Deutscher Wetterdienst
Wetter und Klima aus einer Hand

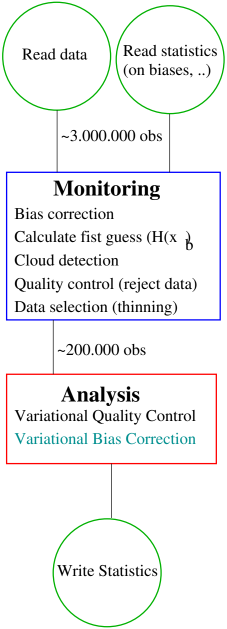
Quality Control and Bias Correction

Andreas Rhodin
FE 12

DWD-HERZ Winterschool on Data Assimilation

2012-02-16

Observation Data Flow in Global Analysis Step



Before entering the assimilation step the data is monitored, quality controlled , bias corrected and thinned.

Only a fraction of the available Data is used in the assimilation.

Statistics on the data is written:
- to be used in subsequent analyses steps (for bias correction).
- For verification and assessment of the data assimilation system (Satmon, TEMP Verification)

Course of this talk

1. General remarks on monitoring, quality control, bias correction,. . .

2. Go through different observation types of the global data assimilation system.

   Describe details of quality control, bias correction,. . .

Monitoring

Monitoring:

- Apply the observation operator to the model background, i.e. calculate $\mathcal{H}(\mathbf{x_b})$

Goal of monitoring:

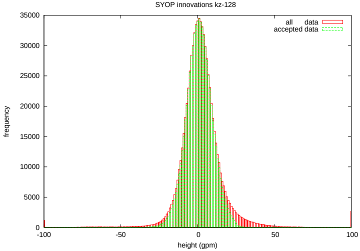
- Gather (innovation) statistics also on data which is currently not assimilated
  in order to assess:
  - errors, biases,
  - overall quality of the data
  - details of data processing.

Applies to:

- New implementations of observation operators
- new instruments to use
- new satellites

Quality Control - Why

- Many assimilation systems assume Gaussian distributions of the observational error.
- However, Observations have outliers. Innovation statistics often show non-Gaussian distributions with fat tails.



- If outliers are assimilated with unreasonable (too small) assumptions on the observation errors this leads to large analysis increments and a degradation of the forecast.

Quality Control

- Blacklisting
  A list of suspicious stations is maintained.
- White-listing
  A list of trusted stations is maintained (here for wind profilers).
- Consider quality information of the data provider
- Observation type specific checks
- Observation minus first guess check
- Variational Quality Control:
  Explicitly account for non-Gaussian observation error distributions in the assimilation step.

Quality Control – Observation minus First Guess Check

Disregard observations if the deviation to the first guess is larger that a multiple of the expected standard deviation:

$$y - \mathcal{H}(\mathbf{x}_b) > c \cdot \sqrt{\sigma_o^2 + \sigma_b^2}$$

$$c = 3 \dots 5$$

If c is chosen too small all observations within a region may be rejected in critical situations (for instance rapid development of a storm which is not captured by the forecast).

Quality Control – Variational Quality Control

Variational methods (3D-Var, 4D-Var) do not require Gaussian error distributions (maximum likelihood estimators).

Method: Assume non-Gaussian observation error distributions p , i.e. non-quadratic cost functionals J_o.

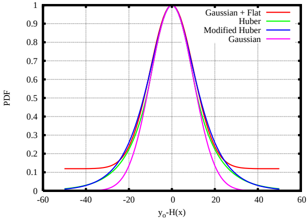
$$(y-\mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(y-\mathcal{H}(\mathbf{x})) \quad \rightarrow \quad J_o(y-\mathcal{H}(\mathbf{x}))$$

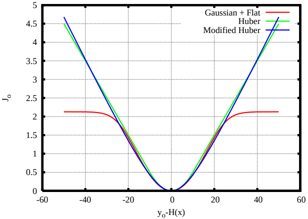
with

$$J_o = -\log\left(p(y-\mathcal{H}(\mathbf{x}))\right)$$

- Advantage compared to first guess check:
  multiple (consistent) observations are able to 'draw' the analysis from the first guess, even in case of large obs - fg.

Variational Quality Control – Choices for non-Gaussian J_o



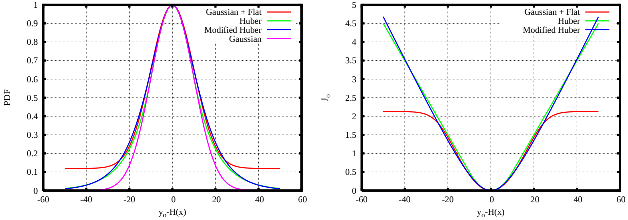


- Superposition of a Gaussian (magenta) and a flat distribution (red)

$$p = e^{\frac{-(y - H(x))^2}{(2\sigma_o)}} + c$$

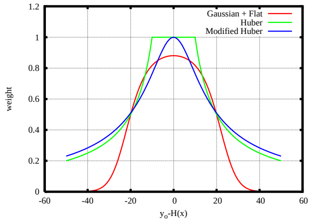
  + large outliers have zero impact on the forecast
  − strong nonlinearities,
    possibly multiple minima of the cost function,
    slow convergence of minimisation algorithm

Variational Quality Control – Choices for non-Gaussian J_o

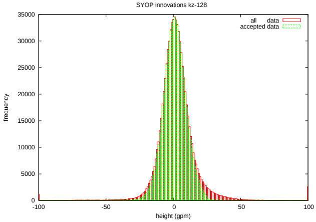


- Gaussian with exponential tails (green)
  'Huber norm':$\approx L_2$ in the centre,$\approx L_1$ in the tail.
  $$p \approx e^{-(y-H(x))^2} \text{ for } |y - H(x)| < c, \quad p \approx e^{-|y-H(x)|} \text{ else}$$
  + better convergence, no multiple minima
  – large outliers have only reduced impact on the forecast
- Our Choice: similar to 'Huber norm',
  but with continuous 2nd derivation. (blue)

Variational Quality Control – Weight of Observations



Weight of an observation: gradient of the cost-function divided by the gradient of the corresponding quadratic cost-function without VQC.



Histogram of innovations (observation - first guess). Observations with a VQC-weight < 0.5 are marked red.

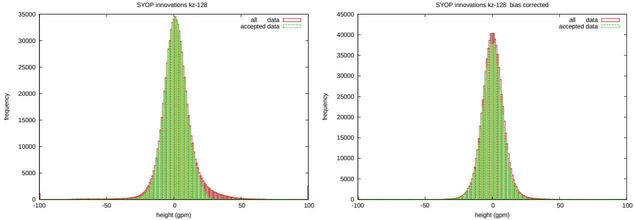
Bias Correction - Why

- Most assimilation systems assume unbiased observations: $E[\epsilon_o] = 0$.
- Various observations have biases (systematic errors):
  - Radio-soundings: temperature & humidity depending on time of the day and radiosonde type.
  - Aircraft observations: temperature bias depending on individual aircrafts.
  - SYNOP: surface pressure observations depending on individual stations (wrong station height reported).
  - Radiances: complex brightness temperature bias depending on:
    - channel,
    - sensor
    - satellite
    - scan angle
    - situation (forward operator error)
- We currently correct for:
  - Radiances (global DA)
  - Vaisala radiosondes (COSMO)
  - Aircrafts (global DA, in preparation)

Bias Correction

- Off-line (operational)
  - Gather statistics on Observation minus first guess and its dependence on the actual situation (predictors) during the assimilation.
  - Based on the above statistics derive the mean bias of the observations (with respect to the first guess) for a period ($\approx$ 1 month). Calculate coefficients for the optimal correction in dependence on predictors.
  - During subsequent assimilation steps correct the observations based on the above coefficients before their use in the analysis step.
- Online (experimental)
  - As above, but calculate the coefficients online based on a running mean of the innovation statistics.
- Variational Bias Correction (under implementation)
  - Directly include the bias correction coefficients as parameters to be optimised in the variational minimisation.

Bias Correction - Effect



Innovation histogram without (left) and with (right) bias correction. Data with a VQC-weight < 0.5 (without bias correction) are marked red.

- The above example is for SYNOP surface pressure. The bias correction is not used operational as it did not improve the forecasts (correction for model bias ?)
- Bias correction for radiances will be described later.

Thinning - Why

- Account for correlated observations.
  - For correlated observations too much weight is given to the data if the correlations are not accounted for.
    (To be specified in the off-diagonal entries of $\mathbf{R}$).
  - Spatial correlated data may be thinned until the remaining data is un-correlated.
- Reduces the computational burden in the assimilation step

Thinning vs. Specification of observation error correlations

- Correct specification of observation error correlations should lead to a more optimal analysis than thinning. But ...
- It is difficult to derive observation error correlations accurately. (cf. talk of Harald Anlauf)
- Effect of spacial error correlations on the analysis:
  - Correlated **background errors**: smoothing of the data.
  - Correlated **observation errors**: amplification of variations on small spatial scales.
- If the length scale of variations in the observations is much smaller than the correlation length scale of the background ($\approx$ 300 km in global data assimilation) the small scale patterns will not be assimilated anyway.
- Thinning is often more practical than observation error specification.

Thinning

- Associate every observation to the nearest grid-point (need not be the model resolution).
- If there is more than one report at a grid-point: Discard excessive observations following certain criteria.
  Default sequence:
  1. status of observation (active, first guess rejected,...)
  2. user defined preferences
  3. time distance to the analysis
  4. distance to the 'grid-point'
  5. amount of data in the report

  May be configured by the respective 'namelist'.
  (Like most quality control parameters)

Sequence of Monitoring and Quality Control

The sequence of monitoring and quality control tasks is important: (dependencies)

1. bias correction
2. application of observation operator
3. cloud detection
4. first guess check
5. gather statistics for bias correction
6. thinning

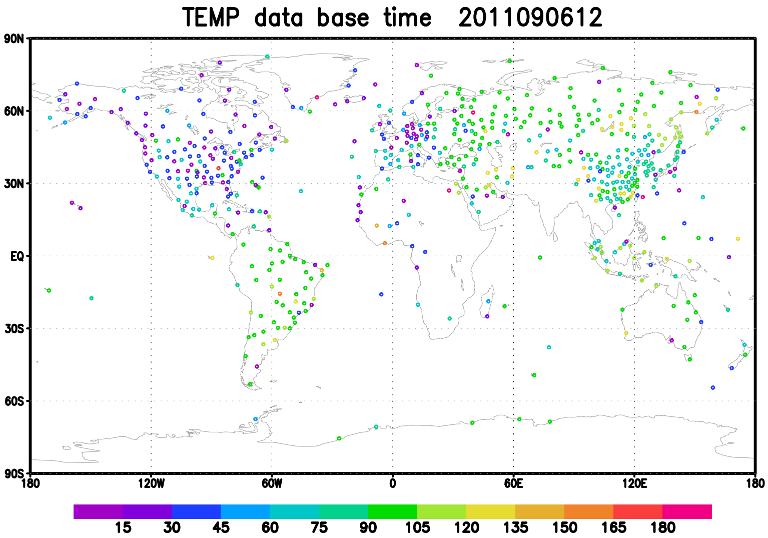
Usage of Monitoring & Data Assimilation Statistics Output

- **bias_. . .** -files
  Statistics on observational biases and bias correction coefficients
  - updated and cycled from assimilation step to assimilation step to provide information for bias correction.
- **mon. . . .nc, cof. . . .nc**-files
  Observed and forecasted values, results of quality checks, observational and background errors dots
  - SATMON, Monitoring of obs, fg in the assimilation cycle
  - (TEMP)-Verification: Verification of forecasts with quality controlled data.

Selected Observation Types used in the Global DA System

- In situ data (conventional data)
  - TEMP (Radiosondes)
  - Aircraft Observations
- Retrievals
  - AMV Atmospheric Motion Vectors (Winds)
- Remote Sensing Observations
  - GNSS Radio Occultations
  - Satellite Radiances

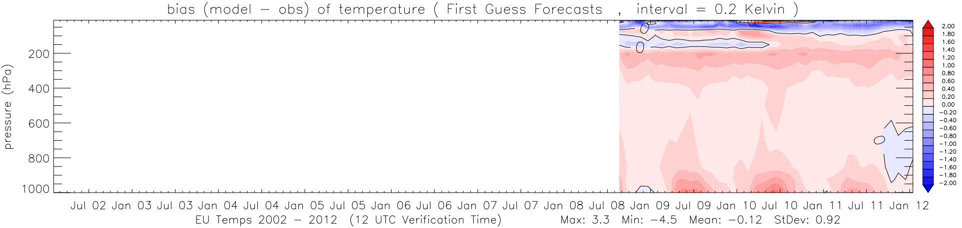
TEMP (Radio-soundings) – observation coverage

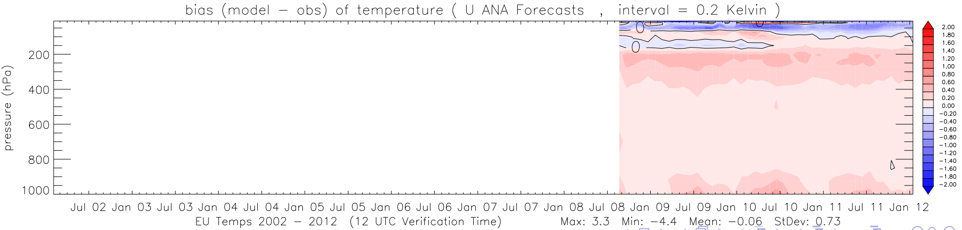


TEMP Characteristics

- Vertical soundings of temperature humidity, and wind.
- Still most reliable source of information in the free atmosphere.
- Used as anchor for observations with biases.
- Problem: uneven distribution (northern hemisphere, continents)
- Used for verification of forecasts and analyses.

TEMP Verification (Temperature Europe 12UT): Mean FG - Obs, Ana - Obs

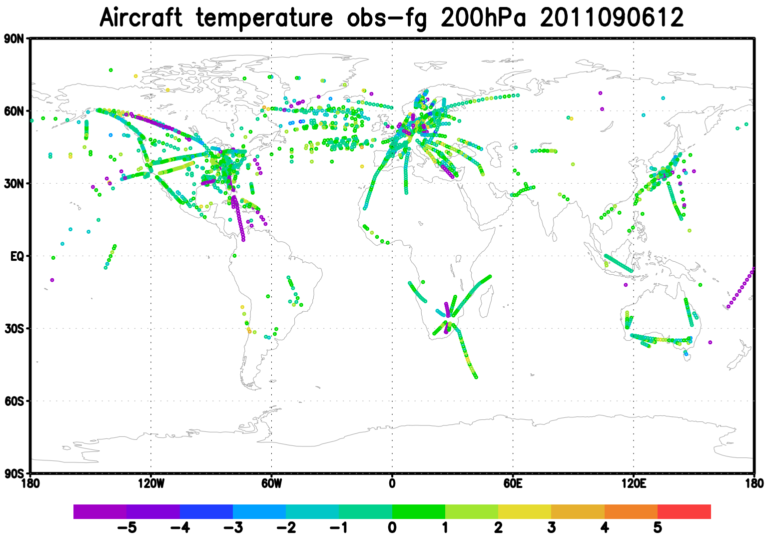




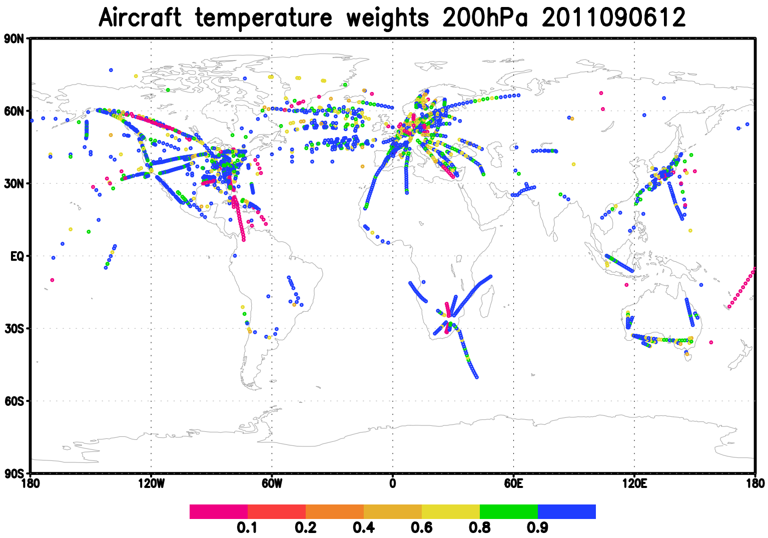
TEMP Verification (Europe 12UT)

- First Guess - Obs
  - Constant bias in the 200 hPa level
  - Bias in the boundary layer in summer, reduced in the last season.
- Analysis - Obs
  - The bias is not fully reduced
- Interpretation
  - The 200 hPa bias is due to biased aircraft observations. This is the dominant flight level. The bias is removed if aircraft temperature is not assimilated.
  - The seasonal bias in the boundary layer is reduced by the SMA (Soil Moisture Assimilation) which became operational last spring.

Aircraft – Innovations



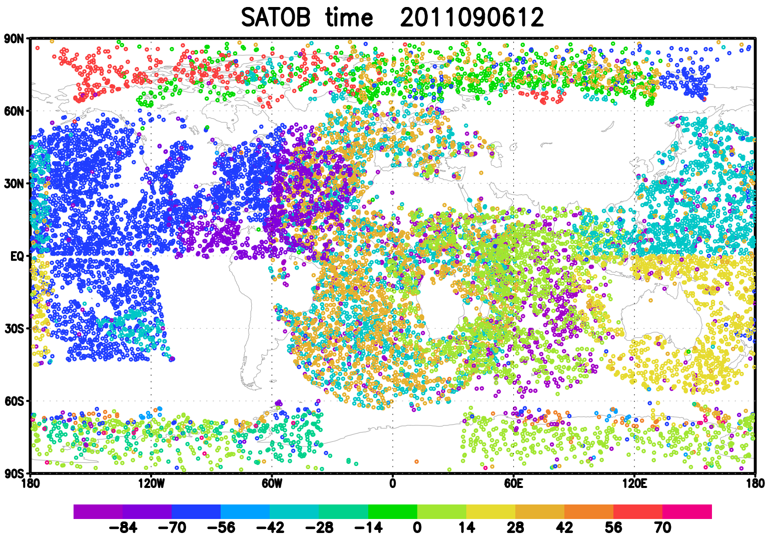
Aircraft – Variational Quality Control Weight



Aircraft – Bias Correction

- Aircraft temperature observations show considerable biases for individual aircrafts.
- The mean effect is a 0.3 K bias of the analysis.
- The bias is absent if aircraft temperature is not assimilated.
- The bias is reduced if aircraft temperature is bias corrected. (individually for each aircraft)
- Assimilation experiments are on the way.

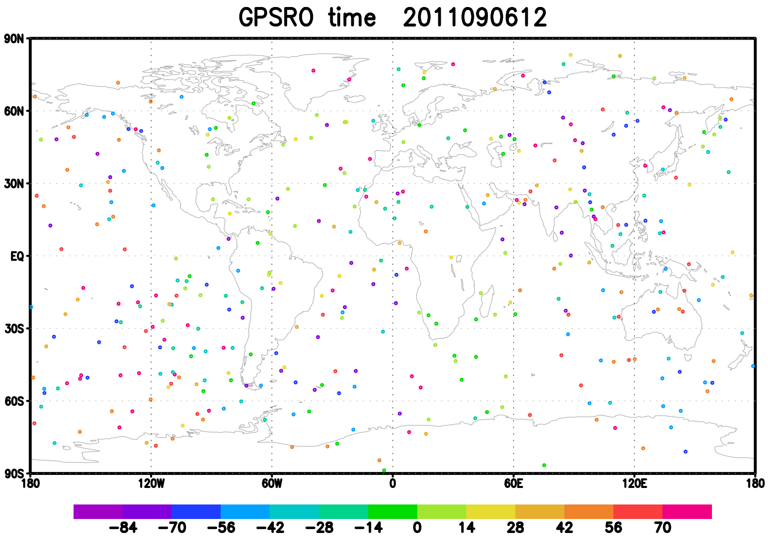
AMV Winds – Observation coverage



AMV Winds – Thinning

- Atmospheric Motion Vectors are subject to thinning.
- AMV observations are spatially correlated.
- A cause for spatial correlation is the height assignment error.
- cf. poster by Kathrin Folger & Martin Weissmann.

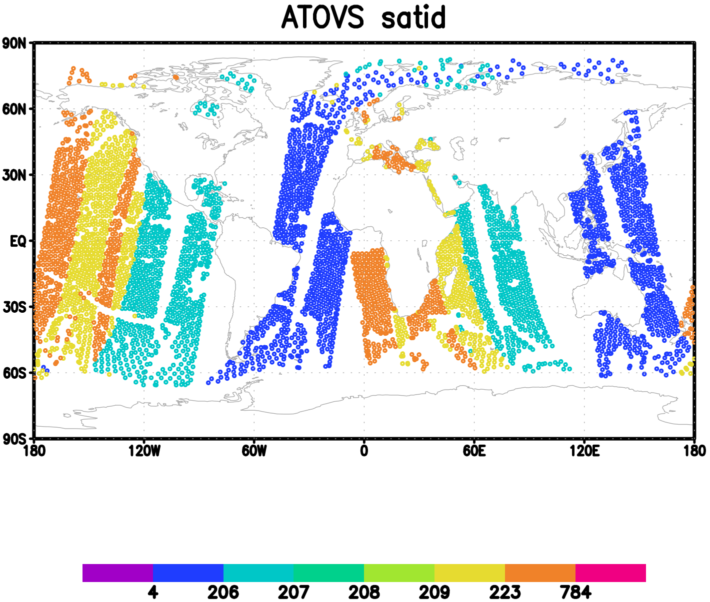
GNSS Radio Occultations – Observation Coverage



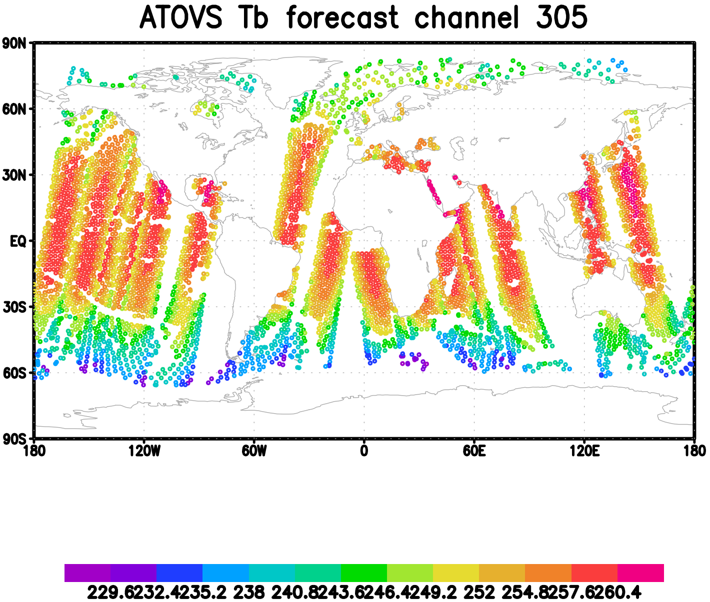
GNSS Radio Occultations

- cf. poster by Harald Anlauf for a detailed description of the method
- GNSS Radio Occultations can be regarded as a very accurate vertical temperature sounding in the upper troposphere and lower stratosphere
- Advantage (compared to TEMP): Uniform distribution over the globe
- Serve as anchor for observations with biases.

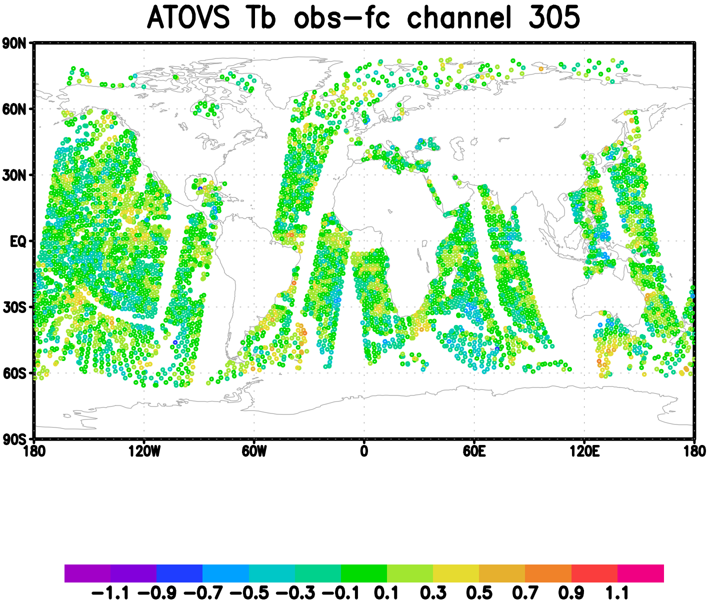
Radiances – Satellites



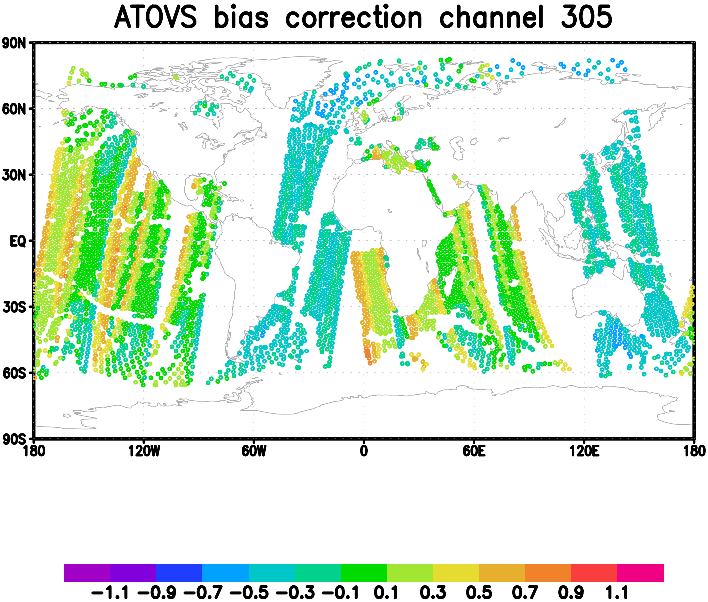
Radiances – AMSU-A chan 5 First Guess



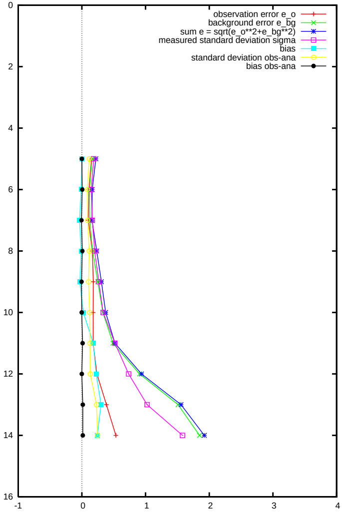
Radiances – AMSU-A chan 5 Obs - First Guess



Radiances – AMSU-A chan 5 Bias Correction



Radiances – AMSU-A Innovation Statistics



Channel 5 to 10 are temperature soundings. (from bottom to top of the atmosphere)

Channel 11 to 14 are effected by stratospheric temperature above the GME model top (5 hPa) and will not be discussed here.

Innovations of temperature sensitive channels in the troposphere are well beyond 0.5 K .

Bias correction has efficiently removed any bias (≈ 1K).

Radiances – Usage

- Bias depends on instrument, scan angle, situation
- Bias is large compared to (bias corrected obs - first guess)
- In our global DA system Radiances are used over land only
  - quality control, cloud detection works better over sea due to known emissivities of the sea surface
- Only few independent observations are available (over sea) to constrain radiance bias correction
  - The model first guess has a bias as well

Radiance Bias Correction – Difficulties

- Radiance (over sea) bias corrections are currently not well constrained by other observations.
- Radiance bias correction tends to correct for the model bias as well.
- As the observation systems over sea and land are mostly disjunct, analysis bias is different over continents and sea.
- Solution:
  - use radiances over land.
  - give more weight to independent observations when estimating bias correction coefficients.
    $->$ Variational Bias Correction

Radiance Bias correction – Predictors

- Scan angle or FOV number
- Observed brightness temperature itself (used in 1dvar)
- Thicknesses (similar to brightness temperatures)
- Other (used for IASI)
  - Surface skin temperature
  - Integrated water vapour

Predictors must be chosen with care. They should account for systematic instrument and observation operator biases, not for model biases.

Radiance Bias Correction – Formulation

The bias correction $\mathbf{b}$ (to be subtracted from observation $\mathbf{y}$ to yield the bias corrected observation $\mathbf{y}_c$) is assumed to depend linearly on the model state $\mathbf{x}$. For a given channel k it is calculated as the sum of products of scalar coefficients β_{ki} and model predictors $p_i(\mathbf{x})$:

$$\mathbf{y}_c = \mathbf{y} - \mathbf{b}(\beta, \mathbf{x}) \quad \text{with} \quad b_k = \sum_{i=0}^{N_p} \beta_{ki}\, p_i(\mathbf{x}) \tag{1}$$

The bias correction term can be regarded as the product of the matrix of model predictors $\mathbf{P}$ times the vector of coefficients β:

$$\mathbf{y}_c = \mathbf{y} - \mathbf{P}(\mathbf{x})\, \beta \tag{2}$$

$\mathbf{P}$ is of size 'number of coefficients' times 'number of observations'.

Radiance Bias Correction – Coefficient estimation

The bias correction coefficient vector β may be estimated by a minimum variance approach or, equivalently, by minimising the quadratic cost functional:

$$
\begin{aligned}
J &= \frac{1}{2}\,(\mathbf{y}_c - \mathcal{H}(\mathbf{x}))^{\mathcal{T}}(\mathbf{y}_\rfloor - \mathcal{H}(\mathbf{x})) \\
&= \frac{1}{2}\,((\mathbf{y} - \mathcal{H}(\mathbf{x})) - \mathbf{P}\,\beta)^{\mathcal{T}}((\mathbf{y} - \mathcal{H}(\mathbf{x})) - \mathbf{P}\,\beta) \\
&= \frac{1}{2}\,(\mathbf{d} - \mathbf{P}\,\beta)^T(\mathbf{d} - \mathbf{P}\,\beta) \qquad (3)
\end{aligned}
$$

Here $\mathbf{x}$ denotes the background state of the atmosphere and $\mathbf{d}$ are the deviations of the uncorrected observations $\mathbf{y}$ from the model background, calculated by the observation operator $\mathcal{H}$.
Setting the gradient of the cost function to zero yields a set of linear equations to be solved for β

$$\frac{dJ}{d\beta} = \mathbf{P}^T(\mathbf{d} - \mathbf{P}\,\beta) = 0 \qquad (4)$$

with the solution:

$$\beta = (\mathbf{P}^T\mathbf{P})^{-1}\,\mathbf{P}^T\mathbf{d} \qquad (5)$$

Radiance Bias Correction – Statistics Required

In order to solve

$$\beta = (\mathbf{P}^T\mathbf{P})^{-1}\,\mathbf{P}^T\mathbf{d}$$

We need:

$\mathbf{P}^T\mathbf{P}$	Matrix of size $N_p \times N_p$	Sum of predictor products
$\mathbf{P}^T\mathbf{d}$	Vector of size N_p	Sum of products predictor x deviation

- The sum is over all observations.
- N_p is the number of predictors used for a given channel.
- These statistics can be calculated in advance,
  or updated on the fly (online), when new data comes in.
- Old coefficients may be rescaled so that old data will have less impact.
  (Running mean with a given time constant)

Radiance Bias Correction – Continuous update of Statistics

Estimate new bias correction coefficients from the updated statistics.

$$\beta = (\mathbf{P}^{b\,T}\mathbf{P}^b + \mathbf{P}^T\mathbf{P})^{-1}(\mathbf{P}^{b\,T}\mathbf{d}^b + \mathbf{P}^T\mathbf{d}) \quad (6)$$

Instead of keeping and updating the statistics used to estimate the bias coefficients it is possible to use the old coefficients β^b as a background constraint.

$$\beta - \beta^b = (\mathbf{B}_\beta^{-1} + \mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T(\mathbf{y} - (\mathbf{h}(\mathbf{x}) + \mathbf{b}(\beta^b, \mathbf{x})) \quad (7)$$

Radiance Bias Correction – Variational

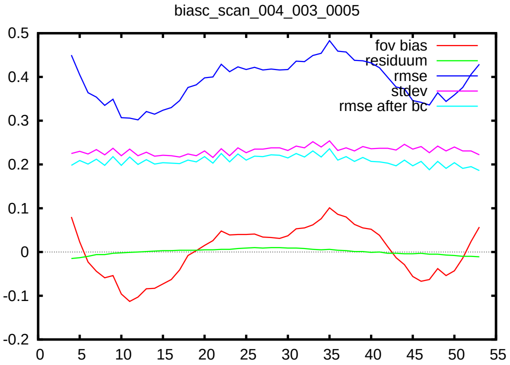
The bias correction term may be included in the cost functional of the variational assimilation scheme:

$$
\begin{aligned}
J &= \frac{1}{2}\,(\mathbf{x}^b - \mathbf{x})^T \mathbf{B}_x^{-1} (\mathbf{x}^b - \mathbf{x}) && (8) \\
&+ \frac{1}{2}\,(\beta^b - \beta)^T \mathbf{B}_\beta^{-1} (\beta^b - \beta) \\
&+ \frac{1}{2}\,(\mathbf{y} - \mathbf{h}(\mathbf{x}) - \mathbf{b}(\mathbf{x},\beta))^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{h}(\mathbf{x}) - \mathbf{b}(\mathbf{x},\beta))
\end{aligned}
$$

Minimising (9) with respect to β (not accounting for the uncertainty in $\mathbf{x}$) is equivalent to solving the following set of linear equations:

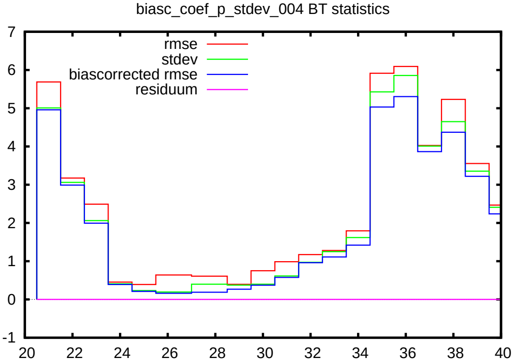
$$\beta - \beta^b = (\mathbf{B}_\beta^{-1} + \mathbf{P}^T \mathbf{R}^{-1} \mathbf{P})^{-1}\, \mathbf{P}^T \mathbf{R}^{-1} (\mathbf{y} - (\mathbf{h}(\mathbf{x}) + \mathbf{b}(\beta^b, \mathbf{x}))) \qquad (9)$$

Radiance Bias Correction – Reduction of RMSE (Scan Angle)



fov bias	field of view dependent bias correction (with global mean removed).
residuum	Residual bias after correction. The residuum is not exactly zero because the mean value of global predictors slightly depends on the field of view index. This behaviour may be changed by estimating scan dependent and global bias coefficients jointly together.
rmse	RMSE error in in the data set.
stdev	Standard deviation of obs-fg in data set after removal of the bias (without using global predictors).
rmse after bc	RMSE after application of bias correction.

Radiance Bias Correction – Reduction of RMSE (Channel)



rmse	RMSE of original data set.
stdev	Standard deviation (after bias removal for each FOV).
biascor.rmse	RMSE after bias correction.
residuum	Residuum after bias correction (here exactly zero).

Summary

- Observations have to undergo a monitoring and quality control procedure before assimilation.
- Quality control is essential.
- Bias correction is advisable for some conventional data.
- Bias correction is essential for satellite radiance data.
- Predictors must be chosen with care.
- Model bias is an issue as well.