Bias Correction

for

Environmental Monitoring

Richard Engelen

European Centre for Medium-Range Weather Forecasts

Many thanks to Antje Dethof, Soumia Serrar, Angela Benedetti, and others.

Outline

- Introduction
- Sources of bias
- Validation data
- Bias correction
- Summary

What is bias?

The main assumption of a Best Linear Unbiased Estimation (BLUE) system is that the expected values of the background and the expected values of the observations are equal to the expected values of the real world.

If we use the mean to represent the expected value of the ensemble, we need to define averaging time scales that are both practical and representative for the expected mean.

If we have for instance slow-varying (seasonal) model biases, this might not be straightforward.

Why correct for bias?

A biased background or biased observations result in a biased analysis. And a biased analysis can result in a biased forecast. And Mrs Jones in Norwich does not like a biased forecast.

Strong biases can also make the analysis system unstable, resulting in incorrect solutions to the minimization problem.

The main goal in environmental data assimilation is to provide unbiased analyses, not good NWP forecasts.

When using tracer analyses fields for surface flux inversions, remaining biases that are not constant in space and time can have disastrous effects.

Bias

Observation bias

Correct whenever possible with bias correction

Model bias

Should not be corrected by bias correction

Can cause rejection of valid observations

Can cause oscillations in analysis

Should be corrected by improving forecast model

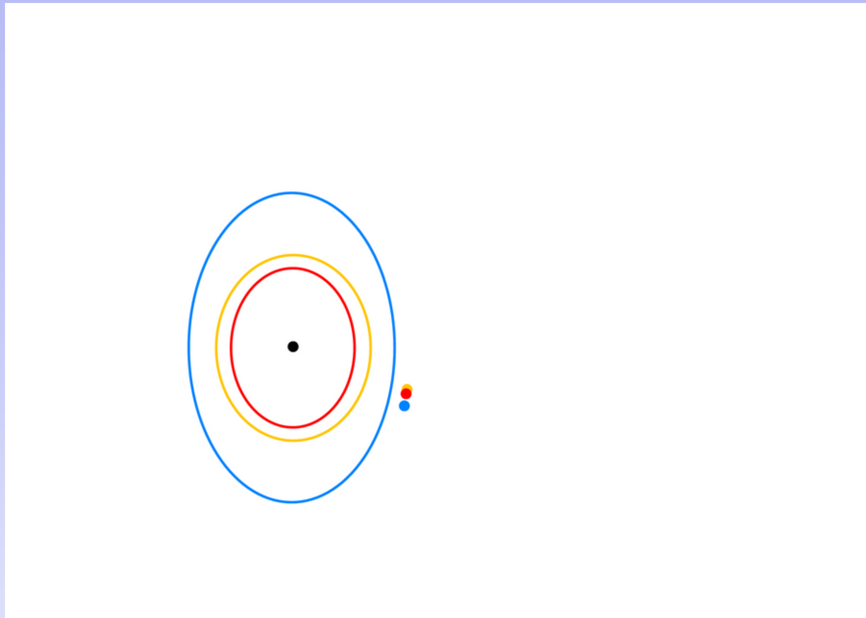
Bias in environment data assimilation

The main target in environmental data assimilation is to provide unbiased analyses, not good NWP forecasts.

This means that model bias is only important as part of the assimilation. It is not a real problem if the model drifts in a 10-day forecast.

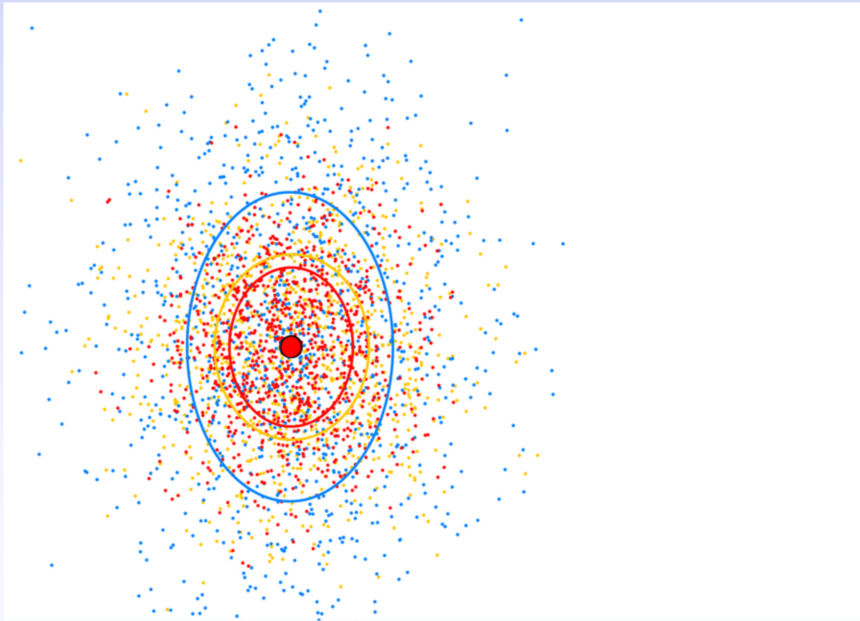
The effect of any model bias on the analysis is controlled by the ratio of the background errors and the observation errors.

What is bias?



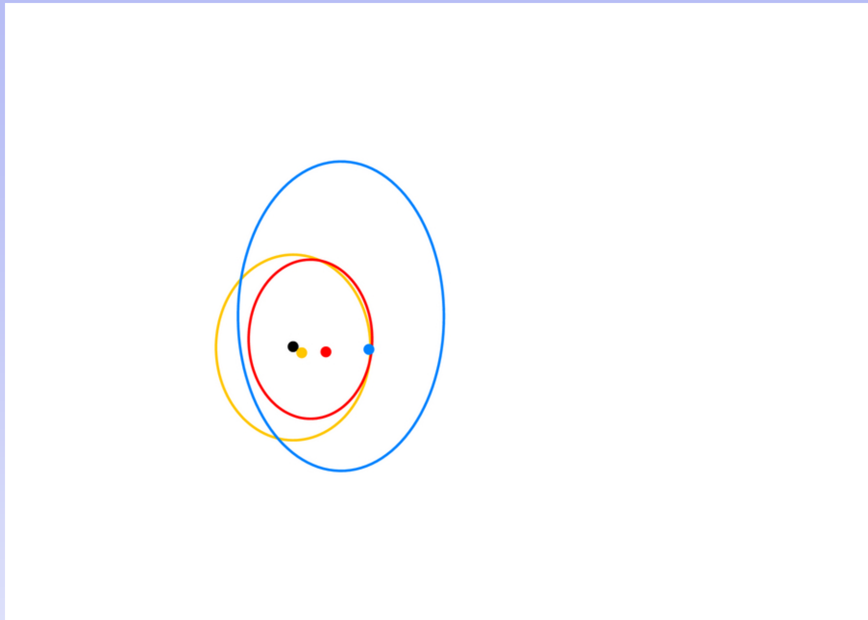
Background bias = 0

No problems



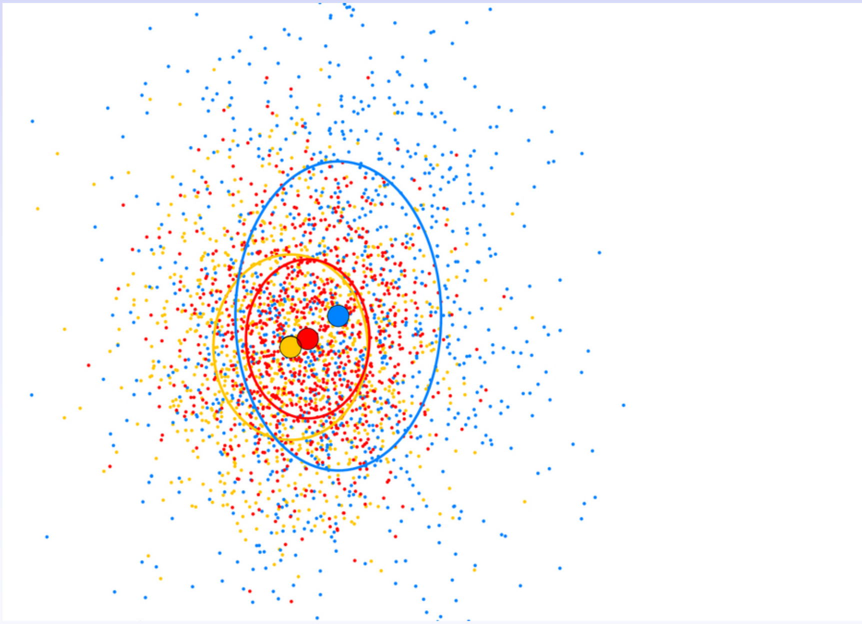
- True value
- Observation
- Background value
- Analysis value

What is bias?



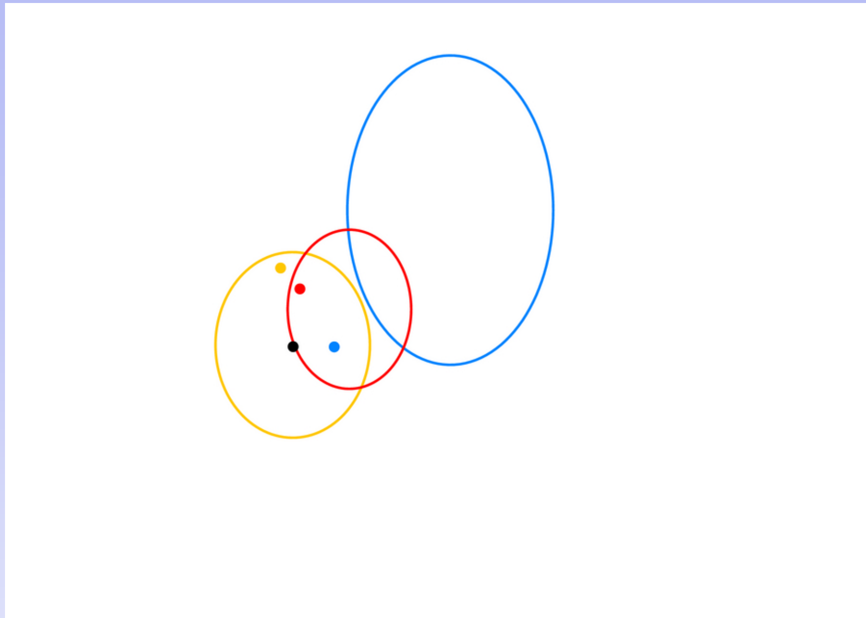
Background bias = O(σ)

Mean analysis is biased, but stability is probably still all right.



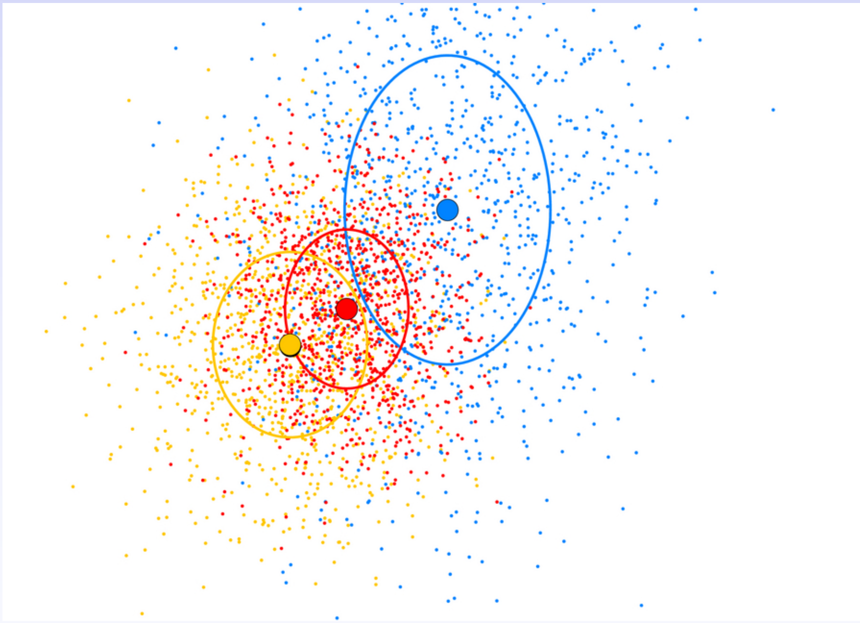
- True value
- Observation
- Background value
- Analysis value

What is bias?



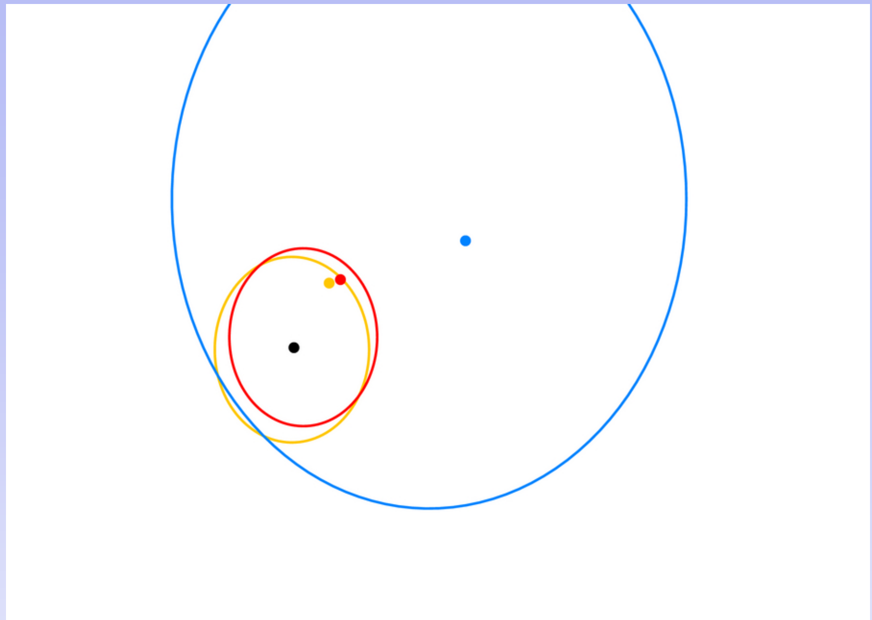
Background bias > σ

Mean analysis is biased, and stability problems can arise.



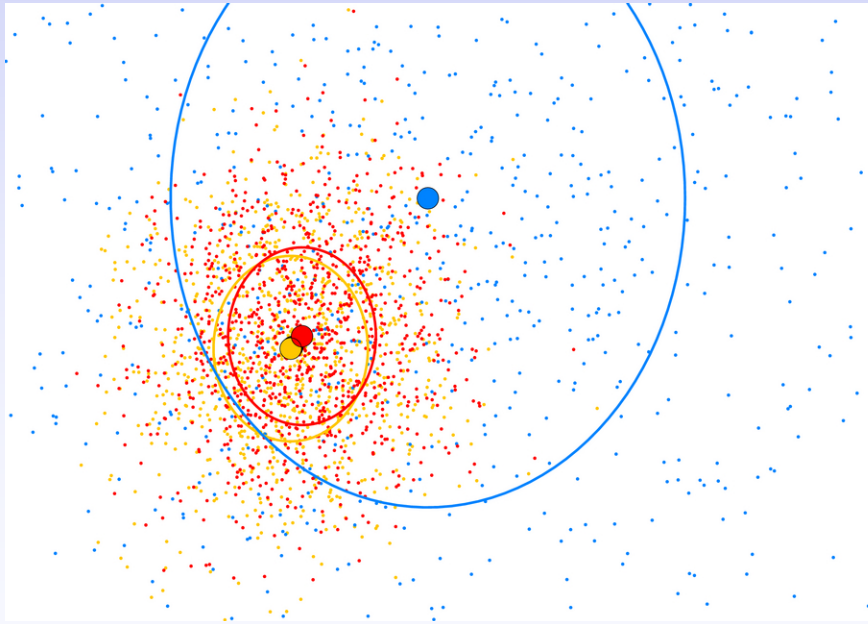
- True value
- Observation
- Background value
- Analysis value

What is bias?



Background bias < σ

Mean analysis is much less biased.



- True value
- Observation
- Background value
- Analysis value

Summary so far

- Effect of bias in background depends on the ratio of the bias to the standard deviation of the error.
- In environment monitoring we would like to correct background bias with observations to get a less biased analysis.
- This could imply the use of large standard deviations in the background covariance matrix.
- Observations should be bias-corrected as best as possible.
- What time scales do you consider for bias correction?

Causes for biases

- Instrument bias (relatively constant)
- Aerosol and cloud effects (regionally varying)
- Radiative transfer biases (air mass dependent)

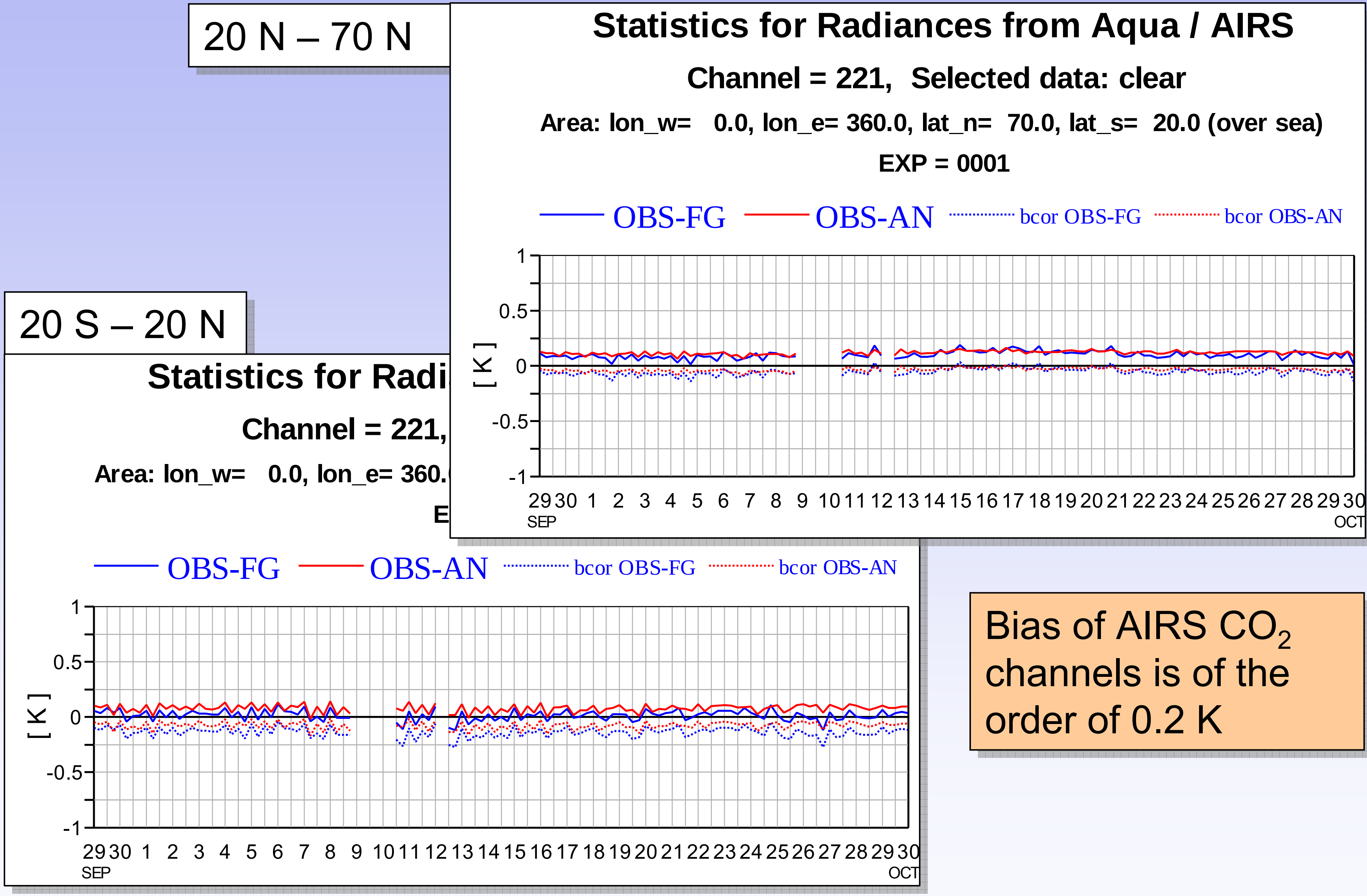
- Forecast model transport
- Forecast model physics
- Forecast model surface fluxes
- Forecast model chemistry

Satellite Instruments in GEMS

- Greenhouse
  - AIRS
  - IASI
  - CrIS
  - Mopitt
  - Sciamachy
  - OCO
  - GOSAT
- Chemistry
  - Sciamachy
  - SBUV
  - OMI
  - TOMS
  - GOME
  - MIPAS
  - MLS
- Aerosol
  - MODIS
  - SAGE

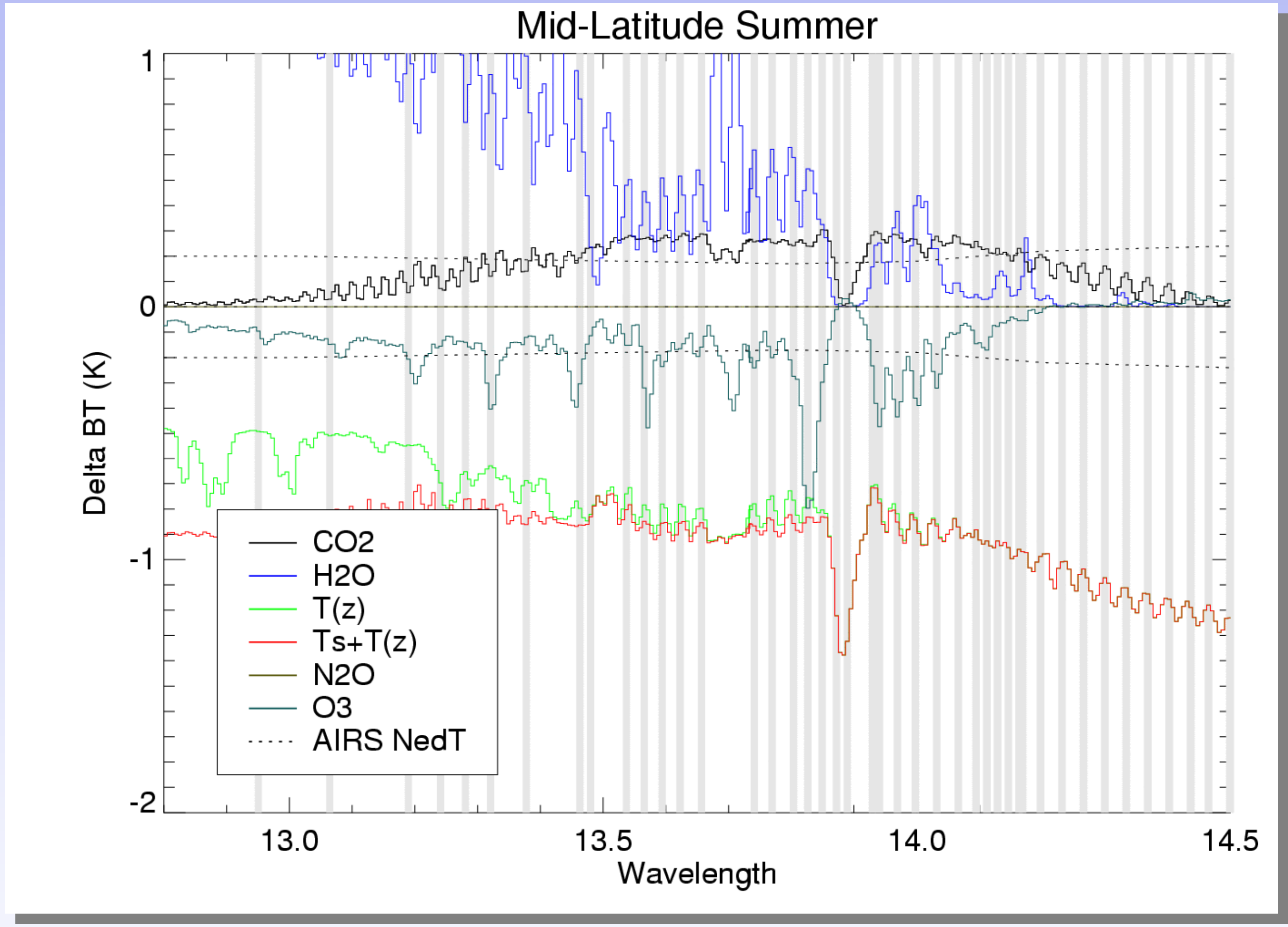
Causes for bias: spectroscopy errors, (undetected) clouds, (undetected) aerosol, surface reflectivity errors, air mass factors, errors in climatological temperature fields.

Observation bias



Bias of AIRS CO_2 channels is of the order of 0.2 K

Observation bias



AIRS CO_2 signal has the same order of magnitude (0.2 – 0.3 K)

Example: γ-correction and CO_2

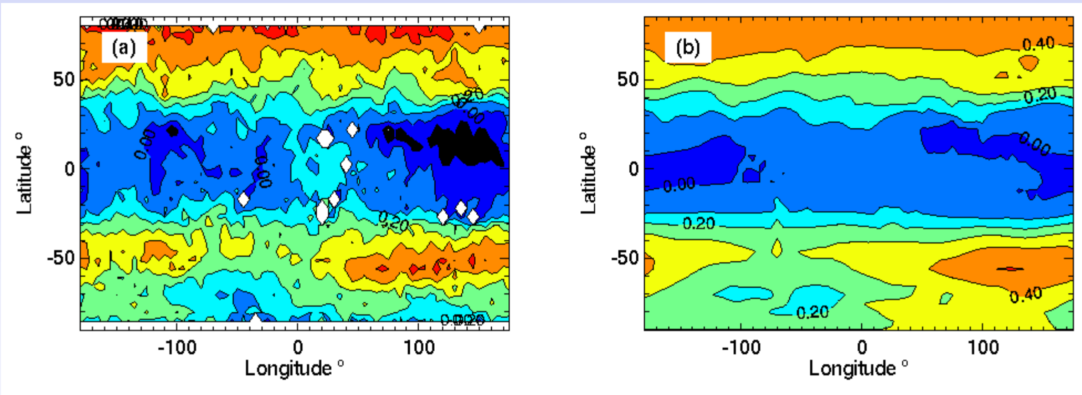
$$\Gamma(p) = \exp\left[-\gamma \int_{p}^{0} \kappa(p)\rho(p)\,dp \right]$$

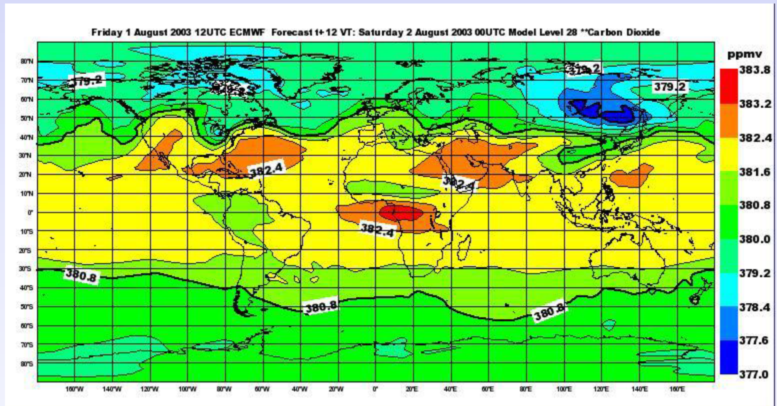
$\Gamma(p) =$ transmittance

$\gamma =$ correction coefficient

$\kappa(p) =$ absorption coefficient

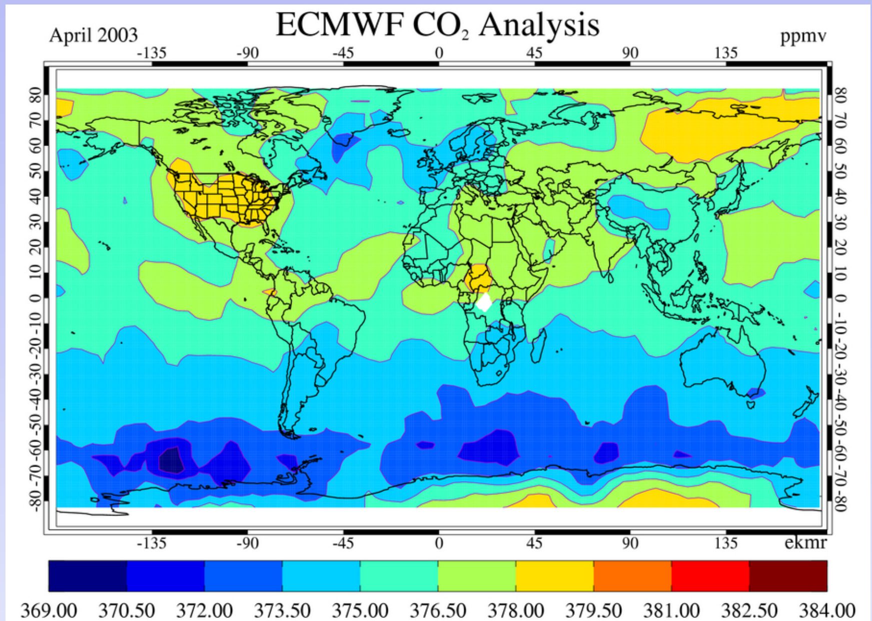
$\rho(p) =$ absorbing gas density



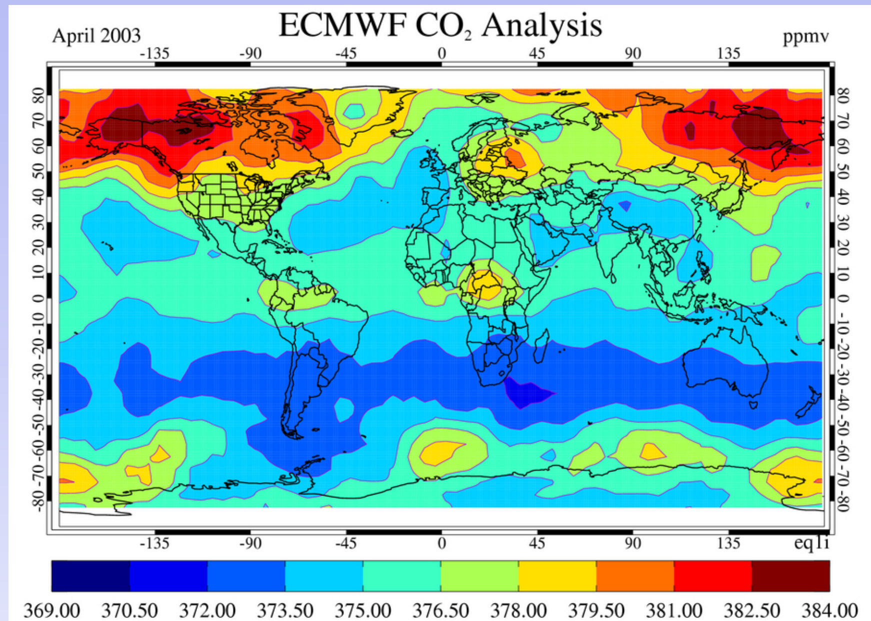


An error in the absorption coefficient will move the weighting function up or down. A gamma correction will therefore produce an air-mass (temperature) dependent bias correction.

Example: γ-correction and CO_2



flat bias correction



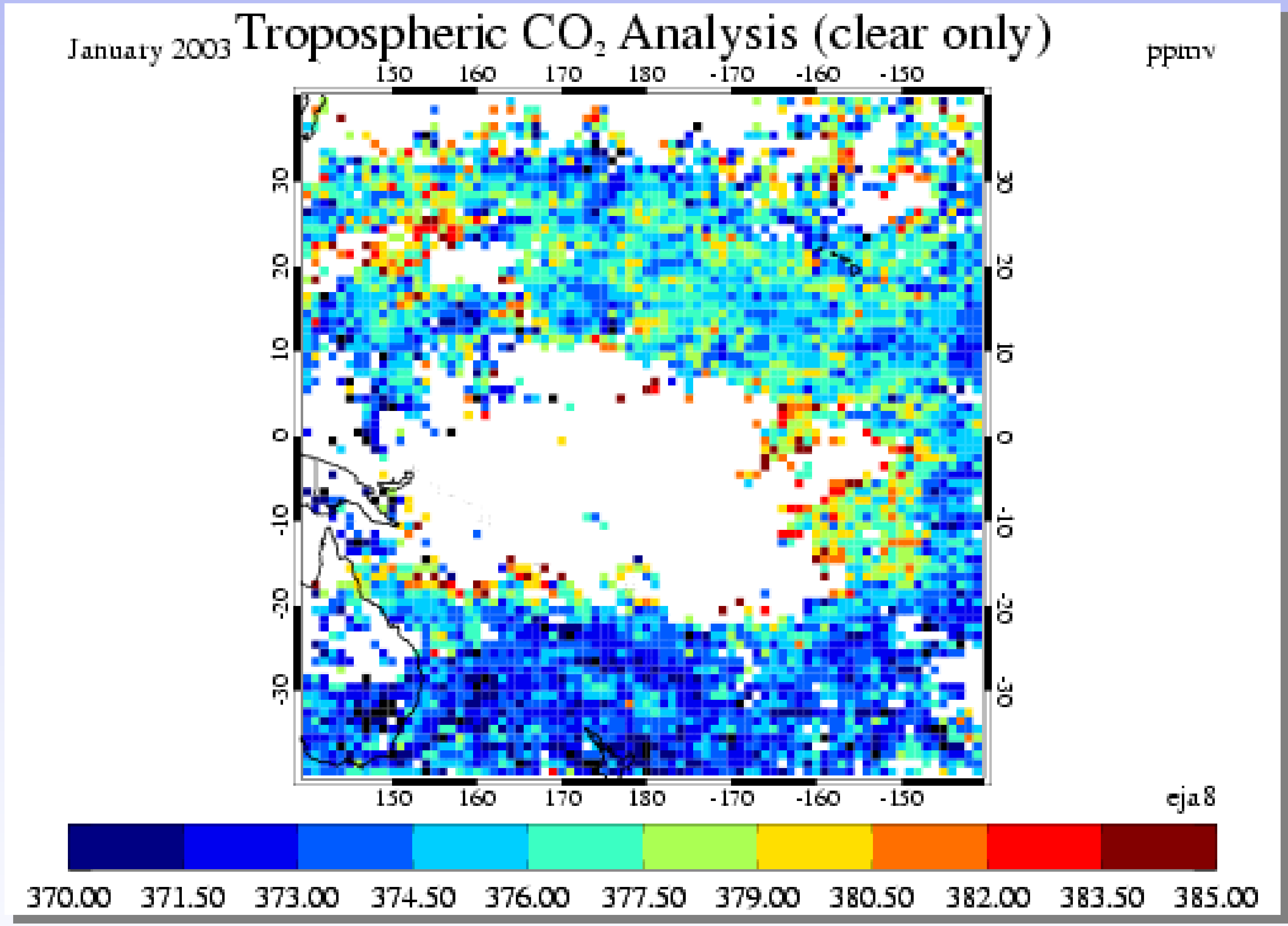
γ bias correction

Because the CO_2 signal in AIRS radiances is small, differences in bias correction can be dramatic.

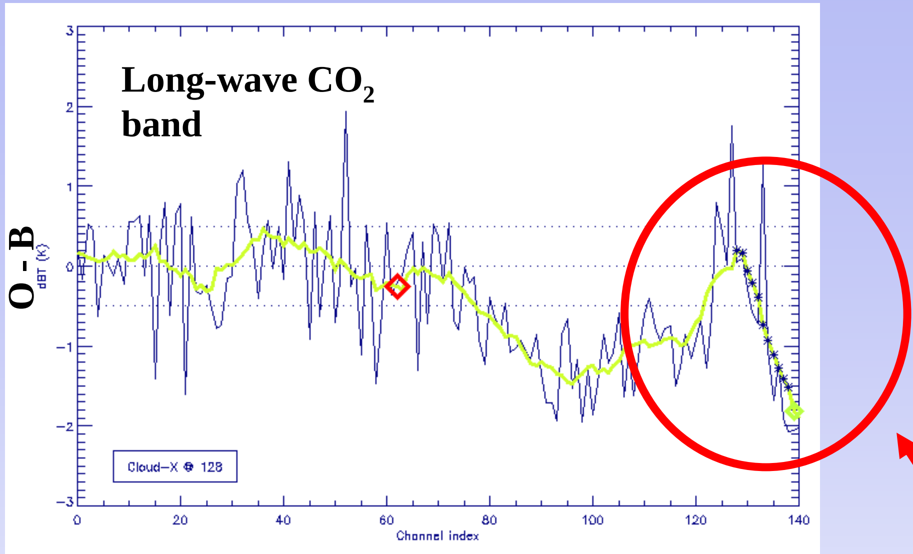
If the CO_2 model bias is spatially correlated to the γ-patterns, some of this model bias ends up in the bias correction.

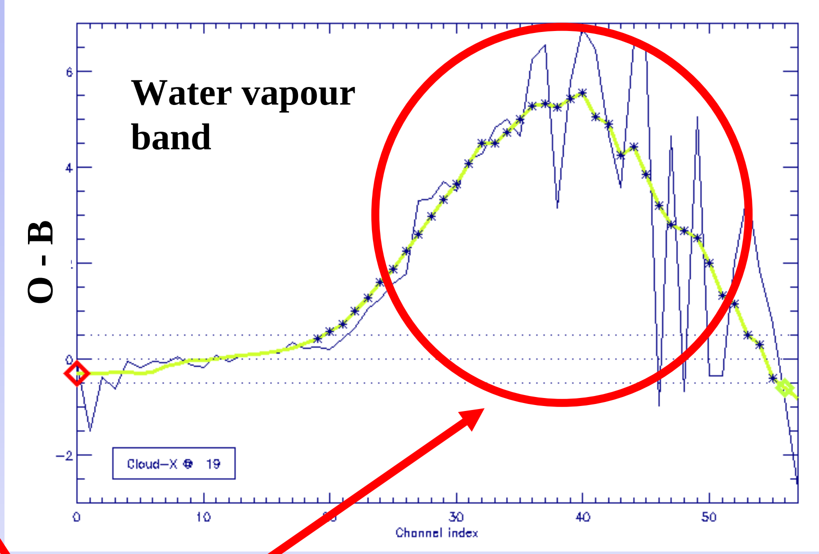
Only proper validation can help to sort things out.

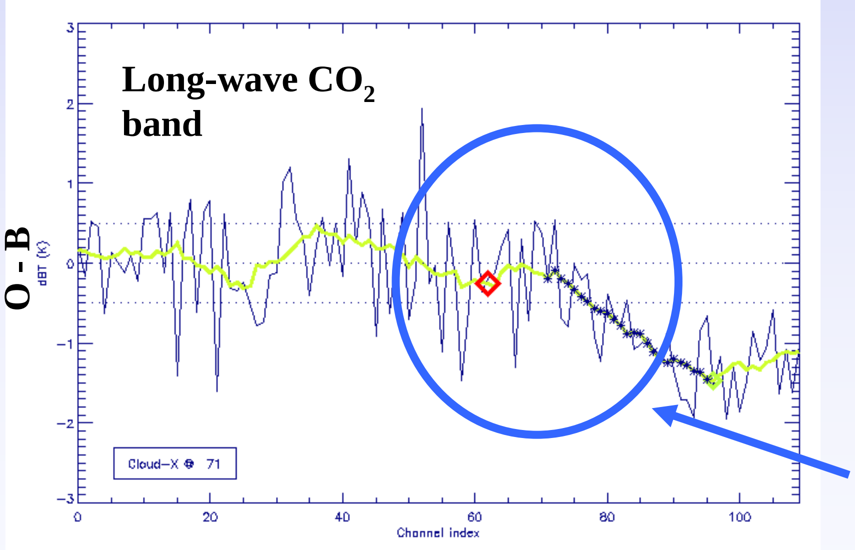
Cloud bias effect on CO_2 estimates



Cloud detection of tropical thin cirrus





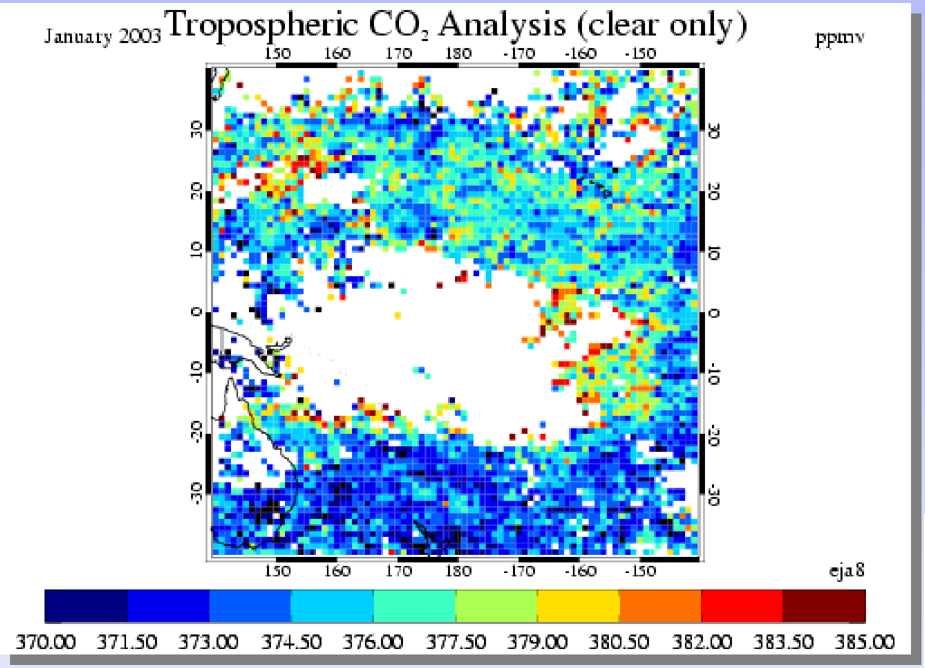


A water vapour background error affects the cloud detection in the water vapour band and the long wave band in cases of blown-off thin cirrus on top of a dry troposphere.

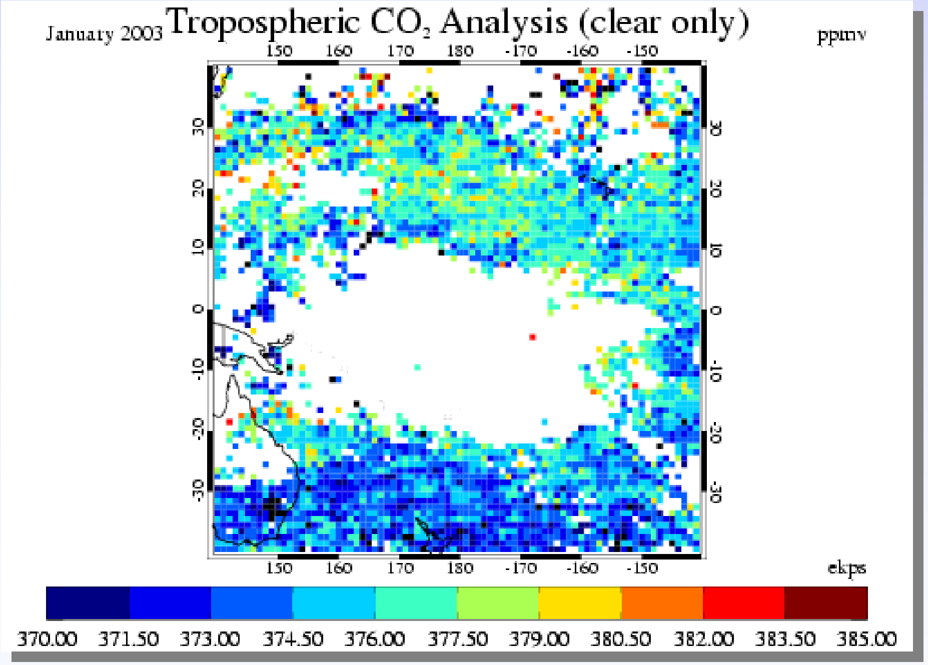
Removing the water vapour sensitive channels from the long-wave cloud detection helps to detect the thin cirrus.

Effect on CO_2 estimates

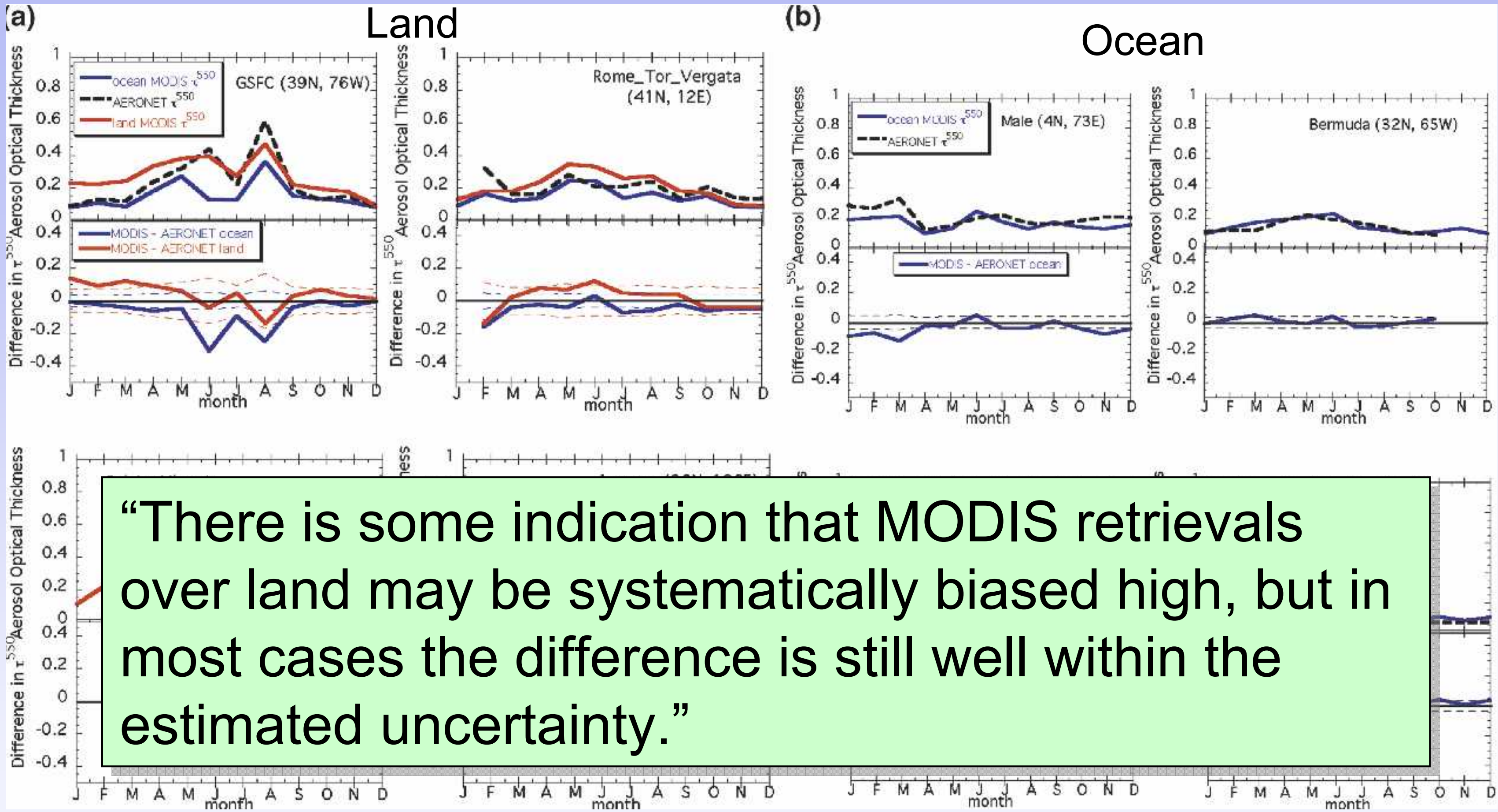
Before



After



Observation bias

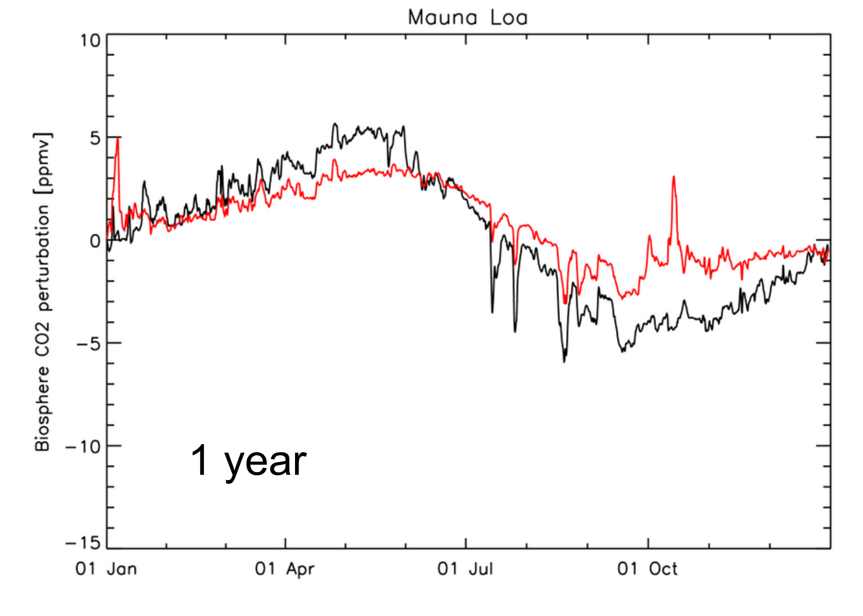


"There is some indication that MODIS retrievals over land may be systematically biased high, but in most cases the difference is still well within the estimated uncertainty."

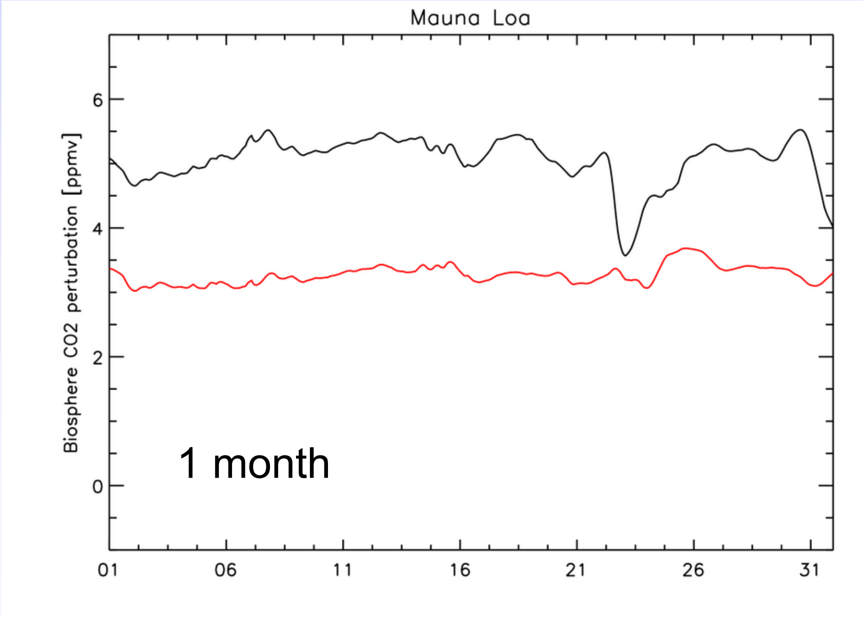
FIG. 10. Monthly mean aerosol optical thickness at 0.55 μm for the year 2001 at (a) four sites with land retrievals and (b) four sites with only ocean retrievals. The top portion of each plot shows the monthly means. The bottom portion shows the difference between MODIS and AERONET values. Also shown by thin dashed lines in the bottom portions are the prelaunch estimated uncertainties of optical thickness retrievals, $\pm 0.03 \pm 0.05\tau$ over ocean and $\pm 0.05 \pm 0.15\tau$ over land. Blue denotes MODIS ocean retrievals, red denotes MODIS land retrievals, and black denotes AERONET. The MODIS values are calculated from level 3 daily statistics and represent a 3° latitude by 3° longitude box centered on the AERONET station.

MODIS aerosol biases from Remer et al., JAS, 2005.

Model bias

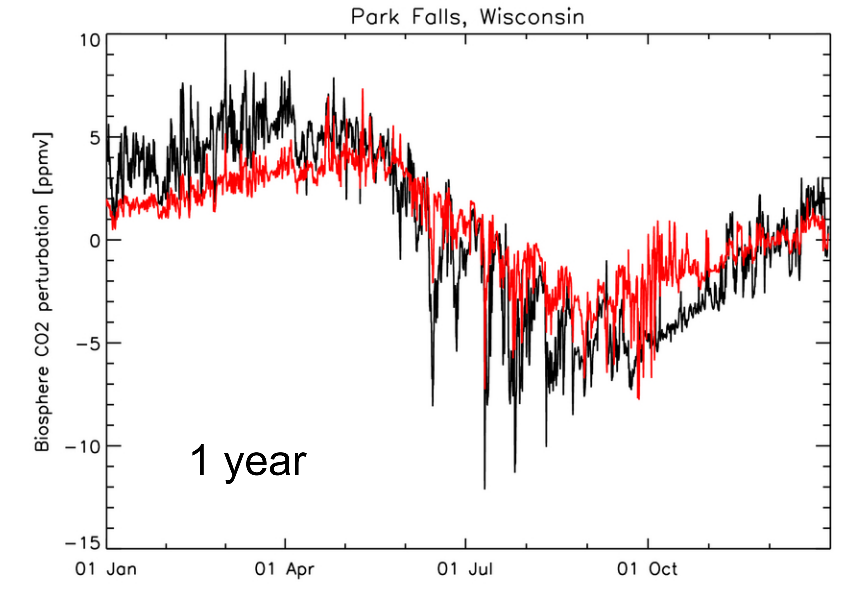


Using the same transport model with 2 different biosphere surface flux climatologies shows considerable (systematic) differences in CO_2 mixing ratios around 500 hPa.

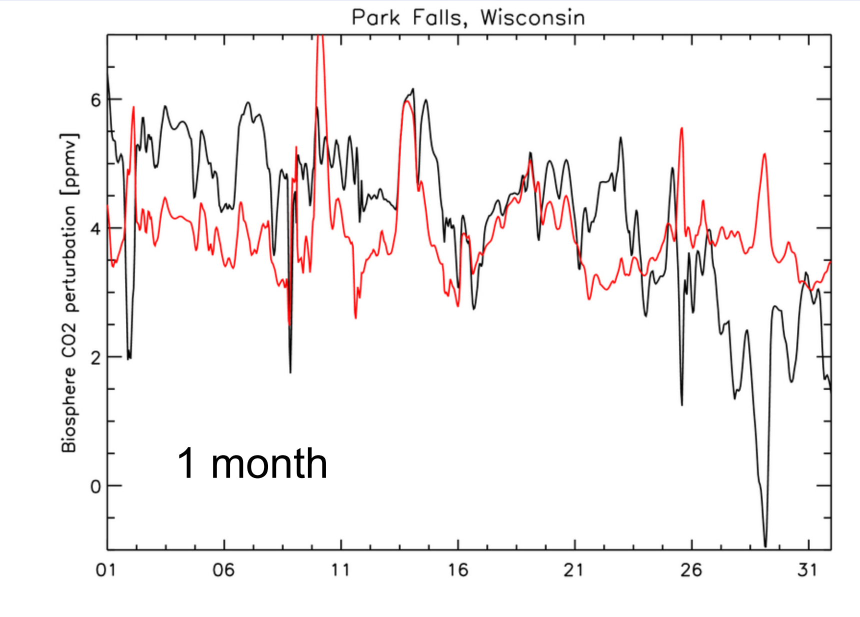


The magnitudes of these systematic differences are regionally dependent.

Model bias

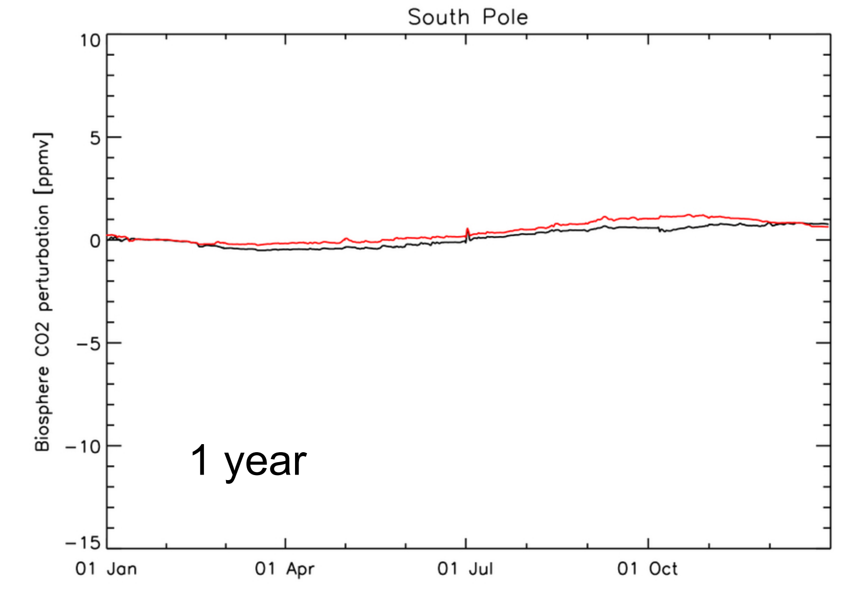


Using the same transport model with 2 different biosphere surface flux climatologies shows considerable (systematic) differences in CO_2 mixing ratios around 500 hPa.

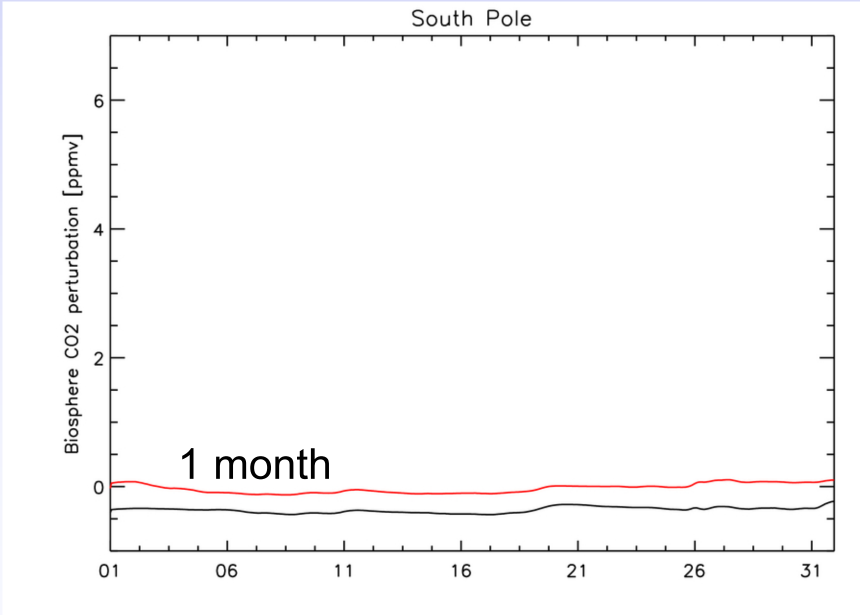


The magnitudes of these systematic differences are regionally dependent.

Model bias

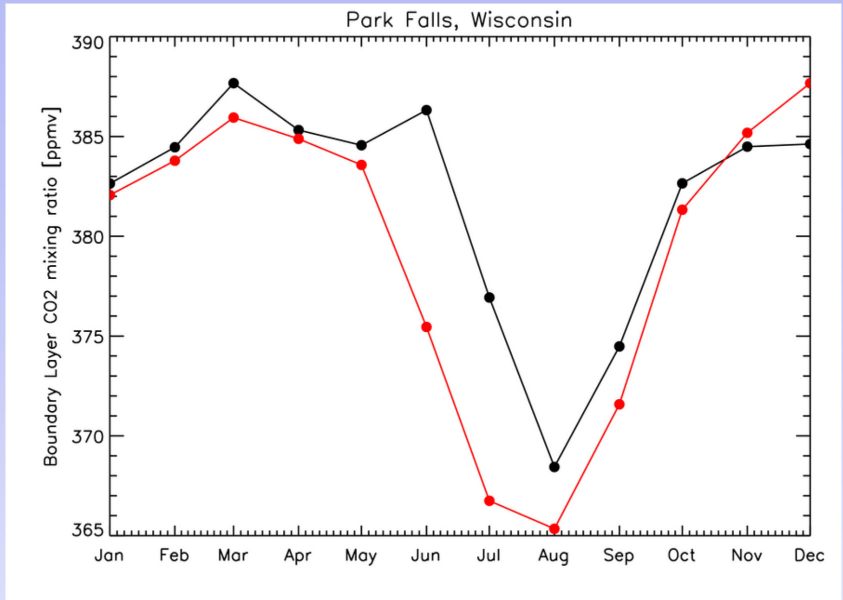


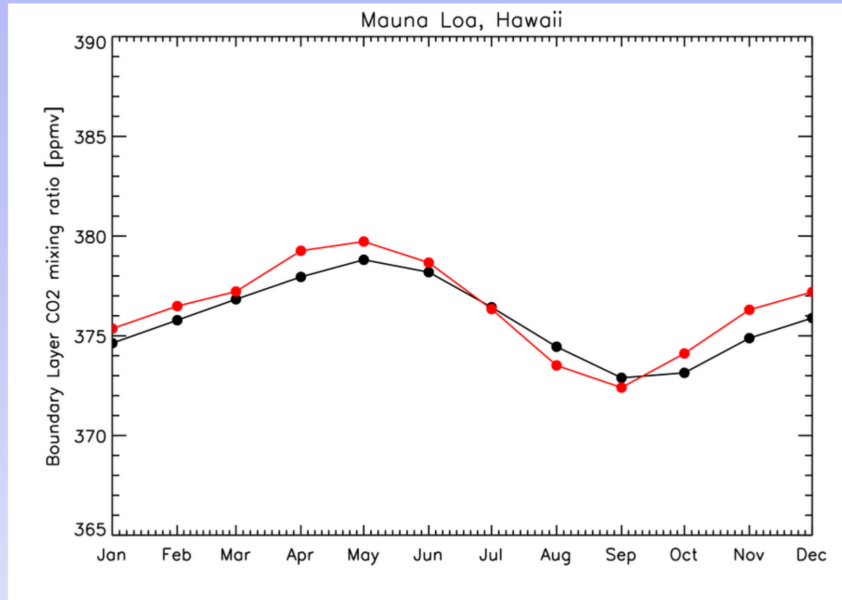
Using the same transport model with 2 different biosphere surface flux climatologies shows considerable (systematic) differences in CO_2 mixing ratios around 500 hPa.

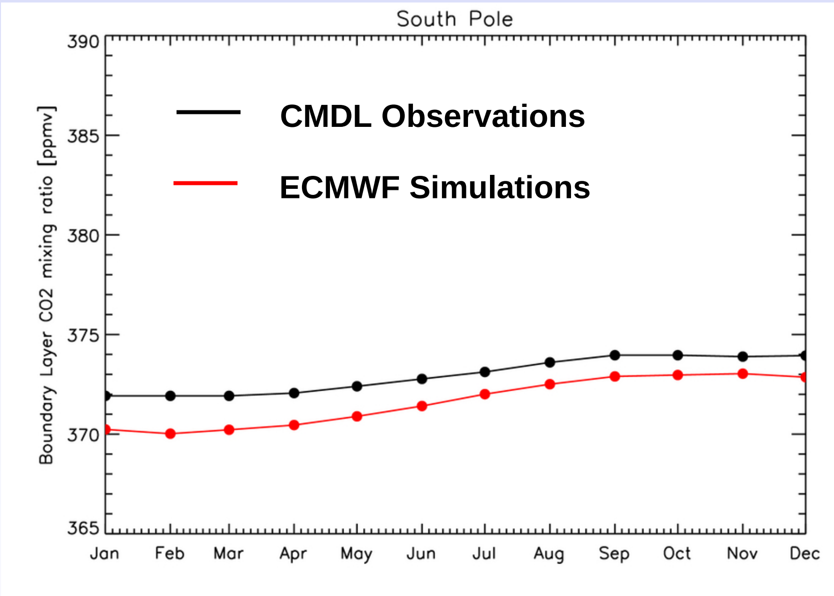


The magnitudes of these systematic differences are regionally dependent.

Model bias

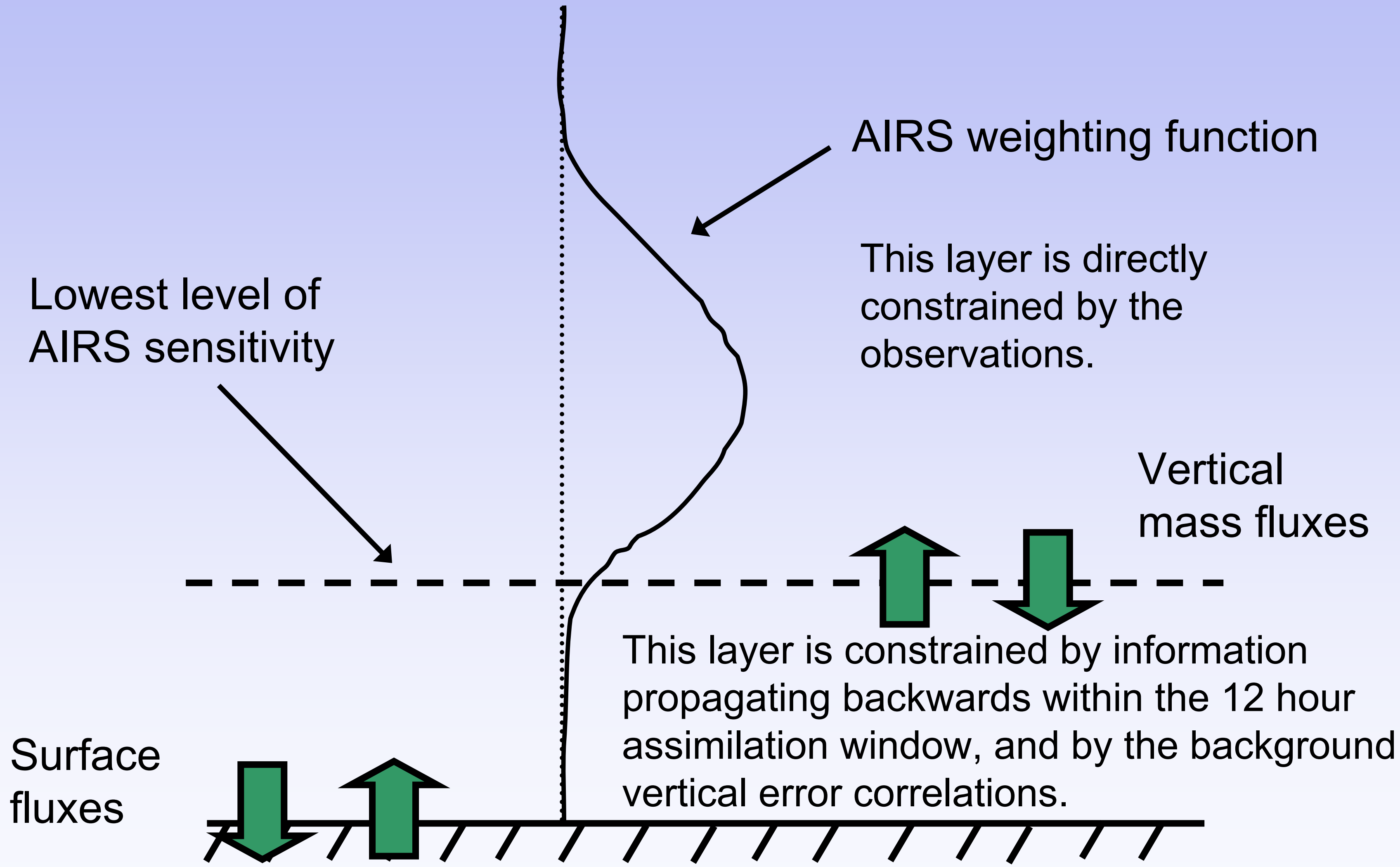




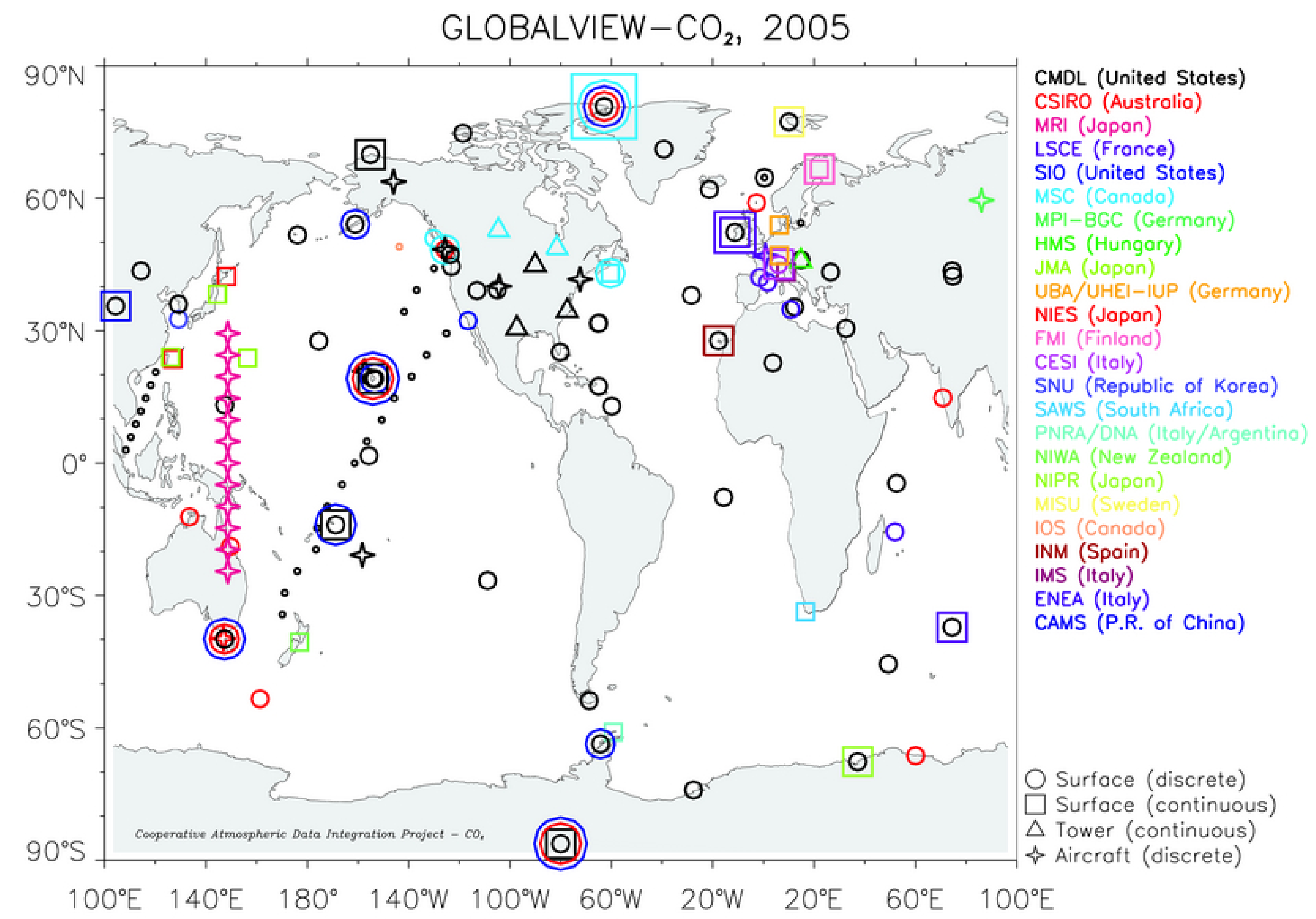


ECMWF model output after 1 year of spin-up shows already nice agreement with surface observations.

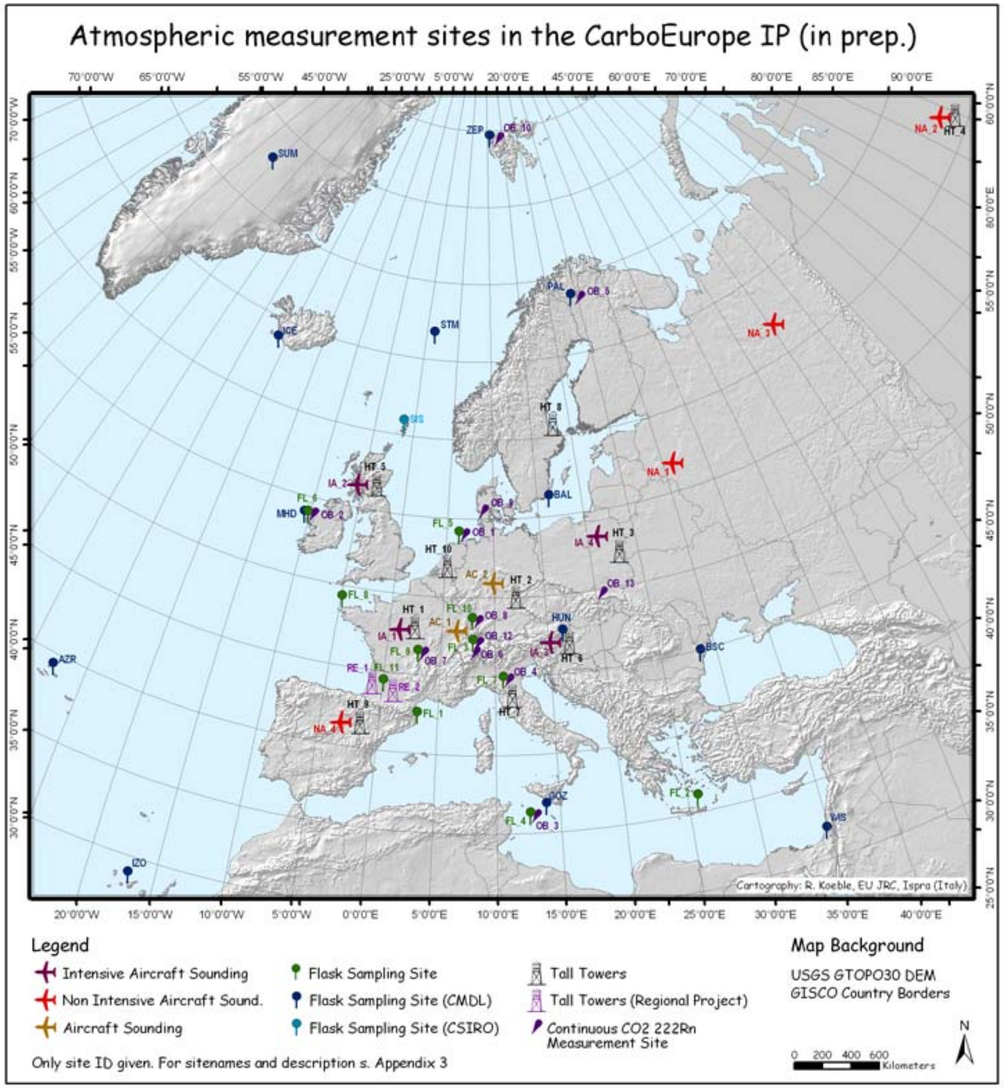
Unconstrained model errors



Greenhouse gas validation

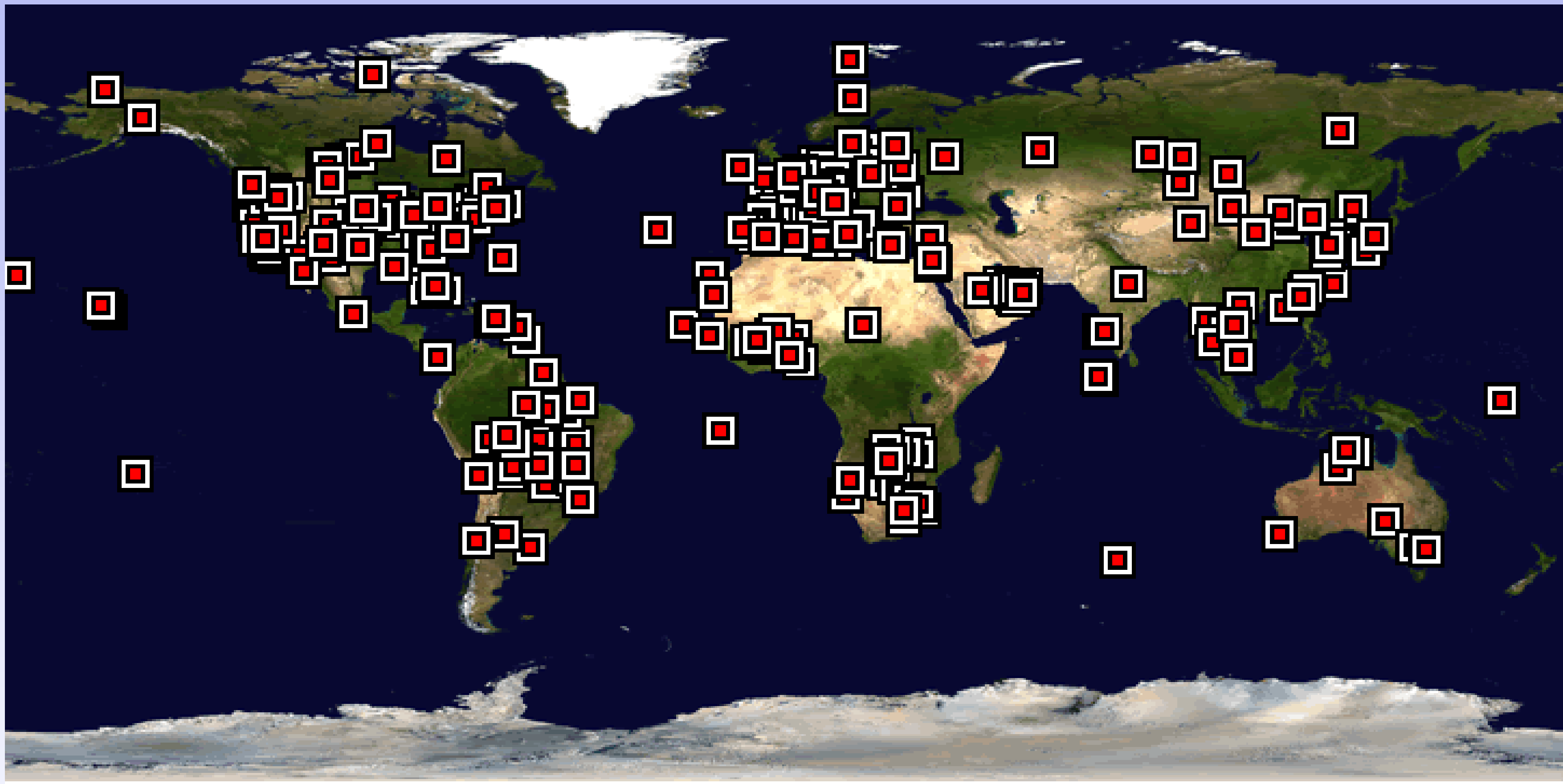


Greenhouse gas validation



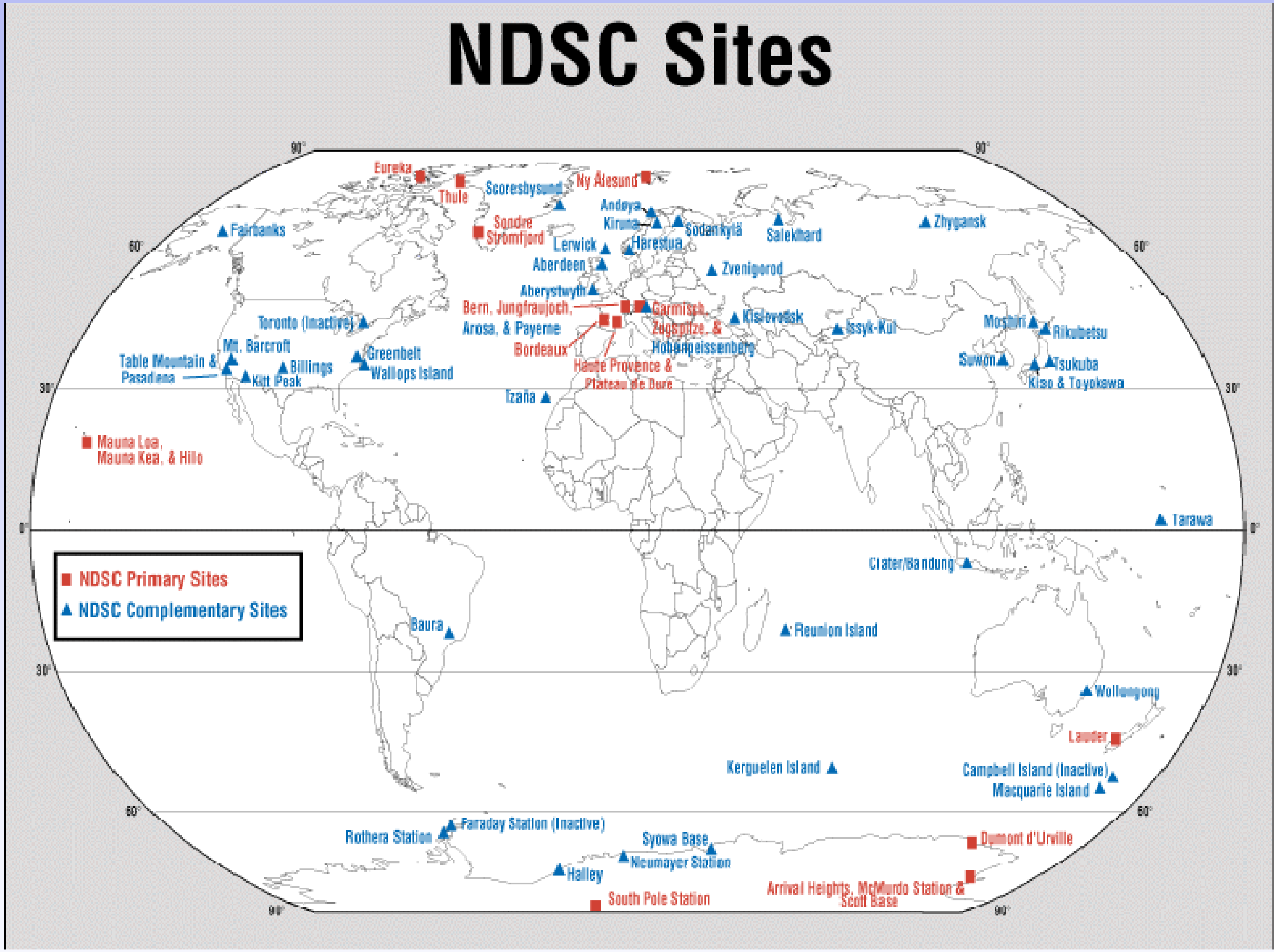
CarboEurope
(2004 – 2009)

Aerosol validation



Aeronet provides aerosol optical thickness observations from various ground-based stations around the world

Reactive gas validation



Example: bias correction for ozone

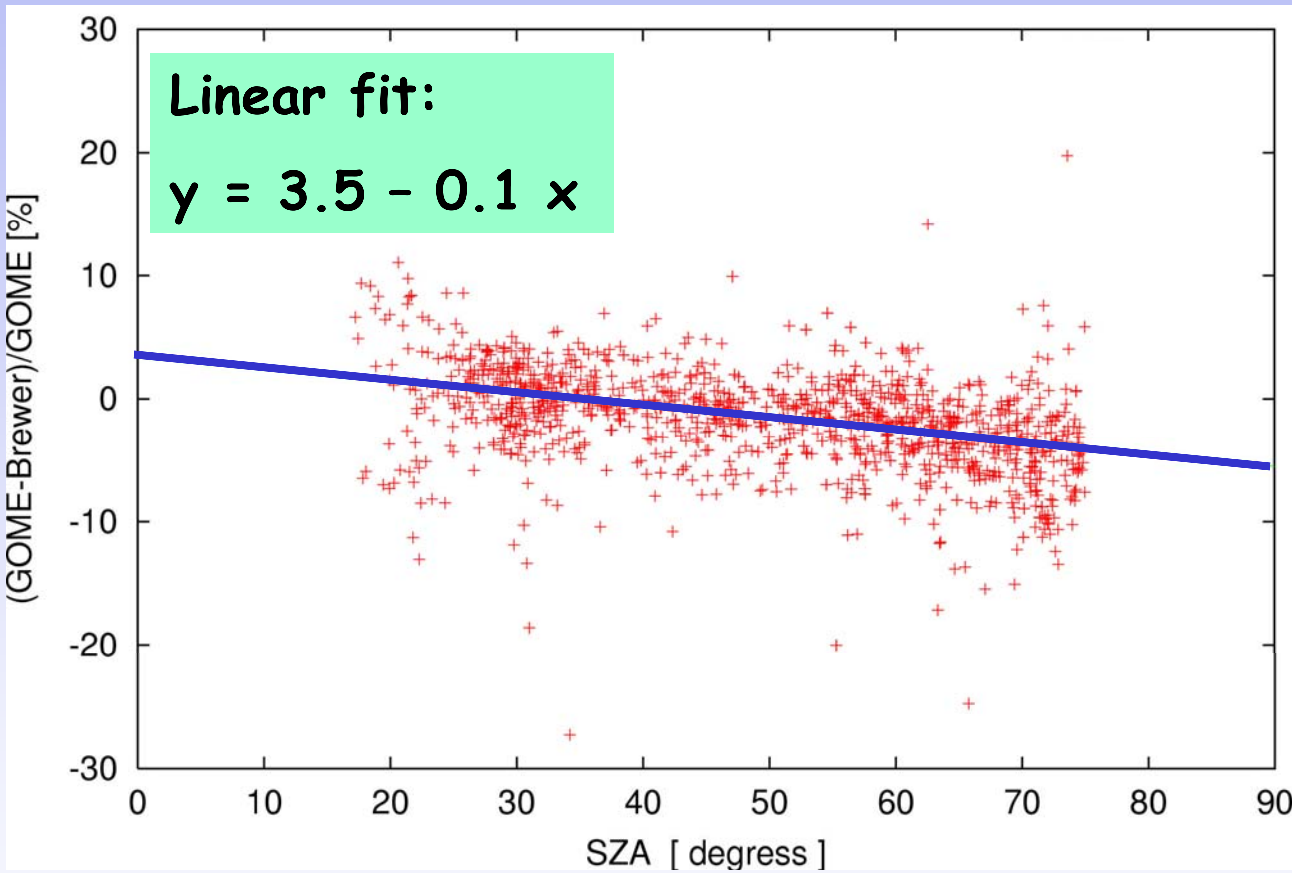
- Bias between model and observations violates underlying assumption of DA that obs and fg are unbiased

- Model **AND** GOME data can have bias

- Understand model bias
- Correct model bias

- Develop a bias correction for ozone data, based on **independent** observations
- Use **ground-based** Brewer and Dobson observations (obtained from WOUDC: http://www.msc-sms.ec.gc.ca/woudc)

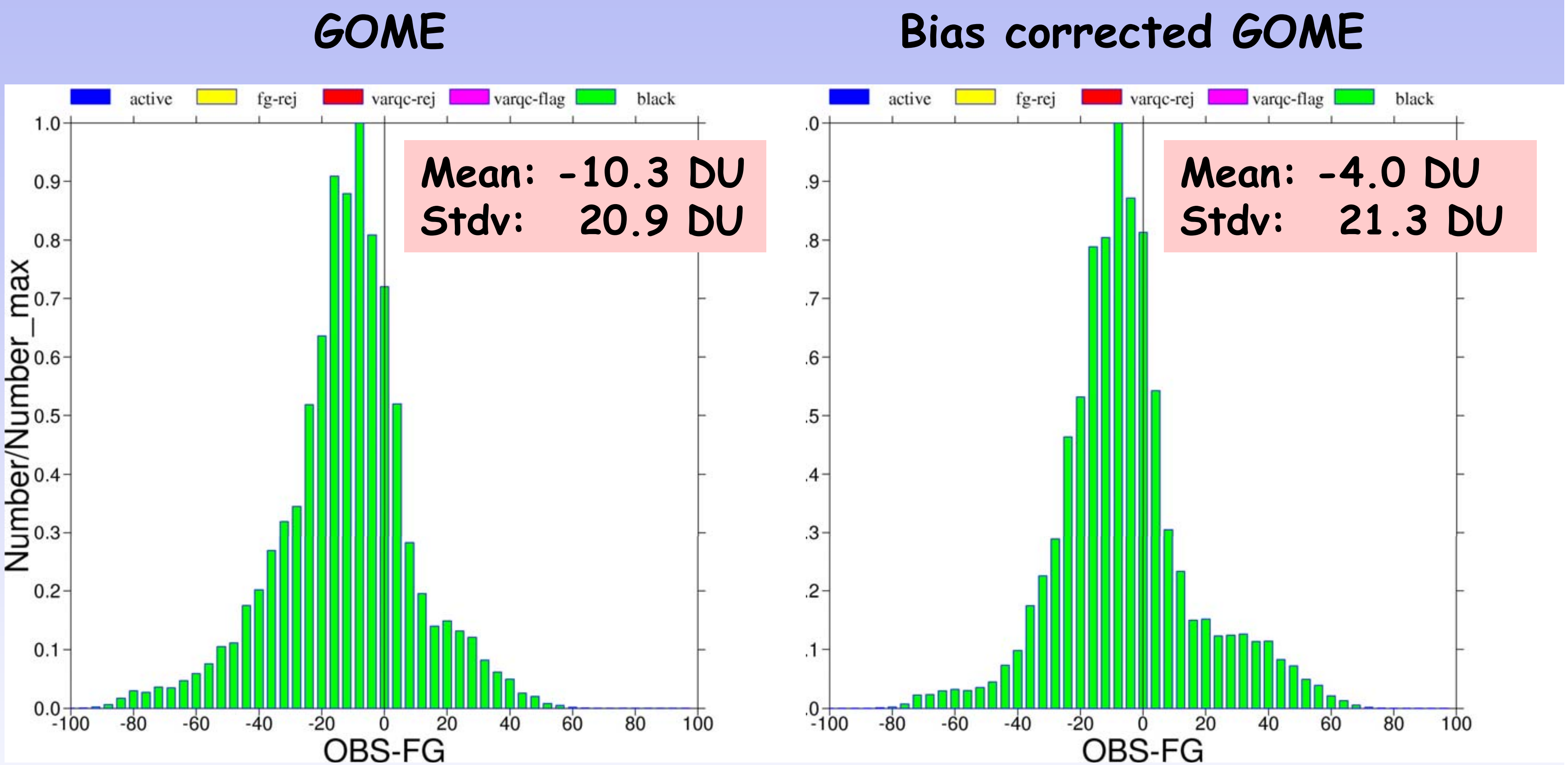
Example: bias correction for ozone



Use **independent** observations to develop bias correction

Relative difference between GOME and Brewer obs. (1999)

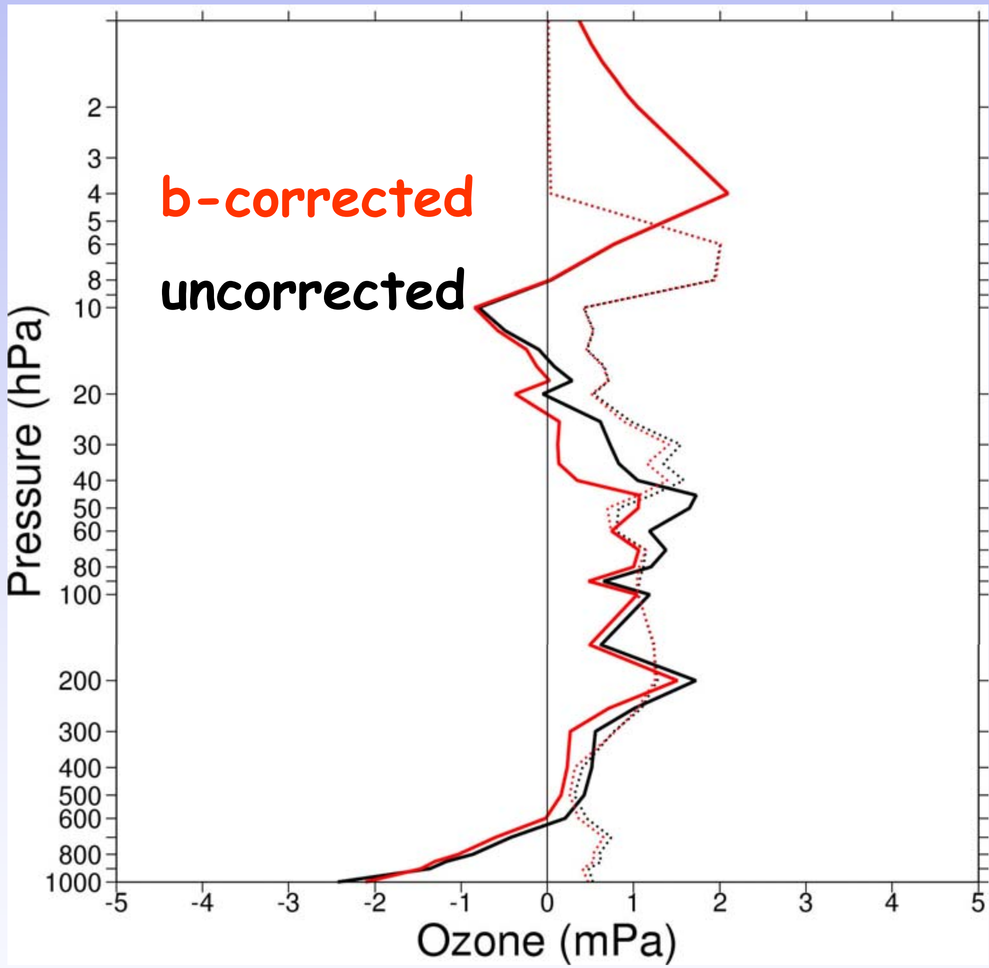
Example: bias correction for ozone



First-guess departures in DU (Period: 20021010 – 20021015)

Example: bias correction for ozone

Ny-Aalesund (78.9 N, 12 E)



October 2002

(9 sondes)

Analysis verified against ozone sondes

Summary

- Observation bias
  - ➢ Bias in retrieval products (e.g., aerosol optical depth) can be large. It is also variable in space, which makes it hard to correct.
  - ➢ Bias in radiance data (AIRS) is small, but signal is small as well. Left-over small biases can affect results.
- Model bias
  - ➢ Model bias can be quite large and is hard to quantify.
  - ➢ Because we are mainly interested in good analyses, it would be desirable to have the observations correct the model bias, even if this has to be done cycle after cycle.

Just a few Questions

- How do we obtain the least-biased analysis?
- How do we estimate correct observation bias corrections, considering the low amount of accurate validation data?
- How do we avoid removing the signal with the observation bias correction?
- How tight do we want to keep the background constraint?