Harmonization of long term data record and bias correction in data assimilation

Hans Hersbach,

Special thanks to

Dick Dee and Adrian Simmons

Dee, ECMWF workshop 2004

Dee and Uppala, QJRMS 2009

Kobayashi et. al., QJRMS 2009

Simmons et. al., QJRMS 2014

Outline

- Motivation for reanalysis being a vehicle for harmonizing data
- The problem of model bias
- The need for variational bias correction: VarBC
- VarBC in ERA-Interim; some examples in the stratosphere
- Conflicts between anchors
- What can we learn from the applied bias corrections?
- Concluding remarks

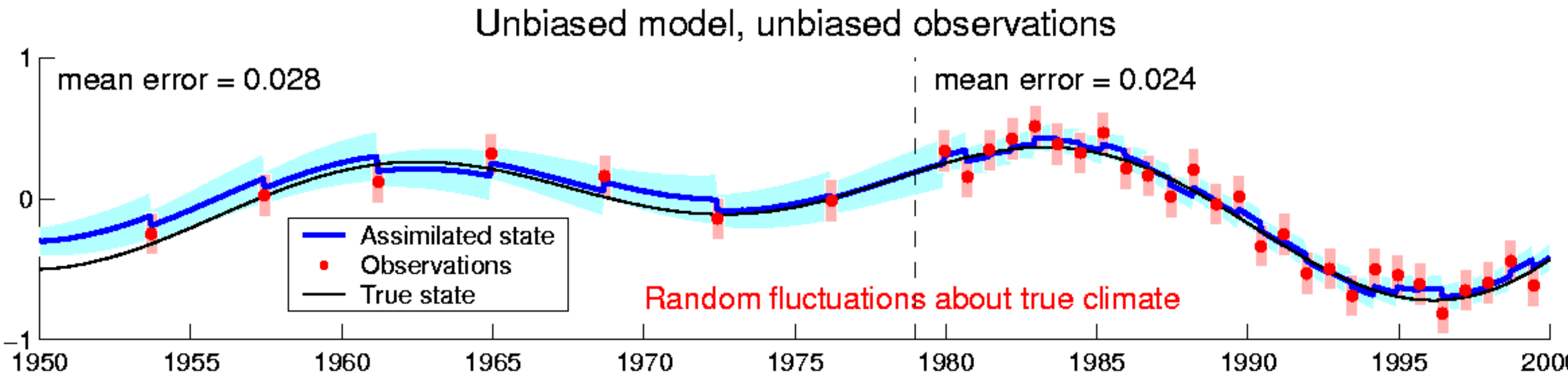
Motivation

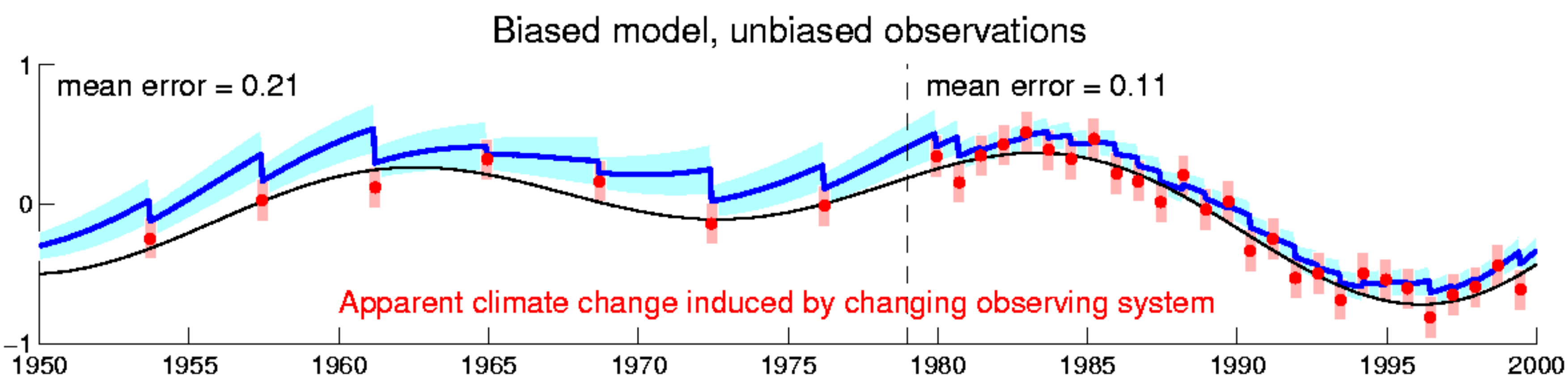
Reanalysis using a variational bias correction scheme is an effective way to 'harmonize' the long-term data record

- Reanalysis glues together a multivariate, complex and evolving observing system.
- Using the laws of physics with a consistent model formulation throughout and an advanced data assimilation system
- A variational bias scheme (VarbC) can be seen as an automatic statistical method for cross-calibration of observing systems
- VarBC relies on:
    - Complementarity of the observing system
    - Independent unbiased reference observations ('anchors')
- If these conditions are not met, interaction between model bias and an evolving observing system can have a negative effect on the estimation of trends.

The problem of model bias

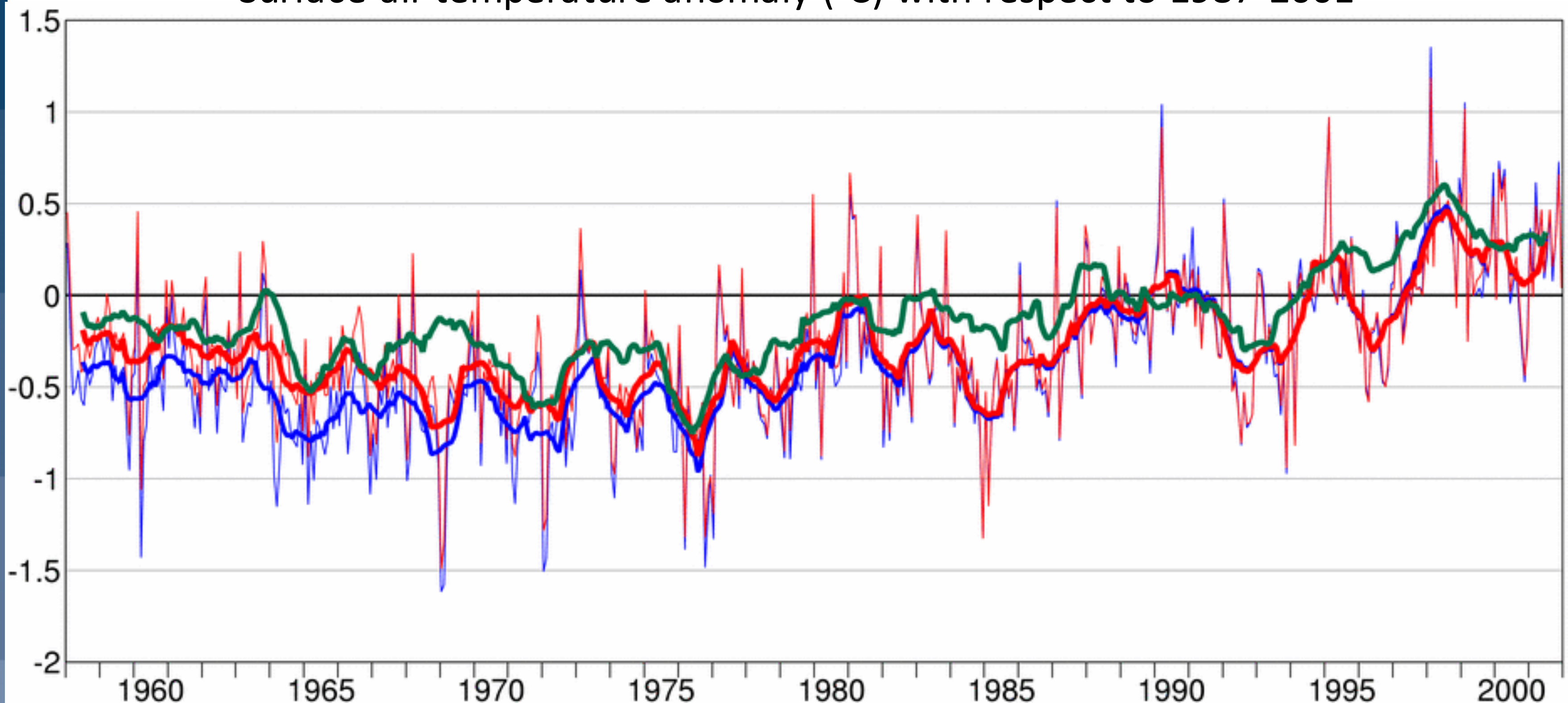
The effect of bias (model or data) on trend estimates





Implications for data assimilation:
ERA-40 surface temperatures compared to land-station values

Surface air temperature anomaly (°C) with respect to 1987-2001



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)

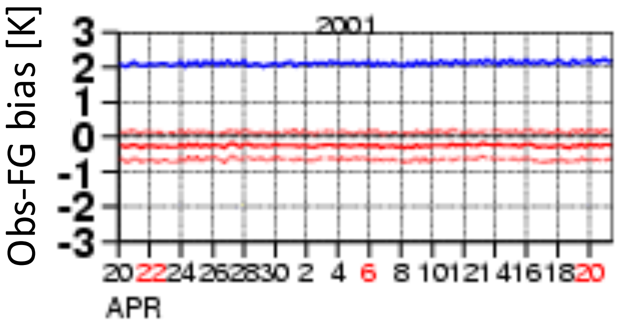
ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

The need for an automatic bias correction scheme

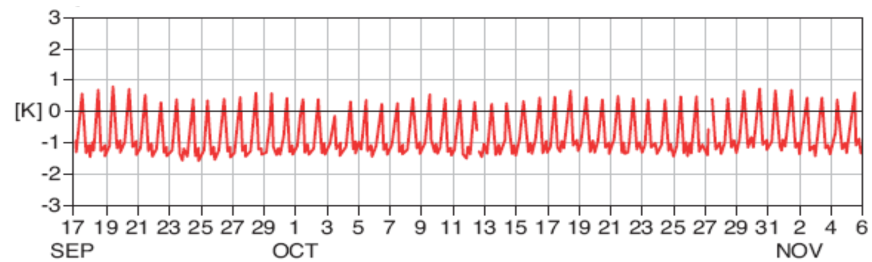
First one needs an adequate bias model

Prerequisite for any bias correction is a good model for the bias
Ideally, guided by the physical origins of the bias.
In practice, bias models are derived empirically from observation monitoring.

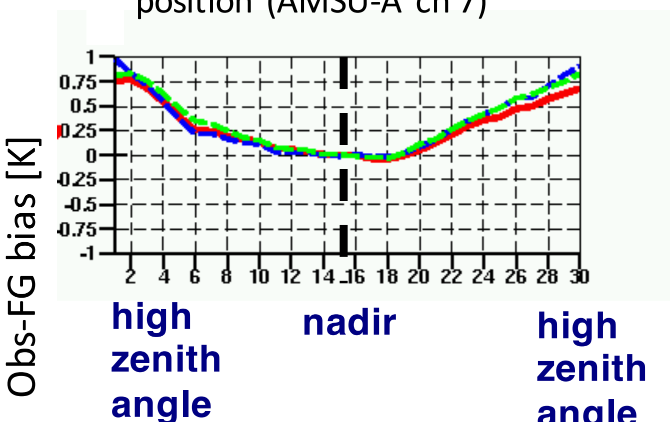
Constant bias (HIRS channel 5)



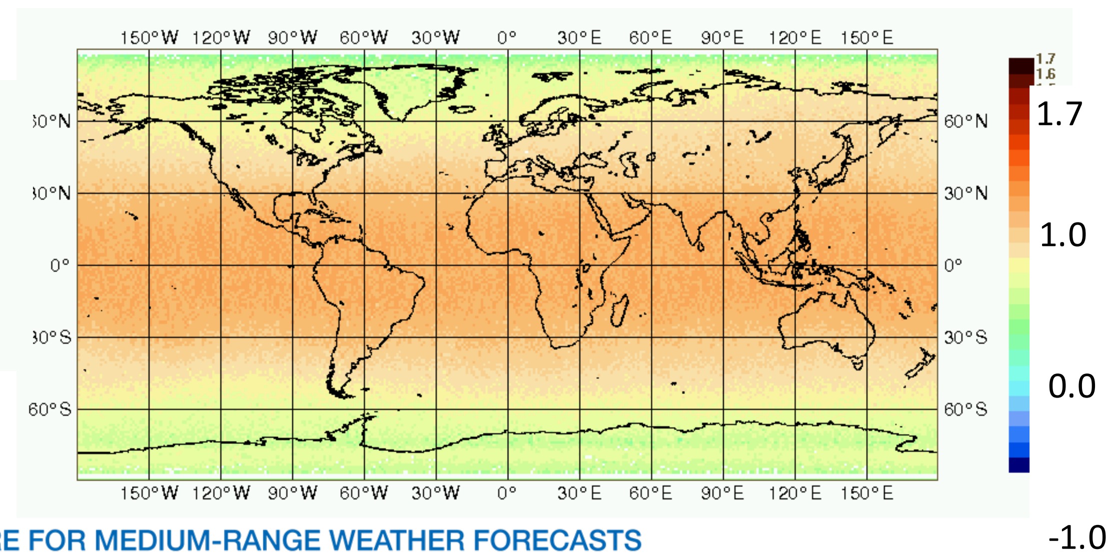
Diurnal bias variation in a geostationary satellite



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



Air-mass dependent bias (AMSU-A ch 10): **issues RTTOV**



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:

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

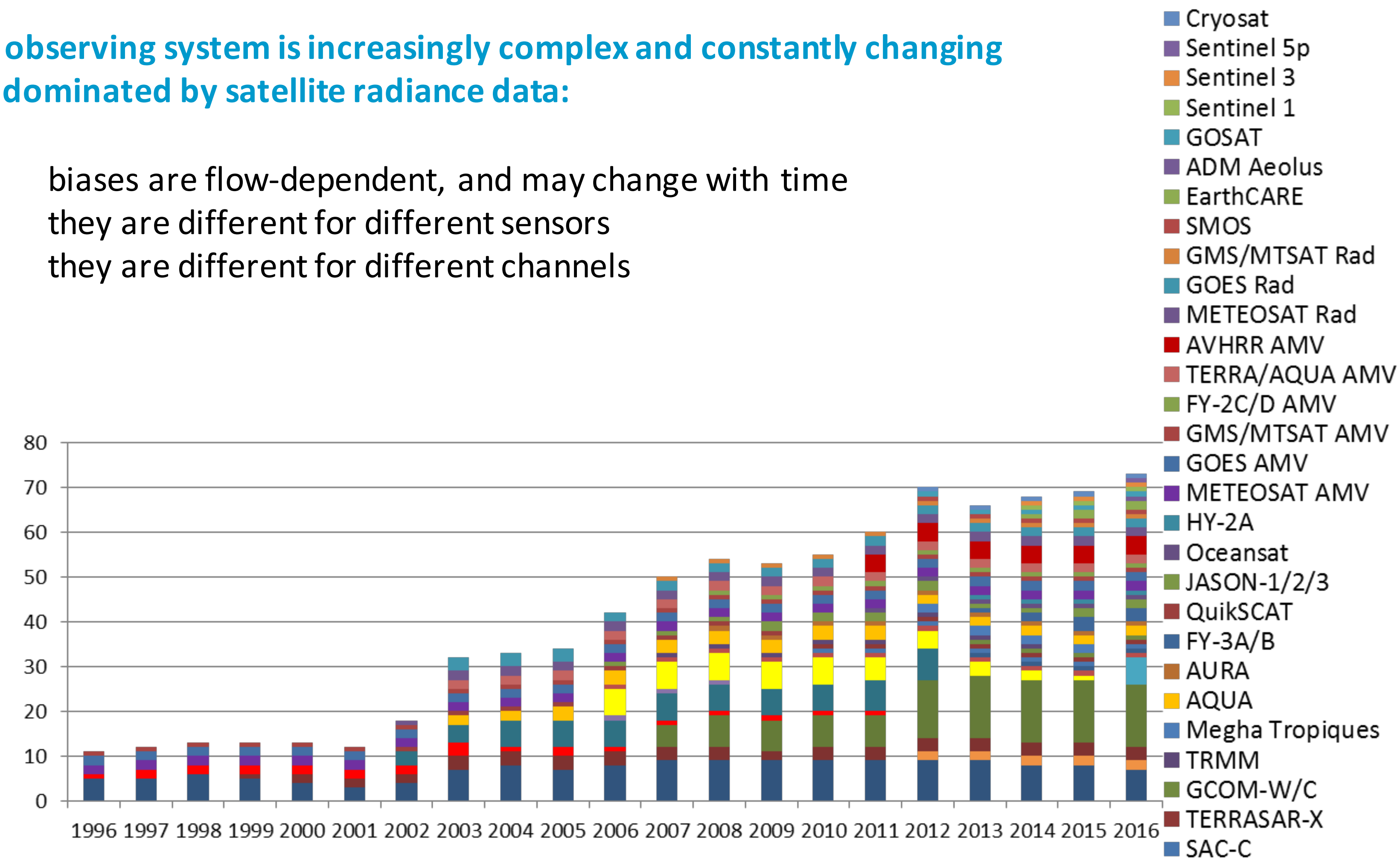
Periodically estimate scan bias and predictor coefficients:

- 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

Variational bias correction of radiance data

- 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, i.e inspired by Harris and Kelly 2001
- 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_b: background constraint for x

J_β: background constraint for β

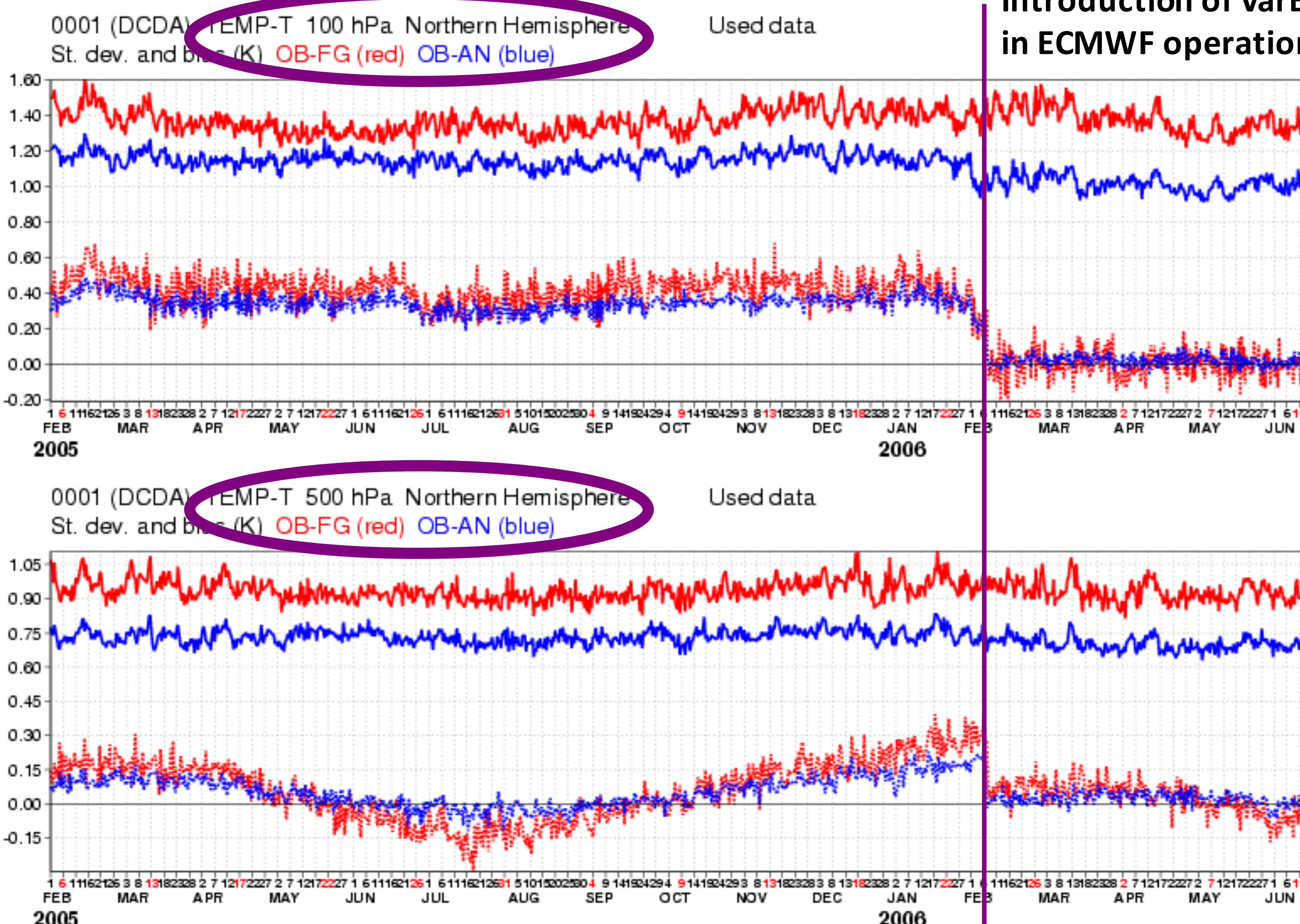
$$J(x, \beta) = \underbrace{(x_b - x)^T B_x^{-1} (x_b - x)}_{J_b} + \underbrace{(\beta_b - \beta)^T B_\beta^{-1} (\beta_b - \beta)}_{J_\beta}$$
$$+ \underbrace{\left[y - b_o(x, \beta) - h(x)\right]^T R^{-1} \left[y - b_o(x, \beta) - h(x)\right]}_{J_o}$$

J_o: bias-corrected observation constraint

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

The ability for anchors to do their job:
fit to conventional data

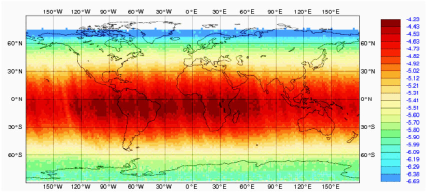
Introduction of VarBC in ECMWF operations



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***: predictor for solar elevation
- ***Aircraft data:*** one group per aircraft
- ***Total column water vapour***: ENVISAT MERIS until April 2012
- ***Ground-based radar precipitation***: one group embracing US stations



Bias estimate Aura/OMI

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, RASE for radiosondes
- ***ERA-20C:*** VarBC surface pressure; one group per station
- ***MACC***: atmospheric composition
- **ERA5**: as in current operational model, but VarBC for surface pressure, RASE or VarBC for radiosondes

Examples of VarBC in ERA interim

Satellite data used in ERA-Interim

Microwave radiances
- temperature sounding
- water vapor sounding

Infrared radiances
- temperature and water vapor sounding
- stratospheric temperature sounding

Imagery — visible, near infrared, water vapor

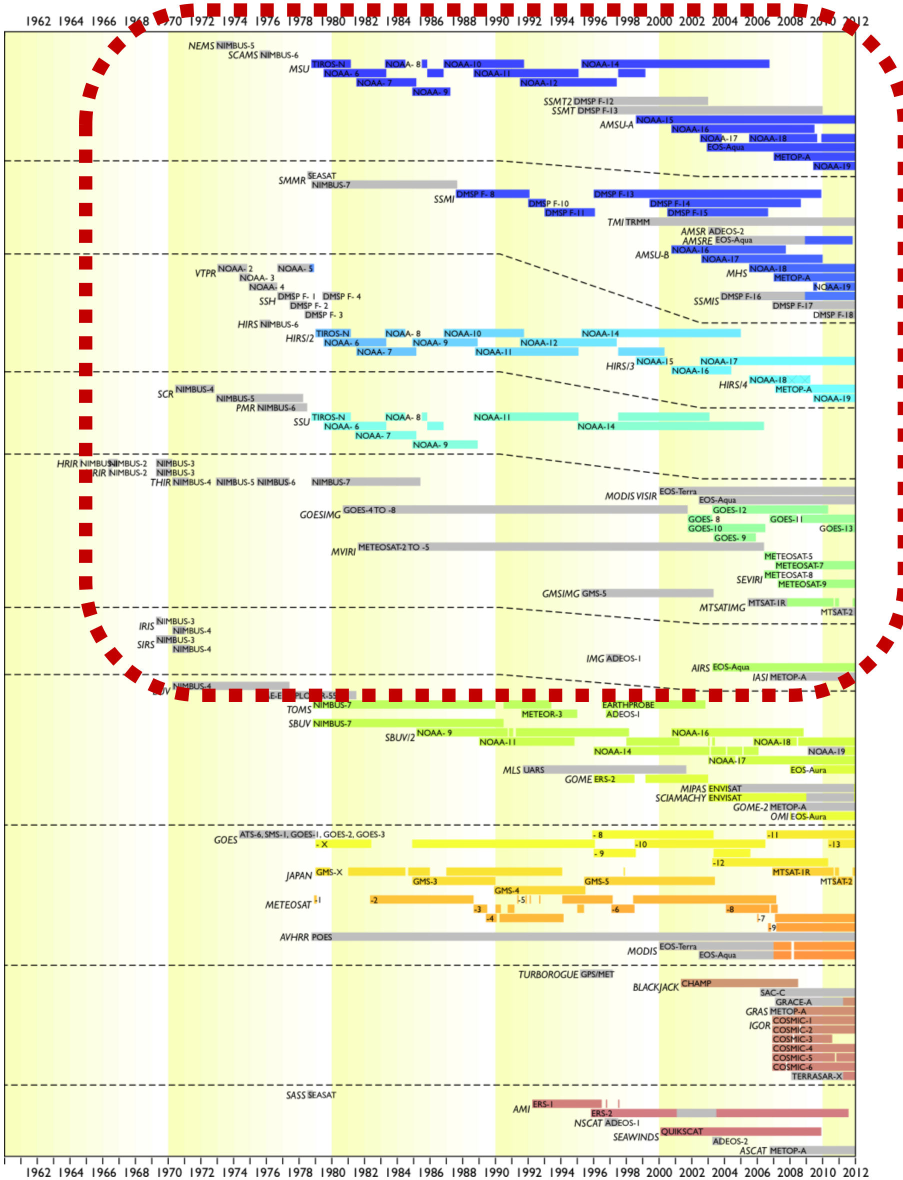
Hyper-spectral infrared

Ozone
- mostly ultra-violet, some limb-viewing infrared

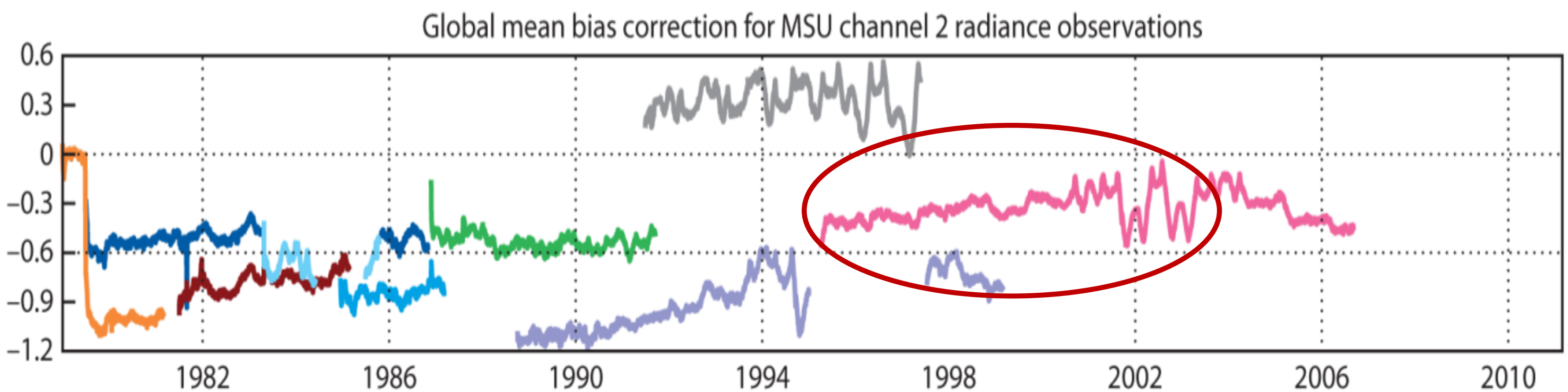
Atmospheric motion vectors
- geostationary (GEO)
- low-earth orbit (LEO)

Bending angles from GPS radio occultation

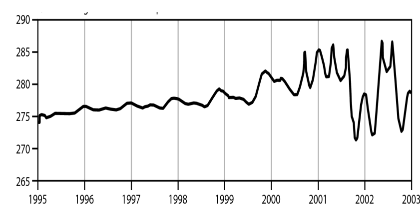
Backscatter — near-surface wind above ocean



Clear example of observation bias: instrument drift for MSU



On-board warm target variations for MSU NOAA-14 *(Grody et al. 2004)*



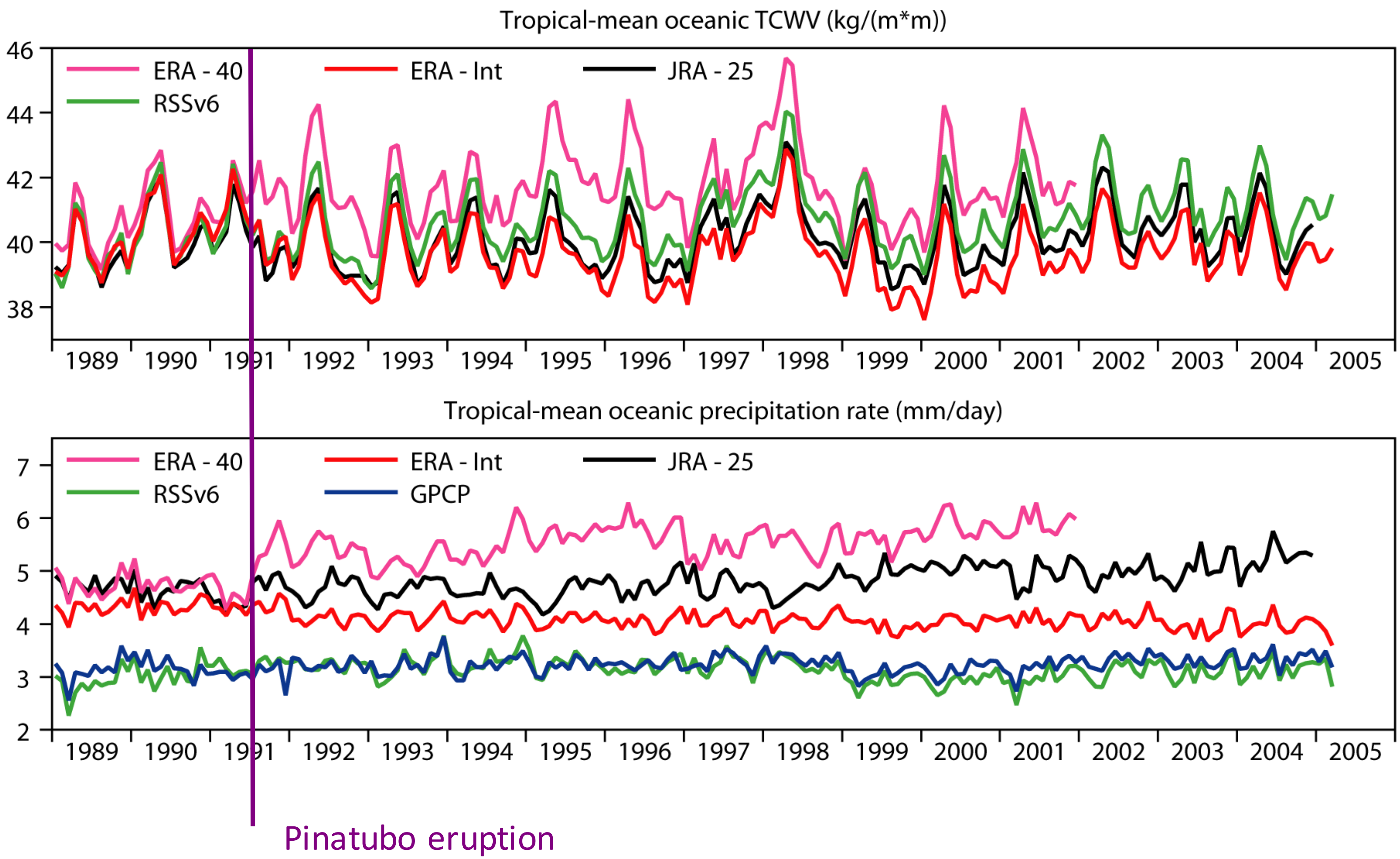
Stratosphere

Topics:

- Response to Pinatubo
- Constraining the upper stratosphere to address model bias

Response to Pinatubo eruption

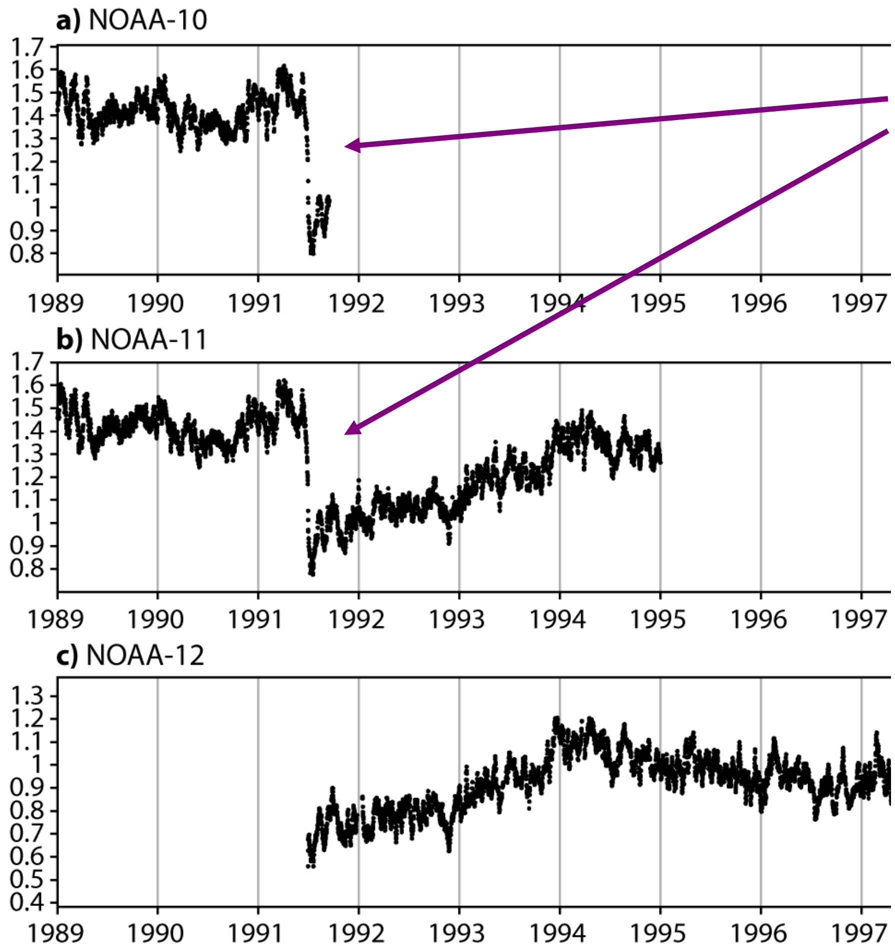
ERA-40: Excessive precipitation over tropical oceans – worse after Pinatubo



ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Response to Pinatubo: HIRS Ch11

Bias corrections for HIRS Ch11 (tropical averages)



Volcanic aerosols in the lower stratosphere:

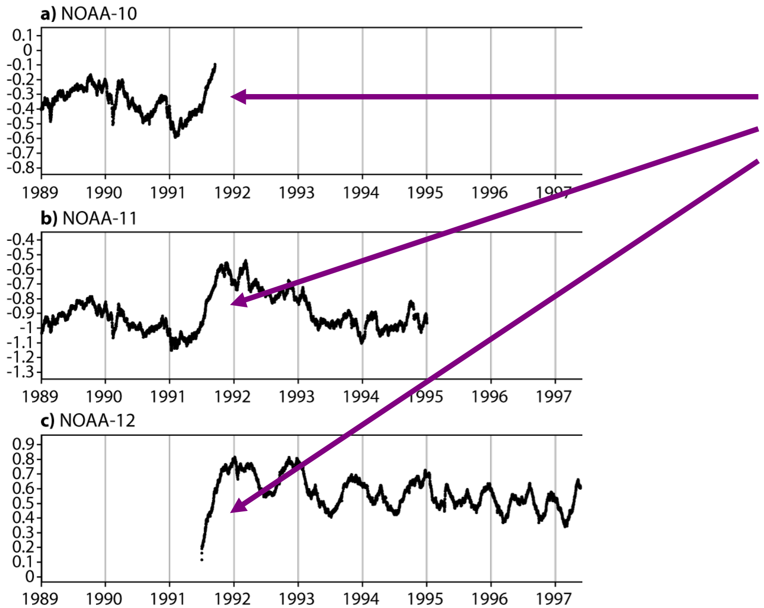
- Cooling effect on radiances
- Not in the radiative transfer model
- ERA-Interim: Change the bias correction
- ERA-40: Change the humidity increments

Bias corrections for NOAA-12:

- In ERA-Interim, correct initialisation followed by gradual recovery
- In ERA-40, bias was held fixed

Response to Pinatubo: MSU Ch4

Bias corrections for MSU Ch4 (tropical averages)



Volcanic aerosols in lower stratosphere:

- Microwave radiances are insensitive to aerosol, but correctly measure warming of the stratosphere
- The effect of aerosol changes on radiation is not accounted for in the forecast model (biased cold)
- This causes a (false) bias adjustment for MSU

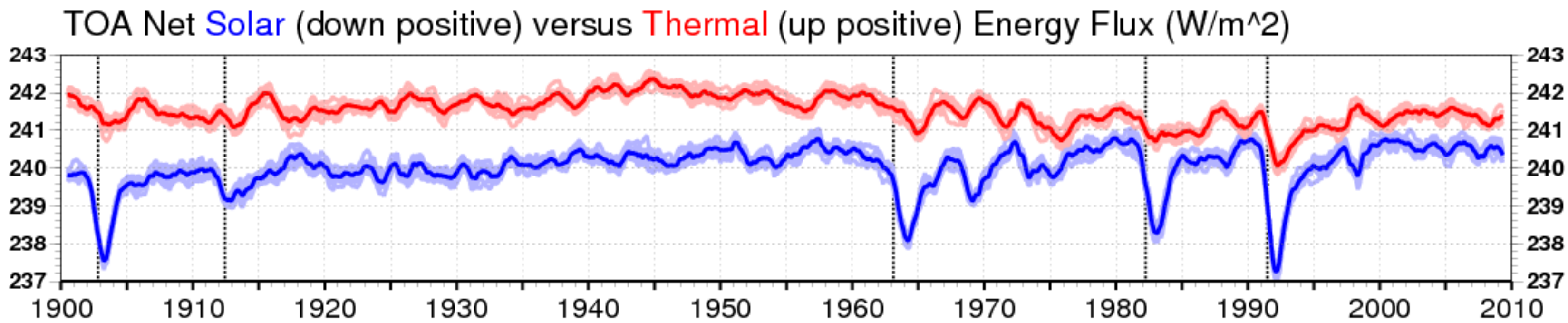
The result is a **slight damping of the Pinatubo signal** in ERA-Interim

Still the best option, given large variations in the MSU biases

Fundamental limitation of variational bias correction:

- bias parameters are used to minimise mean departures, regardless of the cause
- variational bias correction may not work well in poorly observed regions with large model biases

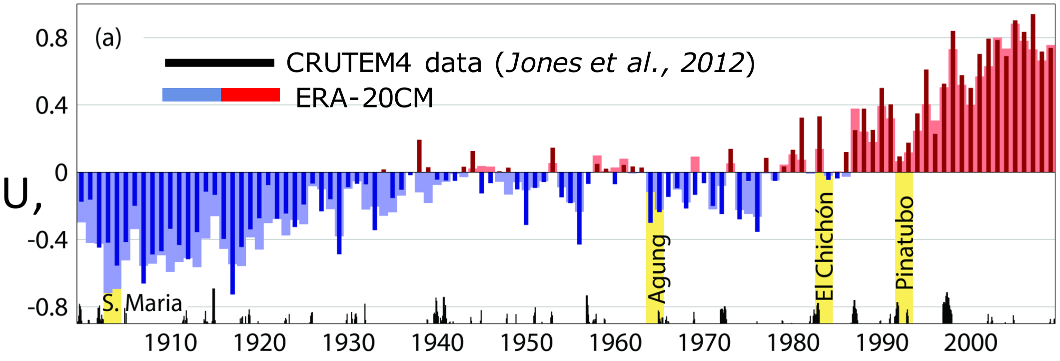
ERA-20CM: model-only integration based on CMIP5 forcing



+ HadISST2

The model knows about Pinatubo

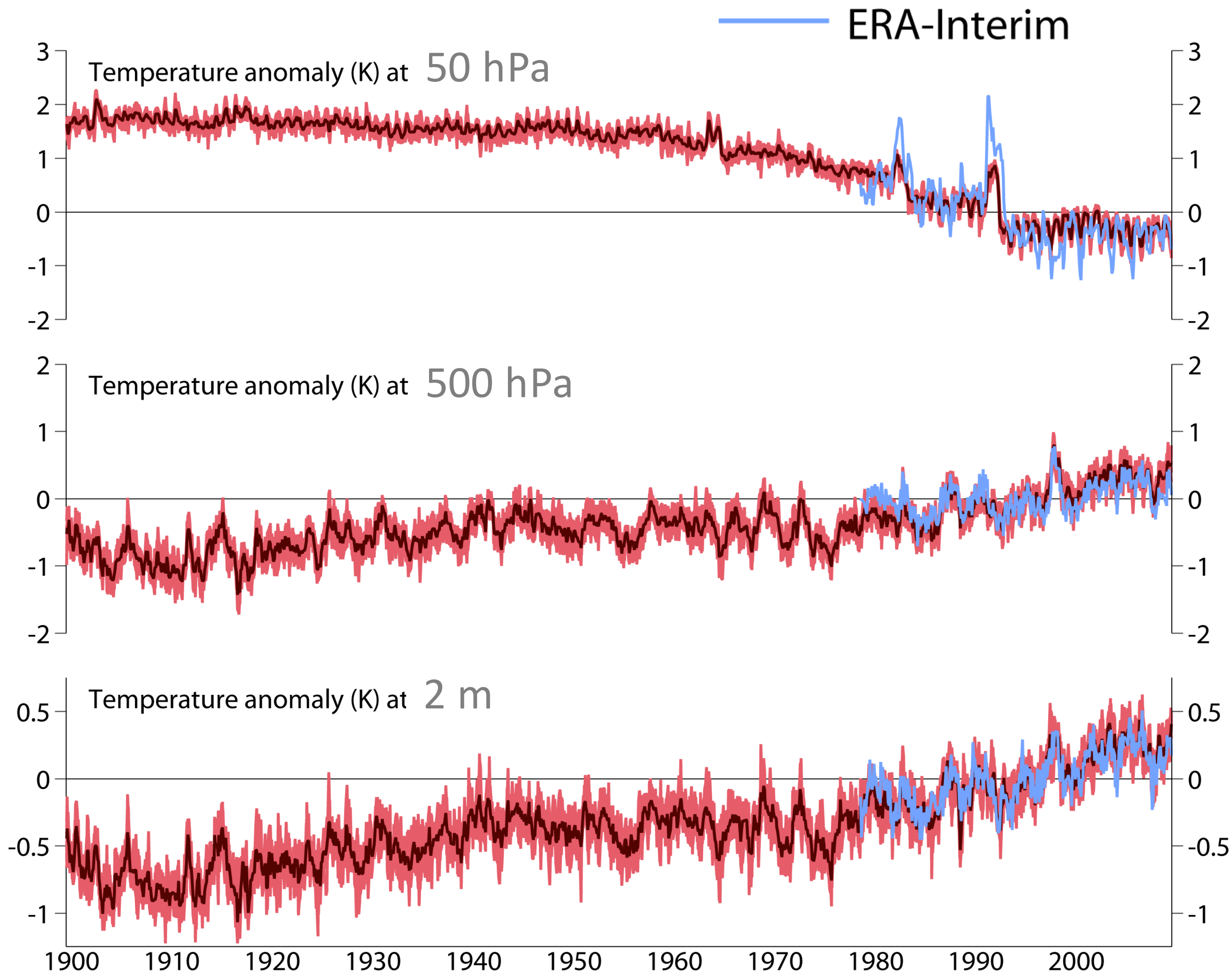
ERA5 will use CMIP5 forcing:
but not for RTTOV, so
should address model issue for MSU,
But not RTTOV issue for HIRS



Hersbach, Peubey et. al. 2015, QJRMS

ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

ERA-20CM global temperature anomalies

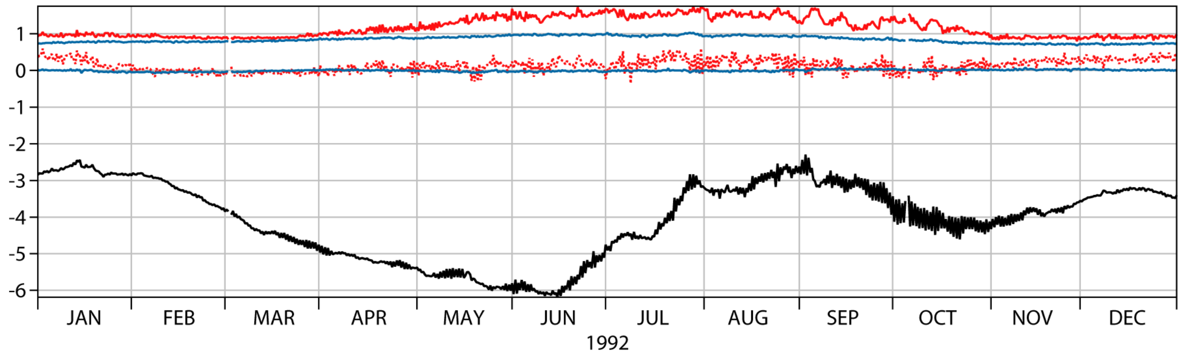


Stratosphere

Topics:

- Response to Pinatubo
- Constraining the upper stratosphere
  to address model bias

How to constrain model biases in the upper stratosphere?

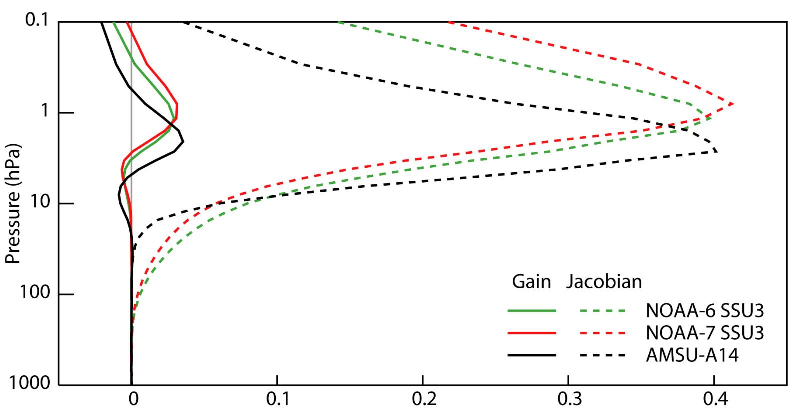


The model is generally too cold (by as much as 20K in polar winter)

Variational bias correction of SSU Ch3 would result in large temperature biases near the stratopause

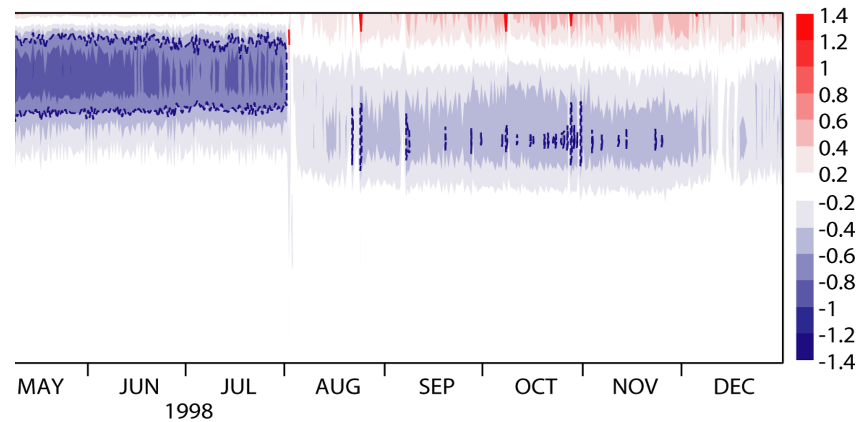
The top of the model must be constrained by uncorrected observations:
SSU Ch3 (available until 2006), AMSU-A Ch14 (available from 1998)

Jacobians for SSU3 and AMSUA-14



The constraints provided by each sensor are fundamentally different

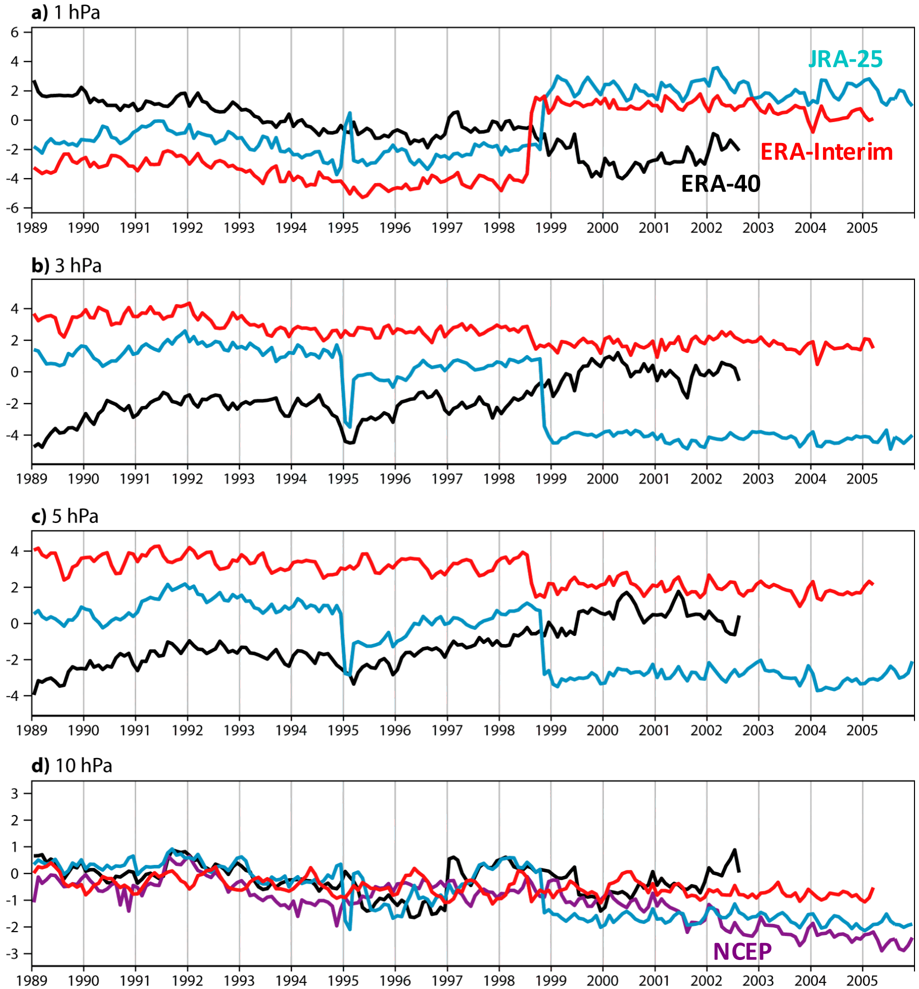
Global mean temperature increments above 40hPa



Both sensors result in systematic (but partial) corrections to the model background

Shifts in upper-stratospheric temperatures

Global mean temperature anomalies in the upper stratosphere



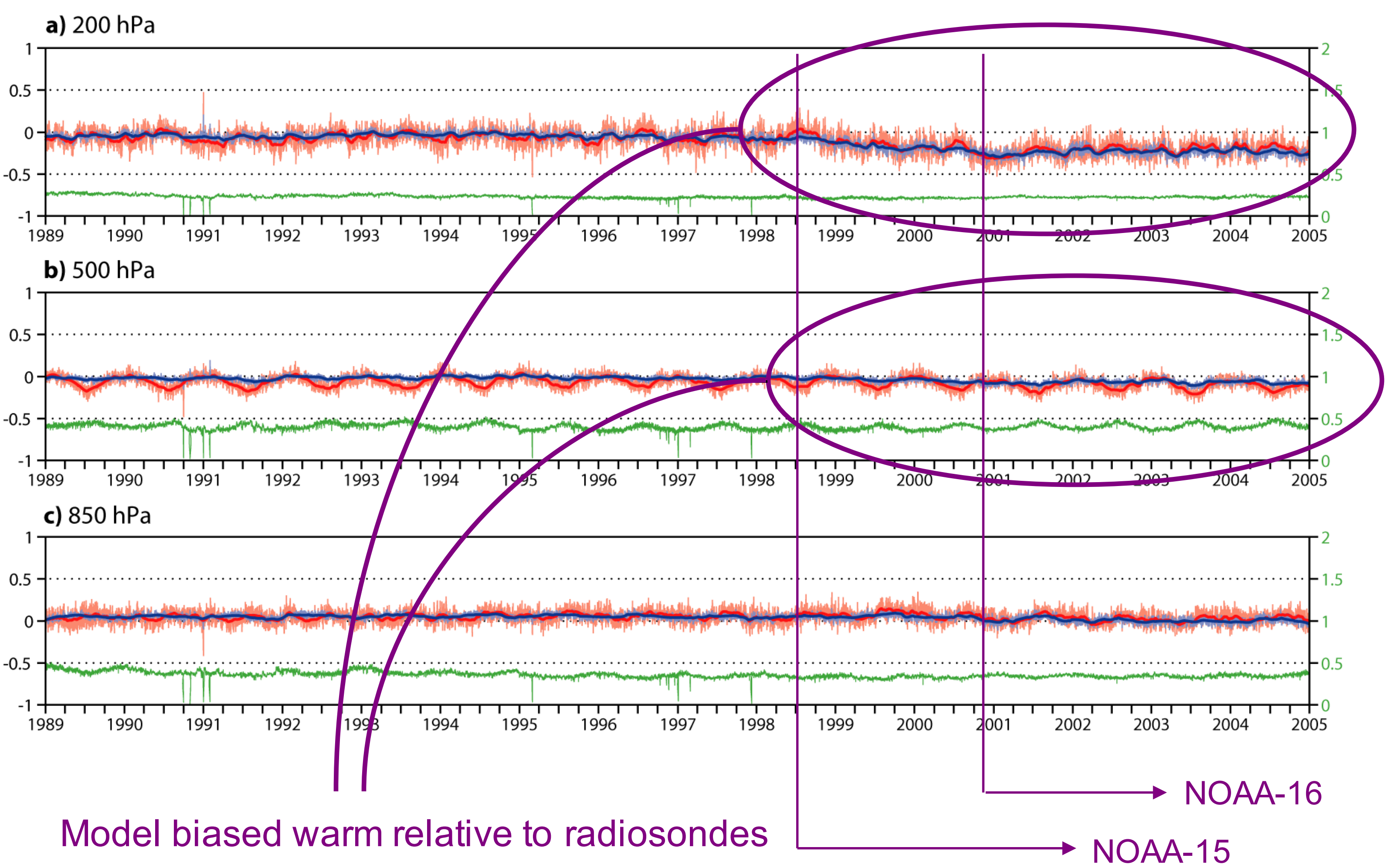
The transition from SSU Ch3 to AMSU-A Ch14 is clearly visible in global mean temperatures at 5hPa and above

This problem cannot be completely solved unless the forecast model is free of bias

Competition between anchors

Anchoring data for the troposphere: Radiosondes

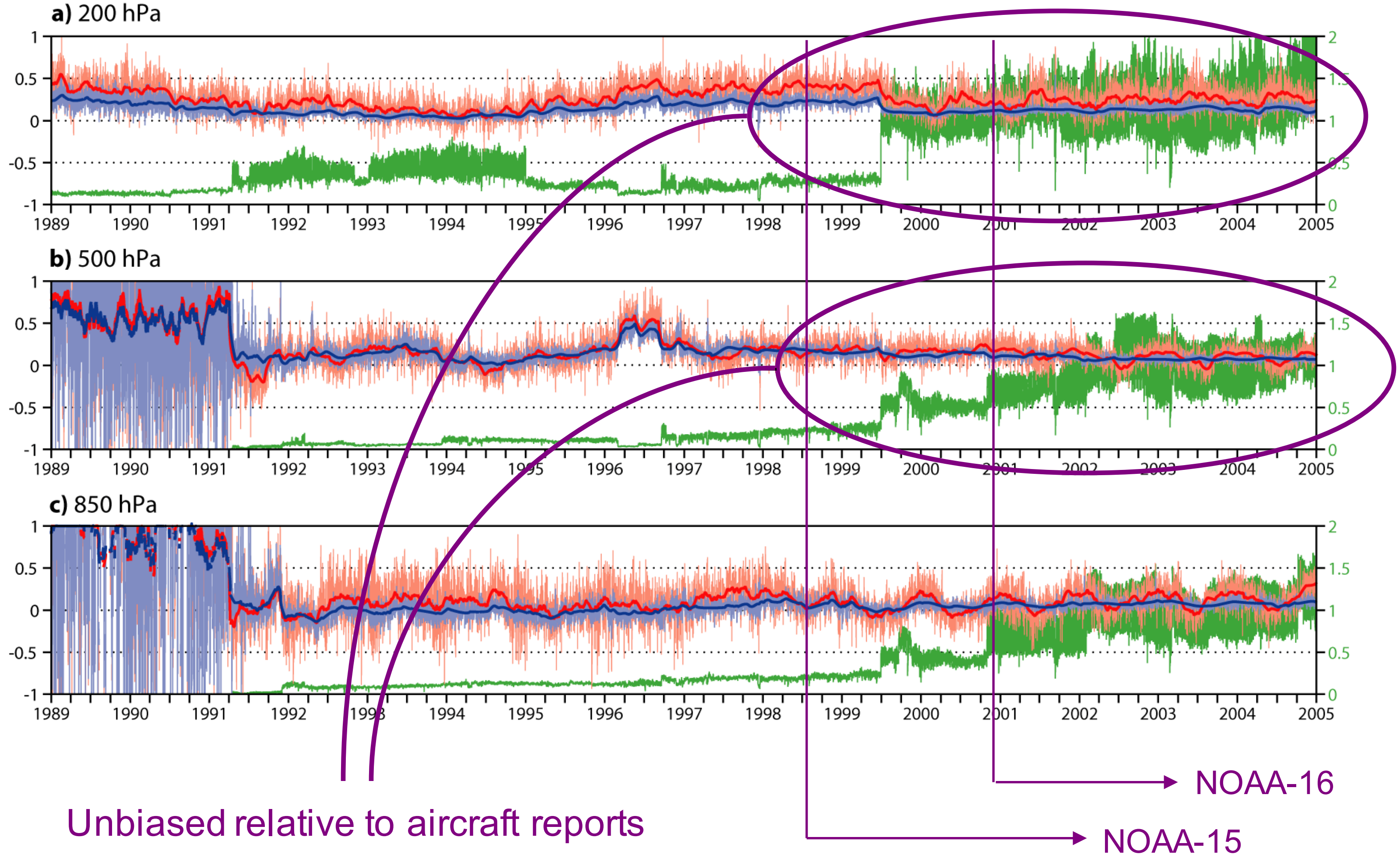
Global mean departures and data counts for radiosonde temperature data



ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Anchoring data for the troposphere: Aircraft reports

Global mean departures and data counts for aircraft temperature data



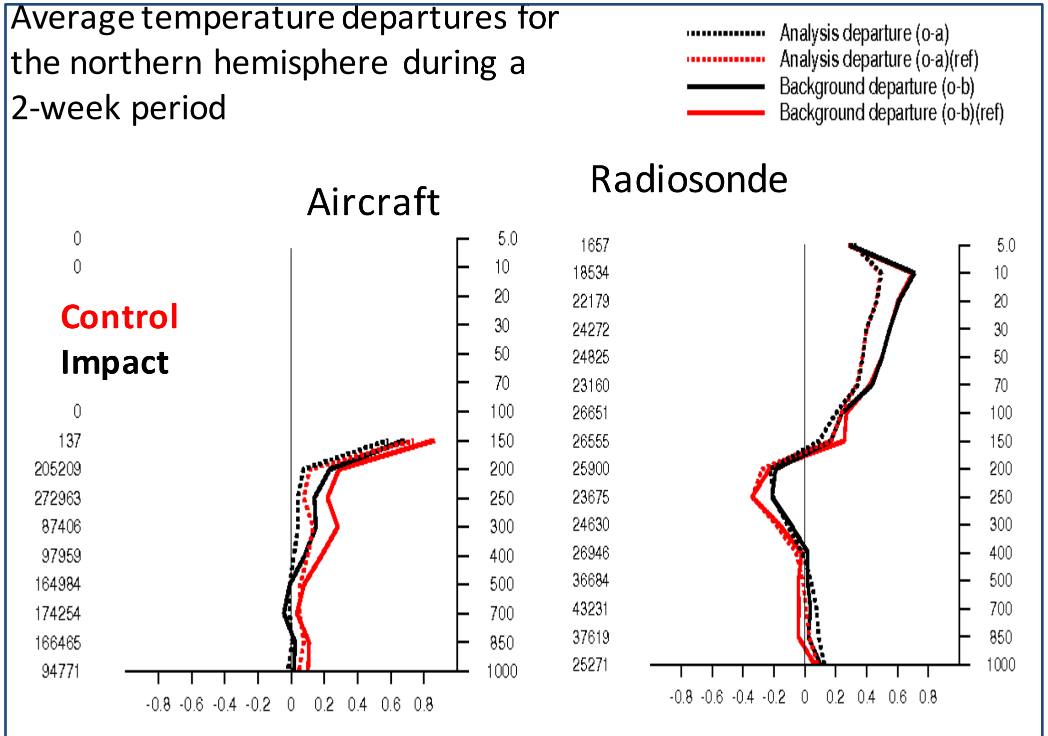
Unbiased relative to aircraft reports

Solution: VarBC for ***aircraft temperature***

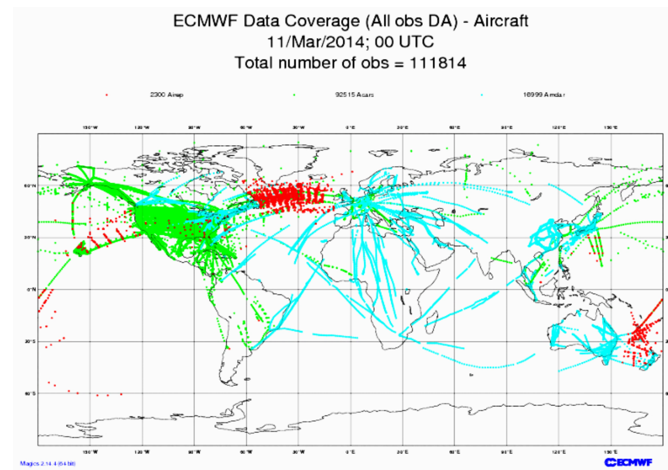
For each aircraft separately (~5000 distinct aircraft)

Anchored to all temperature-sensitive observations

Bias model: $\beta_0 + \beta_1$ x ascent rate + β_2 x descent rate



Average temperature departures for the northern hemisphere during a 2-week period



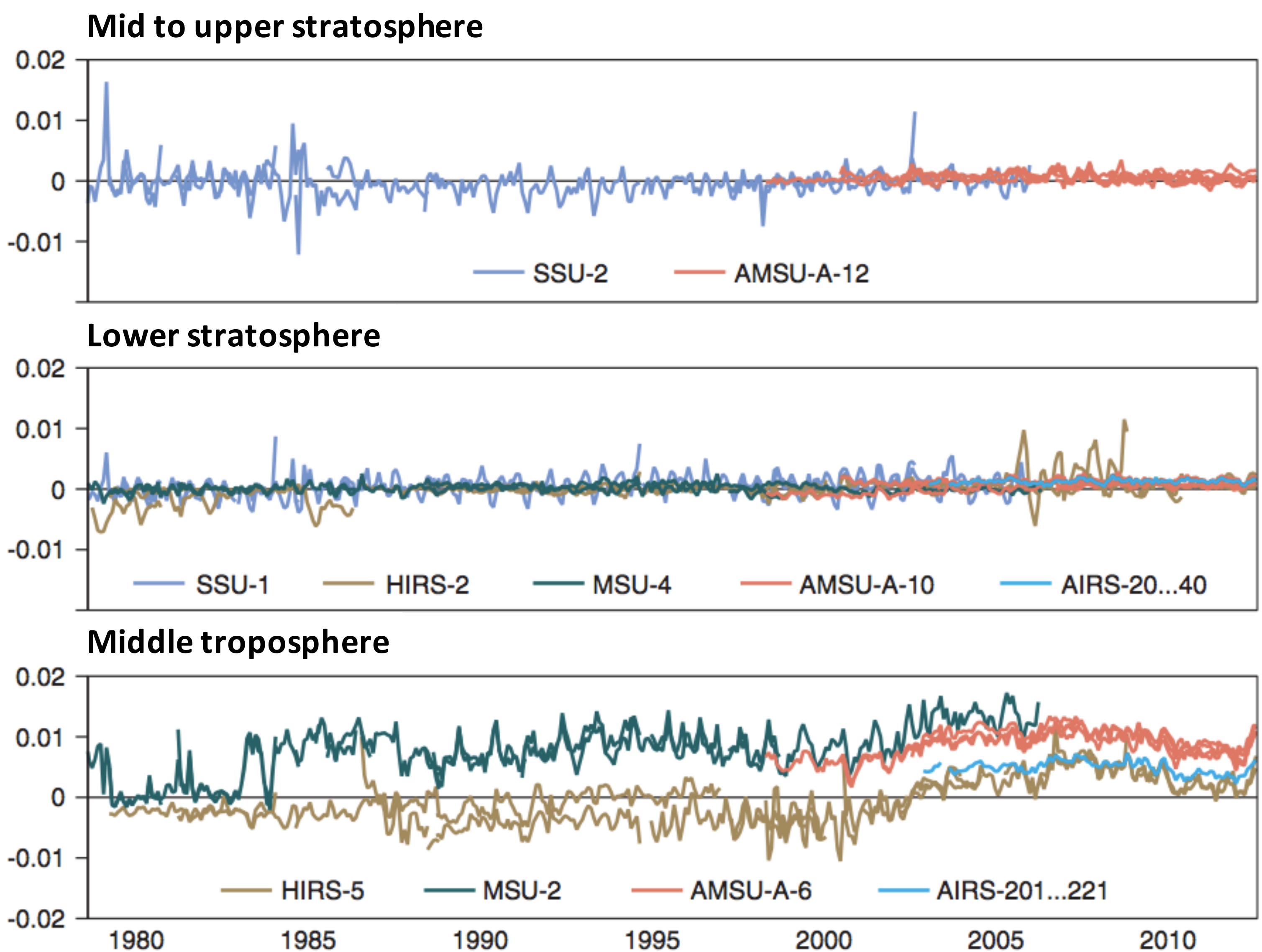
Used in the operational system

And to be used in ERA5

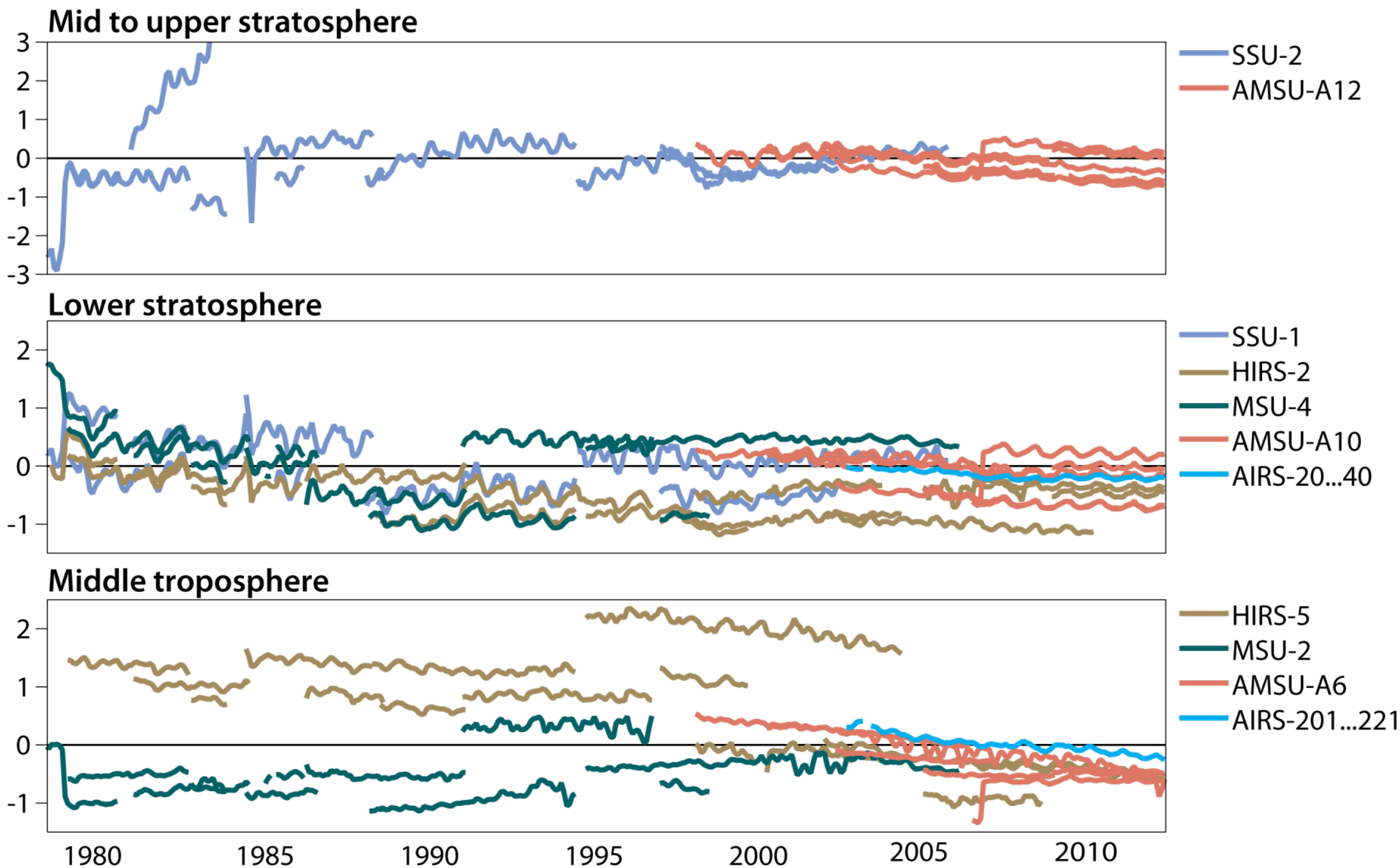
What do we learn from the applied bias corrections?

Do they provide inter-calibrated, homogenized long-term data sets?

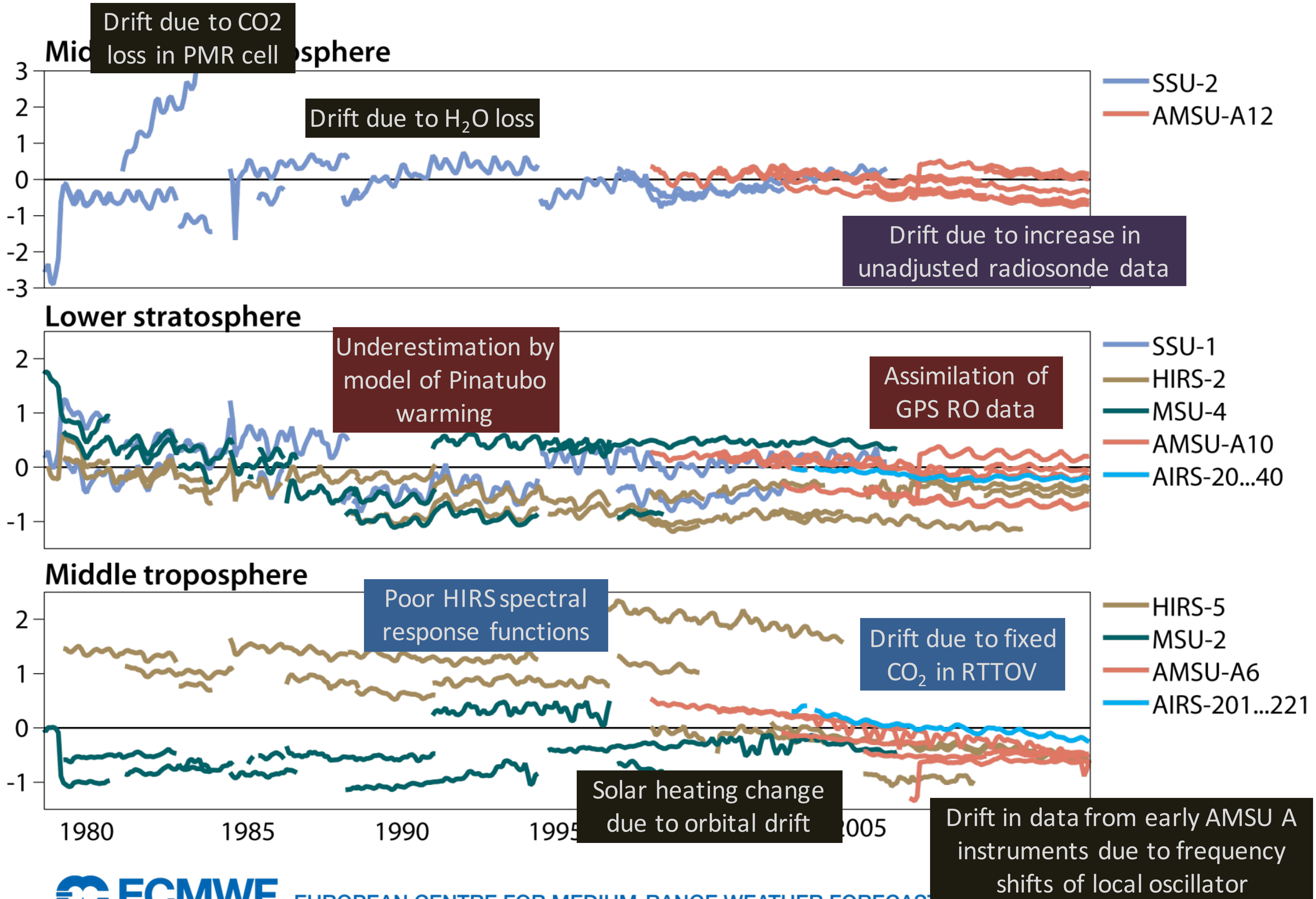
Mean fit to selected sounding channels [K]



Mean bias adjustments [K]

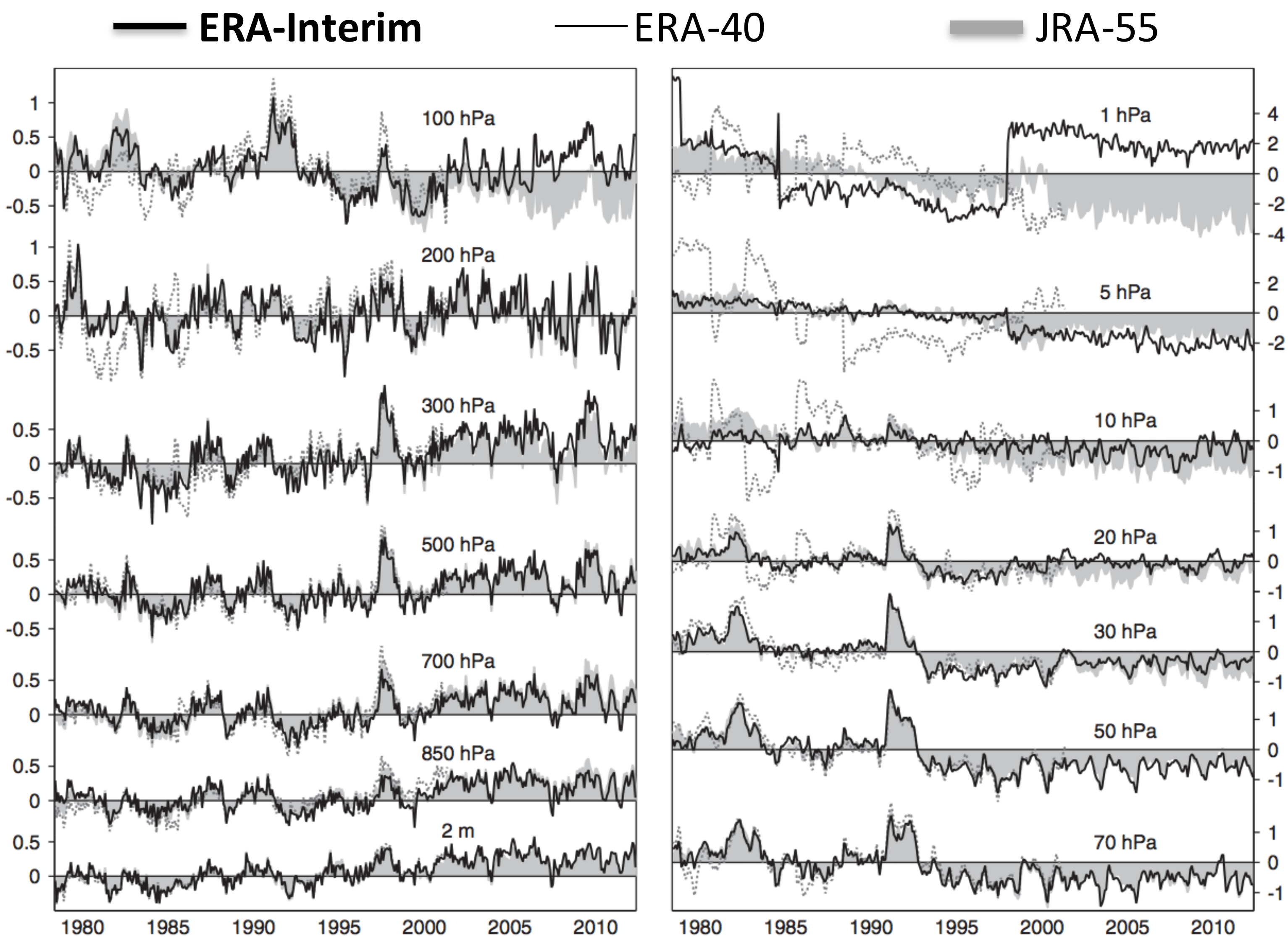


Mean bias adjustments [K]



ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

Global temperature anomalies



Concluding remarks

An adaptive system for dealing with data ingest and bias correction is practically indispensable for reanalysis in the modern satellite era. This was the original motivation for using variational bias correction in ERA-Interim.

From the ERA-Interim experience we have learned that an adaptive bias correction system is in fact a requirement to be able to correct time-varying instrument errors (e.g. MSU), to handle major atmospheric forcing events (e.g. Pinatubo), to detect data drifts (e.g. AMSU-A), and to maintain optimal consistency among all data sources.

The long-term behaviour of the variational bias correction system in ERA-Interim is stable, but it is necessary to constrain model bias in the upper stratosphere with uncorrected SSU and AMSU-A observations.

As long as models have systematic errors it is not possible to completely eliminate false climate signals in a reanalysis or to have confidence in the inter-calibration of long-term datasets.

In the presence of model error there is a conflict of interest between accuracy of the best reanalysis field and fidelity of climate trends.

By not anchoring model bias, trends can still be affected, though in an unnoticeable way.