Bias correction in data assimilation

Hans Hersbach and Dick Dee

Reanalysis Section

ECMWF

Meteorological Training Course
Data Assimilation

17 March 2015

Overview of this lecture

In this lecture the variational bias correction scheme (VarBC) as used at ECMWF is explained. ***VarBC replaced*** the tedious job of estimating observation bias ***off-line*** for each satellite instrument or in-situ network ***by an automatic*** self-adaptive ***system***.

This is achieved by making the bias estimation an ***integral part*** of the ECMWF variational data ***assimilation*** system, where now both the initial model state and observation bias estimates are updated simultaneously.

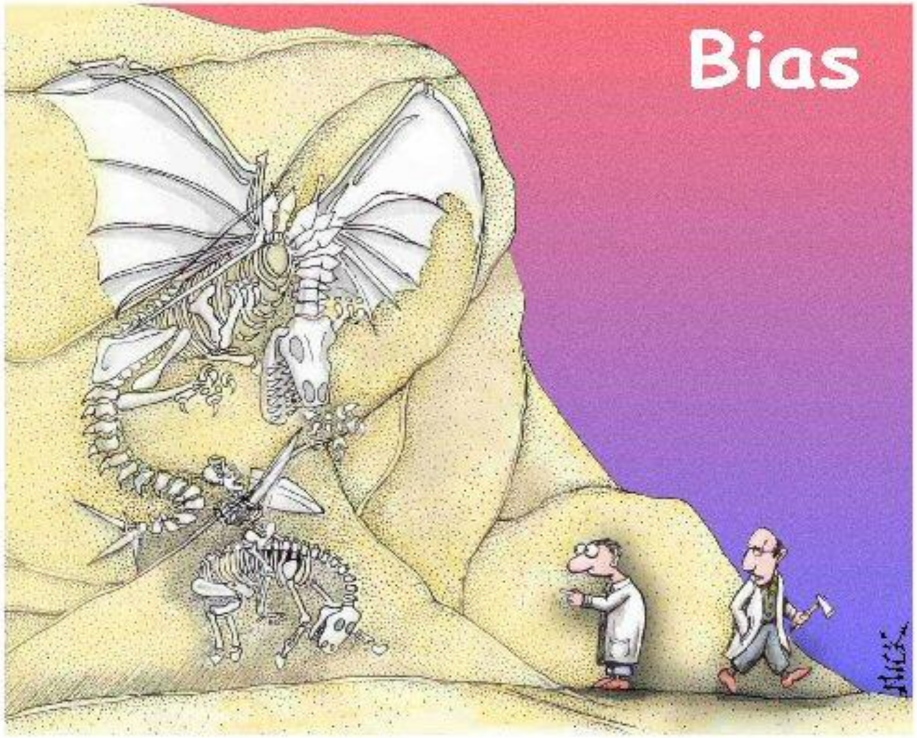
By the end of the session you should be able to realize that:

1. many observations are biased, and that the characteristics of bias ***varies widely*** between types of instruments,
2. ***separation*** between ***model*** bias and ***observation*** bias is often ***difficult***,
3. the success of an adaptive system implicitly relies on a ***redundancy*** in the underlying observing system.

Everyone knows that **models** are biased

Not everyone knows that most **observations** are biased as well

So… **where is the bias** term in this equation**?**



"Ignore it, Jeffries. It's unscientific."

$$J(\mathbf{x}) = \underbrace{(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x})}_{\text{model background constraint}} + \underbrace{[\mathbf{y} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h}(\mathbf{x})]}_{\text{observational constraint}}$$

model background constraint

observational constraint

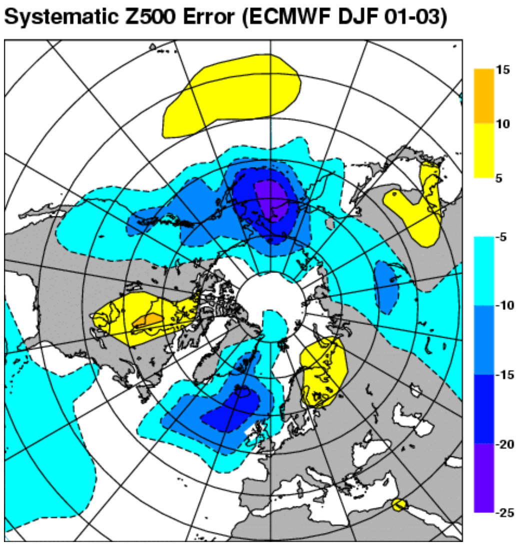
Outline

- Introduction
  - Biases in ***models***, ***observations***, and ***observation operators***
  - ***Implications*** for data assimilation
- Variational analysis and correction of observation bias
  - The need for an adaptive system
  - Variational bias correction (VarBC)
- Extension to other types of observations
- Limitations due to the effects of model bias

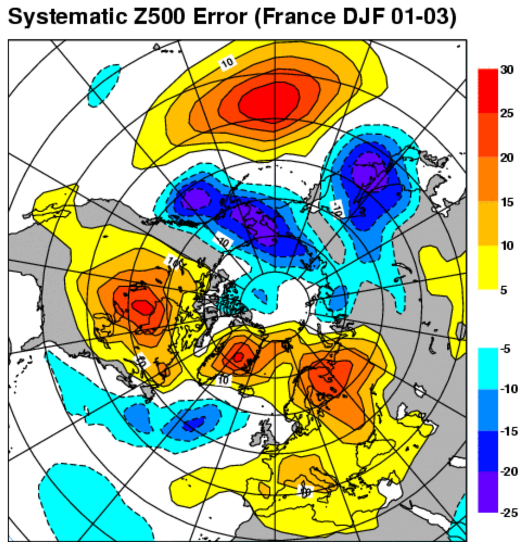
Model bias:

Systematic Day-3 Z500 errors in three different forecast models

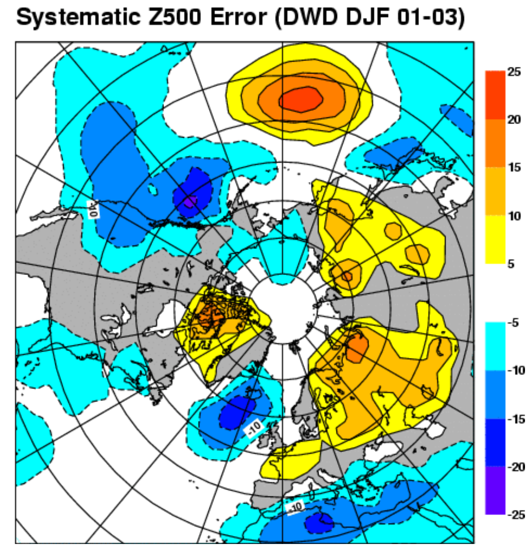
ECMWF



Meteo-France



DWD



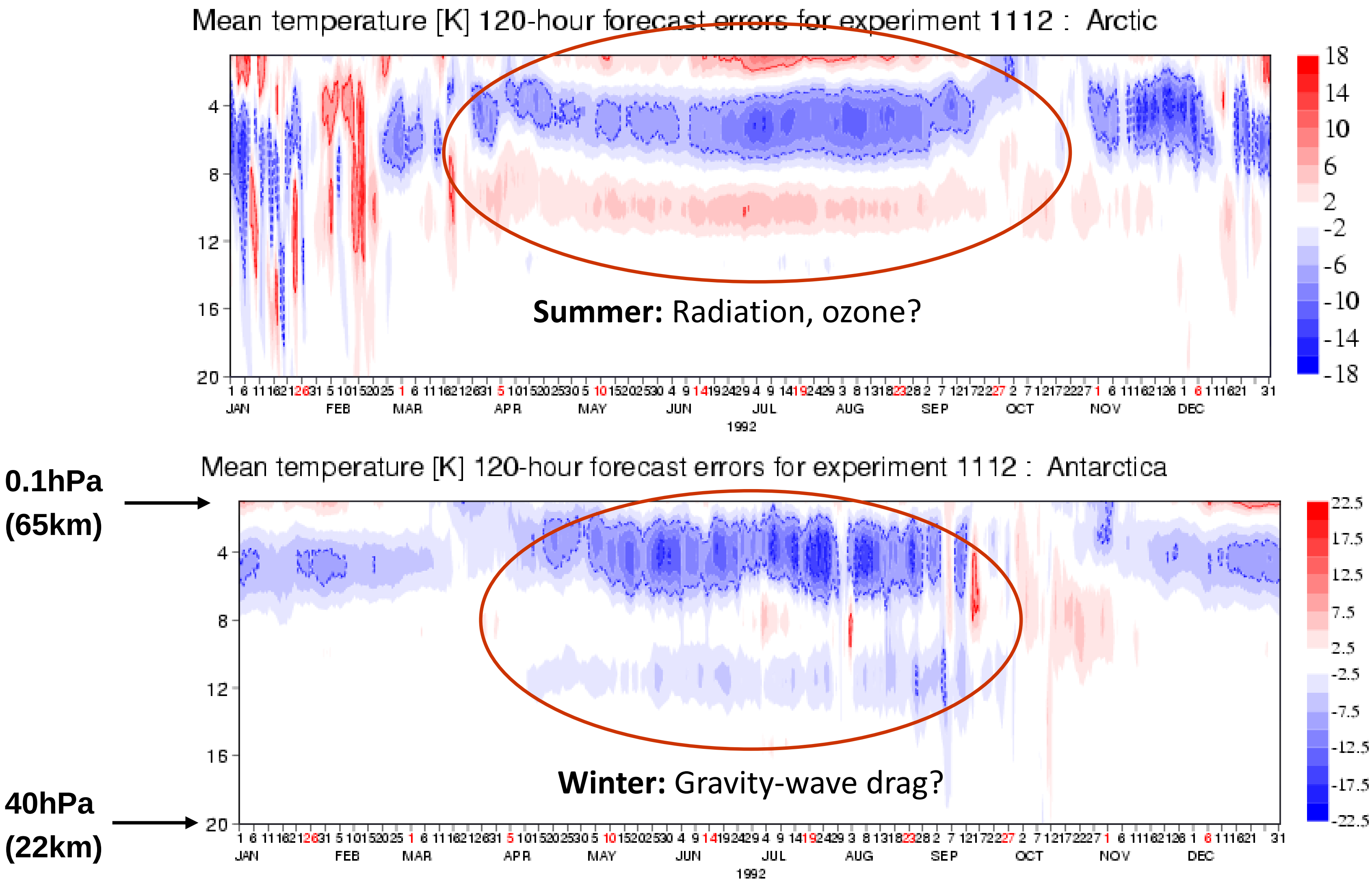
Different models often have similar error characteristics

Period DJF 2001-2003

Model bias:

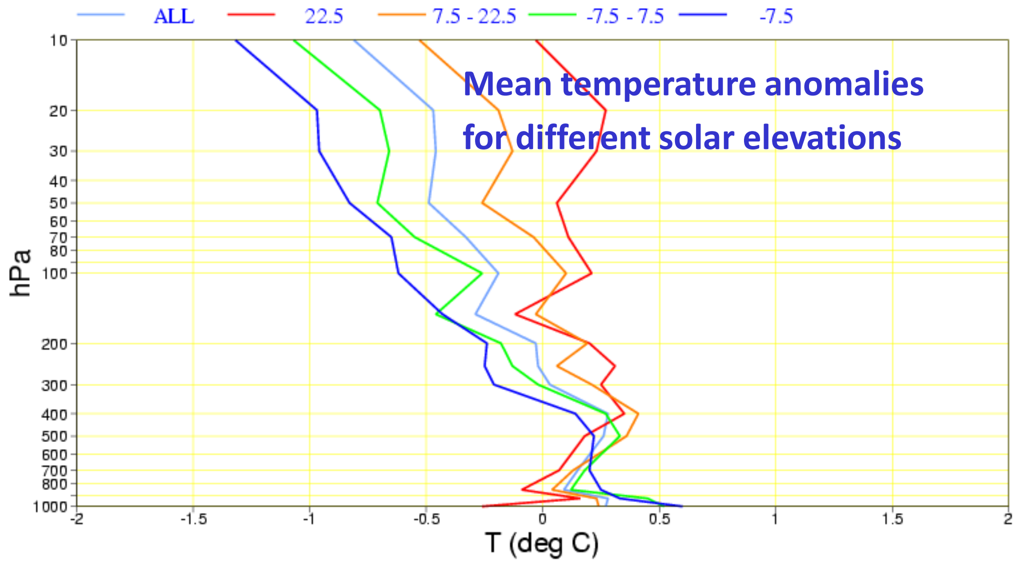
Seasonal variation in upper-stratospheric model errors

T255L60 model currently used for the ***ERA-Interim*** reanalysis

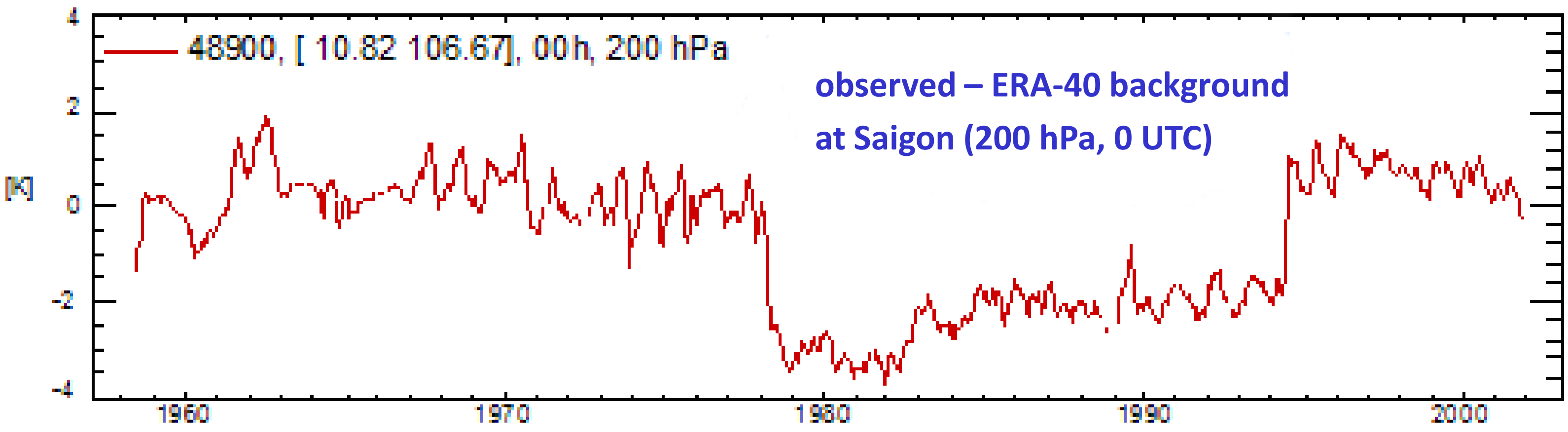


Observation bias:
Radiosonde temperature observations

Daytime warm bias due to radiative heating of the temperature sensor (depends on *solar elevation* and *equipment type*)



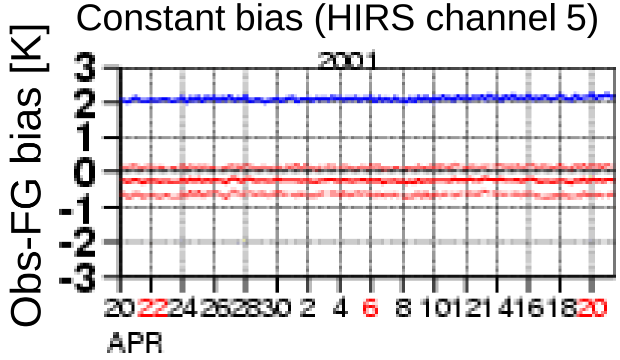
Bias changes due to change of equipment



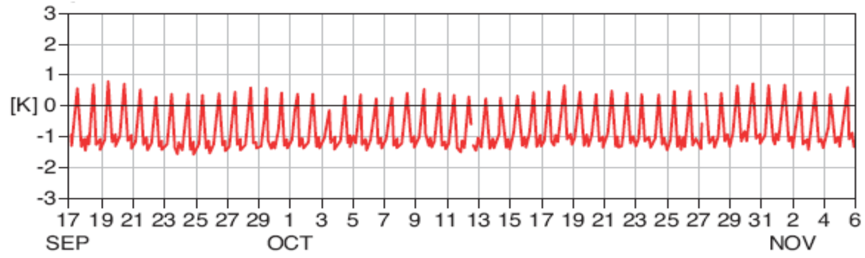
Observation and observation operator bias:
Satellite radiances

Monitoring the background departures (averaged in time and/or space):

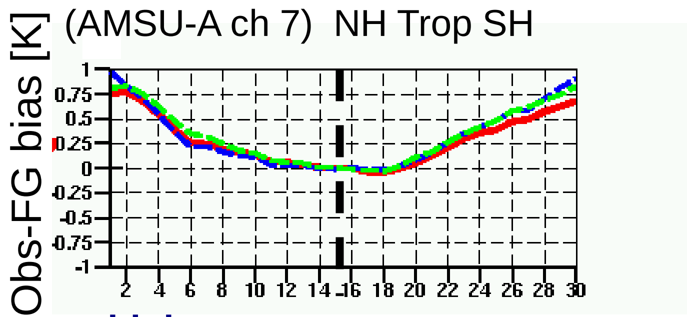
Constant bias (HIRS channel 5)



Diurnal bias variation in a geostationary satellite

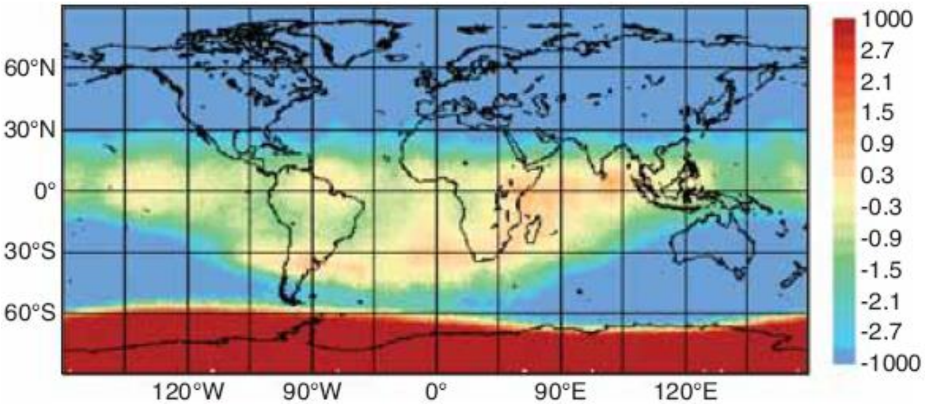


Bias depending on scan position (AMSU-A ch 7) NH Trop SH



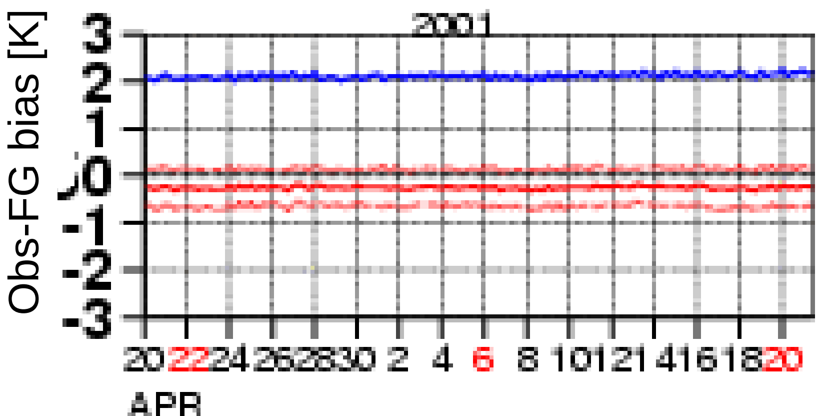
high zenith angle **nadir** **high zenith angle**

Air-mass dependent bias (AMSU-A channel 14)

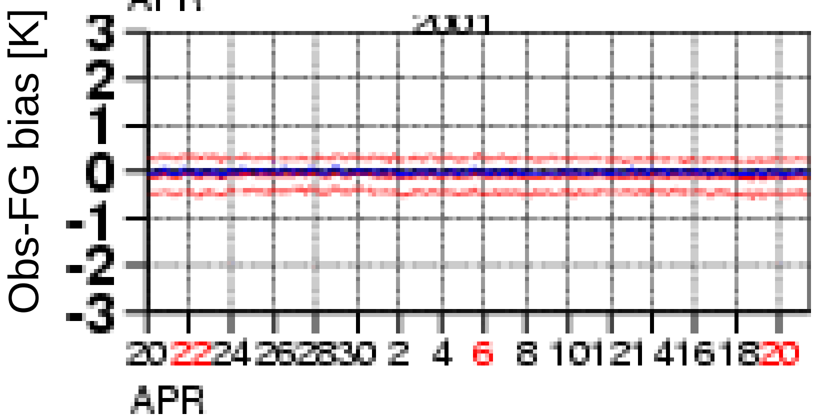


Observation and observation operator bias:
Satellite radiances

Monitoring the background departures (averaged in time and/or space):



HIRS channel 5 (peaking around 600hPa) on NOAA-14 satellite has +2.0K radiance bias against FG.

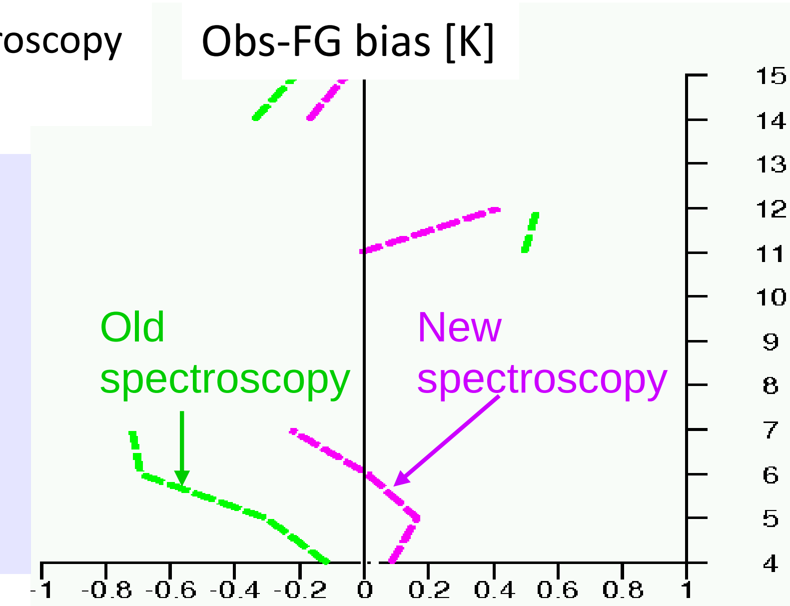


Same channel on NOAA-16 satellite has no radiance bias against FG.

→ NOAA-14 channel 5 has an instrument bias.

Observation and observation operator bias:
Satellite radiances

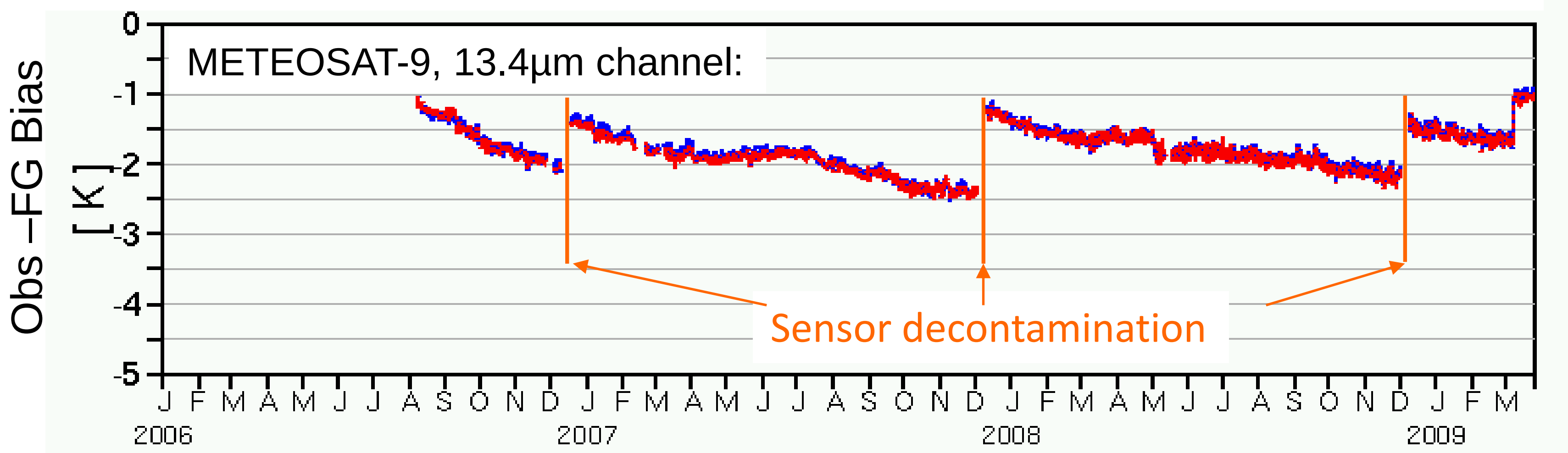
Different bias for HIRS due to change in spectroscopy used in the ***radiative transfer model***:



Other common causes for biases in radiative transfer:

- Bias in assumed concentrations of atmospheric gases (e.g., CO_2, aerosols)
- Neglected effects (e.g., clouds)
- Incorrect spectral response function
- ….

Drift in bias due to ice-build up on sensor:

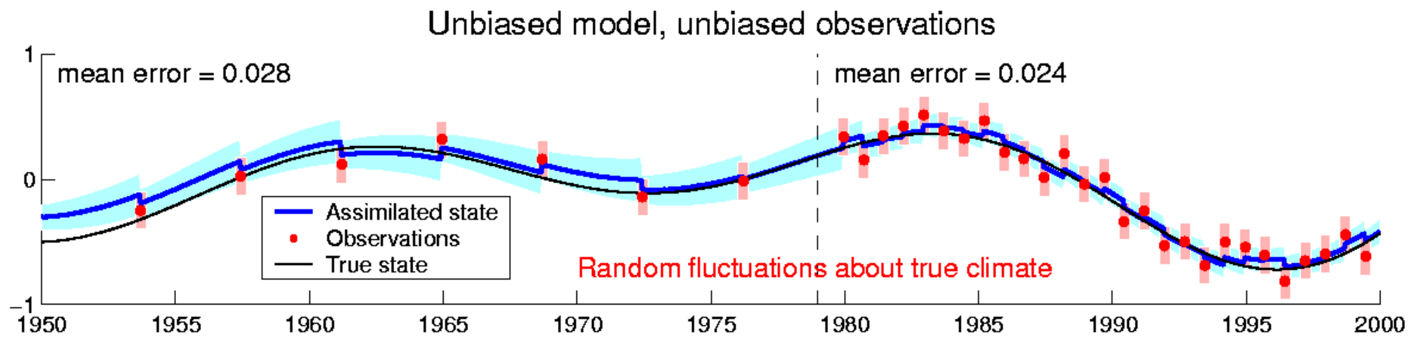


Implications for data assimilation:
Bias problems in a nutshell

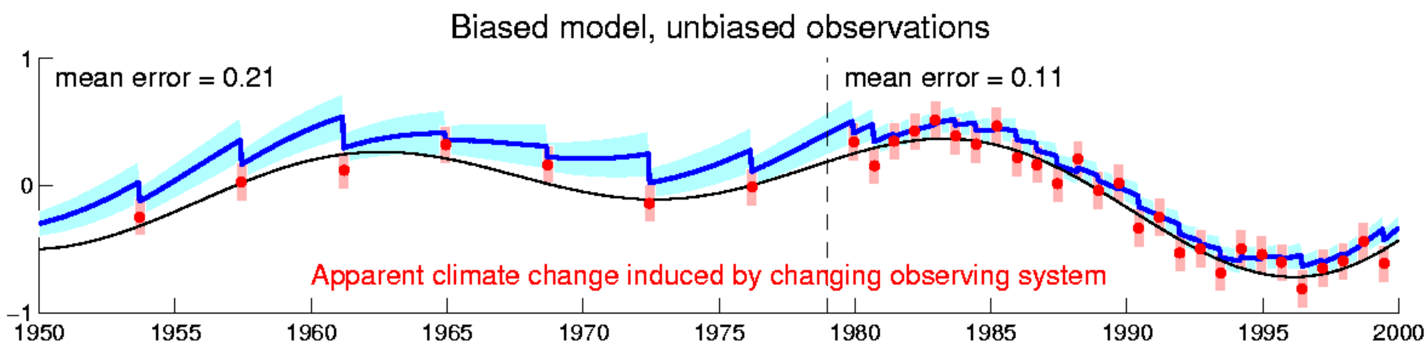
- **Observations and observation operators have biases, which may change over time**
    - Daytime warm bias in radiosonde measurements of stratospheric temperature; radiosonde equipment changes
    - Biases in cloud-drift wind data due to problems in height assignment
    - Biases in satellite radiance measurements and radiative transfer models
- **Models have biases, and changes in observational coverage over time may change the extent to which observations correct these biases**
    - Stratospheric temperature bias modulated by radiance assimilation
    - This is especially important for reanalysis (trend analysis)
- **Data assimilation methods are primarily designed to correct *small random errors* in the model background**
    - Systematic inconsistencies among different parts of the observing system lead to all kinds of problems

Implications for data assimilation:
The effect of model bias on trend estimates

Most assimilation systems assume unbiased models and unbiased data

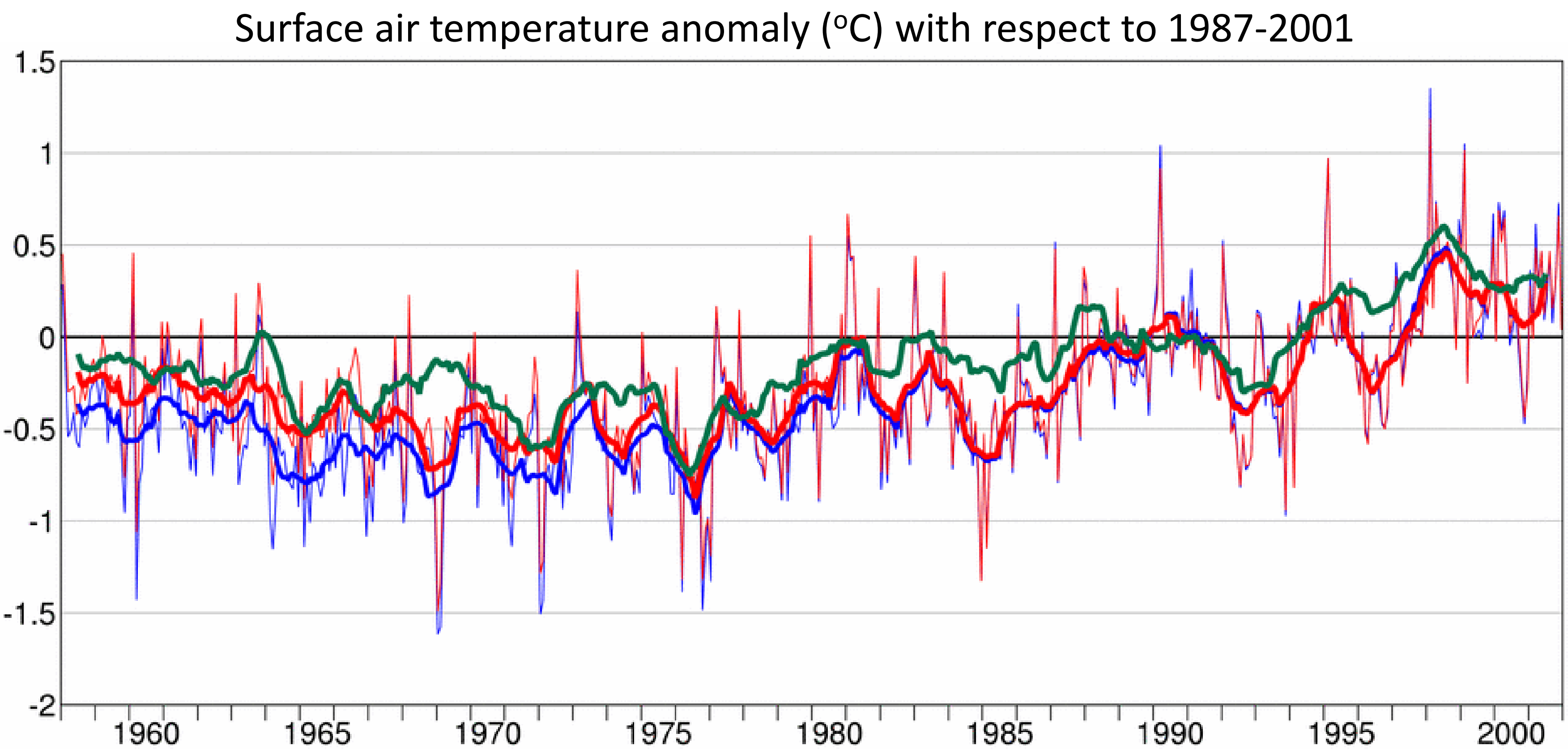


Biases in models and/or data can induce spurious trends in the assimilation



Implications for data assimilation:

ERA-40 surface temperatures compared to land-station values



Based on monthly CRUTEM2v data (Jones and Moberg, 2003)

Based on ERA-40 reanalysis

Based on ERA-40 model simulation (with SST/sea-ice data)

Outline

- Introduction
  - Biases in models, observations, and observation operators
  - Implications for data assimilation
- Variational analysis and correction of observation bias
  - The ***need*** for an adaptive system
  - The variational bias correction scheme: ***VarBC***
- Extension to other types of observations
- Limitations due to the effects of model bias

Variational analysis and bias correction:
A brief review of variational data assimilation

Minimise $$\mathbf{J(x)} = \underbrace{(\mathbf{x_b} - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x_b} - \mathbf{x})}_{\text{background constraint } (J_b)} + \underbrace{[\mathbf{y} - \mathbf{h(x)}]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h(x)}]}_{\text{observational constraint } (J_o)}$$

- The input x_b represents past information propagated by the forecast model

  (the model background)

- The input $[y - h(x_b)]$ represents the new information entering the system

  (the background departures)

- The function h(x) represents a model for simulating observations

  (the observation operator)

- Minimising the cost function J(x) produces an adjustment to the model background based on all used observations

  (the analysis)

Variational analysis and bias correction:
Error sources in the input data

Minimise $$\mathbf{J(x)} = \underbrace{(\mathbf{x_b} - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x_b} - \mathbf{x})}_{\text{background constraint } (J_b)} + \underbrace{[\mathbf{y} - \mathbf{h(x)}]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h(x)}]}_{\text{observational constraint } (J_o)}$$

- **Errors in the input [y – h(x_b)] arise from:**
  - errors in the actual observations
  - errors in the model background
  - errors in the observation operator

- **There is no general method for separating these different error sources**
  - we only have information about differences
  - there is no true reference in the real world!

- **The analysis does not respond well to conflicting input information**

  A lot of work is done to remove biases prior to assimilation:
    - ideally by removing the cause
    - in practise by careful comparison against other data

Satellite radiance bias correction at ECMWF, prior to 2006

Scan bias and **air-mass dependent bias** for each satellite/sensor/channel were estimated off-line from background departures, and stored in files **(Harris and Kelly 2001)**

Error model for brightness temperature data: $y = h(x) + b^{scan} + b^{air}(x) + e^{obs}$

where

$$b^{scan} = b^{scan}(\text{latitude, scan position})$$

$$b^{air} = \beta_0 + \sum_{i=1}^{N} \beta_i p_i(x)$$

$$e^{obs} = \text{random observation error}$$

Predictors, for instance:

- 1000-300 hPa thickness
- 200-50 hPa thickness
- surface skin temperature
- total precipitable water

Average the background departures:

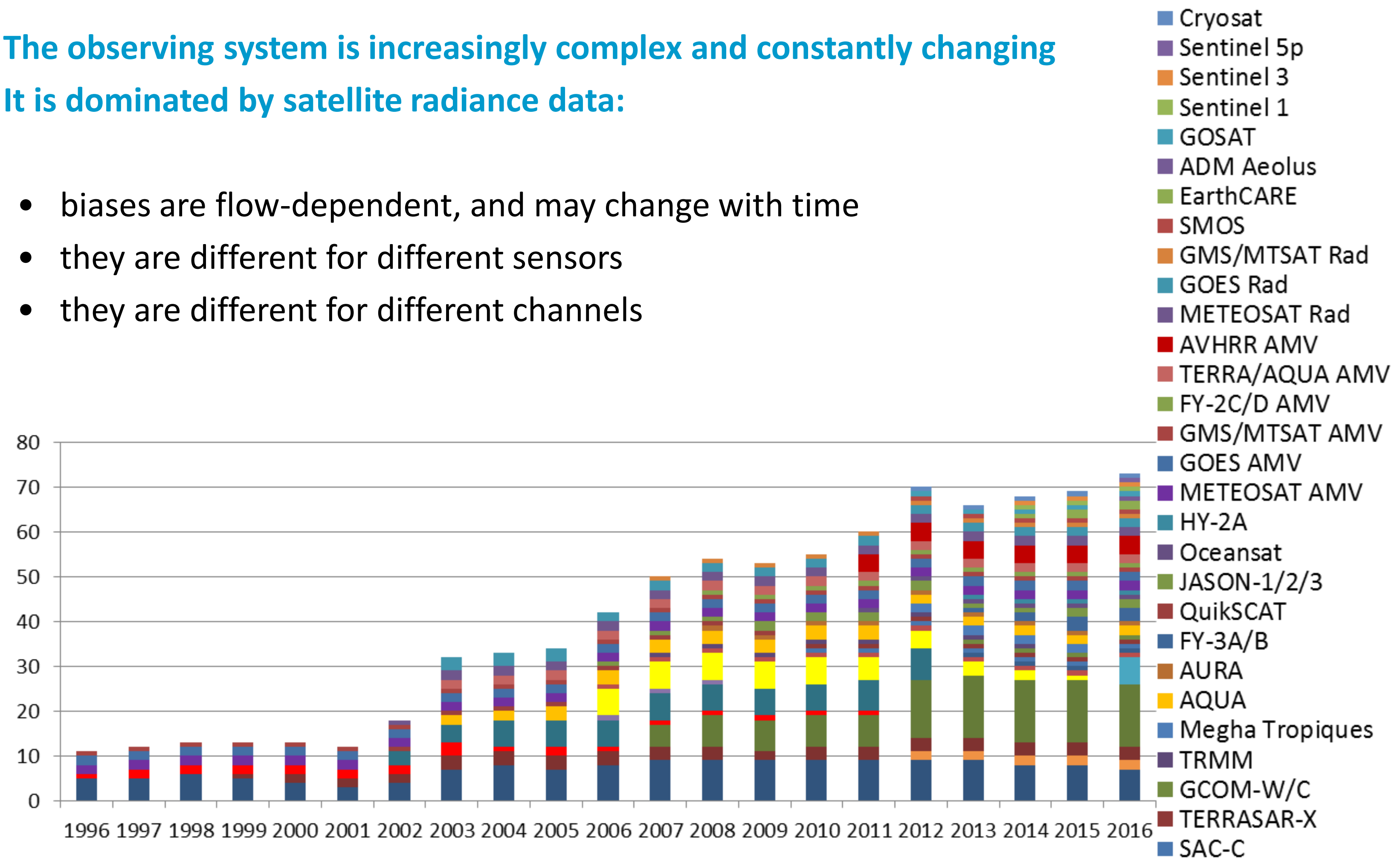
$$\langle y - h(x_b) \rangle = b^{scan} + b^{air}(x)$$

Periodically estimate scan bias and predictor coefficients:

- typically 2 weeks of background departures
- 2-step regression procedure
- careful masking and data selection

The need for an adaptive bias correction system

- **The observing system is increasingly complex and constantly changing**
- **It is dominated by satellite radiance data:**
  - biases are flow-dependent, and may change with time
  - they are different for different sensors
  - they are different for different channels



- **How can we manage the bias corrections for all these different components?**
- **This requires a consistent approach and a flexible, automated system**

The Variational bias correction scheme:

The general idea

The **bias** in a given instrument/channel (**bias group**) is described by (a few) **bias parameters:** typically, these are functions of air-mass and scan-position (the **predictors**)

These parameters can be estimated in a variational analysis along with the model state **(Derber and Wu, 1998 at NCEP, USA)**

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: $z^T = [\, x^T \;\; \beta^T \,]$

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

What is needed to implement this:

1. The modified operator $\tilde{h}(x, \beta)$ and its TL + adjoint
2. A cycling scheme for updating the bias parameter estimates
3. An effective preconditioner for the joint minimization problem

Variational bias correction:
The modified analysis problem

The original problem:

J_b: background constraint

$$J(\mathbf{x}) = (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}) + [\mathbf{y} - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{h}(\mathbf{x})]$$

J_o: observation constraint

The modified problem:

J_b: background constraint for x

J_β: background constraint for β

$$\mathbf{J}(\mathbf{x}, \boldsymbol{\beta}) = (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}_x^{-1} (\mathbf{x}_b - \mathbf{x}) + (\boldsymbol{\beta}_b - \boldsymbol{\beta})^T \mathbf{B}_\beta^{-1} (\boldsymbol{\beta}_b - \boldsymbol{\beta}) + [\mathbf{y} - \mathbf{b}_o(\mathbf{x}, \boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{b}_o(\mathbf{x}, \boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]$$

Parameter estimates from previous analysis

J_o: bias-corrected observation constraint

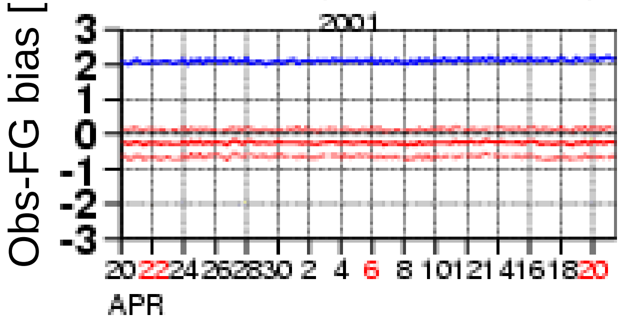
Form inspired by Harris and Kelly

The need for an adequate bias model

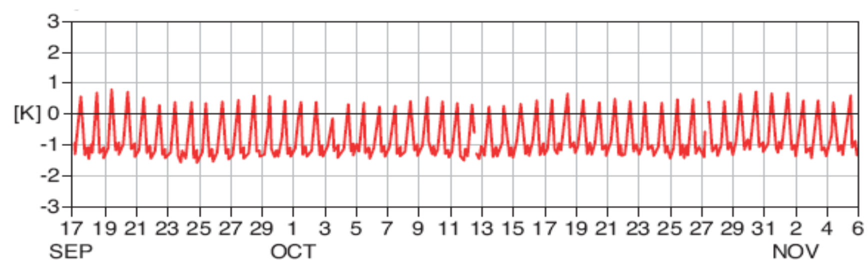
Prerequisite for any bias correction is a good model for the bias (b(x,β)):

- Ideally, guided by the physical origins of the bias.
- In practice, bias models are derived empirically from observation monitoring.

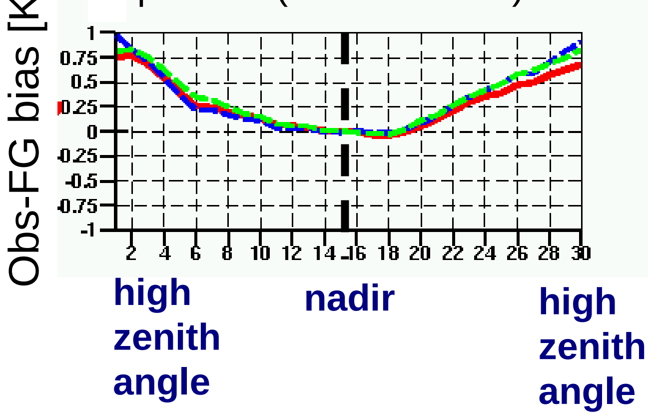
Constant bias (HIRS channel 5)



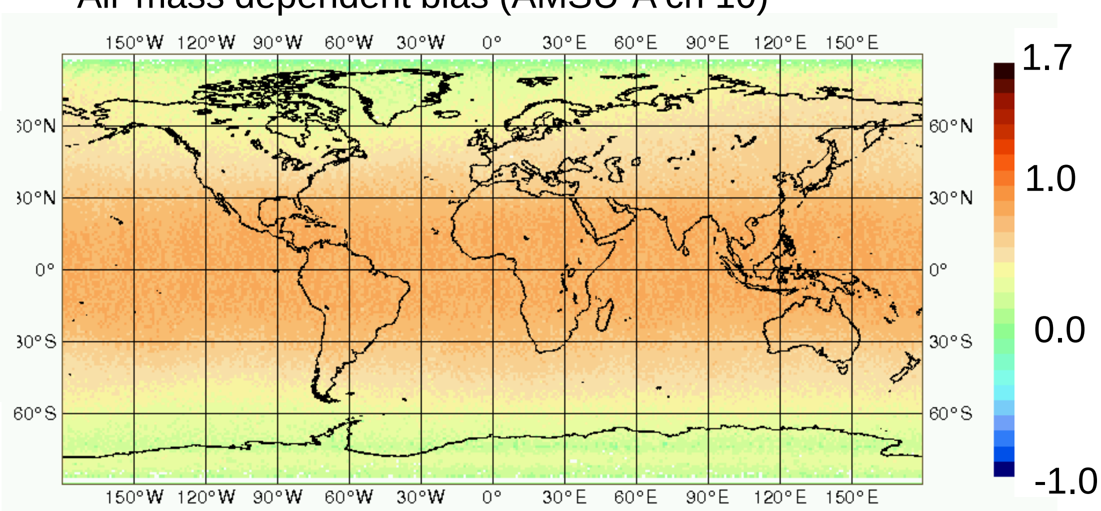
Diurnal bias variation in a geostationary satellite



Bias depending on scan position (AMSU-A ch 7)

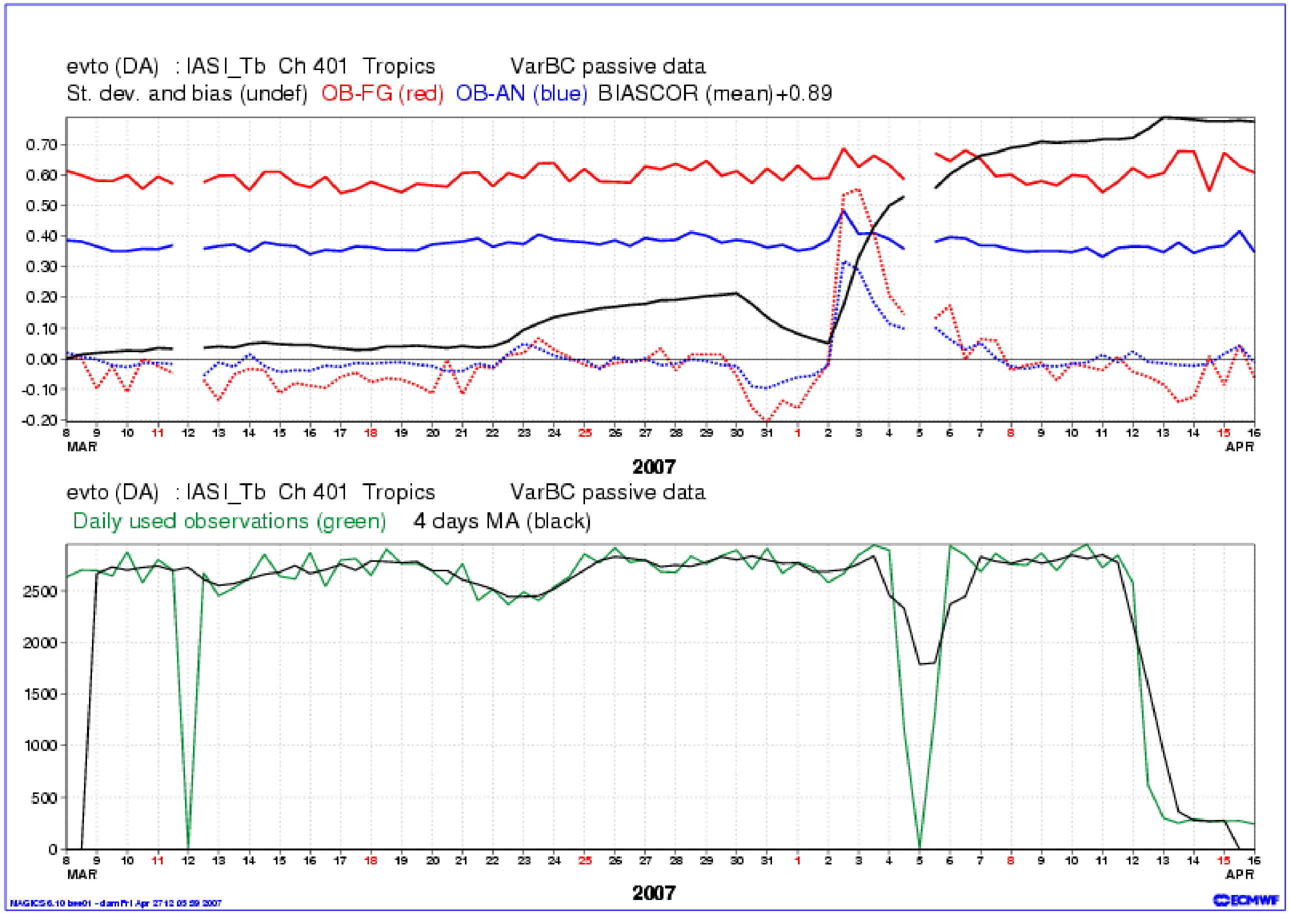


Air-mass dependent bias (AMSU-A ch 10)



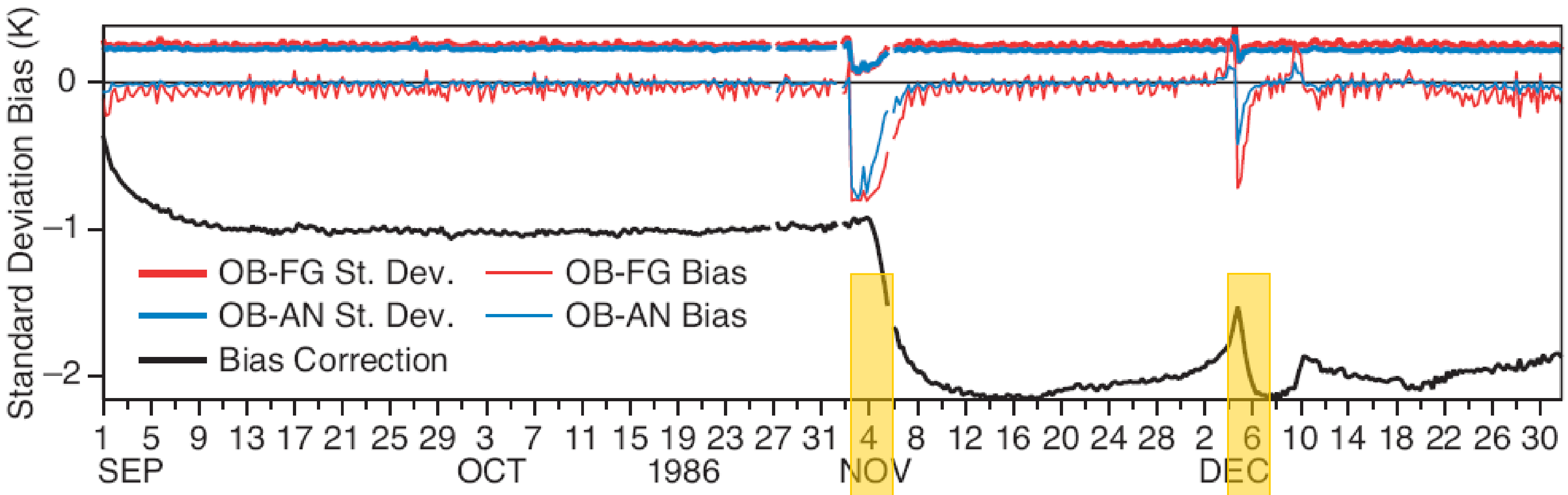
Example 1:
Spinning up new instruments – IASI on MetOp A

- IASI is a high-resolution interferometer with 8461 channels
- Initially unstable – data gaps, preprocessing changes

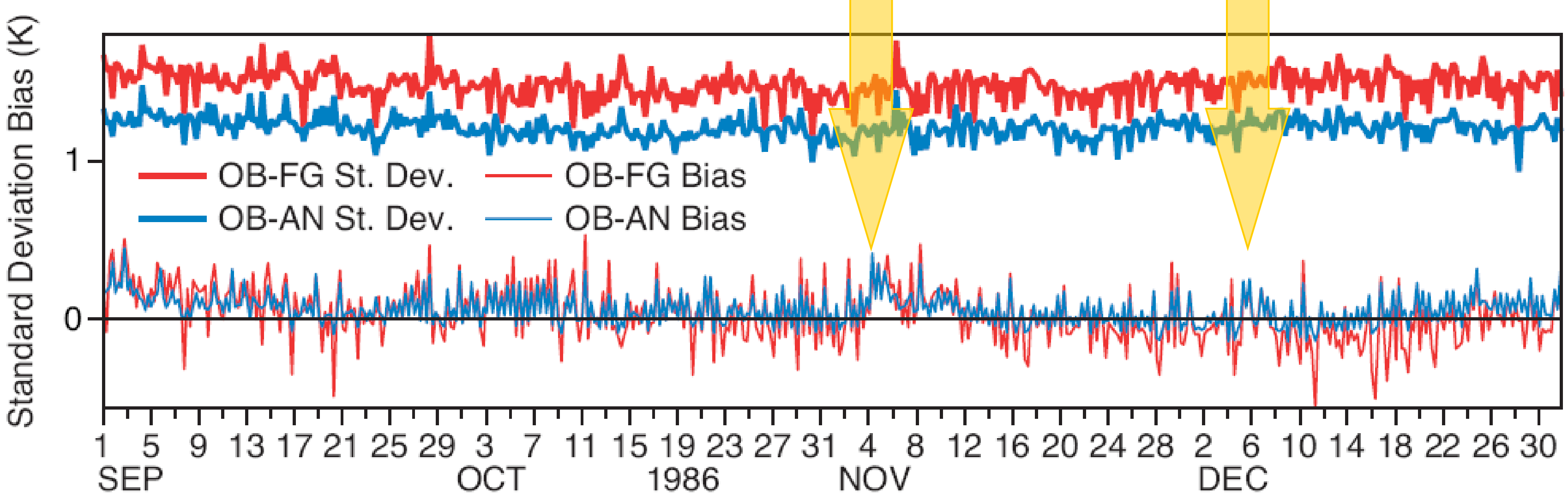


Example 2:

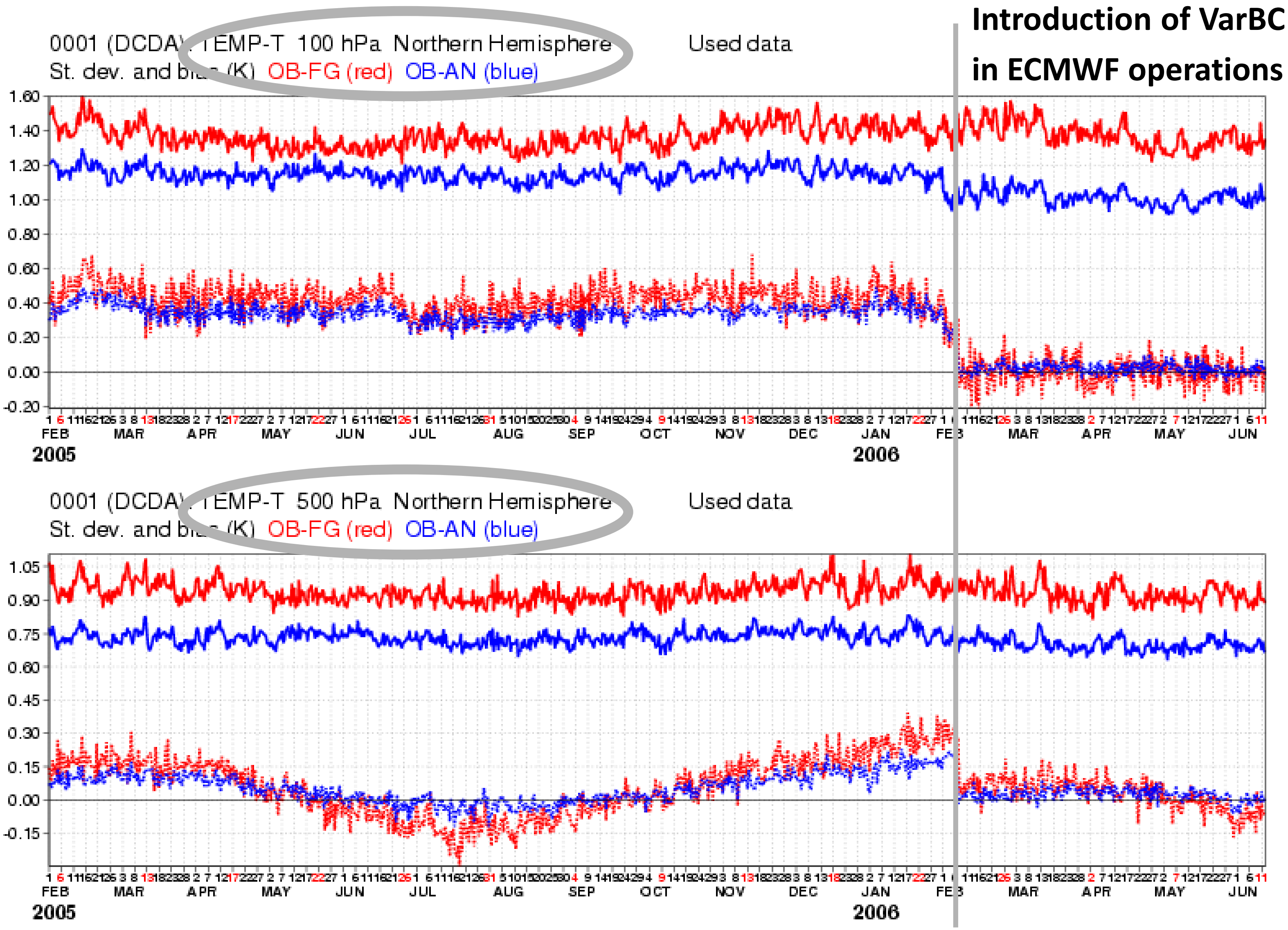
NOAA-9 MSU channel 3 bias corrections (cosmic storm)



200 hPa temperature departures from radiosonde observations



Example 3:
Fit to conventional data



Outline

- Introduction
  - Biases in models, observations, and observation operators
  - Implications for data assimilation
- Variational analysis and correction of observation bias
  - The need for an adaptive system
  - Variational bias correction (VarBC)
- ***Extension*** to other types of observations
- Limitations due to the effects of model bias

Extension to other types of observations

Current bias 'classes' in the ECMWF operational system:

- ***Radiances***: clear sky/all sky, infrared/microwave, polar/geostationary
- ***Total column ozone***: currently only OMI
- ***Aircraft data:*** one group per aircraft
- ***Total column water vapour***: ENVISAT MERIS until April 2012
- ***Ground-based radar precipitation***: one group embracing US stations

Other automated bias corrections, but outside 4D-Var:

- Surface pressure
- Radiosonde temperature and humidity
- Soil moisture (in SEKF surface analysis)

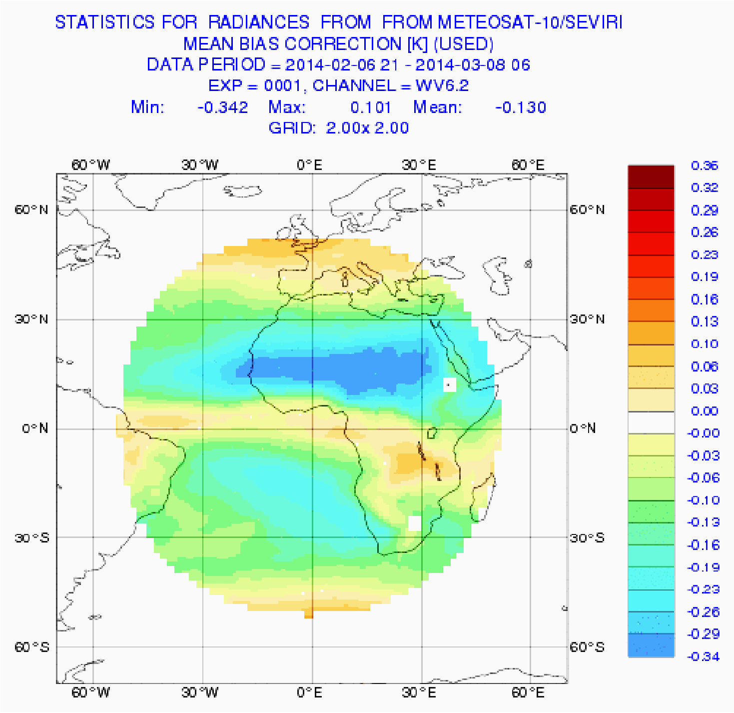
Specific:

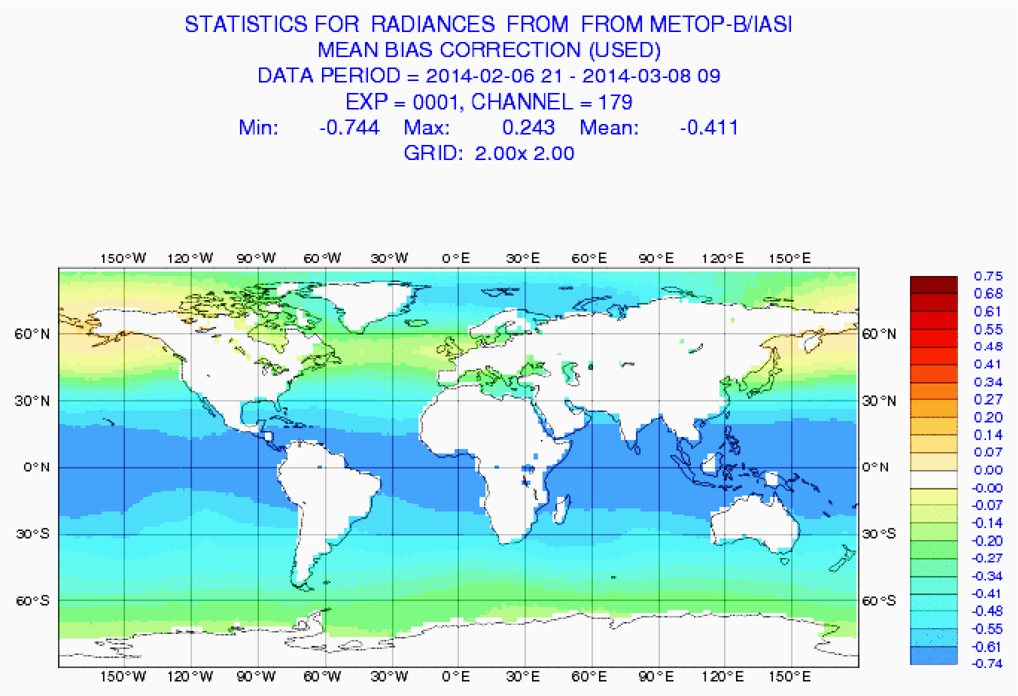
- ***ERA-Interim***: VarBC for radiances only
- ***ERA-20C:*** the 20th century reanalysis using surface observations only
- ***MACC***: atmospheric composition

VarBC for satellite *radiances*

- ***~1,500*** channels (~40 sensors on ~25 different satellites)
- ***Anchored*** to each other, GPS-RO, and all conventional observations
- Bias model: $\beta_0 + \sum\beta_i\, p_i(\text{model state}) + \sum\beta_j\, p_j(\text{instrument state})$

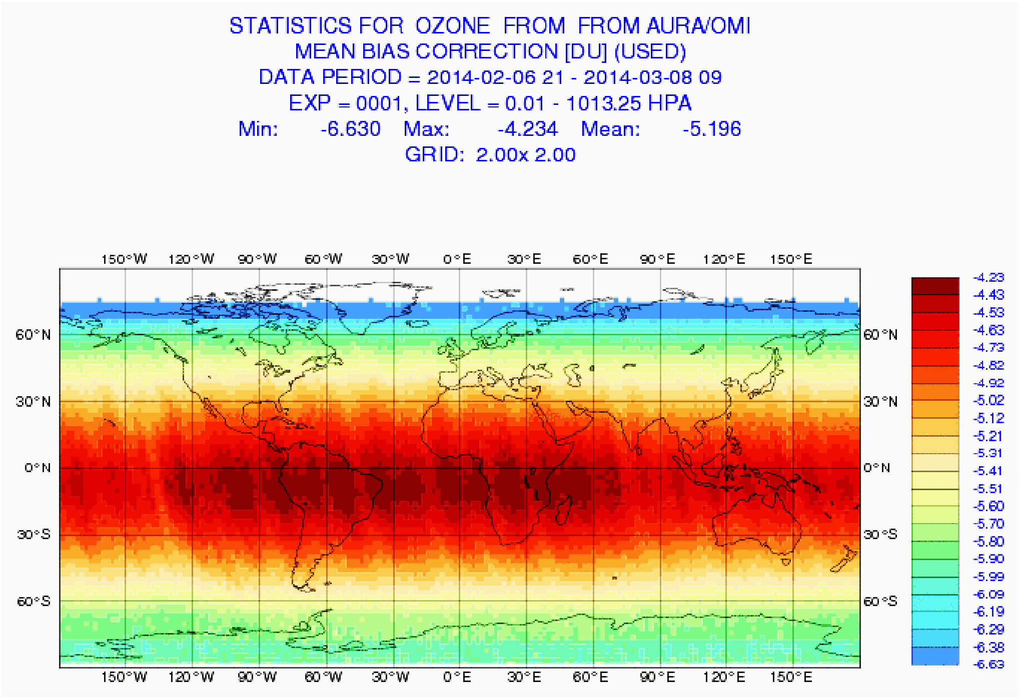
(~11,400 parameters in total)





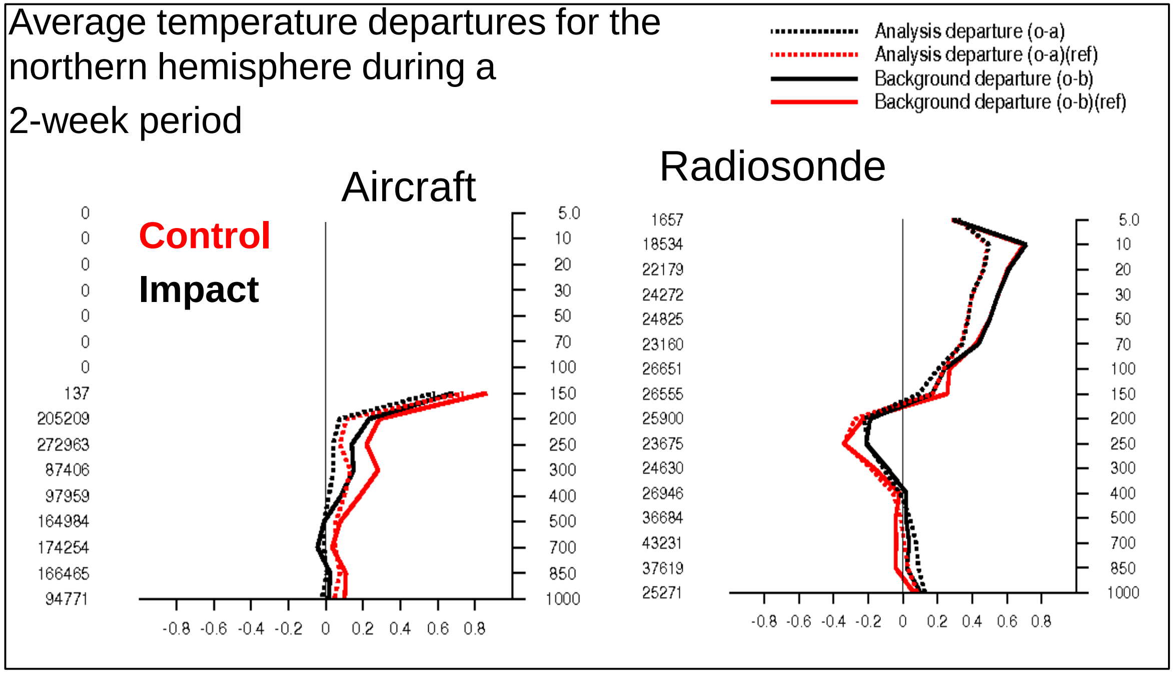
VarBC for ***ozone***

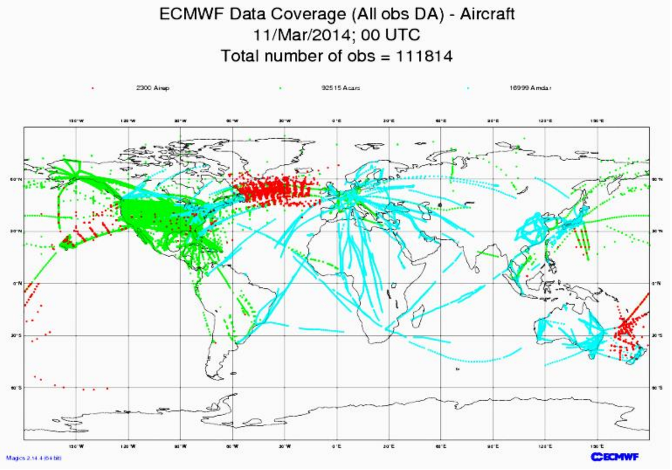
- OMI, (SCIAMACHY, GOMOS, SEVIRI, GOME2, GOME in past)
- ***Anchored*** to SBUV/2
- Bias model: $\beta_0 + \beta_1$ x solar elevation



VarBC for ***aircraft temperature***

- For each aircraft separately (~5000 distinct aircraft)
- ***Anchored*** to all temperature-sensitive observations
- Bias model: $\beta_0 + \beta_1 \times \text{ascent rate} + \beta_2 \times \text{descent rate}$





Outline

- Introduction
  - Biases in models, observations, and observation operators
  - Implications for data assimilation
- Variational analysis and correction of observation bias
  - The need for an adaptive system
  - Variational bias correction (VarBC)
- Extension to other types of observations
- Limitations due to the effects of model bias

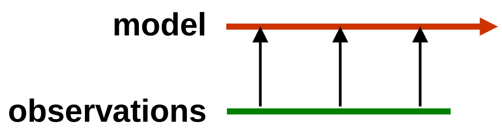
Limitations of VarBC:
Interaction with model bias

VarBC introduces extra degrees of freedom in the variational analysis, to help improve the fit to the (bias-corrected) observations:

$$\mathbf{J}(\mathbf{x},\boldsymbol{\beta}) = (\mathbf{x}_b - \mathbf{x})^T \mathbf{B}_x^{-1} (\mathbf{x}_b - \mathbf{x}) + (\boldsymbol{\beta}_b - \boldsymbol{\beta})^T \mathbf{B}_\beta^{-1} (\boldsymbol{\beta}_b - \boldsymbol{\beta}) + [\mathbf{y} - \mathbf{b}(\mathbf{x},\boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{b}(\mathbf{x},\boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]$$

It works well (even if the model is biased) when the analysis is strongly constrained by observations:

It does not work as well when there are large model biases and few observations to constrain them:



VarBC is not designed to correct model biases: Need for a **weak-constraints 4D-Var (Trémolet)**

Summary

Biases are everywhere:

- Most observations cannot be usefully assimilated without bias adjustments

- Off-line bias tuning for satellite data is practically impossible
- Bias parameters can be estimated and adjusted during the assimilation, using all available information

- Variational bias correction works best in situations where:
  - there is sufficient redundancy in the data; or
  - there are no large model biases

Challenges:

- How to develop good bias models for observations
- How to separate observation bias from model bias

Additional information

Harris and Kelly, 2001: **A satellite radiance-bias correction scheme for data assimilation.** *Q. J. R. Meteorol. Soc., 127, 1453-1468*

Derber and Wu, 1998: **The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system.** *Mon. Wea. Rev., 126, 2287-2299*

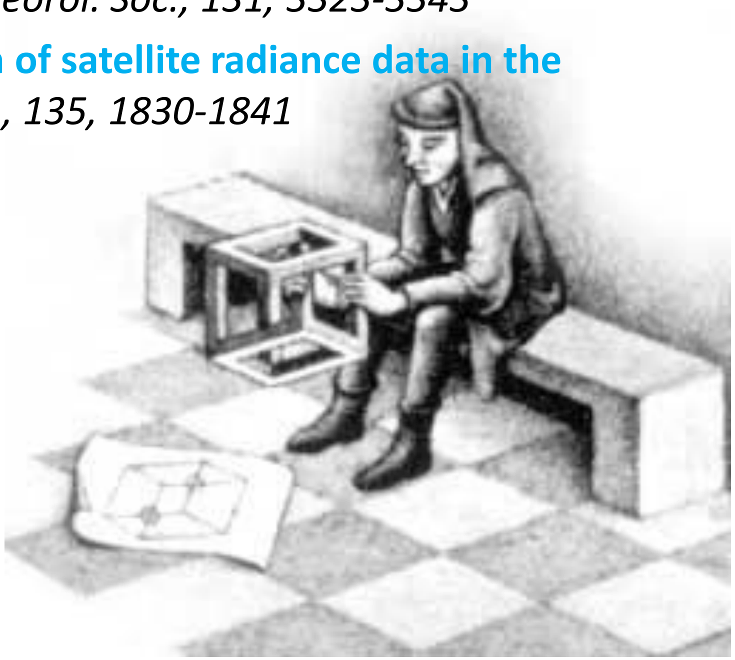
Dee, 2004: **Variational bias correction of radiance data in the ECMWF system.** *Pp. 97-112 in Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP, 28 June-1 July 2004, Reading, UK*

Dee, 2005: **Bias and data assimilation.** *Q. J. R. Meteorol. Soc., 131, 3323-3343*

Dee and Uppala, 2009: **Variational bias correction of satellite radiance data in the ERA-Interim reanalysis.** *Q. J. R. Meteorol. Soc., 135, 1830-1841*

Feel free to contact me with questions:

Hans.Hersbach@ecmwf.int



VarBC for ***total column water vapour***

- ENVISAT/MERIS until April 2012
- Anchored to all other humidity-sensitive observations
- Bias model: $\beta_0 + \beta_1$ x TCWV(model state)