Disentangling biases in observations and models over the years

Dick Dee
JCSDA

Many thanks to my past colleagues at NASA/GMAO and ECMWF

Outline

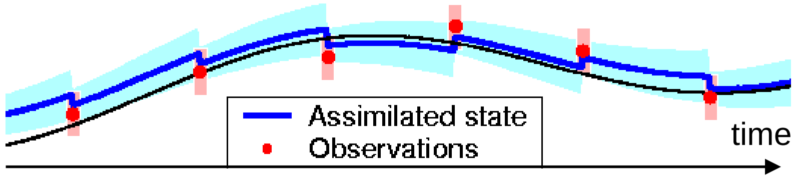
- **Bias-blind data assimilation**
  - Biases are not explicitly accounted for in ‘standard’ methods
  - Suboptimal use of observations, biased analyses, false climate signals
- **Bias-aware data assimilation**
  - Assumptions about the source
  - Variational correction of observation biases (VarBC)
  - Sequential methods for model bias correction
  - Weak-constraint 4D-Var
- **Disentangling sources of bias**
  - Model errors at major airports?
  - Some lessons learned in reanalysis

Outline

- **Bias-blind data assimilation**
  - Biases are not explicitly accounted for in ‘standard’ methods
  - Suboptimal use of observations, biased analyses, false climate signals
- **Bias-aware data assimilation**
  - Assumptions about the source
  - Variational correction of observation biases (VarBC)
  - Sequential methods for model bias correction
  - Weak-constraint 4D-Var
- **Disentangling sources of bias**
  - Model errors at major airports?
  - Some lessons learned in reanalysis

Bias and data assimilation

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



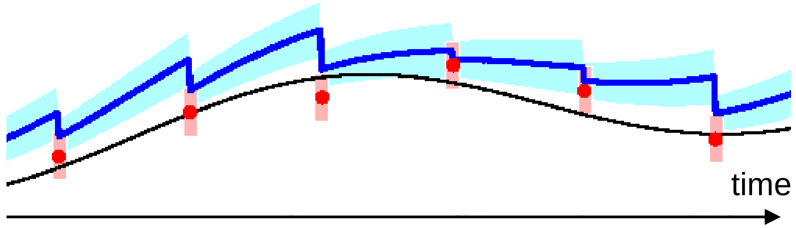
Conventional data assimilation methods are **bias-blind**: designed to correct random errors only.

Background departures and/or analysis increments provide information about biases.

In the absence of biases: $\langle dy \rangle \approx 0$ (observed-minus-background departures)

$\langle dx \rangle \approx 0$ (analysis increments)

Persistent model bias:

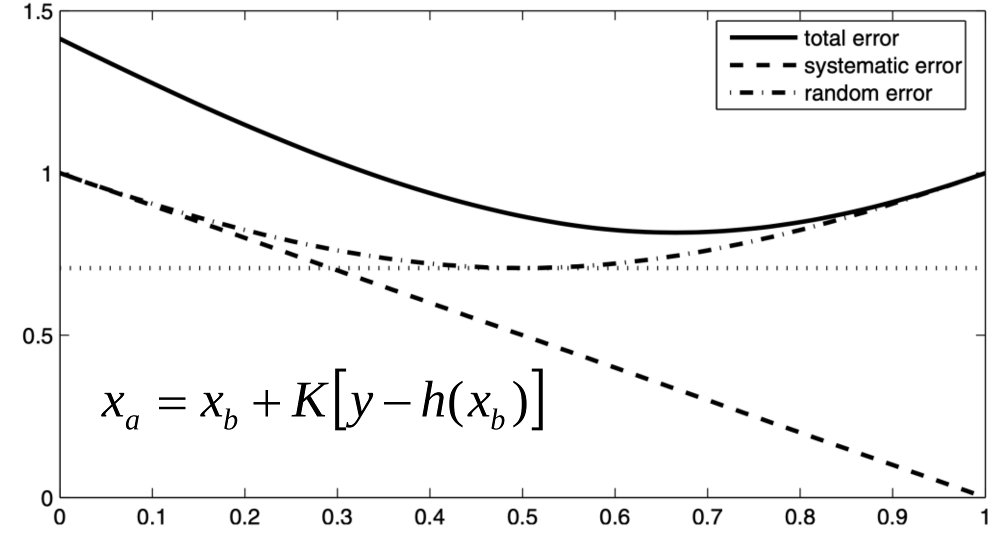


Bias and data assimilation

Systematic errors in models and observations lead to:

- Suboptimal use of observations
- Biases in the assimilated fields
- Non-physical features in the analysis
- Convolution of biases due to multivariate background covariances
- Jumps and other artifacts due to changes in observation coverage

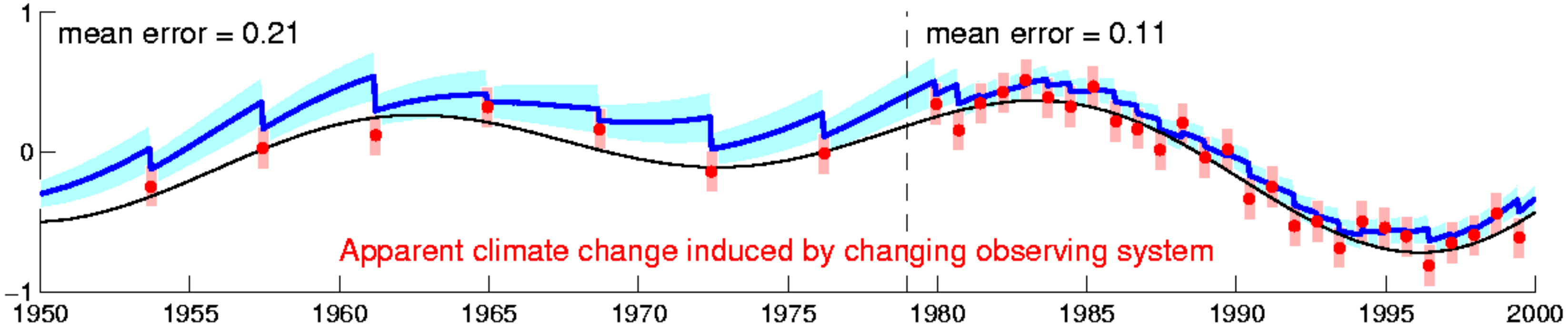
Rms error as a function of K



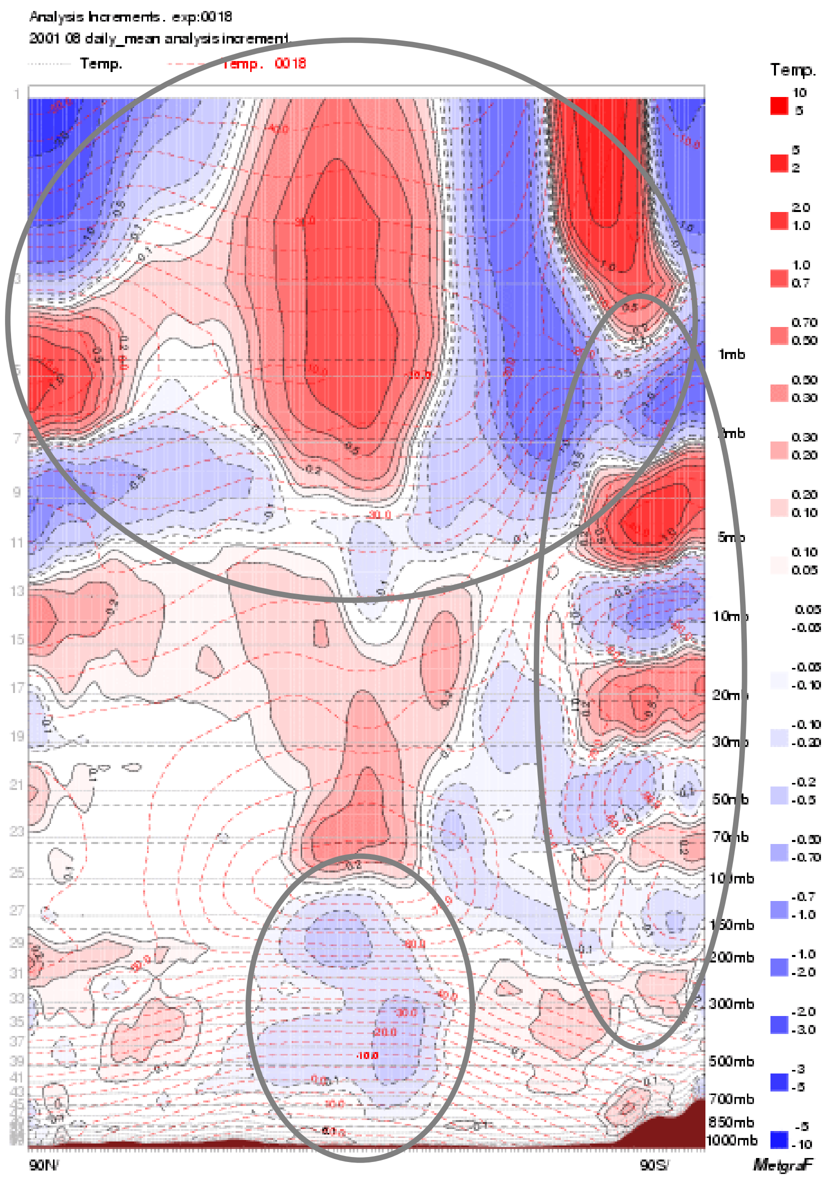
Model error ~N(1,1)

Observation error ~N(0,1)

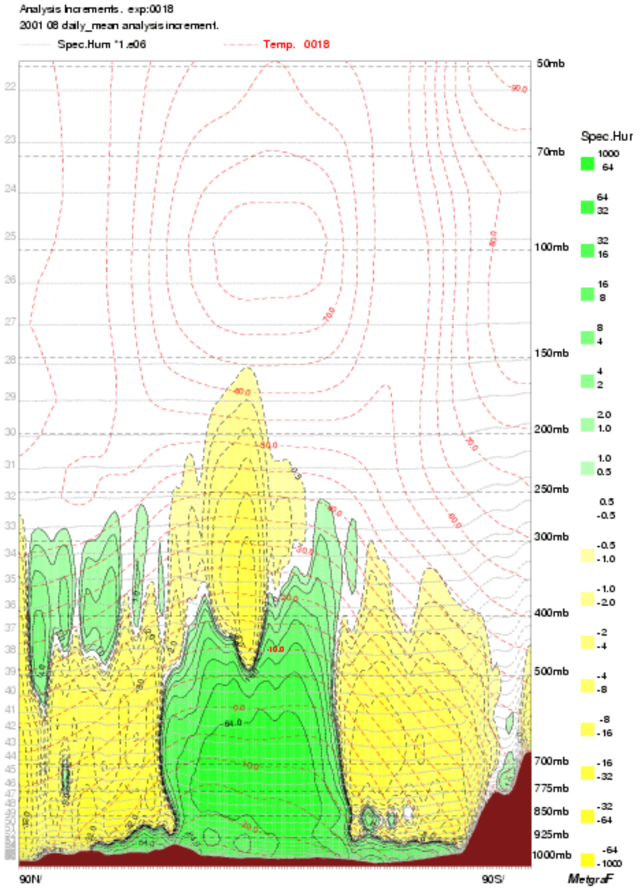
Biased model, unbiased observations



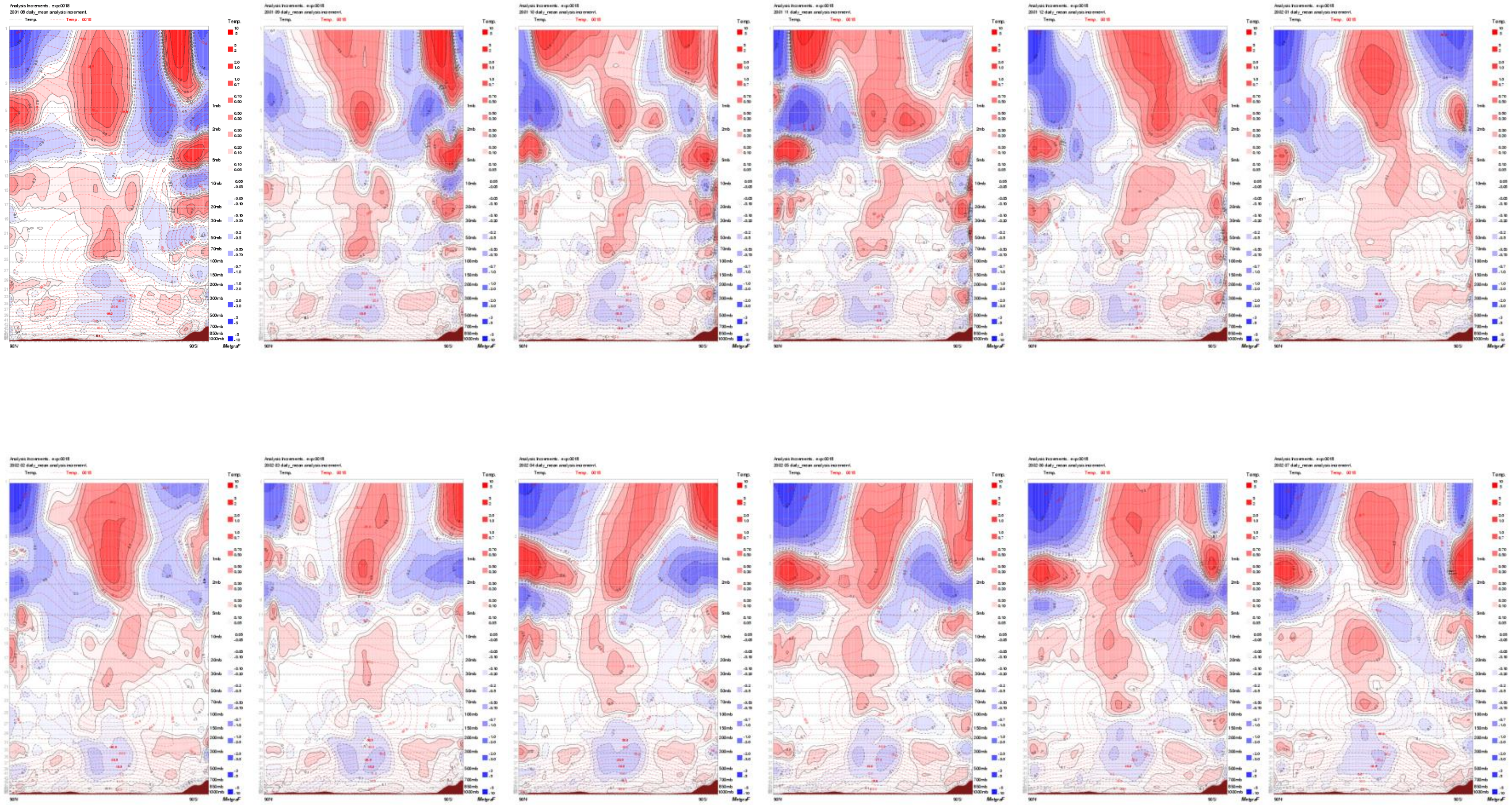
ERA-40 monthly averaged analysis increments: August 2001 zonal mean T, q



- Large upper stratospheric temperature bias
- Vertical structure of the increments reflect B
- Tropospheric mean temperature increments are large as well, especially in tropics
- Persistent dry bias in tropics, wet bias in high latitudes



ERA-40 monthly averaged analysis increments: August 2001 - July 2002 zonal mean T



Outline

- **Bias-blind data assimilation**
  - Biases are not explicitly accounted for in ‘standard’ methods
  - Suboptimal use of observations, biased analyses, false climate signals
- **Bias-aware data assimilation**
  - Assumptions about the source
  - Variational correction of observation biases (VarBC)
  - Sequential methods for model bias correction
  - Weak-constraint 4D-Var
- **Disentangling sources of bias**
  - Model errors at major airports?
  - Some lessons learned in reanalysis

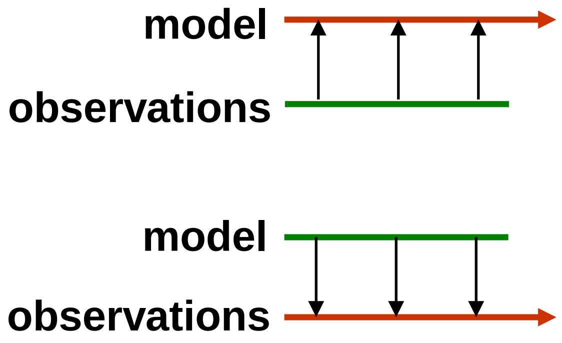
Bias-aware data assimilation

- Can we estimate and correct biases during data assimilation?
- This requires information and/or 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



- Using bias models reduces degrees of freedom in the estimation problem
  - E.g. persistent bias; use of predictors; physically-based (parameterized) models
  - Estimation requires a relationship between bias model parameters and the observations

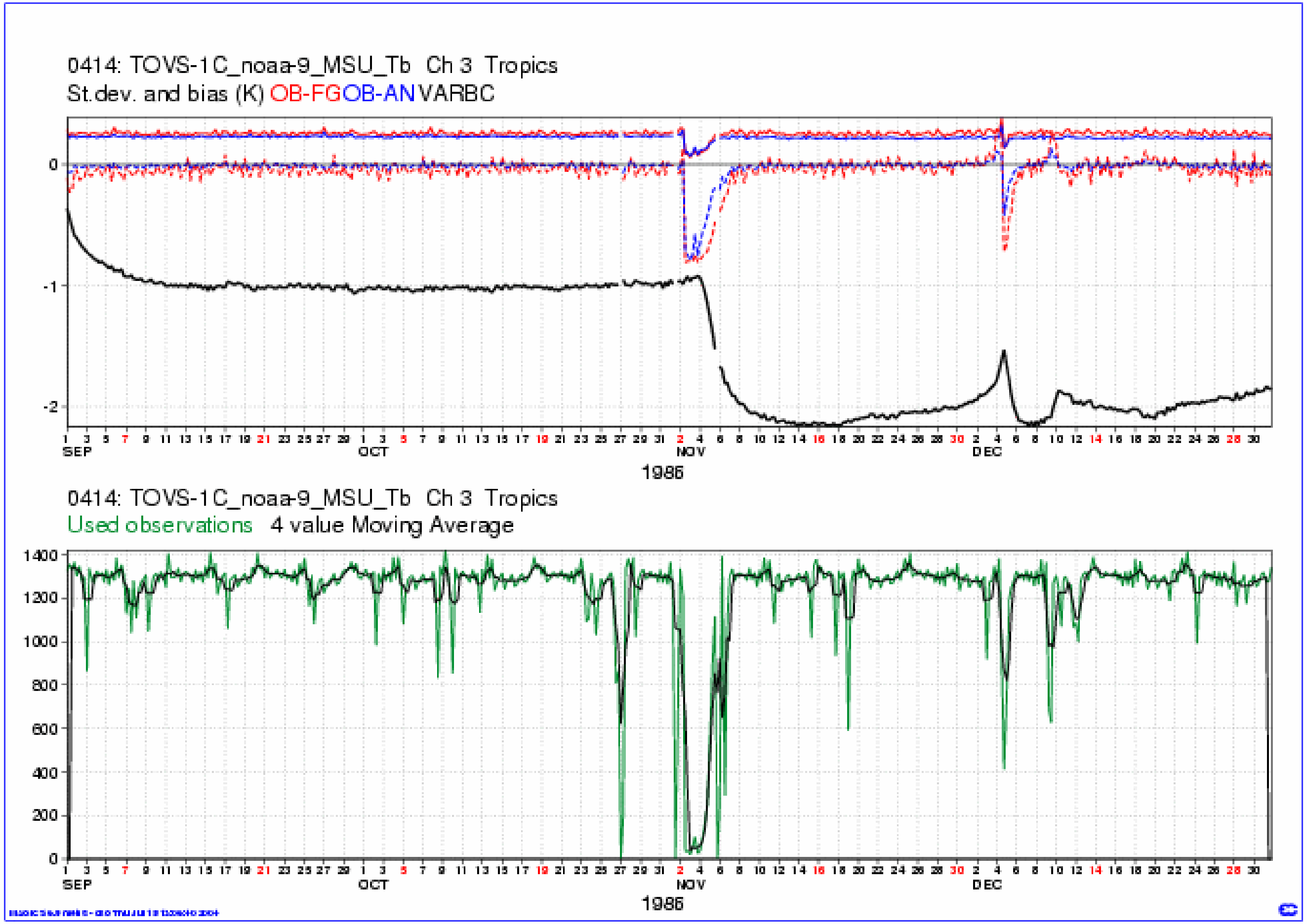
Variational correction of observation biases (VarBC)

- Radiance bias expressed in terms of a small number of parameters:
  - A constant offset
  - Predictors depending on instrument scan position
  - Predictors depending on the atmospheric state x
- Separately for each satellite/sensor/channel: $b(\beta, x) = \beta_0 + \sum_i \beta_i\, p_i$
- Add the bias parameters to the control vector in the variational analysis

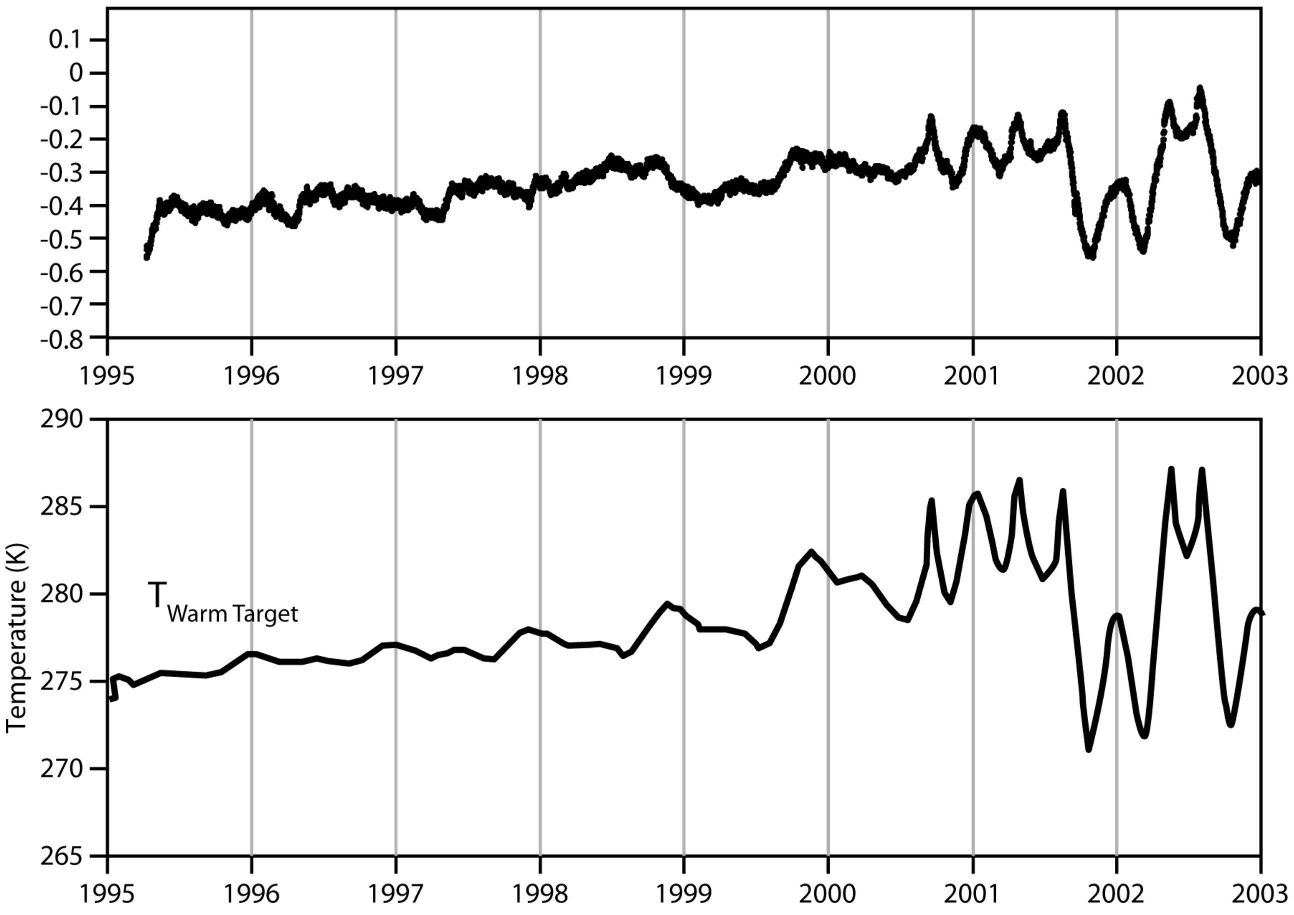
$$J(x,\beta) = \underbrace{(x_b - x)^T B_x^{-1}(x_b - x)}_{J_b\text{: background constraint for } x} + \underbrace{(\beta_b - \beta)^T B_\beta^{-1}(\beta_b - \beta)}_{J_\beta\text{: background constraint for } \beta}$$
$$+ \underbrace{[y - b_o(x,\beta) - h(x)]^T R^{-1}[y - b_o(x,\beta) - h(x)]}_{J_o\text{: bias-corrected observation constraint}}$$

- The analysis then estimates bias parameters jointly with model state variables (Derber and Wu 1998)

VarBC in ERA-Interim: NOAA-9 MSU Ch 3 (solar flare)



VarBC in ERA-Interim: NOAA-14 MSU (orbit drift)



Variational bias estimates for NOAA-14

Actual warm-target temperatures on board NOAA-14 (Grody *et al.* 2004)

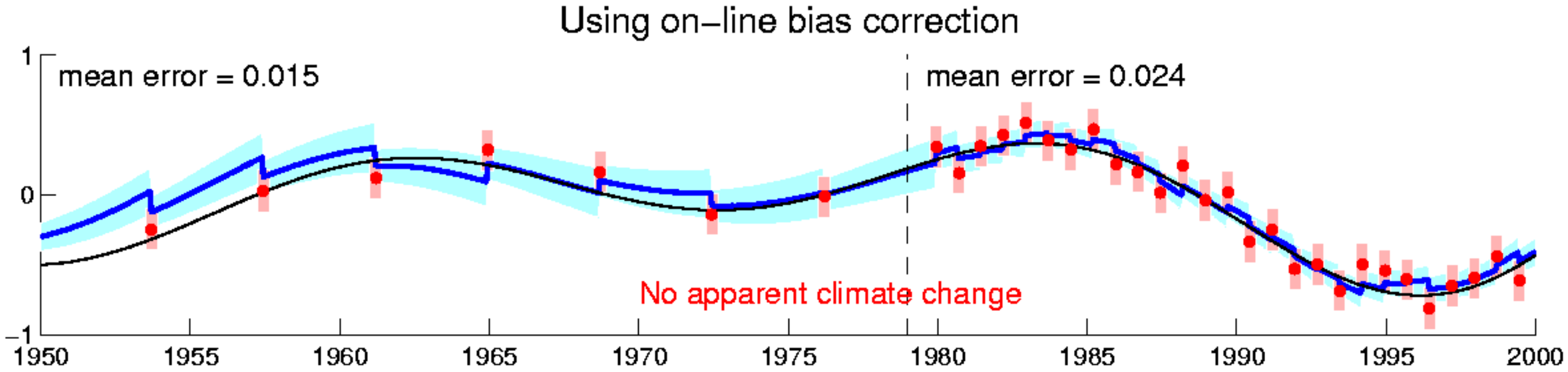
A simple sequential scheme for correcting persistent bias in the model background

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

$$\mathbf{dy} = \mathbf{y}_k - \mathbf{H}\widetilde{\mathbf{x}}$$
$$\mathbf{dx} = \mathbf{K}\mathbf{dy}$$
$$\mathbf{x}_k^a = \widetilde{\mathbf{x}} + \mathbf{dx}$$

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

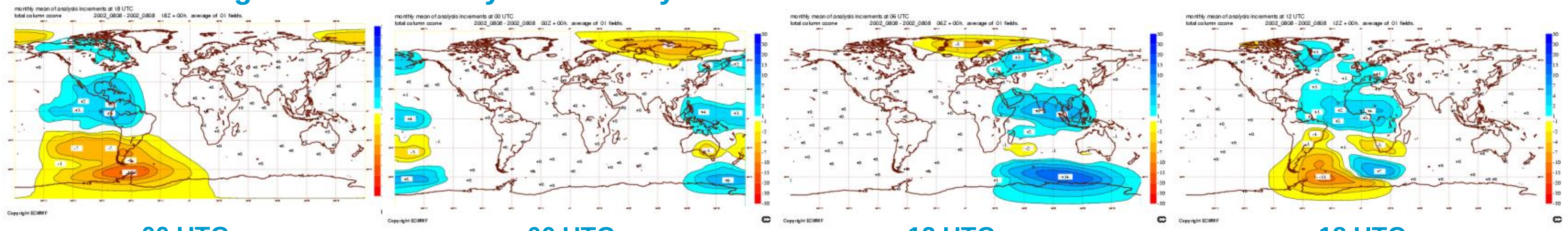
- This simple scheme is a special case of separate-bias estimation (Friedland 1969)
- Provides the Best Linear Unbiased Estimate (BLUE) in case of constant bias parameters
- Can handle observation bias parameters as well (in principle)
- Virtually cost-free and easy to implement
- BUT: purely statistical; does not attempt to correct bias at the source

Some 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)

Outline of a more general scheme for sequential model bias correction

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



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

If we can predict systematic features in the analysis increments, then we can use it to correct model bias:

Predict the analysis increment: $$dx_k^p = f_k(dx_{k-L}, \cdots, dx_{k-1})$$

Bias-aware assimilation: $$dx_k = dx_k^p + K_k(y_k - h(x_k^b + dx_k^p))$$

Assuming unbiased observations, this scheme can be shown to produce unbiased analyses.

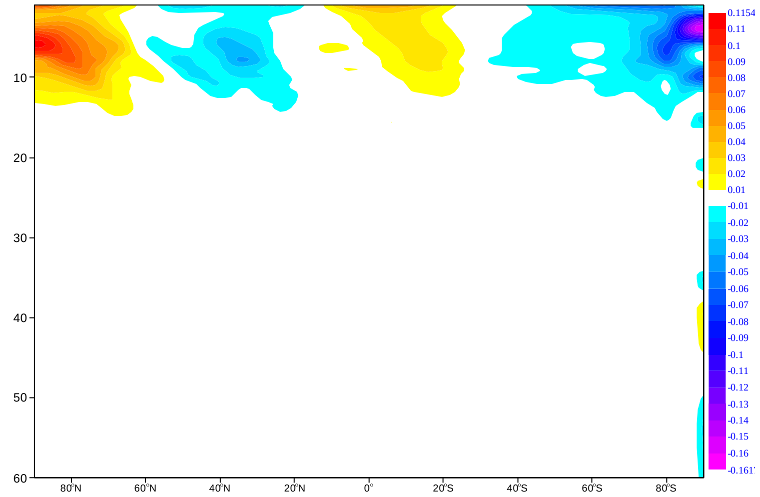
Model bias correction in weak-constraint 4D-Var

Trémolet, ECMWF

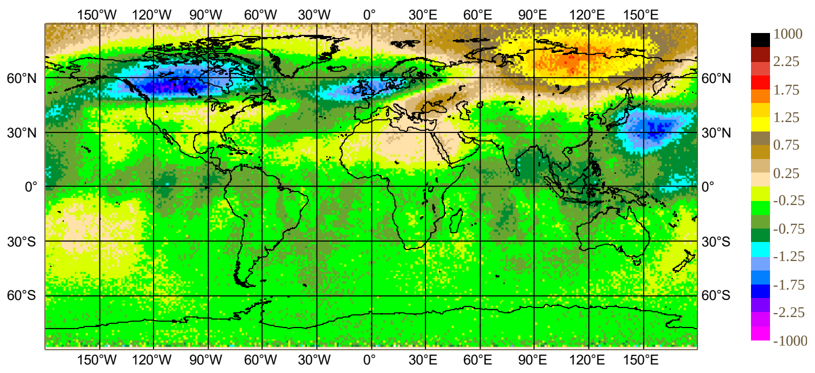
- Extend 4D-Var by including model forcing in the control vector (Derber 1989; Zupanski 1997)
- Reduce size by assuming that model error is constant for the length of the assimilation window
- Model error constraints (Q) are obtained from time series of tendency differences

- 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)

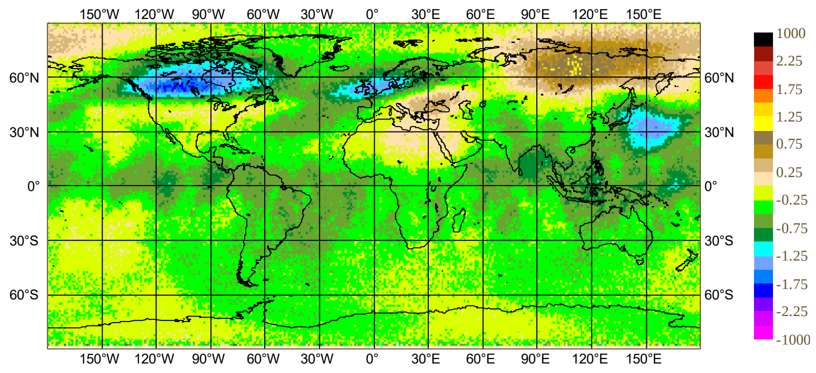
Mean Model Error Forcing - Experiment=ej5j
Dates=2004020100-2004021512 - Param=T



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



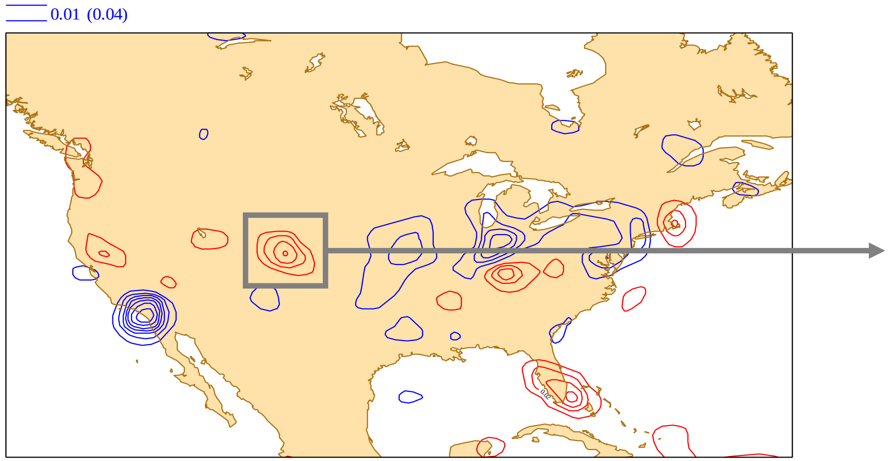
Outline

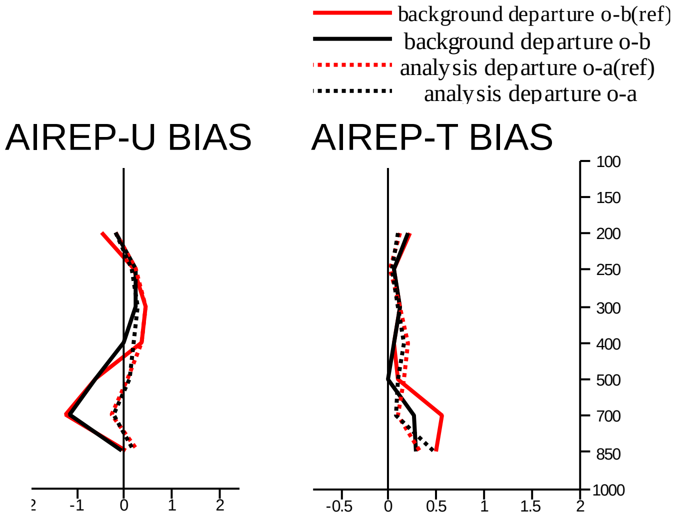
- **Bias-blind data assimilation**
  - Biases are not explicitly accounted for in ‘standard’ methods
  - Suboptimal use of observations, biased analyses, false climate signals
- **Bias-aware data assimilation**
  - Assumptions about the source
  - Variational correction of observation biases (VarBC)
  - Sequential methods for model bias correction
  - Weak-constraint 4D-Var
- **Disentangling sources of bias**
  - Model errors at major airports?
  - Some lessons learned in reanalysis

Observation bias attributed to model bias

Trémolet & Isaksen

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



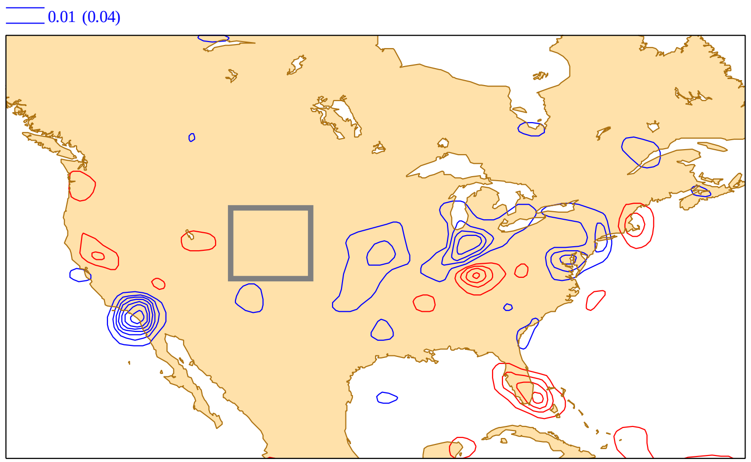


Persistent model error forcing at lower levels in the vicinity of major airports

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

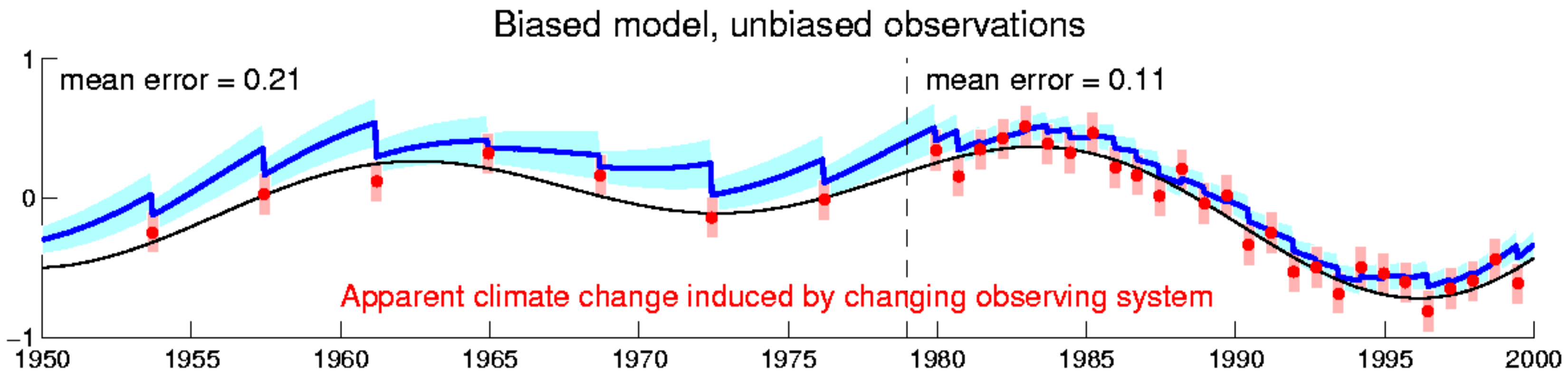
Local model error forcing disappears when all aircraft reports near Denver airport are withheld from the assimilation

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



The fundamental problem:

- Observation coverage changes over time
- Models biases are partly corrected by observations

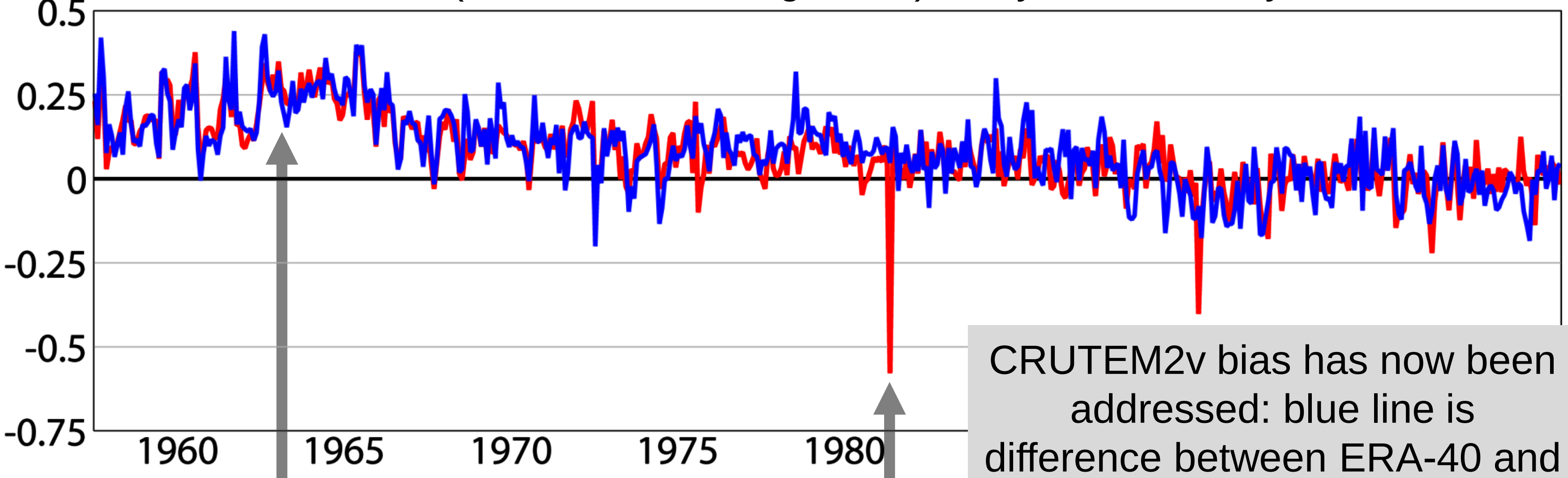


- Observations are also biased
- Data assimilation may exacerbate the problem

ERA-40 surface air temperature anomalies

Simmons et al, 2004

Red line is difference over Europe between ERA-40 analysis of synoptic data and CRUTEM2v (Jones and Moberg, 2003) analysis of monthly station data



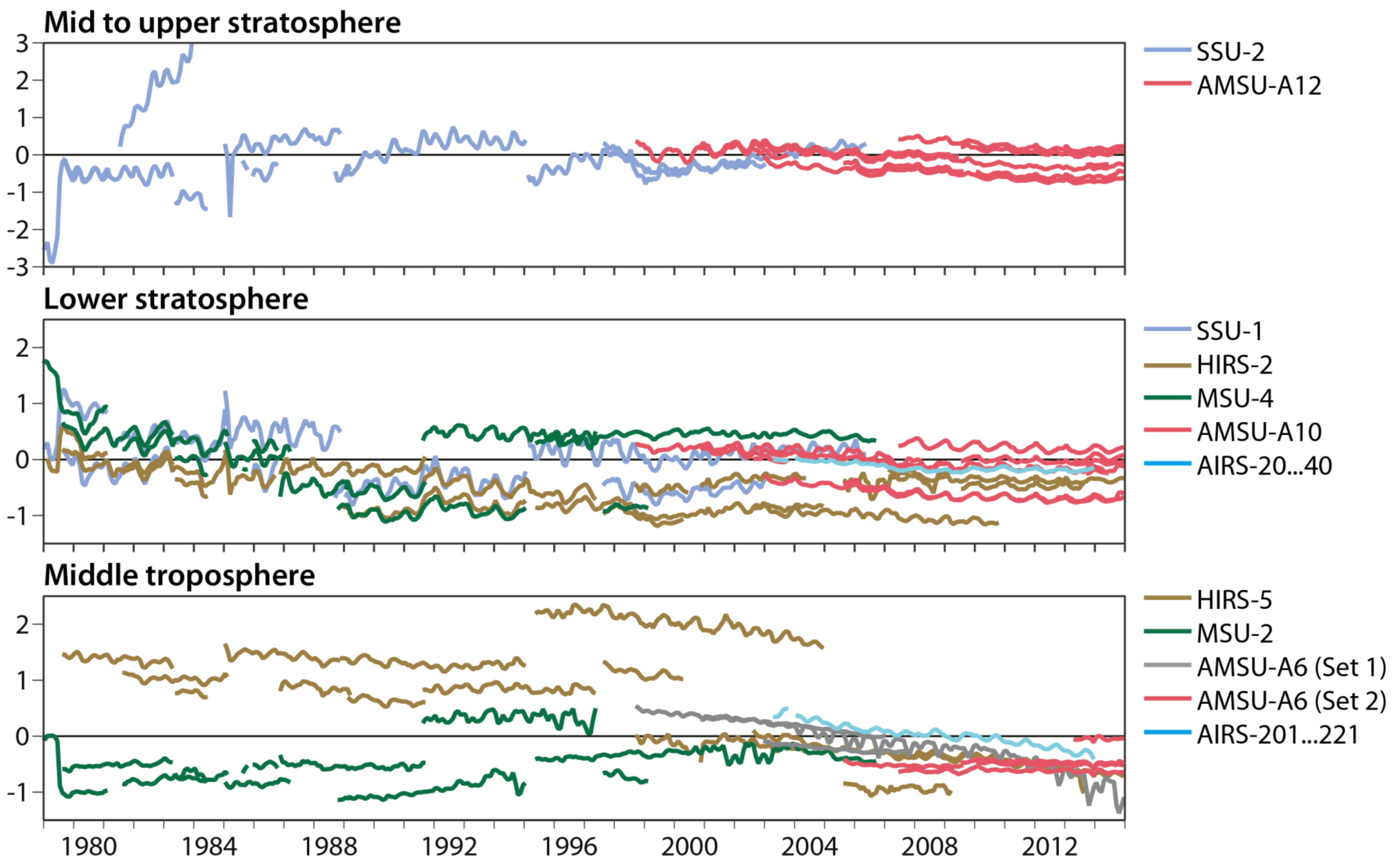
CRUTEM2v bias has now been addressed: blue line is difference between ERA-40 and CRUTEM4 (Jones *et al*., 2012)

Warm bias in ERA-40 due to model bias and insufficient SYNOP coverage to correct it

Warm bias in CRUTEM2v due to erroneous station data

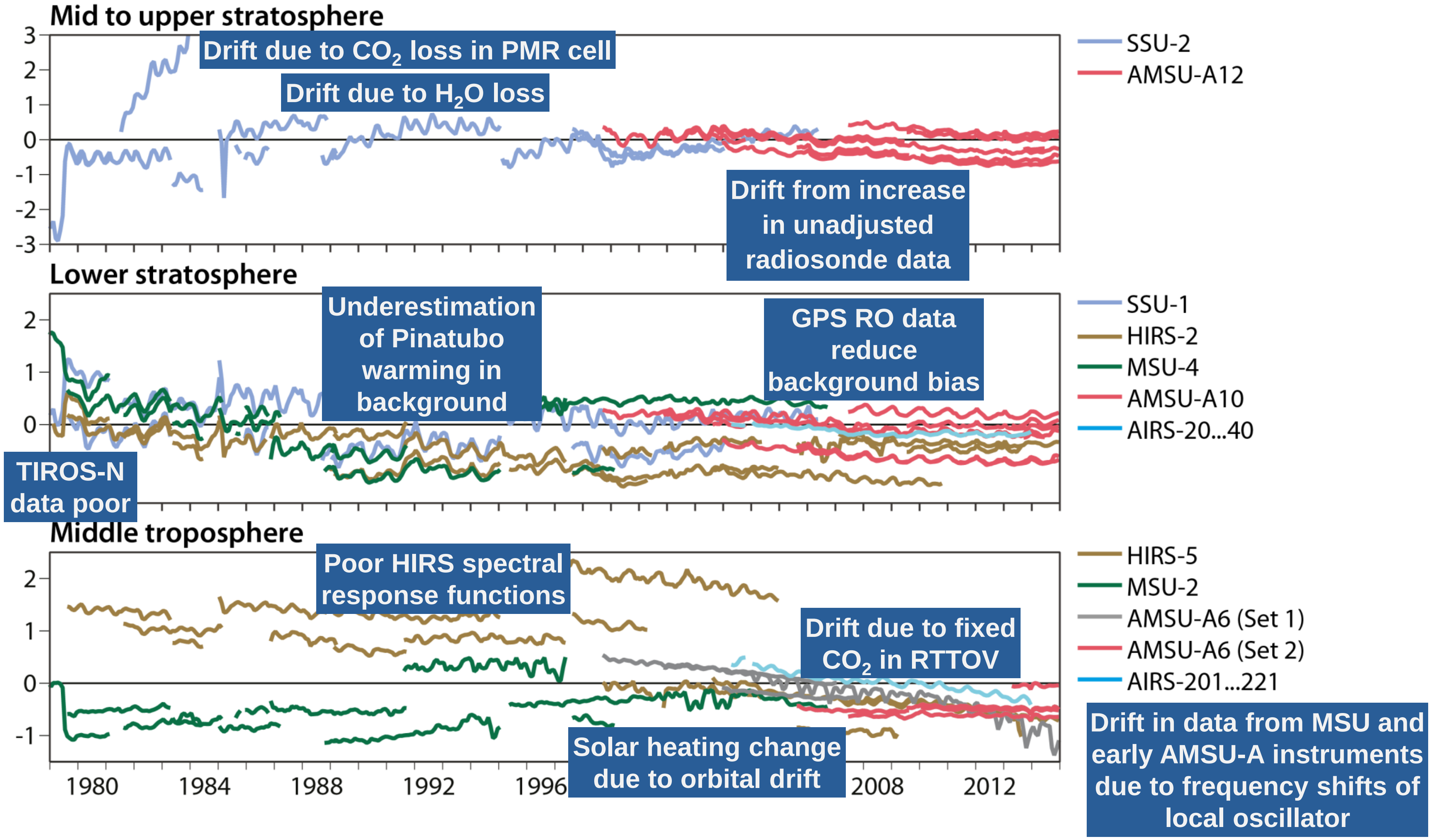
Variational bias corrections (K) in ERA-Interim

Simmons et al, 2014



Variational bias corrections (K) in ERA-Interim

Simmons et al, 2014



Cao et al. (2009), Dee and Uppala (2009), Kobayashi et al. (2009), Chung and Soden (2011), Nash and Saunders (2013), Saunders et al. (2013), Lu and Bell (2014), Simmons et al. (2014), …

Questions for the working groups

Can we use Machine Learning techniques to:

- Detect sources of biases?
- Identify bias models and predictors?

How can we improve representation of climate signals in reanalysis:

- Model bias correction methods that work in sparsely observed situations?
- Lack of constraints to anchor variational bias correction?