# Observational bias correction in data assimilation

Niels Bormann,
Hans Hersbach and Dick Dee

Data Assimilation Training Course

# Outline

1. **Introduction**
   - Biases in **models**, **observations**, and **observation operators**
   - Implications for **data assimilation**
2. **Variational analysis and correction of observation bias**
   - The need for an **adaptive** system
   - Variational bias correction (**VarBC**)
3. **Limitations of VarBC and how to address them**
   - Interaction with **model bias**
   - Adding further **constraints**
4. **Summary**

# Outline

1. **Introduction**
   – Biases in **models**, **observations**, and **observation operators**
   – Implications for **data assimilation**
2. **Variational analysis and correction of observation bias**
   – The need for an **adaptive** system
   – Variational bias correction (**VarBC**)
3. **Limitations of VarBC and how to address them**
   – Interaction with **model bias**
   – Adding further **constraints**
4. **Summary**

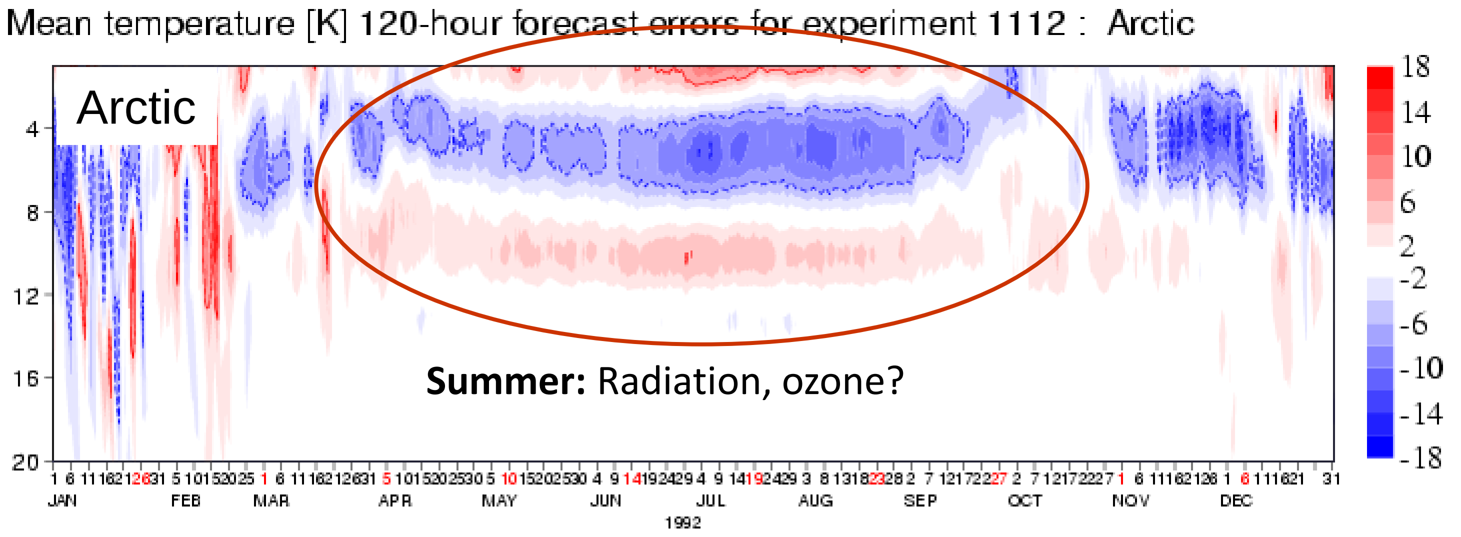
# Biases are everywhere – in models, observations, observation operators

**Example of a <u>model bias</u>:**

Seasonal variations in temperature biases in the upper-stratosphere (T255L60 model used for the *ERA-Interim* reanalysis)

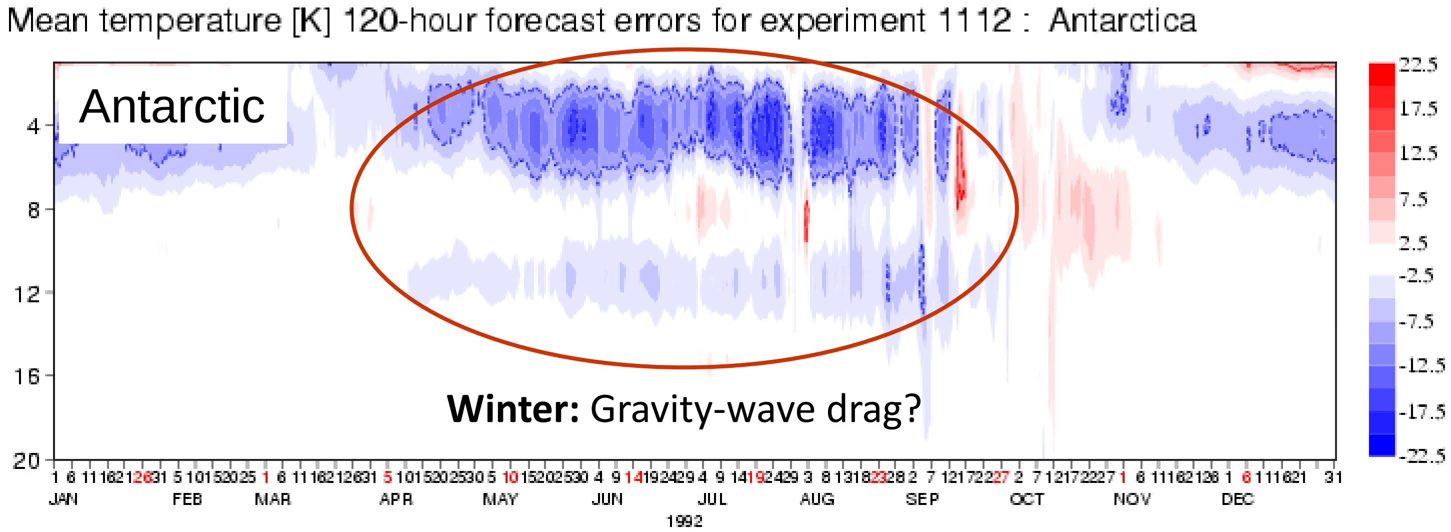
**0.1hPa (65km)**

**40hPa (22km)**



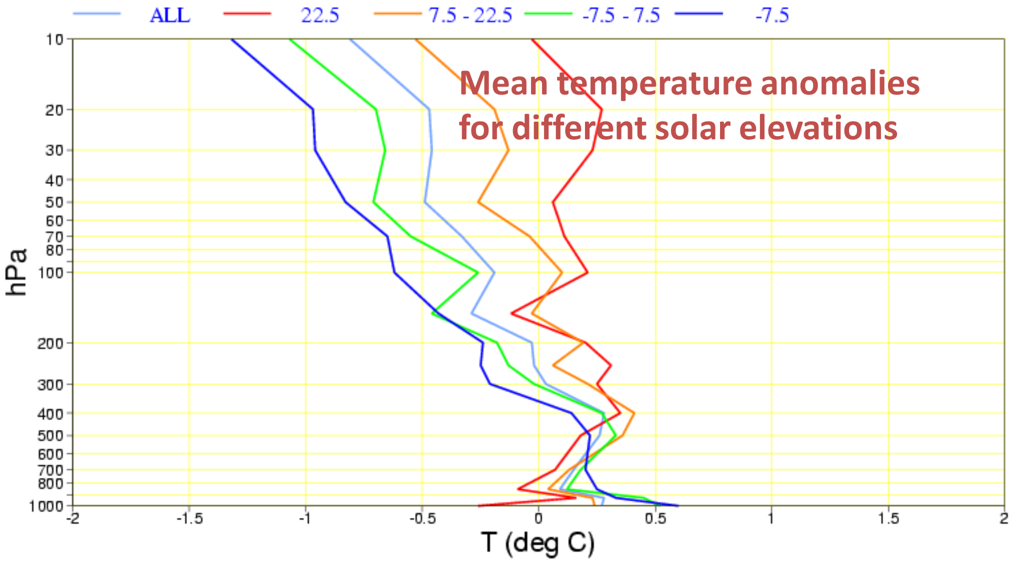

**0.1hPa (65km)**

**40hPa (22km)**

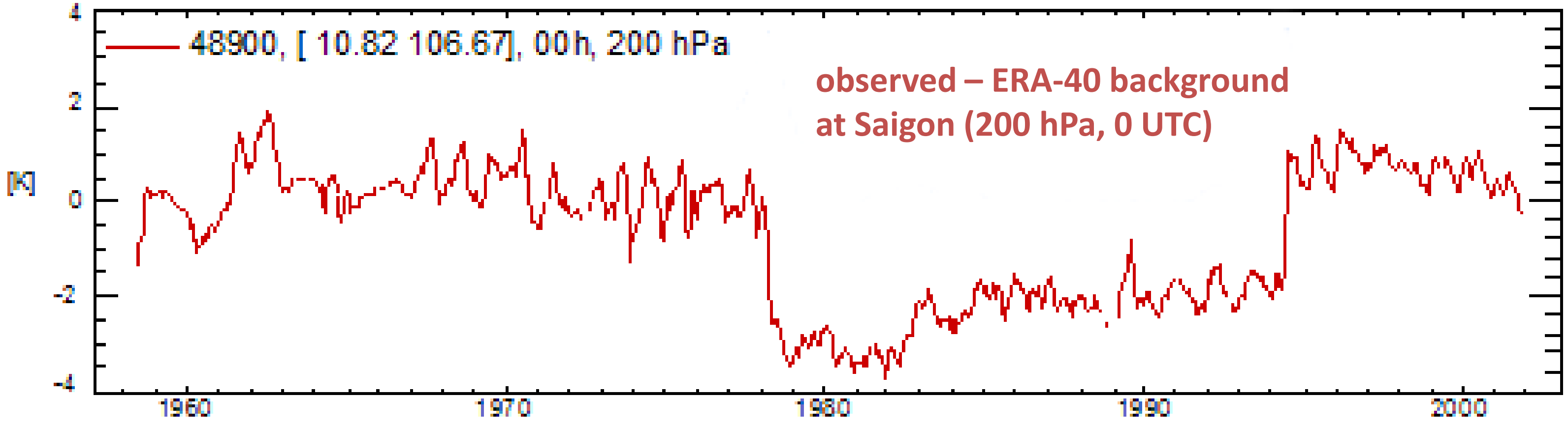


# Observation bias
# E.g., : Radiosonde temperature observations

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



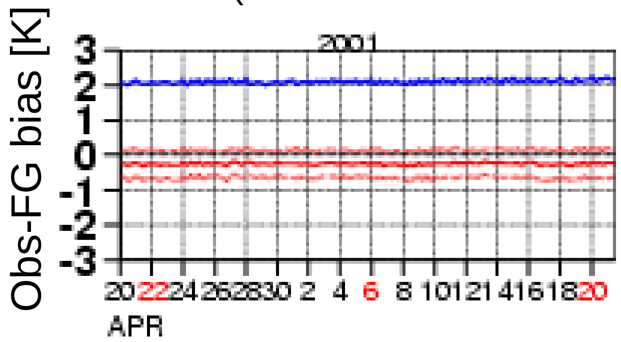

**Bias changes due to change of equipment**



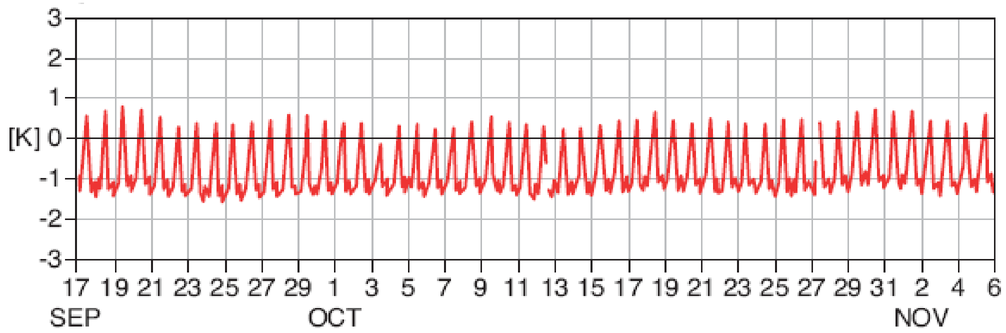
# Observation and observation operator bias (or bias in the background?): Satellite radiances

**Monitoring the background departures o-b (averaged in time and/or space):**

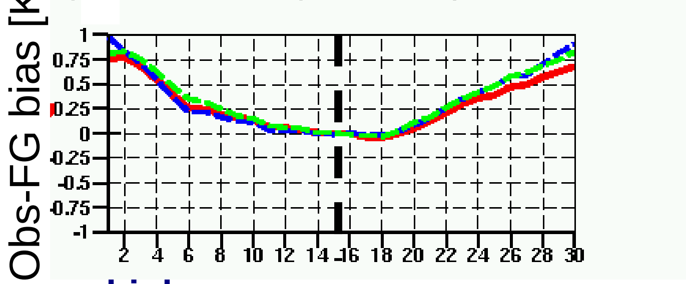
Constant bias (NOAA-14 HIRS channel 5)




Diurnal bias variation in a geostationary satellite

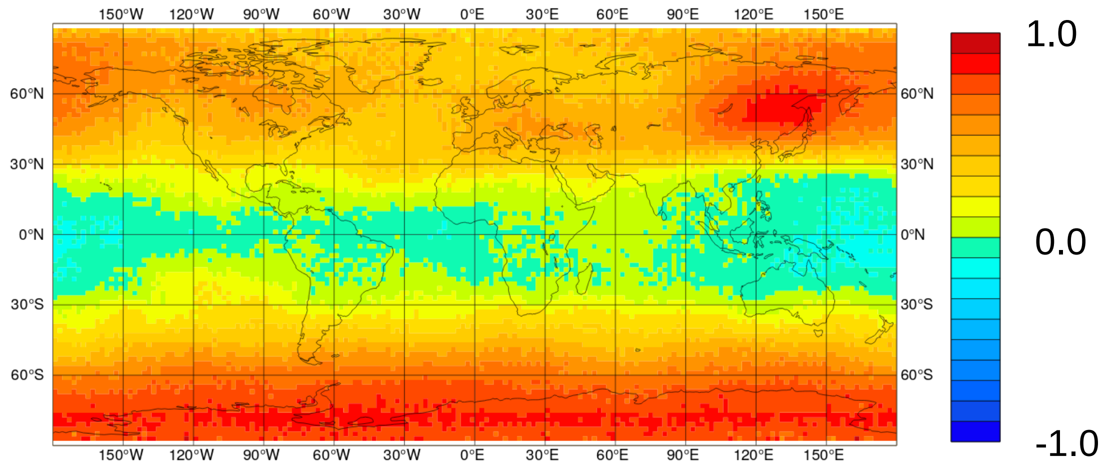



Bias depending on scan position
(AMSU-A ch 7) NH Trop SH



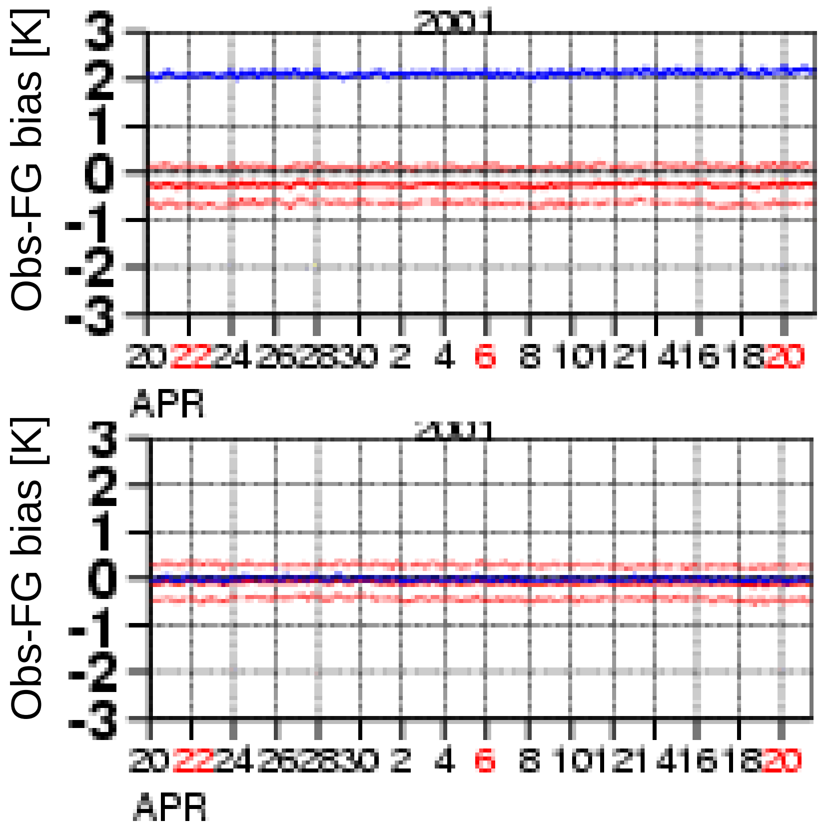

**high zenith angle** **nadir** **high zenith angle**

Air-mass dependent bias (AMSU-A ch 8)



# Observation and observation operator bias:
# Satellite radiances – identifying sources of bias

**Monitoring the background departures o-b (averaged in time and/or space):**



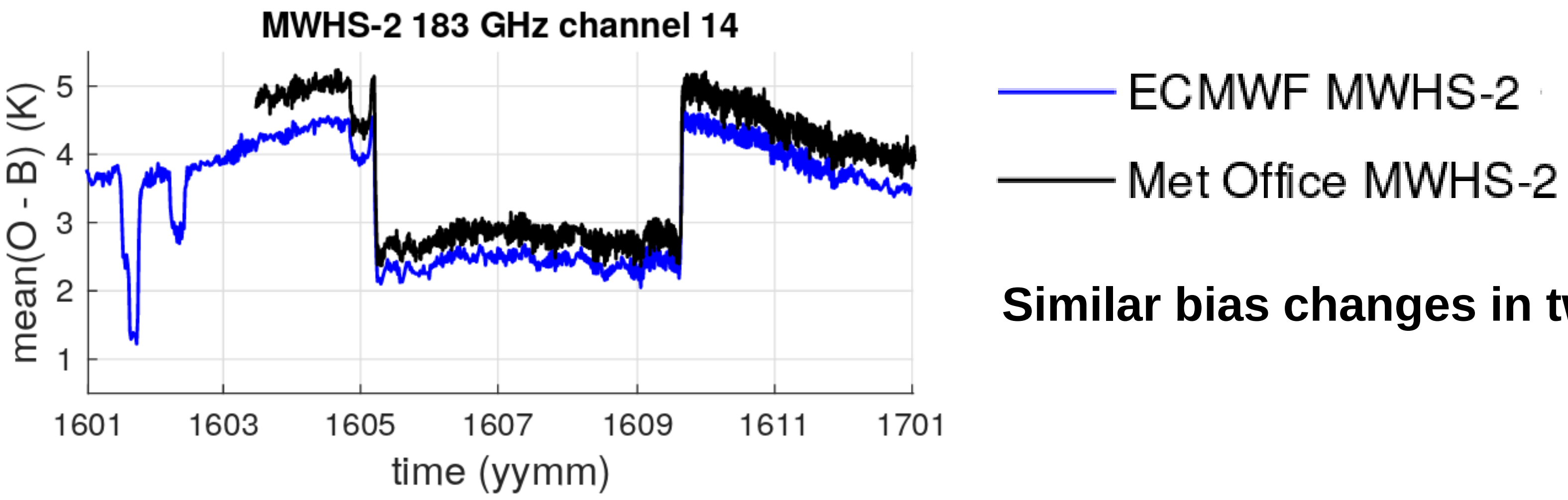

HIRS channel 5 (peaking around 600hPa) on NOAA-14 satellite has +2.0K radiance bias against the background.

The same channel on the NOAA-16 satellite and other similar radiances have no bias against the background.

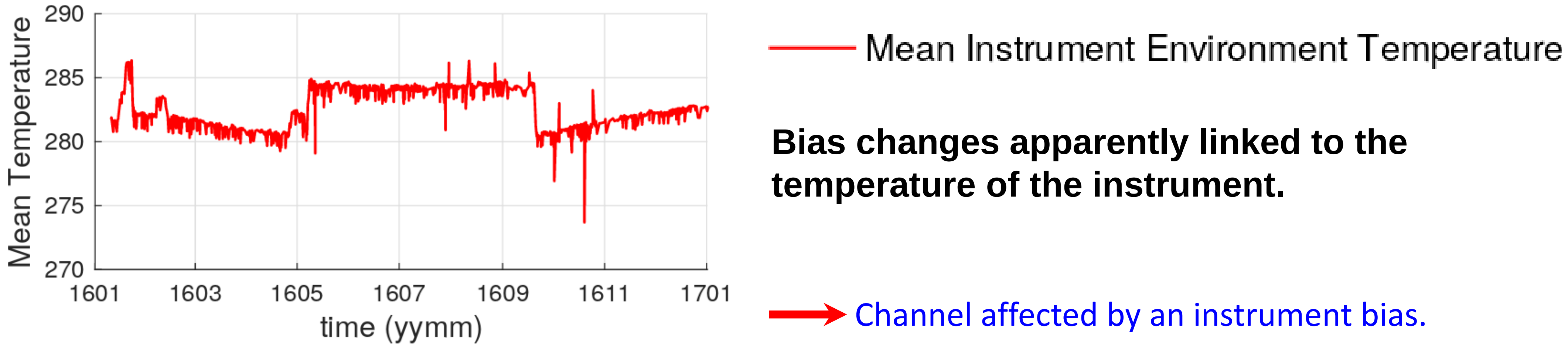
→ NOAA-14 channel 5 has an instrument bias.

# Observation and observation operator bias: Satellite radiances – identifying sources of bias

**A time-varying bias:**




**Similar bias changes in two NWP systems.**




**Bias changes apparently linked to the temperature of the instrument.**

→ Channel affected by an instrument bias.

# Observation and observation operator bias: Radiative transfer bias for satellite radiances

Examples of causes for biases in radiative transfer:

Bias in assumed concentrations of atmospheric gases (e.g., $CO_2$, aerosols)

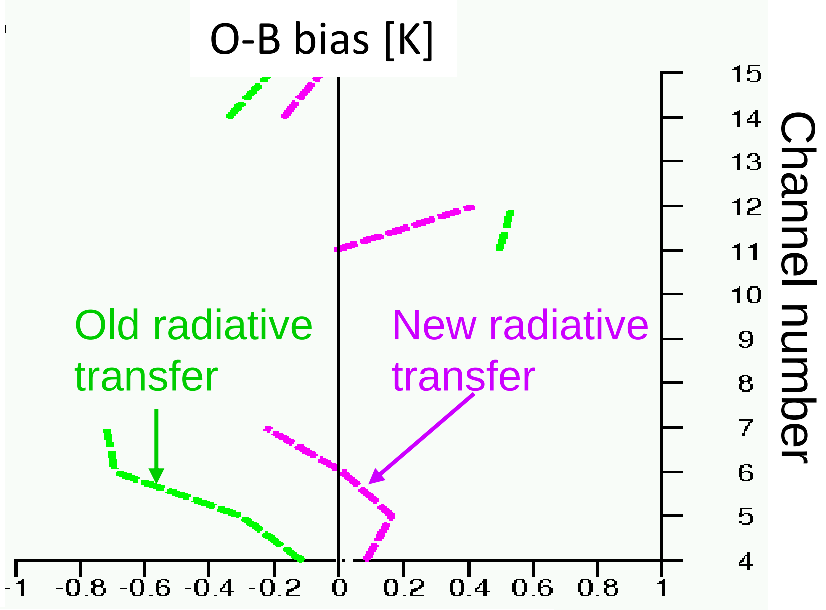
Biases in the spectroscopy

Neglected effects (e.g., clouds)

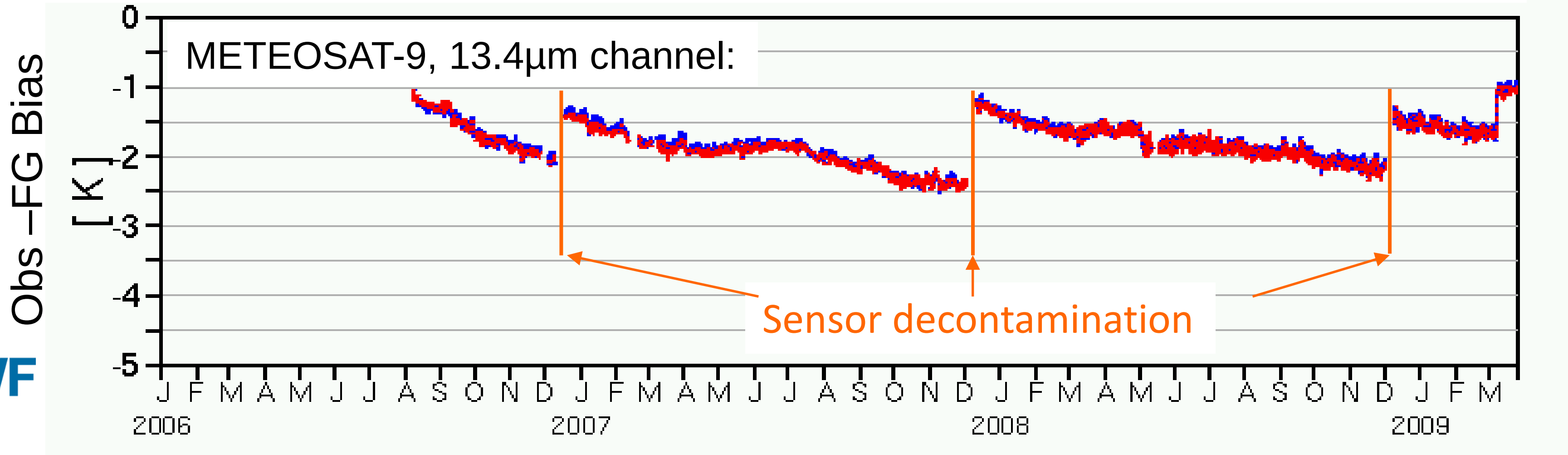
Incorrect spectral response function

…

Change in bias for HIRS resulting from an update of the Radiative Transfer model:




Drift in bias due to ice-build up on sensor, altering the spectral response of the channel:



## Bias problems in a nut-shell
# Implications for data assimilation

- **Observations and observation operators have biases, which may change over time.**
- **Models have biases, and changes in observational coverage over time may change the extent to which observations correct these biases.**
- **Where do these biases matter in data assimilation?**

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

- **Standard 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
  - Need to correct for biases prior/during assimilation.

# Outline

1. **Introduction**
   - Biases in **models**, **observations**, and **observation operators**
   - Implications for **data assimilation**
2. **Variational analysis and correction of observation bias**
   - The need for an **adaptive** system
   - Variational bias correction (**VarBC**)
3. **Limitations of VarBC and how to address them**
   - Interaction with **model bias**
   - Adding further **constraints**
4. **Summary**

# Variational analysis and bias correction
## Recall variational analysis and error sources…

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

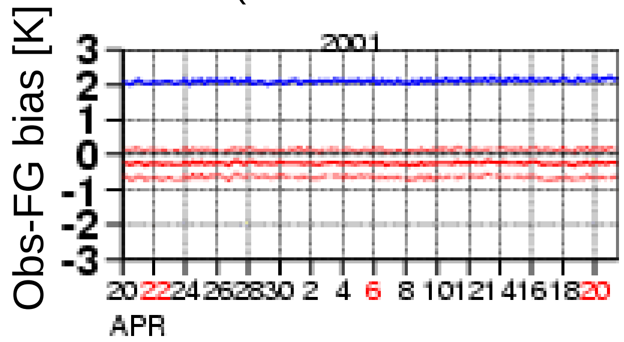
- **Errors in the input $[y - h(x_b)]$ arise from:**
    - Errors in the actual observations
    - Errors in the model background
    - Errors in the observation operator
- **In the above, all errors are assumed to have zero mean. But this is rarely the case.**
    - There is no true reference in the real world!
    - The only information available are differences.
- **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 practice by careful comparison against other data

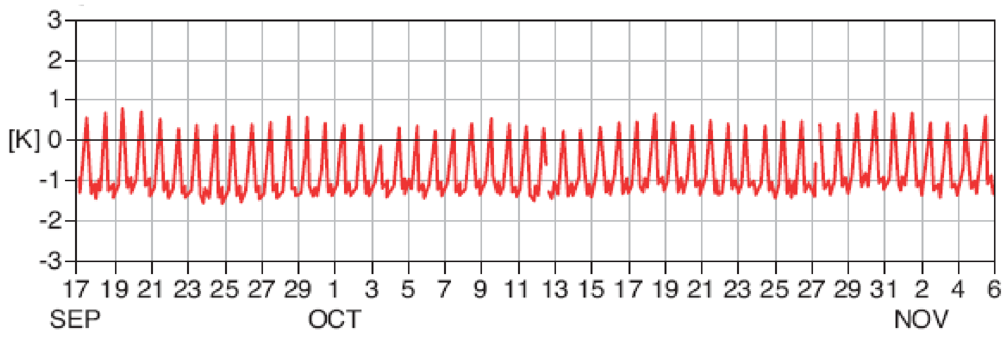
# How to address systematic errors?
# The need for an adequate bias model

Prerequisite for any bias correction is a 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 after careful diagnosis of the bias.

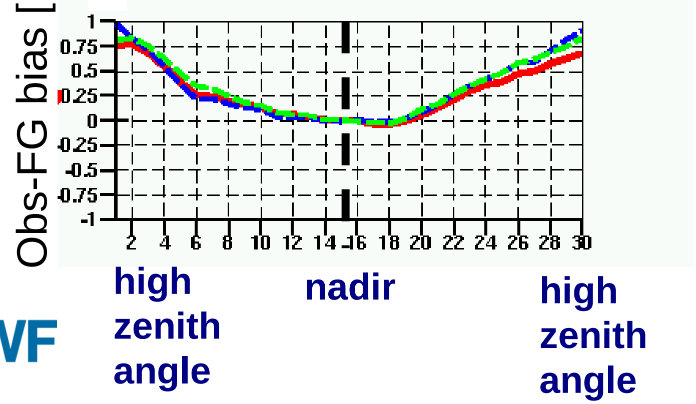
Constant bias (NOAA-14 HIRS channel 5)



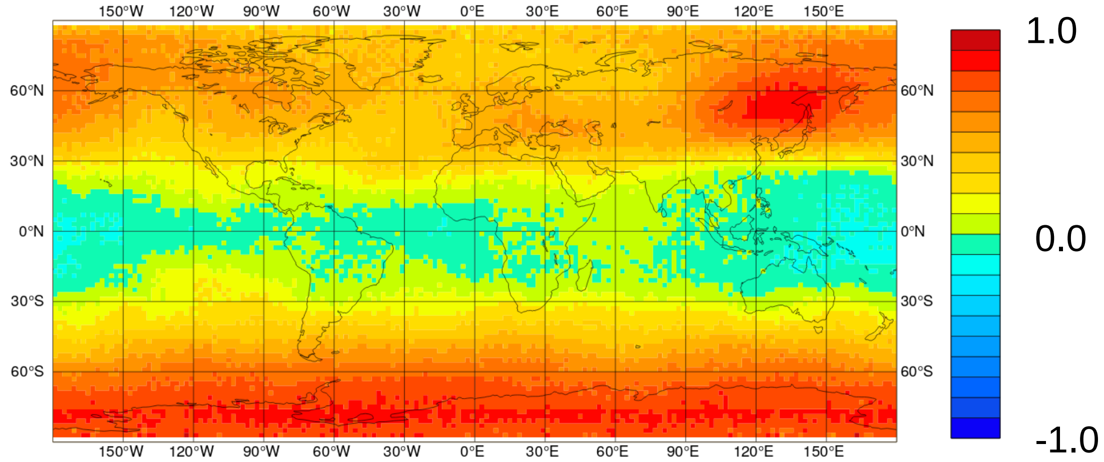

Diurnal bias variation in a geostationary satellite




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




Air-mass dependent bias (AMSU-A ch 8)



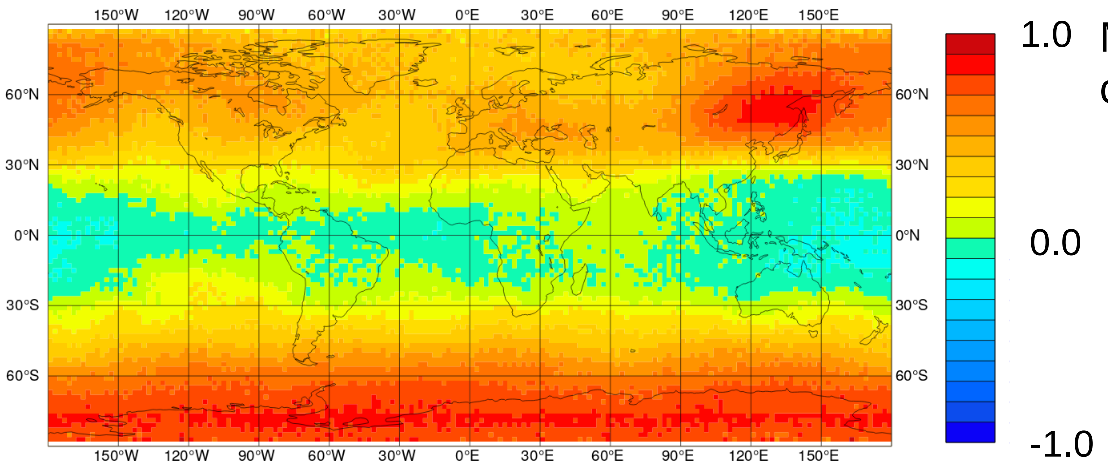

ECMWF

# How to address systematic errors?
# The need for an adequate bias model

Prerequisite for any bias correction is a model for the bias ($b(x,\beta)$):

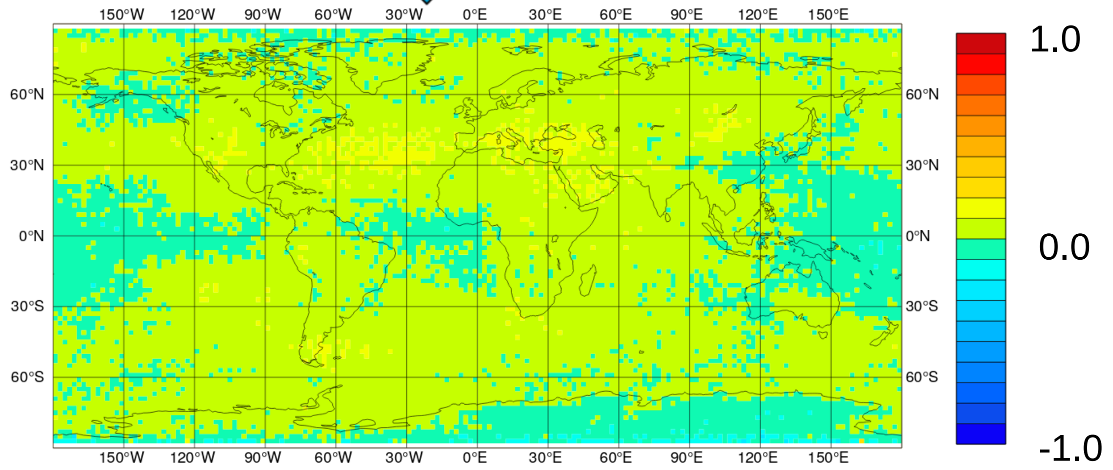
- For instance, a linear model with some predictors $p_1, p_2, \ldots p_n$, and free parameters $\beta_0, \beta_1, \beta_2, \ldots \beta_n$ :

  $b(\mathbf{x},\boldsymbol{\beta}) = \beta_0 + \beta_1\, p_1 + \beta_2\, p_2 + \ldots + \beta_n\, p_n$




Mean o-b before bias correction

After bias correction



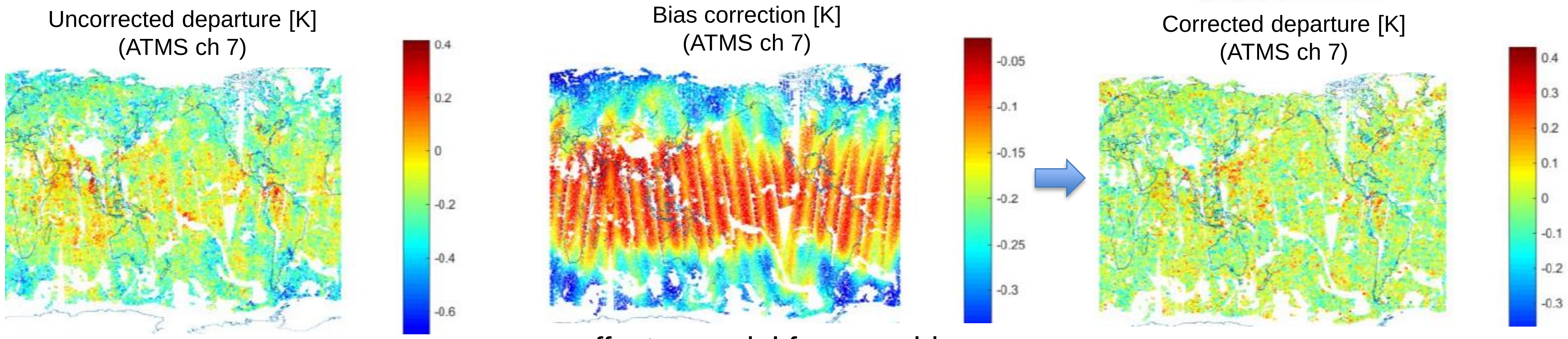

The example uses a linear bias model with a constant $\beta_0$ and four layer thicknesses as predictors (1000-300hPa, 200-50hPa, 50-5hPa,10-1hPa thickness) + a model for scan-bias

# How to address systematic errors?
# The need for an adequate bias model

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

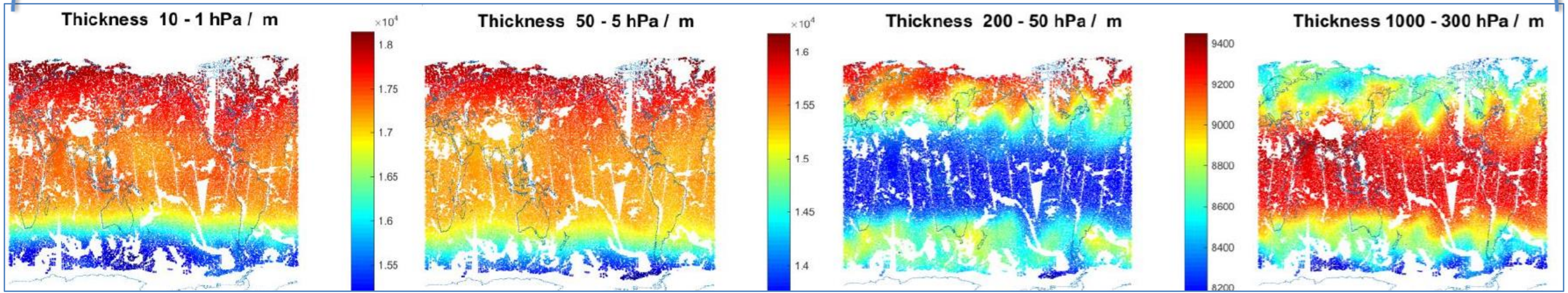
- For instance, a linear model with some predictors $p_1, p_2, \ldots p_n$, and free parameters $\beta_0, \beta_1, \beta_2, \ldots \beta_n$ :

  $b(\mathbf{x},\boldsymbol{\beta}) = \beta_0 + \beta_1\, p_1 + \beta_2\, p_2 + \ldots + \beta_n\, p_n$




+ offset + model for scan-bias

Airmass predictors $p_i$



# Offline bias correction
# (as used for satellite radiances at ECMWF before 2006)

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

- Error model for brightness temperature data: $y = h(x) + \underbrace{b^{scan} + b^{air}(x)}_{\text{Bias model}} + e^{obs}$

  – With: $e^{obs}$ = random observation error

  Separate models for scan-bias and airmass bias:

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

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

  **Predictors, for instance:**
  - Layer thicknesses
  - Surface skin temperature
  - Total precipitable water

  **Bias coefficients**

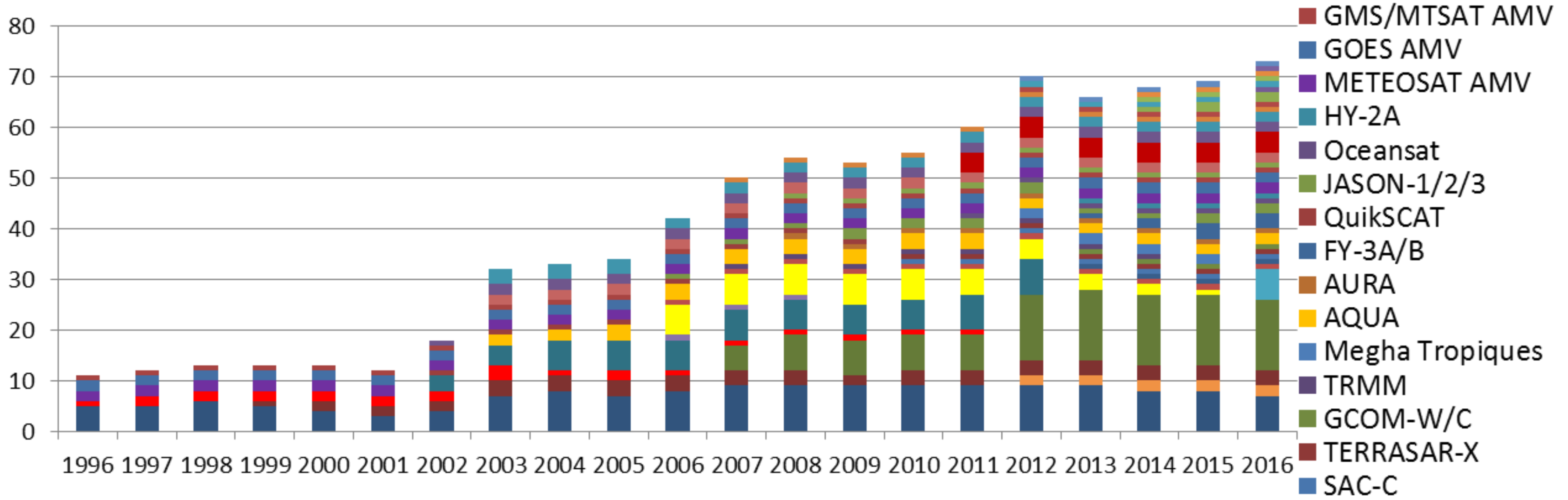
  – Averaging the background departures gives:

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

- The bias coefficients (scan-bias look-up-table regression coefficients) were estimated periodically:
  - typically 2 weeks of background departures
  - 2-step regression procedure
  - careful masking and data selection

# The need for an adaptive bias correction system

- **The global observing system is increasingly complex and constantly changing.**
- **It is dominated by satellite radiance observations for which**
  - 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?**
  - **Requires a consistent approach and a flexible, automated system**



# Variational bias correction: General idea

The **bias** in a given instrument/channel 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: $z^T = [\, x^T \quad \beta^T \,]$

**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: Modified analysis problem

**The original problem:**

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

**The modified problem:**

$$J(\mathbf{x},\boldsymbol{\beta}) = \overbrace{(\mathbf{x}_b - \mathbf{x})^T \mathbf{B}_x^{-1} (\mathbf{x}_b - \mathbf{x})}^{J_b\text{: background constraint for x}} + \overbrace{(\boldsymbol{\beta}_b - \boldsymbol{\beta})^T \mathbf{B}_\beta^{-1} (\boldsymbol{\beta}_b - \boldsymbol{\beta})}^{J_\beta\text{: background constraint for } \beta}$$

$$+ \underbrace{[\mathbf{y} - \mathbf{b}_o(\mathbf{x},\boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathbf{b}_o(\mathbf{x},\boldsymbol{\beta}) - \mathbf{h}(\mathbf{x})]}_{J_o\text{: bias-corrected observation constraint}}$$

Parameter estimates from previous analysis

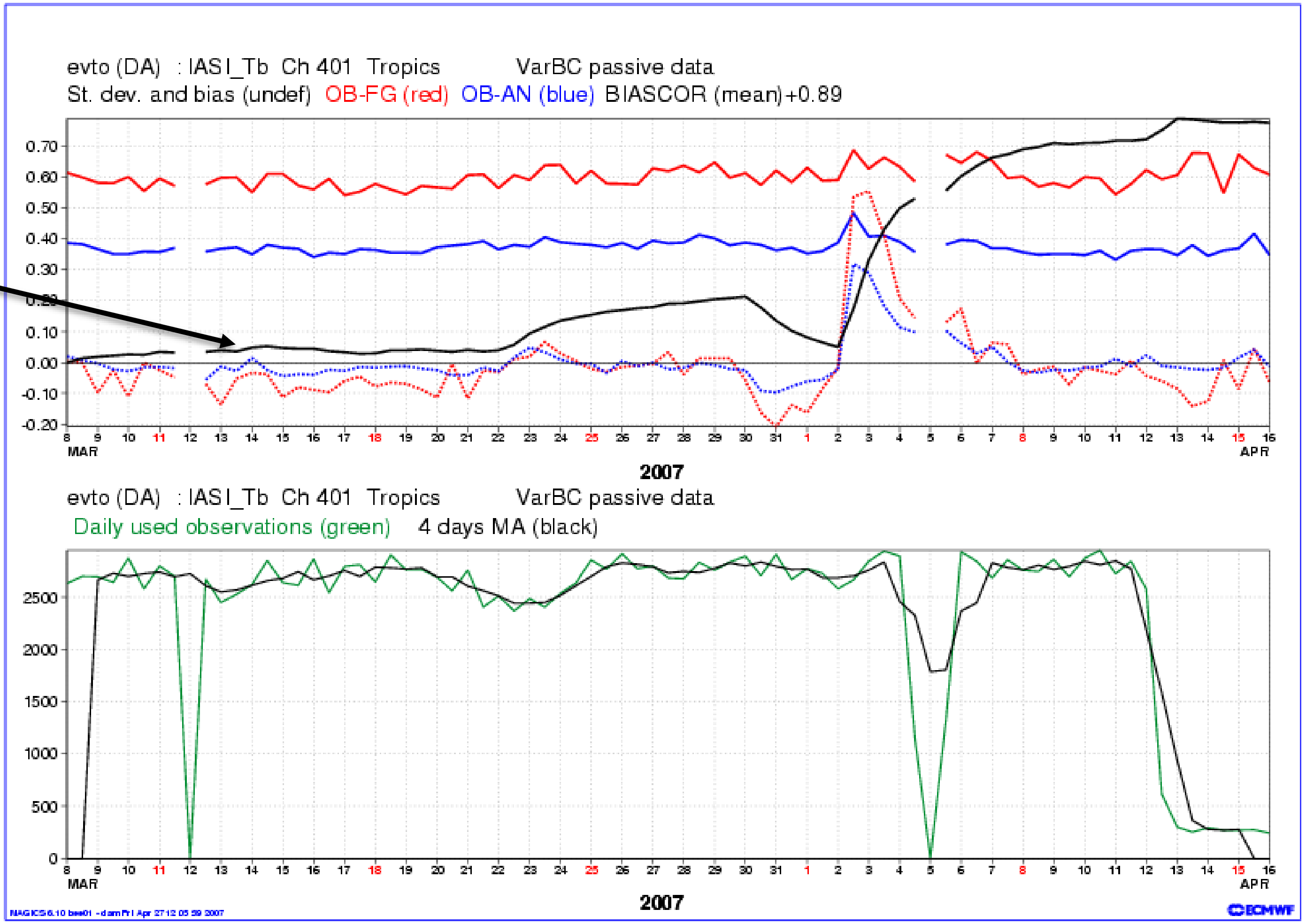
A model for the observation bias

# Example of using VarBC (I):

## Spinning up a new instrument – IASI on MetOp A

- IASI is an interferometer with 8461 channels
- Initially unstable – data gaps, preprocessing changes

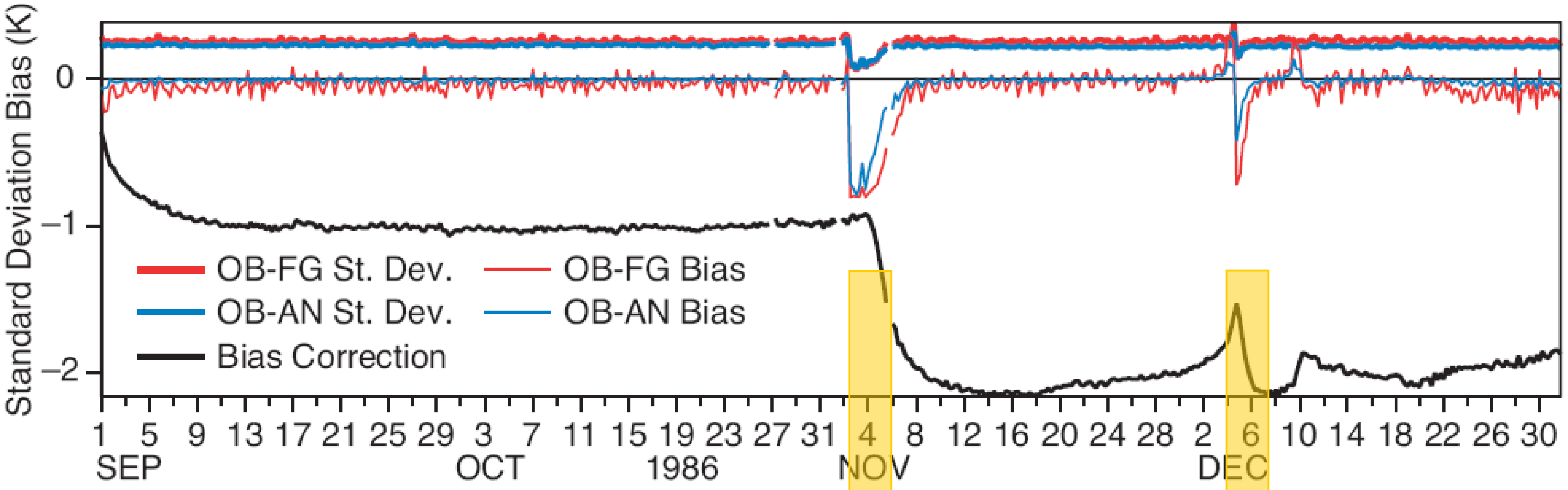
Bias correction
+ 0.89



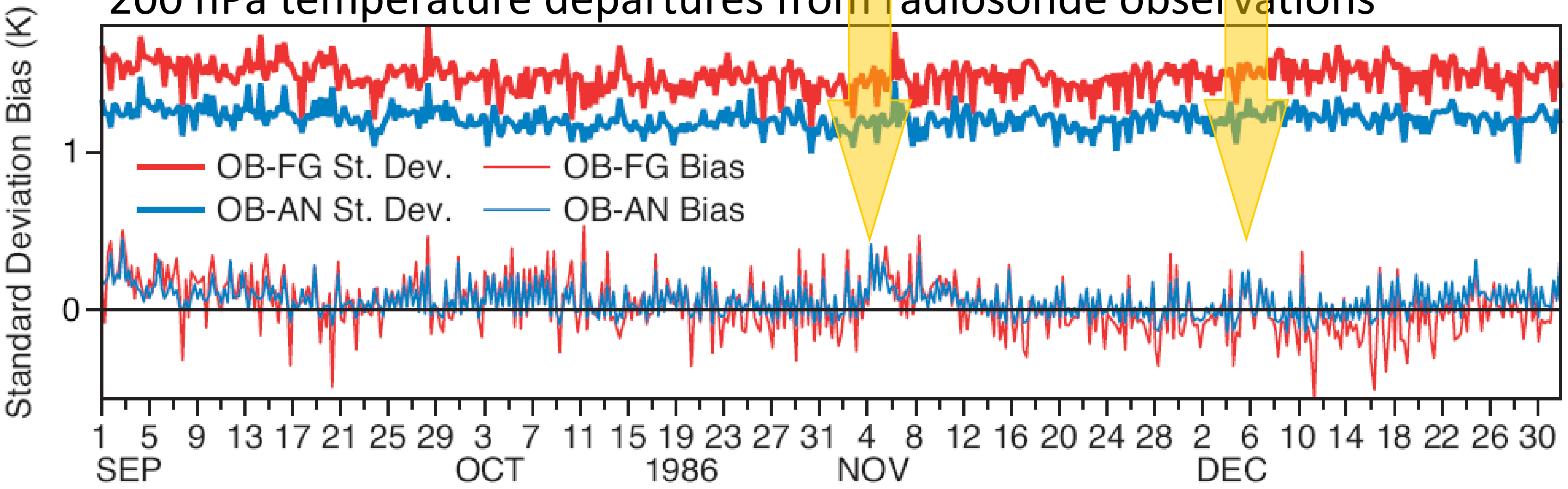

ECMWF

# Example of using VarBC (II):

Reaction of NOAA-9 MSU channel 3 bias corrections following a cosmic storm

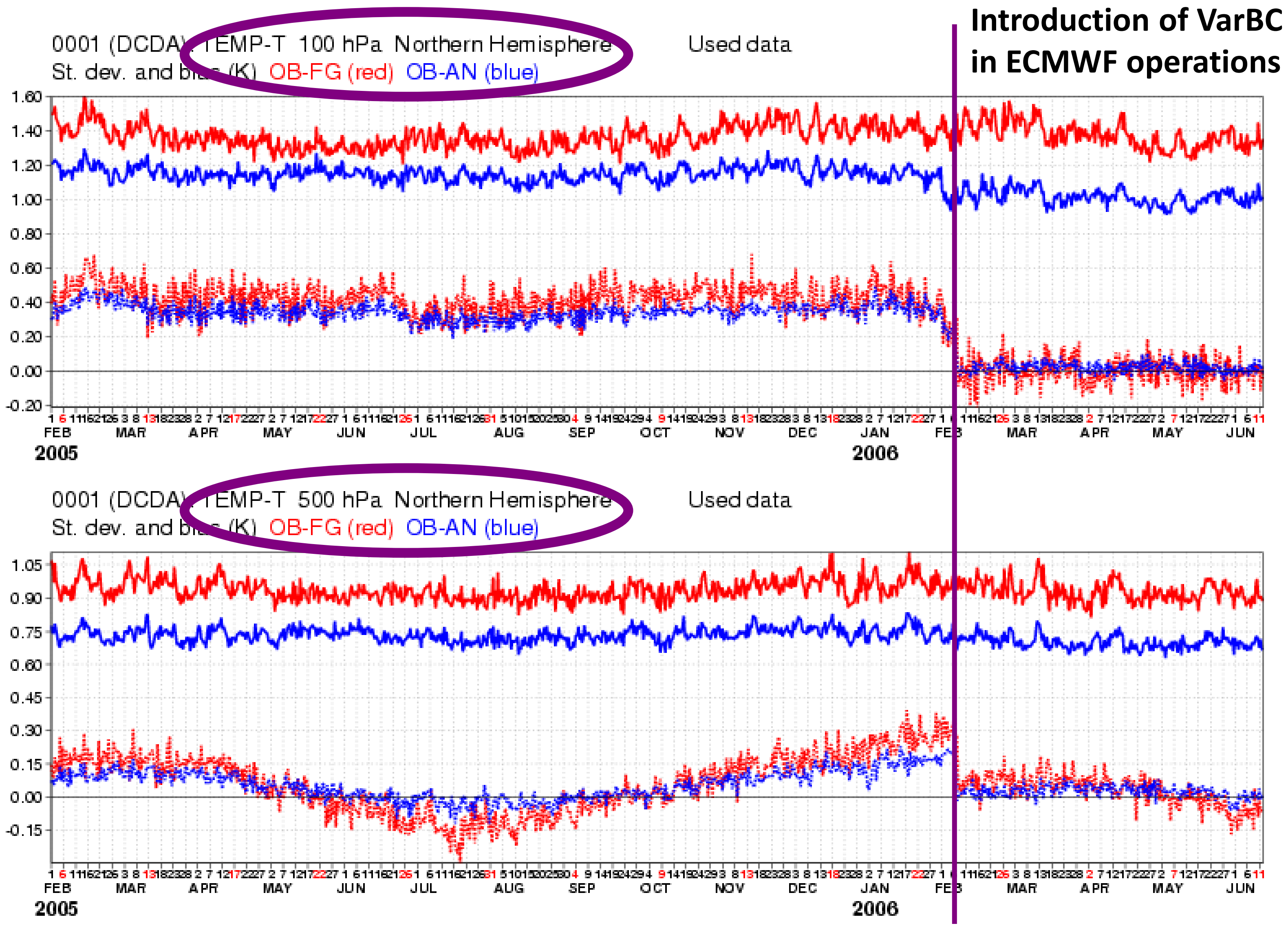



200 hPa temperature departures from radiosonde observations



# Example of using VarBC (III):

## Better Fit to conventional data



# Current use of observational bias correction at ECMWF

**Observations treated by VarBC in the operational ECMWF system:**

- Radiances
- Ozone
- Aircraft data
- Ground-based radar precipitation

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

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

# Outline

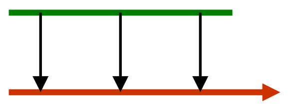
1. **Introduction**
   - Biases in **models**, **observations**, and **observation operators**
   - Implications for **data assimilation**
2. **Variational analysis and correction of observation bias**
   - The need for an **adaptive** system
   - Variational bias correction (**VarBC**)
3. **Limitations of VarBC and how to address them**
   - Interaction with **model bias**
   - Adding further **constraints**
4. **Summary**

# 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.

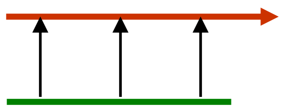
It works well (even if the model is biased) when the analysis is strongly constrained by observations:

model
abundant observations



It does not work as well when there are large model biases and observation biases are poorly constrained (e.g., few anchoring observations; many bias-corrected observations with similar characteristics):

model
observations



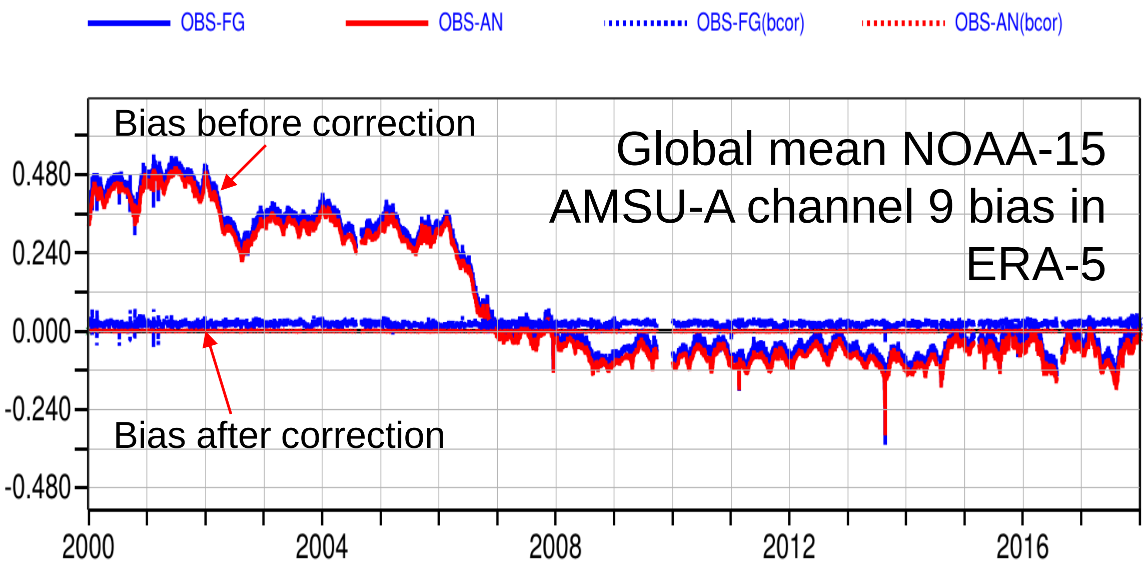
VarBC is not designed to correct model biases: Need for a **weak-constraint 4D-Var (see lecture by Patrick Laloyaux)**

# Limitations of VarBC: Interaction with model bias and the role of anchor observations

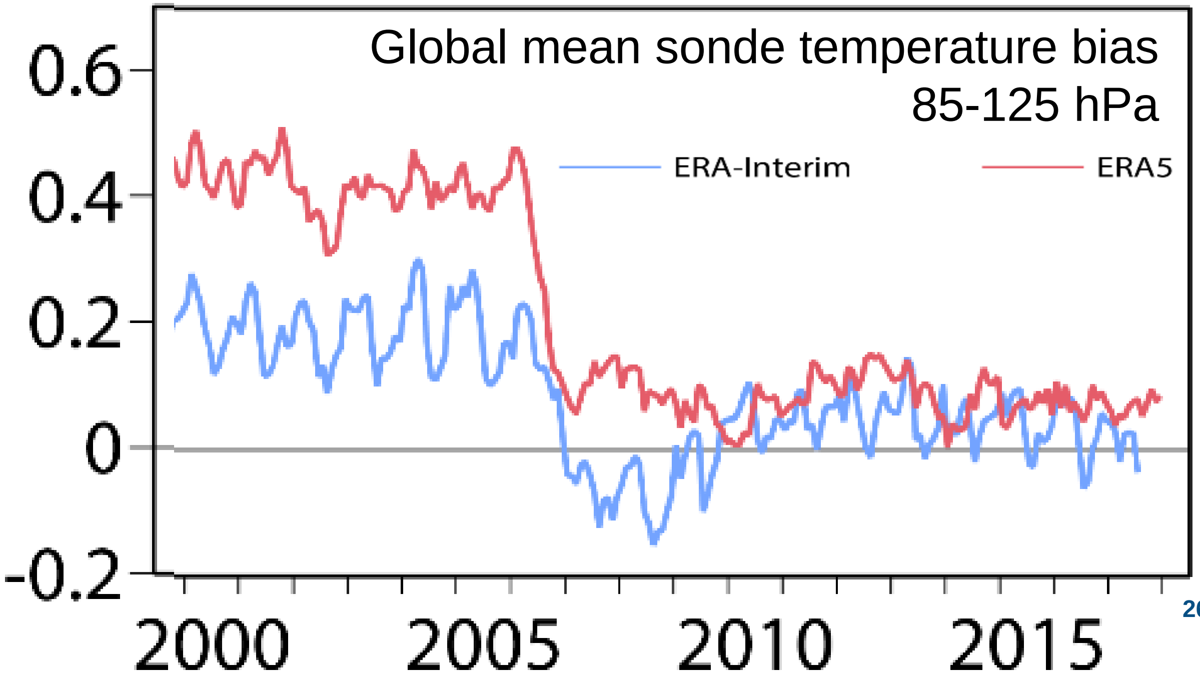
***Example: Stratospheric temperature biases***

- Model biases affect the bias correction in the absence of sufficient anchor observations.
- GNSS-RO provides a good anchor from mid-2006.

- The solution of the bias correction is also affected by other aspects, including the background error covariance.




Increased availability of GNSS-RO data



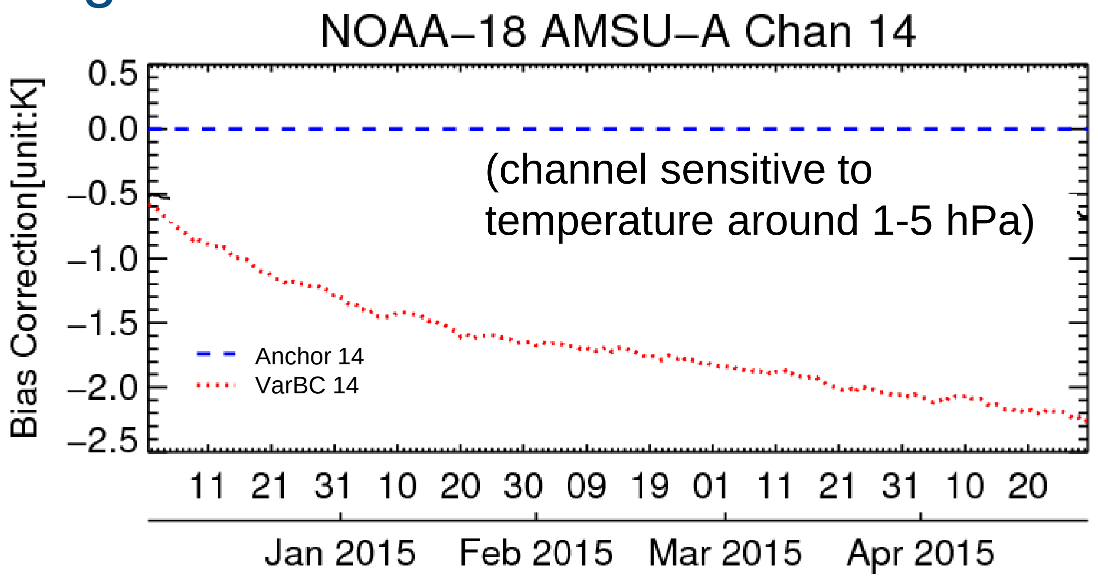

ECMWF

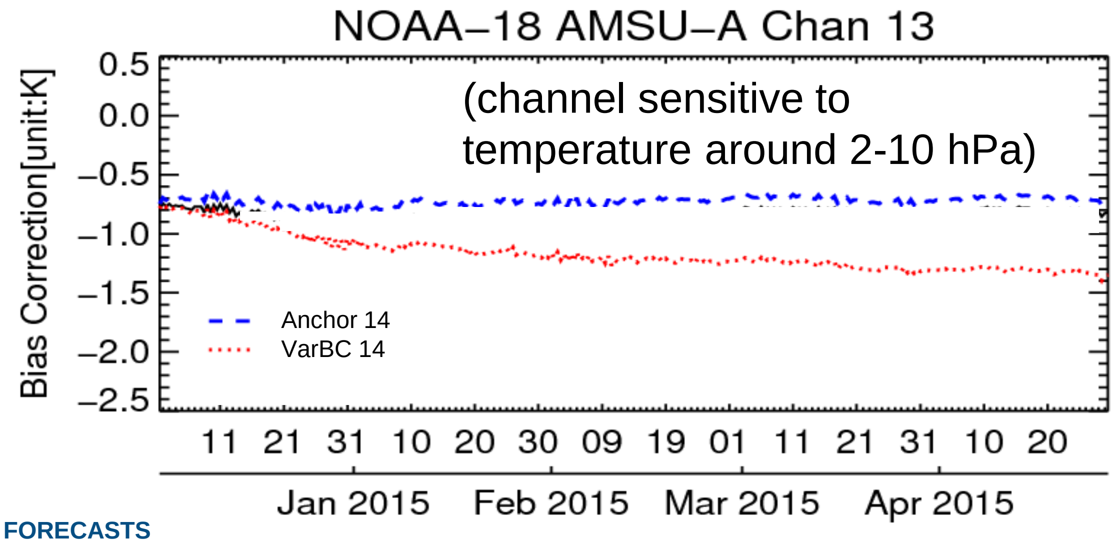
# Limitations of VarBC:
# Interaction with model bias - selecting an anchor observation

***Example: Upper stratospheric temperature biases***

- Unrealistic drift in the bias corrections due to model bias (red line)
- Additional ***anchoring*** can be imposed through assimilating AMSU-A channel 14 without a bias correction (blue line)

- Other anchoring in the ECMWF system: selected ozone-sensitive IR channels



# Extending VarBC:
# Constrain the bias correction to counter-act model bias

- Alternative concept to constrain the size of bias corrections:

  - **Constrained VarBC** (Han and Bormann 2016):
    - Penalise large bias corrections through an ***additional term*** in the cost function:

$$
\begin{aligned}
J(\mathbf{x},\boldsymbol{\beta}) &= \frac{1}{2}(\mathbf{x}_b-\mathbf{x})^T\mathbf{B}_x^{-1}(\mathbf{x}_b-\mathbf{x}) \\
&+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_b)^T\mathbf{B}_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_b) \\
&+\frac{1}{2}[\mathbf{y}-H(\mathbf{x})-b(\mathbf{x},\boldsymbol{\beta})]^T\mathbf{R}^{-1}[\mathbf{y}-H(\mathbf{x})-b(\mathbf{x},\boldsymbol{\beta})] \\
&+\frac{1}{2}\gamma^2[b(\mathbf{x},\boldsymbol{\beta})-\mathbf{b}_0]^T\mathbf{R}_b^{-1}[b(\mathbf{x},\boldsymbol{\beta})-\mathbf{b}_0]
\end{aligned}
$$

$\mathbf{b}_0$: Priori estimate of observation bias

$\mathbf{R}_b$: Priori estimate (or on orbit estimation) of radiometric uncertainty

$\boldsymbol{\beta}_b$: Background predictor coefficients

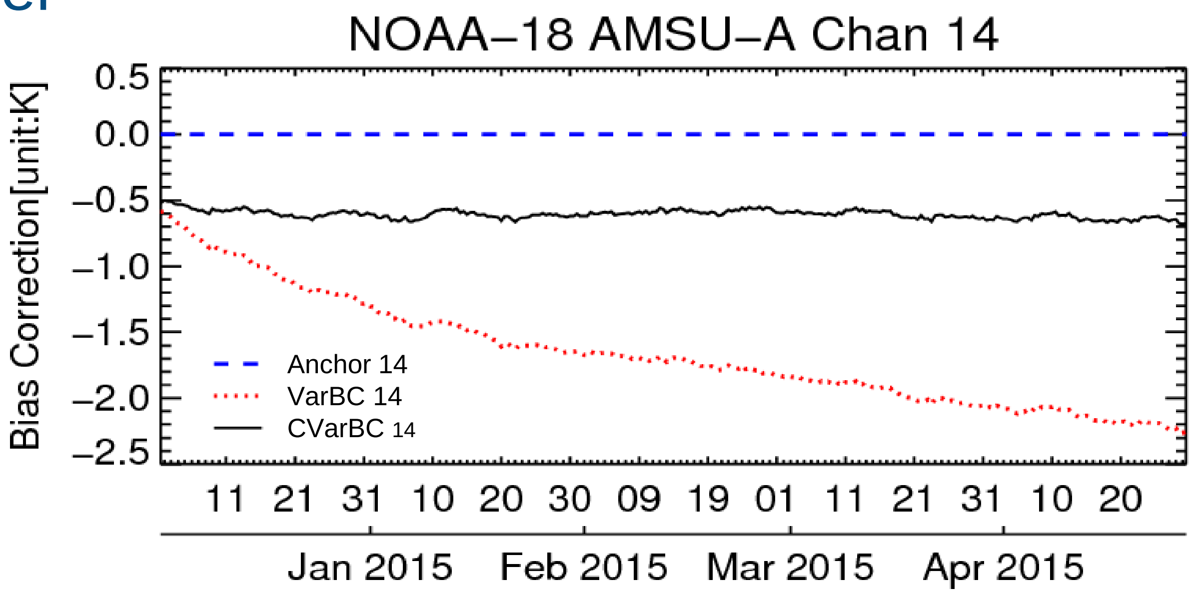
$\mathbf{B}_\beta$: Background predictor coefficients uncertainty

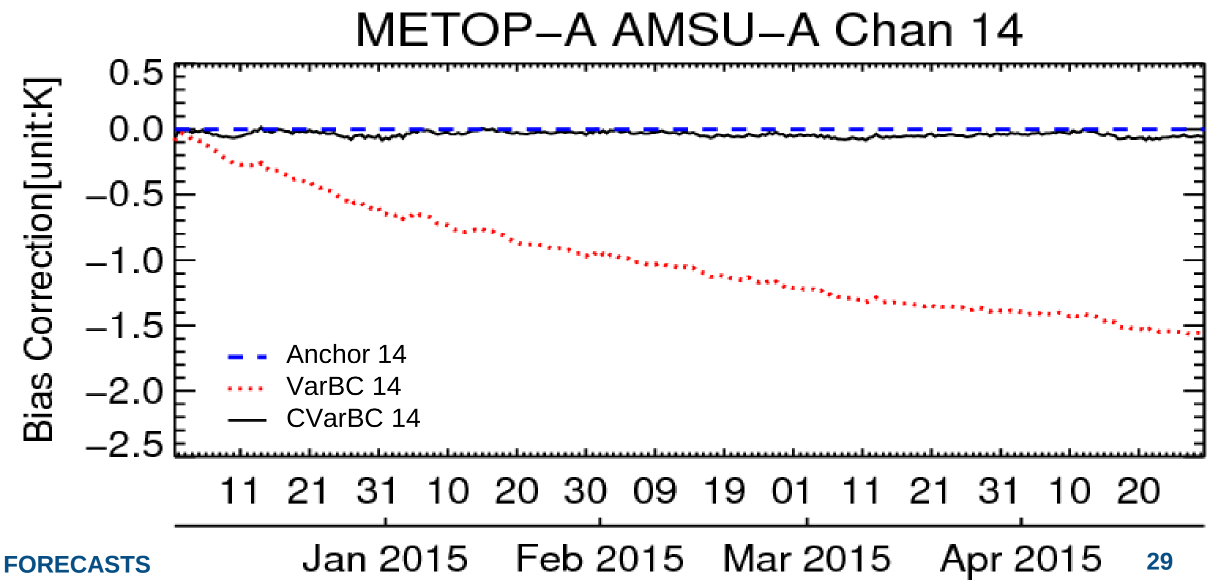
# Extending VarBC:
# Constrain bias correction to counter-act model bias

***Example: Upper stratospheric temperature biases***

- ***Constrained VarBC*** is now used operationally for AMSU-A ch 14 and ATMS ch 15
  - Different bias characteristics for different satellites are now corrected. They were previously ignored when these channels were assimilated without bias correction.

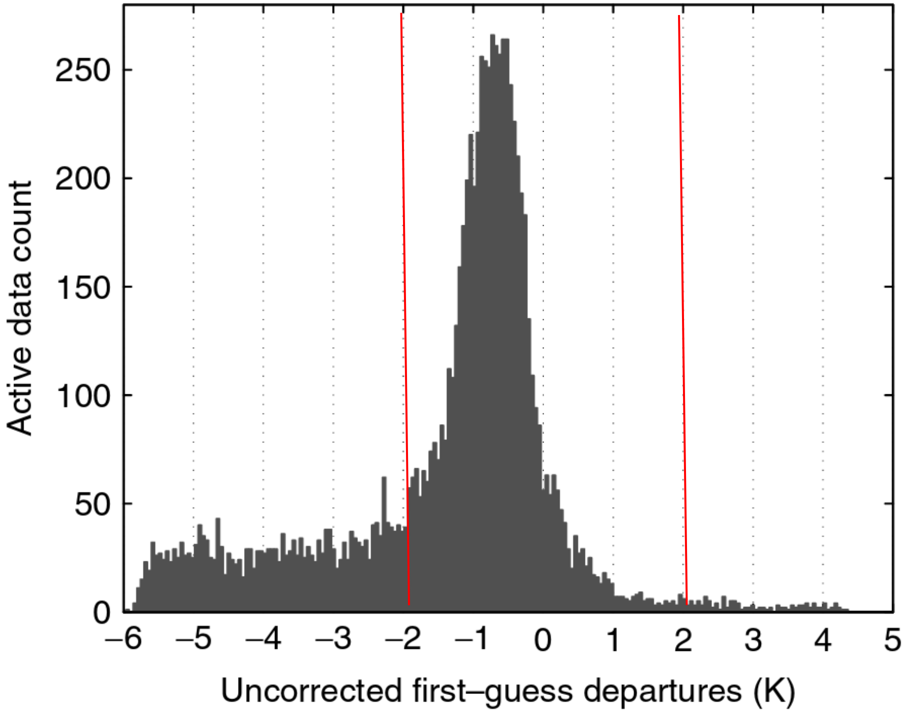
- Further constraints could be introduced by using a more restrictive bias model (e.g., no air-mass component in bias model)



# Limitations of VarBC:
# Other pit-falls: Removing the signal

- **Avoid** bias correction models with **too many predictors**, to avoid correcting for situation-dependent background errors/biases to be incorrectly removed.
- Beware of interaction between VarBC and **departure-based quality control** and asymmetric distributions:
  - Can lead to unwanted drifts in the population after QC




Histogram of IR window channel departures with cold cloud tail.

# Summary

- **Biases are everywhere:**
    - Most observations cannot be usefully assimilated without bias adjustments.

- **Manual estimation of biases in satellite data is practically impossible.**
- **Bias estimates can be updated automatically during data assimilation.**

- **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.
    - Potential for machine learning?
- 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
- Han and Bormann, 2016: Constrained adaptive bias correction for satellite radiance assimilation in the ECMWF 4D-Var system. ECMWF Technical Memorandum 783.


Feel free to contact me with questions: Niels.Bormann@ecmwf.int