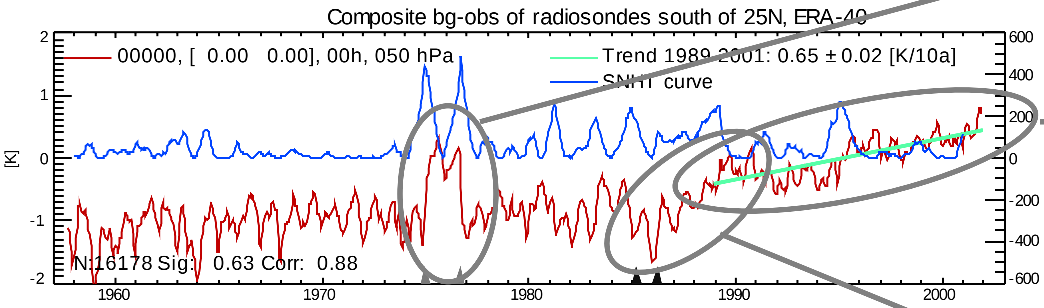
Corrected bg-obs temperature at 50hPa

**Major challenge:** To account for jumps and trends in background temperature field that are due to changes in the observing system (particularly satellites) and associated bias corrections

# Some examples of identifiable artificial jumps and trends in mean background temperatures

## Uncorrected bg-obs temperature at 50hPa:

Mean over all stations south of 25N

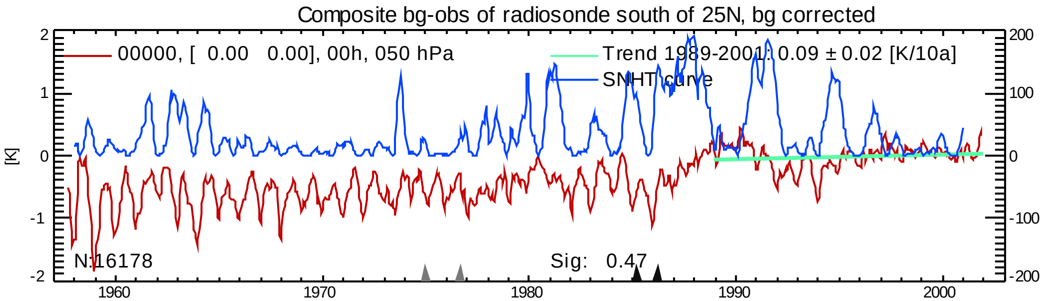



Errors in NOAA-4 bias correction

Excessive tropical precipitation associated with increase in humidity observations

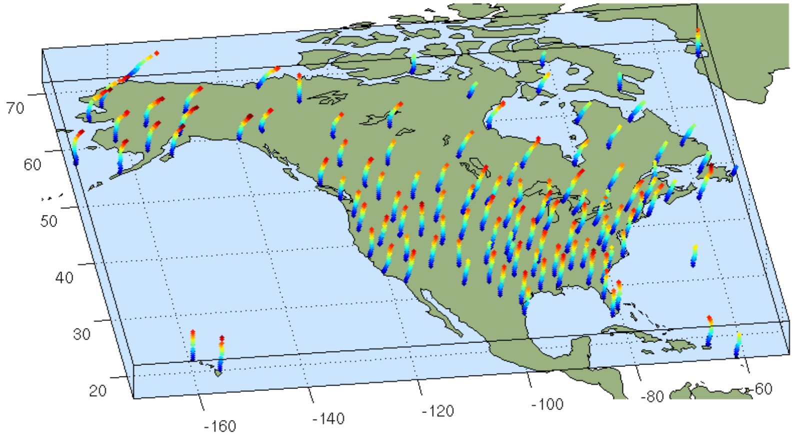
Gradual replacement of radiosondes in Australia and Pacific Islands (not global)

## After global bg corrections:




**Leo Haimberger, 2005**: Homogenization of radiosonde temperature time series using ERA-40 analysis feedback information. ERA-40 Project Report No. 22, ECMWF

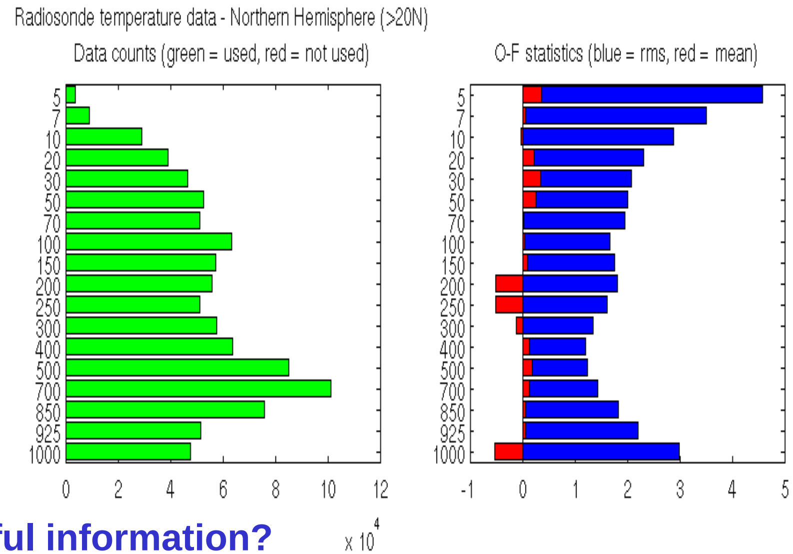
# Spectral analysis of observed-minus-background differences (weather time scales)



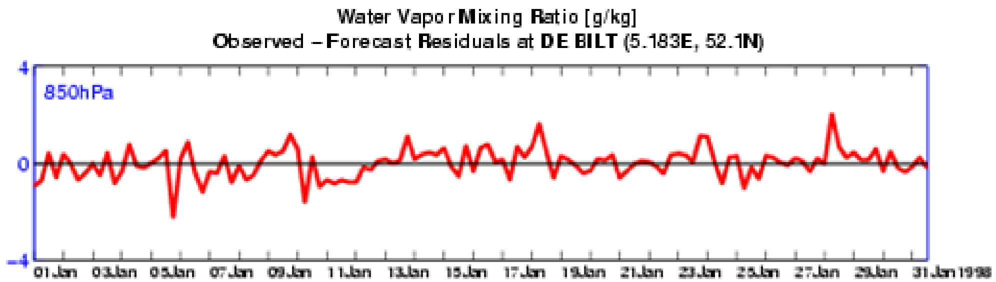

Background fit to radiosonde observations is commonly used to assess the impact of changes to the assimilation system

We usually monitor certain basic statistics:

- Data counts
- Mean departures
- Rms departures



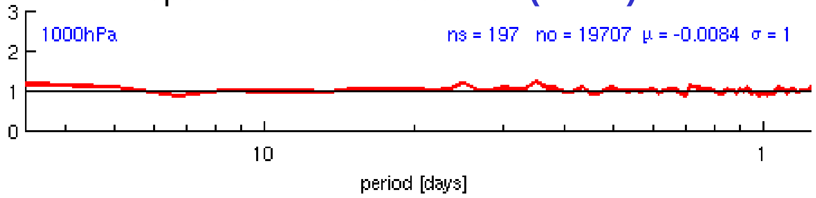

**Do station time series contain additional useful information?**



## Spectral analysis of observed-minus-background differences

***Innovation property:*** If the assimilation is optimal then the obs-bg time series will be white

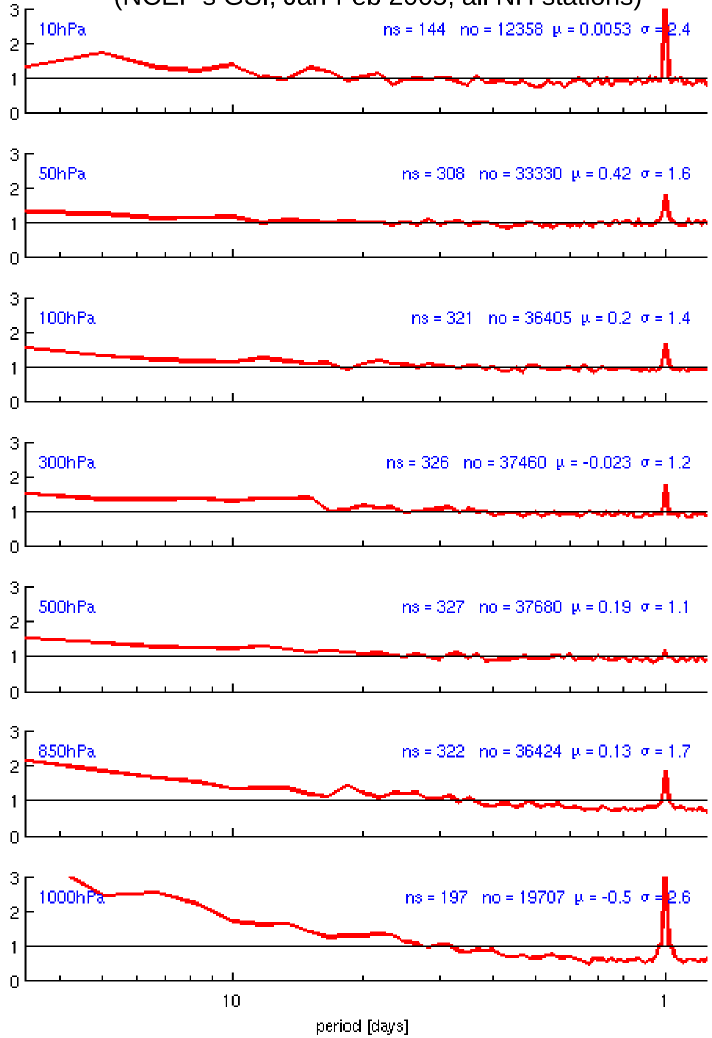
Normalized spectrum of white obs-bg with station-dependent bias and std dev **(FAKE)**




**A great deal of useful information in the observations is not extracted by the analysis**

Using 2 months of actual temperature data:

(NCEP's GSI, Jan-Feb 2005, all NH stations)



## Outline

- Bias-blind data assimilation
  - Standard assumptions
  - OSSE example
  - ERA-40 analysis increments
- Time series analysis of station data
  - Long timescales (climate)
  - Short timescales (weather)
- Bias-aware data assimilation
  - Variational correction of observation bias
  - Weak-constraint 4D-Var
  - Sequential model bias correction schemes
- Summary

# Bias-aware assimilation

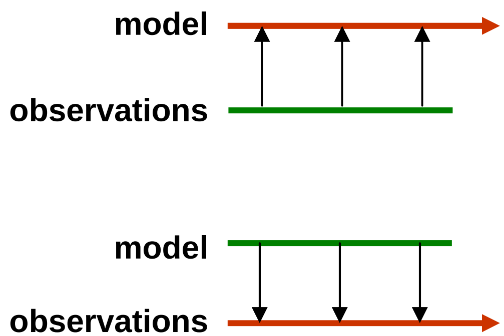
## Analysis methods designed to correct (some) biases during data assimilation

- These methods always rely on **assumptions about the sources of bias**

Correct for observation bias

$$x_a = x_b + K\left[y - h(x_b)\right]$$

Correct for model bias




- They require meaningful **bias models:**
  - Essentially a way to reduce the number of degrees of freedom
  - Persistent bias; use of basis functions; physically-based (parameterized) models
  - Estimation requires a relationship between bias model parameters and the observations
  - Fundamentally: The **bias parameters must be observable**

## Variational correction of observation biases

The **bias** in a given instrument/channel is usually modeled in terms of a relatively small number of parameters – e.g. **linear predictor model** for radiance bias **(Harris and Kelly 2000)**

It is natural to add these parameters to the control vector and correct the observations during the analysis **(Derber and Wu 1998; Dee 2004)**

The **standard variational analysis** minimizes

$$J(x) \;=\; (x_b - x)^T B_x^{-1}(x_b - x) \;+\; [y - h(x)]^{\,T} R^{-1}\,[y - h(x)]$$

Modify the observation operator to account for bias: $\tilde{h}(z) = \tilde{h}(x, \beta)$

Include the bias parameters in the control vector: $z^T \;=\; [\; x^T \quad \beta^T \;]$

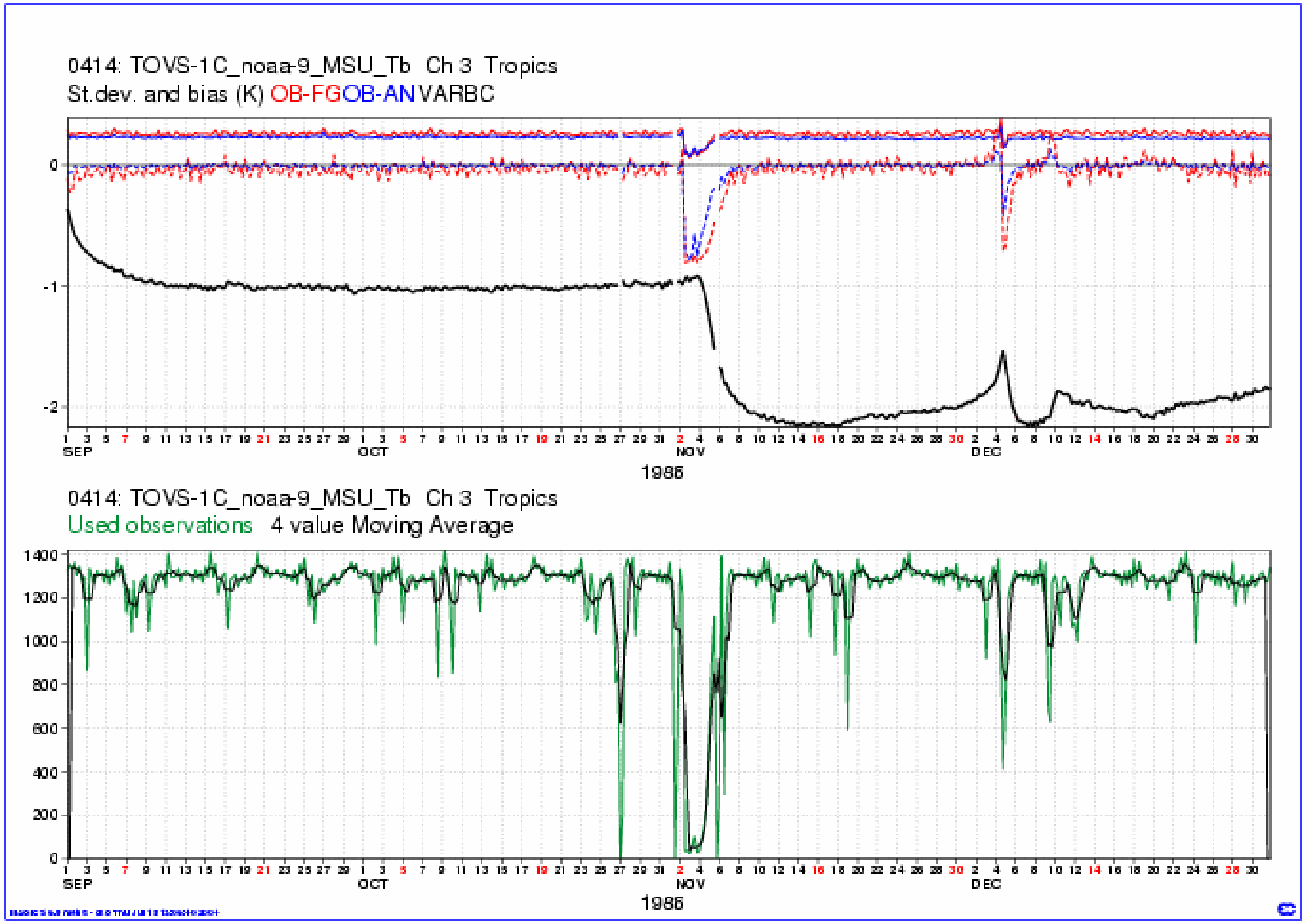
**Minimize instead** $\boxed{J(z) \;=\; (z_b - z)^T B_z^{-1}(z_b - z) \;+\; [y - \tilde{h}(z)]^{\,T} R^{-1}\,[y - \tilde{h}(z)]}$

**What is needed to implement this:**

1. A modified operator $\tilde{h}(x, \beta)$ and its TL + adjoint
2. Background error covariances for the bias parameters
3. An effective preconditioner for the joint minimization problem

# Observed-background statistics and adaptive bias correction

## NOAA-9 MSU Ch 3



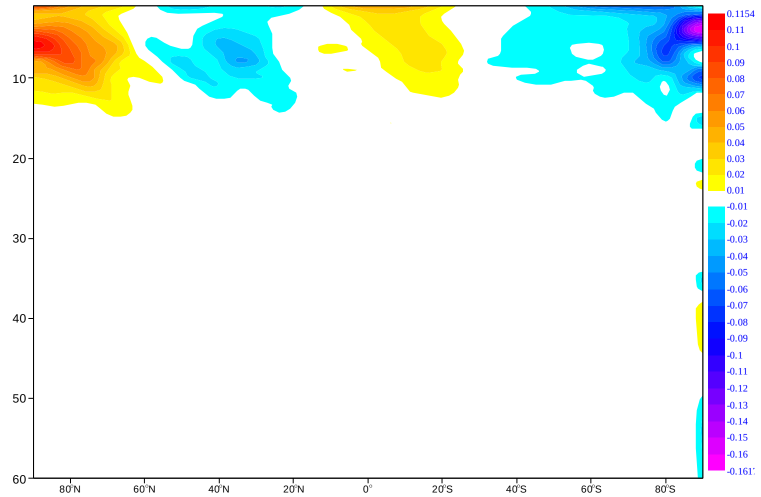
# Model bias correction in weak-constraint 4D-Var

Yannick Trémolet, ECMWF

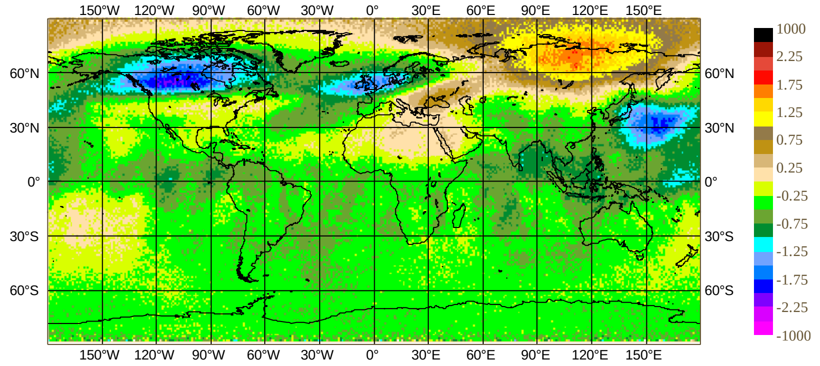
- Extend 4D-Var by including model forcing in the control vector (Derber 1989; Zupanski 1997)
- Reduce size by assuming that model error is constant for the length of the assimilation window
- Model error constraints (Q) are obtained from time series of tendency differences

- Estimated model errors in the stratosphere are consistent with large stratospheric temperature bias
- Improved agreement with observed radiances in stratospheric temperature sounding (AMSU-A Ch13)

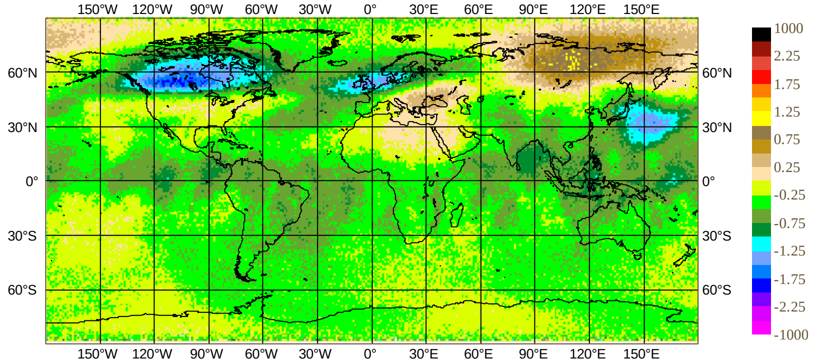
**Mean Model Error Forcing - Experiment=ej5j**
**Dates=2004020100-2004021512 - Param=T**




STATISTICS FOR RADIANCES FROM NOAA-15 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004013118 - 2004021512 , HOUR = ALL
EXP = EJ5H
Min: -2.1916 Max: 1.8620 Mean: -0.309585




STATISTICS FOR RADIANCES FROM NOAA-15 / AMSU-A - 13
MEAN FIRST GUESS DEPARTURE (OBS-FG) (BCORR.) (CLEAR)
DATA PERIOD = 2004013118 - 2004021512 , HOUR = ALL
EXP = EJ5J
Min: -1.9278 Max: 1.2571 Mean: -0.333228

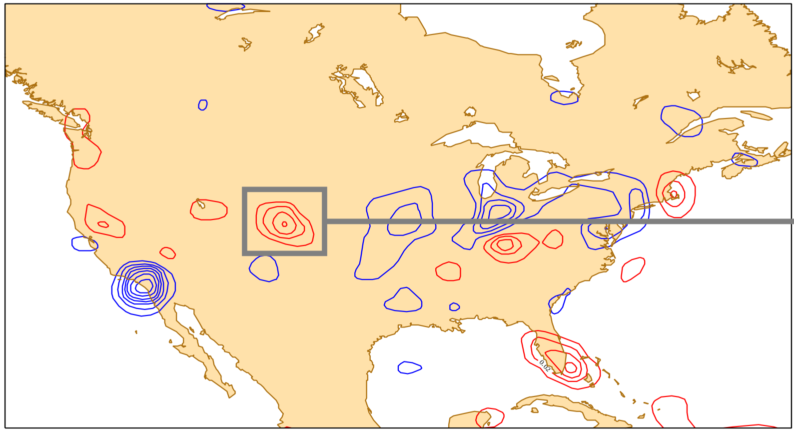


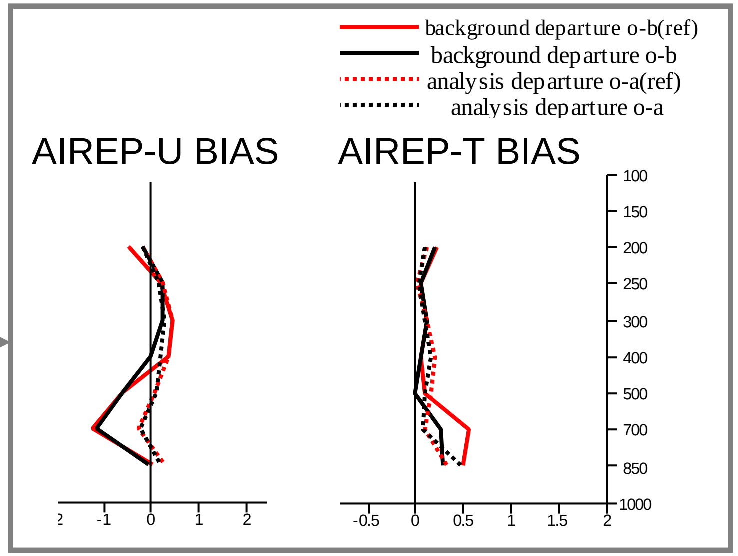
# Model bias confused with observation bias

Yannick Trémolet, Lars Isaksen (ECMWF)

**Mean Model Error Forcing - Experiment=ej6a**
**Dates=2004050100-2004051012 - Param=T - Level=60**

0.01 (0.04)





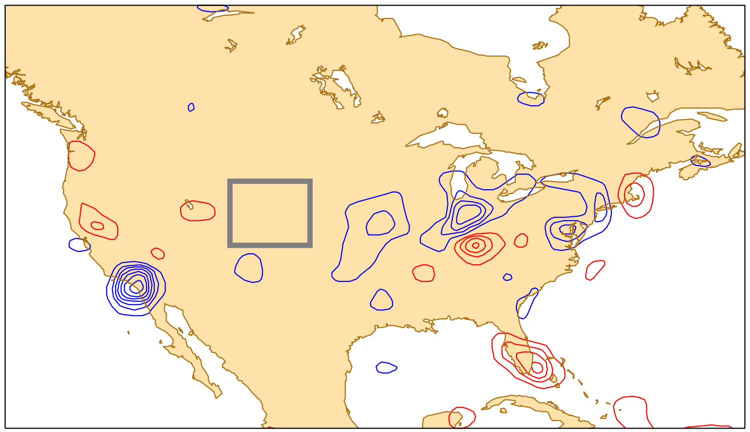

Persistent model error forcing at lower levels in the vicinity of major airports

Explained by observation bias due to slight delay in reports during ascents/descents?

Local model error forcing disappears when all aircraft reports near Denver airport are withheld from the assimilation

**Mean Model Error Forcing - Experiment=ej8k**
**Dates=2004050100-2004051012 - Param=T - Level=60**

0.01 (0.04)

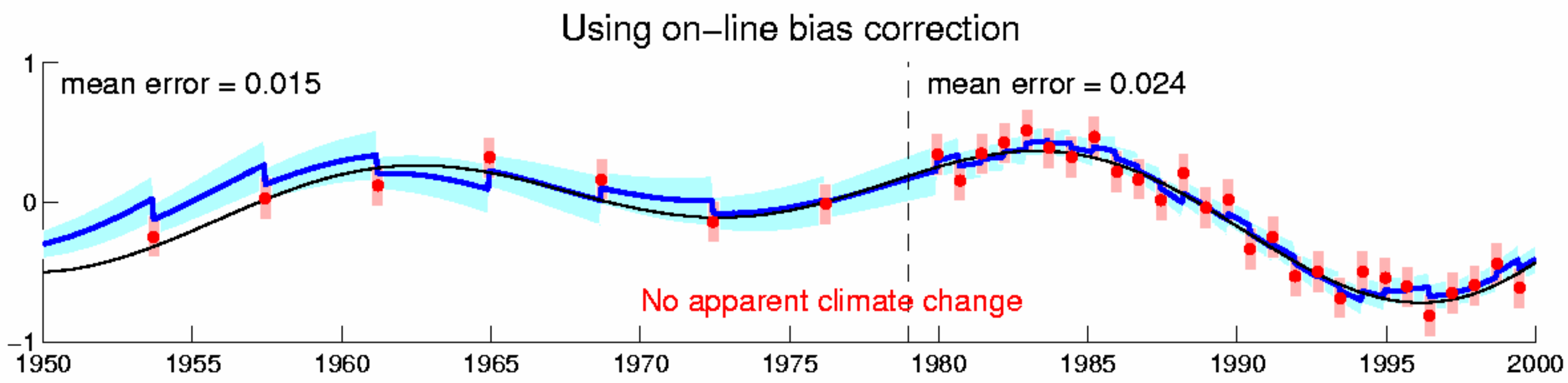


## A simple sequential scheme for correcting bias in the model background

$$\widetilde{\mathbf{x}} = \mathbf{x}_k^f - \hat{\mathbf{b}}_{k-1}$$ bias correction

$$\left.\begin{aligned} \mathbf{dy} &= \mathbf{y}_k - \mathbf{H}\widetilde{\mathbf{x}} \\ \mathbf{dx} &= \mathbf{K}\mathbf{dy} \\ \mathbf{x}_k^a &= \widetilde{\mathbf{x}} + \mathbf{dx} \end{aligned}\right\}$$ the usual bias-blind analysis

$$\hat{\mathbf{b}}_k = \hat{\mathbf{b}}_{k-1} - \alpha\mathbf{dx}$$ bias estimation



## Sequential schemes for correcting model bias correction

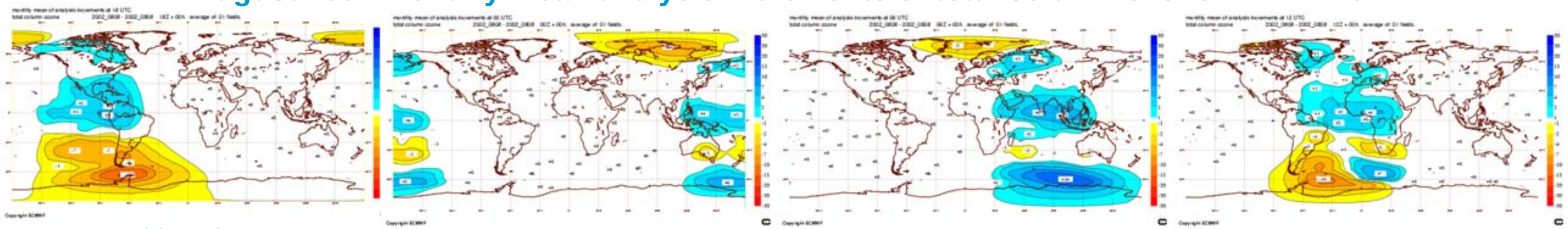
- This simple scheme is a special case of separate-bias estimation (Friedland 1969)
- Provides the Best Linear Unbiased Estimate (BLUE) in case of constant bias parameters
- Can be used to estimate observation bias parameters as well
- Virtually cost-free and very easy to implement
- BUT: the approach is purely statistical; no attempt to correct bias at the source

### Applications and enhancements:

- Atmospheric humidity analysis (Dee and Todling 2001)
- Sequential estimation of model bias parameters (Dee 2003)
- Bias correction via model forcing (Nichols et al.; Bell et al. 2004)
- Skin temperature analysis (Radakovich et al. 2004)
- Constituent assimilation (Lamarque et al. 2004)
- Ocean data assimilation (Balmaseda 2005; Chepurin et al. 2005)

## Another simple sequential scheme: Predictability of analysis increments

**August 2002 monthly mean analysis increments of total column ozone in ERA-40**



**00 UTC** **06 UTC** **12 UTC** **18 UTC**

It seems clear that certain aspects of the analysis increments are very predictable…

Can we take advantage of this to improve the data assimilation?

### Outline of a method:

Bias-blind assimilation:

$$dx_k = K_k(y_k - h(x_k^b))$$

Prediction of the analysis increment:

$$dx_k^p = f_k(dx_{k-L}, \cdots, dx_{k-1})$$

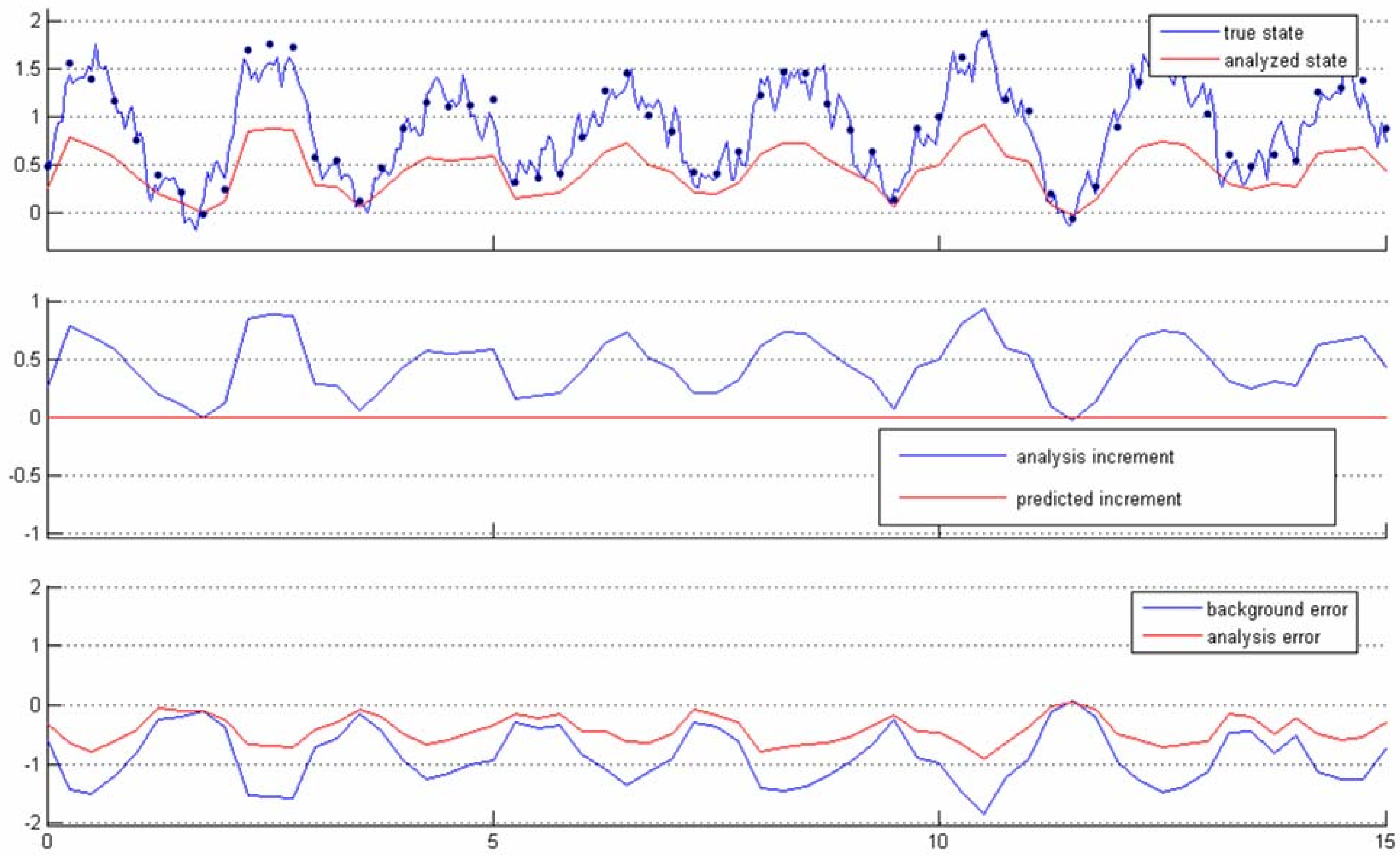
Bias-aware assimilation:

$$dx_k = dx_k^p + K_k(y_k - h(x_k^b + dx_k^p))$$

## Demonstration with a simple statistical estimator

Background error = slowly varying bias + quasi-diurnal cycle + serially correlated noise

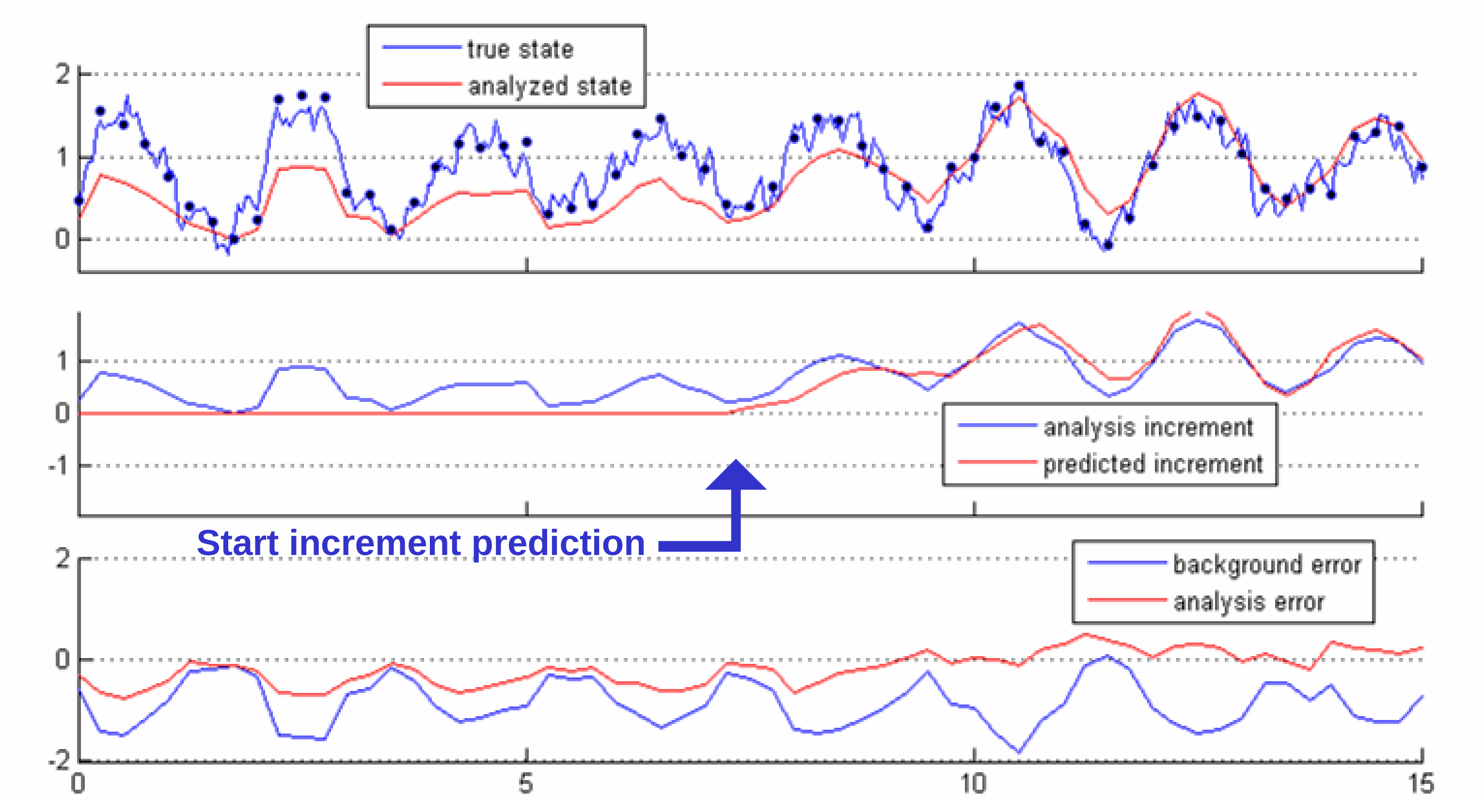
**Bias-blind assimilation:**



## Demonstration with a simple statistical estimator

Prediction of the analysis increment using a linear autoregressive model (2-day lag)

**Bias-aware assimilation:**



# Summary

## Traditional data assimilation methods are not designed to handle biases

- It is always preferable to correct bias at the source, if it can be identified
- A lot of work goes into bias correction of observations before they can be usefully assimilated

  (presentations by many at this workshop)
- Data assimilation systems provide excellent tools for identifying biases

  (presentations by L. Haimberger, D. Vasiljevic)

## All assimilation systems show evidence of residual biases in both models and data

- Persistent spatial patterns in analysis increments; temporal aspects of departures
- ***Impact on NWP:*** Loss of information; obstacles to proper utilization of satellite data

  (presentation by T. McNally)
- ***Impact on reanalysis:*** Difficult to separate real climate changes/trends from spurious signals

  (presentation by S. Uppala)

## Need for adaptive methods to correct systematic errors during the assimilation

- It is not reasonable to assume that errors are strictly random (either in models or data)
- There are compelling practical reasons for implementing adaptive, bias-aware systems
- ***Major challenge:*** To develop meaningful error models that can separate model bias from observation bias
- Many bias-aware techniques (variational and sequential) are available and are being implemented

  (presentations by T. Auligné, Y. Trémolet)