Operational bias correction of satellite radiances at DWD

by Olaf Stiller[(1)] and Andreas Rhodin[(2)]
(1)German weather service (DWD), data assimilation section
(2)formerly DWD

March 30, 2021

1 Online bias correction

At DWD we are currently using an online bias correction scheme. This is an adaptive scheme which estimates the observation biases as part of the DA cycle adjusting the bias correction continuously. The scheme is implemented as part of the data assimilation code (DACE) and can be configured to estimate the time varying bias correction coefficients based on a linear regression using i) observation minus background or ii) observation minus analysis. A variational bias correction scheme (iii) is also available. Currently observation minus background statistics are used [option i)], but work is underway to replace this by using observation minus analysis statistics [option ii)].

A static bias correction scheme operating on statistics accumulated over a fixed time period is available as well. Before describing the adaptive scheme, we start with a short summary of the underlying static scheme.

Static bias correction

Static BC schemes minimize a cost function for the departures of bias corrected observation from the background or analysis equivalents.

$$J_{BC} = \left(\boldsymbol{y}^o - \boldsymbol{b}^o\left(\boldsymbol{x},\beta\right) - H\left(\boldsymbol{x}\right)\right)^T\left(\boldsymbol{y}^o - \boldsymbol{b}^o\left(\boldsymbol{x},\beta\right) - H\left(\boldsymbol{x}\right)\right) \tag{1}$$

where H is the observation operator which computes the model equivalents to the observation $\boldsymbol{y}^o$ from the model states $\boldsymbol{x}$. As usual the bias correction increments

$$\boldsymbol{b}^o\left(\boldsymbol{x},\beta\right) = \sum_i \beta_i\,\boldsymbol{p}_i\left(\boldsymbol{x}\right) \tag{2}$$

are defined as linear combination of model state dependent predictors $\boldsymbol{p}_i\left(\boldsymbol{x}\right)$. Note that $\boldsymbol{y}^o$, $\boldsymbol{b}^o$, $H\left(\boldsymbol{x}\right)$ as well as $\boldsymbol{p}_i$ are vectors in observation space. As bias correction coefficient estimates are based on statistics of periods of typically one month, the vectors actually extend over observations from this period. At DWD, different satellite channels are treated independently, i.e., they have their own independent set of predictors which correspond to a subset of all possible β_i and only those components of $\boldsymbol{p}_i$ that apply to the corresponding channel are non-zero, respectively.

Minimizing the cost function (1) for the bias correction coefficients β_i leads to the equation

$$\beta_i = \sum_j \left\{C_p^{-1}\right\}_{ij}\left\{V_p\right\}_j \tag{3}$$

with

$$\left\{C_p\right\}_{ij} = \boldsymbol{p}_i^T\boldsymbol{p}_j\left[\equiv \sum_\alpha \left\{\boldsymbol{p}_i\right\}_\alpha\left\{\boldsymbol{p}_j\right\}_\alpha\right] \tag{4a}$$

$$\left\{V_p\right\}_j = \boldsymbol{p}_j^T\left(\boldsymbol{y^o} - H\left(\boldsymbol{x}\right)\right). \tag{4b}$$

Here, C_p and V_p are a matrix and a vector in predictor space. Their elements are computed as scalar products in observation space (as illustrated in Eq.4a by including the sum over all observations α explicitly). These computations involve a large number of observations, the learning data from which the bias correction coefficients β_i are computed. Particular care has to be taken when inverting the matrix C_p and, as explained below, some regularization may be applied to ensure that C_p is positive definite.

The online scheme

In our online BC the terms $\{C_p\}_{ij}$ and $\{V_p\}_j$ obtained from the learning data are continuously updated in each assimilation cycle using the new observations by replacing Eqs. 4a and 4b by

$$\{C_p\}_{ij} = f_\tau \left\{C_p^b\right\}_{ij} + \boldsymbol{p}_i^T \boldsymbol{p}_j \tag{5a}$$

$$\{V_p\}_j = f_\tau \left\{V_p^b\right\}_j + \boldsymbol{p}_j^T \left(\boldsymbol{y}^o - H\left(\boldsymbol{x}\right)\right) \tag{5b}$$

where $\left\{C_p^b\right\}_{ij}$ and $\left\{V_p^b\right\}_j$ are the old statistics used at the previous cycle while with the second terms on the right hand sides, statistics related to the new observations $\boldsymbol{y}^o$ (with the corresponding predictors $\boldsymbol{p}_i$) are added.

The weight of the old statistics stemming from the previous cycles is reduced by a factor

$$f_\tau = exp\left(-\frac{\Delta t}{\tau}\right) \tag{6}$$

(where Δt is the time since the last statistics where taken and τ is the chosen decay time for the memory of the system). This means that the influence of the collected statistics decays with a characteristic time τ.

Regularization

As for many statistical schemes, robustness critically depends on the inversion (or invertibility) of matrices which are obtained from averaging over real world data and which therefore are vulnerable to statistical fluctuations that might prevent their invertibility. To this end, provision has been made to inflate the diagonal elements of the matrix C_p (Eq. 4a or 5a) by a factor $(1+\nu)$ where ν is a number between zero and one. This would penalize large changes of the bias correction coefficients. So far, however, no problems with the inversion of this matrix have been encountered with the operational system but forecast impact studies gave small detrimental results when error inflation is applied. Therefore $\nu = 0$ is used which means that there is currently no regularization applied in the DWD online bias correction scheme.

2 Technical setup at DWD

Available Predictors

For each satellite and instrument a different set of predictors may be chosen. The bias correction scheme can be configured to either i) ignore a predictor, ii) gather statistics but not use the predictor, or iii) gather statistics and use the predictor. This way statistics can be accumulated online for predictors, instruments or channels that will be used actively at a later time.

Currently the following predictors may be specified:

thickness	thickness between two pressure levels
iwv	integrated water vapor
bt	modelled brightness temperature (for specified channel and instrument)
obs_bt	observed brightness temperature (for specified channel and instrument)
T_SurfSkin	surface skin temperature
scan.ang	scan angle (as an alternative to the field of view predictors, see below)
v10	10 m wind speed
Y_lm	spherical harmonics (to specify an explicit spatial dependence)
product	product of any 2 predictors presented above

Currently only meridional wave numbers are implemented for the spherical harmonics predictor. By specifying the product of a predictor with the spherical harmonics predictor, a spatial dependence of the predictor is achieved (if not present so far).

Predictors currently used at DWD

At DWD the dependence of biases on the scan angle of the satellite instrument is taken into account by having a different constant term for each field of view. For an instrument for which N_{sc} scan positions are assimilated,

	τ
AMSU-A and ATMS	30 days
IASI, HIRS,	14 days
SEVIRI, ABI, AHI	14 days
MHS, SSMIS, GMI, SAPHIR	7 days

Table 1: Operationally employed values for the memory decay time τ

the first N_{sc} predictors therefore take the form $(1 \leq n \leq N_{sc})$

$$p_n = \begin{cases} 1 & \text{at scan position } n \\ 0 & \text{otherwise} \end{cases} .$$

(Note that this choice of predictors makes a globally constant term redundant).

From the list above, only four air mass related bias correction predictors are used for satellite sounding channels. These involve layer thicknesses between:

- 1 hPa and 10 hPa
- 5 hPa and 50 hPa
- 50 hPa and 200 hPa
- 300 hPa and 900 hPa

For updating the bias correction statistics, only observations are used which have passed all the quality checks (including cloud detection where applicable) that are required for the assimilated data. Only the thinning employed for selecting these data is less restrictive than for the operationally assimilated observations. In practice, depending on the observation type and weather situations, five to ten times more data than actively assimilated are typically used for updating the bias correction statistics at each assimilation step. The current operational setting for the memory decay time τ is given in table 1.

Apart from satellite radiances, scatterometer data are also bias corrected but with a single predictor (apart from a constant term) which is the 10m wind speed. Furthermore, a simple form of bias correction is also applied to aircraft temperature measurements which for each aircraft and each phase of flight (ascending, descending or cruise level, respectively) involves an independent linear regression with only a constant (i.e., model state independent) predictor.

3 Comparison between online BC and VarBC

The biggest difference between our current use of the online BC and the variational bias correction (Var BC) used at some other centers is the fact that we are currently correcting the biases with respect to the background and not to the analysis state. Otherwise, when using the online BC with respect to the analysis, we expect our scheme to bear strong similarity with an appropriate Var BC scheme. In this context the work by Eyre (2016) is instructive who investigated the impact of a continuous relaxation towards the model bias by the short term forecast. He considered the case where the system has reached its asymptotic quasi-stationary state in which the bias correction is fully converged to an asymptotic constant value. He showed that this asymptotic value is a linear combination of the model bias and the bias of the anchoring observations. Following his arguments, it is clear that the asymptotic stationary bias is largely identical for the online scheme and Var BC as long as the bias correction is performed with respect to the analysis (which is always the case for the Var BC). The corresponding stationary bias is actually also quite similar for the online BC versus background although slightly less optimal (i.e., more influenced by the model bias) than for the other two schemes.

The main question therefore is how the different schemes react when away from their asymptotic values which comprises

- the behavior in the vicinity of this quasi stationary state (this includes the response to slow variations of the model bias, e.g., with changes of season and/or weather regimes).
- the schemes' ability to compensate for sudden changes of observational or model bias (compare cases discussed in Auligné *et al.* (2007)).

Below it is shown that the equations for computing the analysis increments (for model state and bias correction coefficients) are basically the same for the online scheme versus analysis and an appropriate Var BC scheme. The main difference is that the Var BC scheme is implicit (both equations for model state and bias coefficient increments are solved simultaneously in a single step) while in the online scheme the increments are computed sequentially. This is not expected to make a big difference for the first point above where changes are generally slow enough, but for the fast response listed under the second point the more implicit Var BC scheme has a better scope. To this end it has to be pointed out that the rapidity of the schemes' response to changes in biases is in principle governed by tunable parameters in these schemes (the decay time τ for the online scheme and the background uncertainty $\mathbf{B}_\beta$ for the estimated coefficients in Var BC). The more implicit nature of the variational scheme, however, allows a more aggressive tuning in this respect. The parameters τ and $\mathbf{B}_\beta$ can be configured so that the schemes are fully equivalent, with the only exception that the VarBC scheme already uses the deviations of the current analysis step to estimate the bias correction parameters whereas the adaptive scheme only utilizes the statistics of the past. Consequently these two configurations should behave differently only if i) the system is actually configured to be able to capture sudden changes in the bias (τ not much larger than the time Δt between two assimilation steps and, accordingly, $\mathbf{B}_\beta$ sufficiently large) and ii) sudden changes actually occur. However, small reaction times will be specified with care as this implies a smaller statistical base and may involve regularization issues.

Below we show that the equations for the online scheme (versus analysis) and an appropriate variational scheme are indeed largely identical. To show this, we first reformulate the equations for the online BC (introduced above) and then briefly summarize Var BC focusing on the Var BC method which matches our online BC most closely.

Var BC schemes estimate the bias correction coefficients β using the background term β^b from the previous cycle, its background error covariance $\mathbf{B}_\beta$ and data from the current assimilation cycle. For a straightforward comparison we rewrite the equation for β^b from the online bias correction scheme in terms of the contributions from the different cycles. For this we insert equations (5a) and (5b) into (3):

$$\beta_i = \sum_j \left\{ f_\tau \left\{ C_p^b \right\}_{ij} + \boldsymbol{p}_i^T \boldsymbol{p}_j \right\}^{-1} \left\{ f_\tau \left\{ V_p^b \right\}_j + \boldsymbol{p}_j^T \left(\boldsymbol{y}^o - H\left(\boldsymbol{x}\right)\right) \right\} \tag{7}$$

$$= \sum_j \left\{ f_\tau \left\{ C_p^b \right\}_{ij} + \boldsymbol{p}_i^T \boldsymbol{p}_j \right\}^{-1} \left\{ f_\tau \left\{ \beta^b\, C_p^b \right\}_j + \boldsymbol{p}_j^T \left(\boldsymbol{y}^o - H\left(\boldsymbol{x}\right)\right) \right\} \tag{8}$$

where in the last line we used $V_p^b = \beta^b\, C_p^b$ which follows directly from Eq.3.

Each term in this equation corresponds to a sum over a large number of observations, respectively. For comparing the online method with Var BC below, it is useful to introduce normalized quantities

$$\begin{aligned} \hat{C}_p &= m_n^{-1} C_p \\ \hat{C}_p^b &= \left(m_n^b\right)^{-1} C_p \end{aligned}$$

where the scaling factor m_n is the effective number of observations used for computing C_p which is updated with each assimilation cycle as

$$m_n = f_\tau m_n^b + m$$

where m is the number of assimilated observations (i.e., the dimension of $\boldsymbol{y}^o$). Assuming for simplicity constant numbers m at each cycle, m_n has the saturation value $m_n^{asymp} = m/\left(1 - f_\tau\right)$ which m_n approaches exponentially with the characteristic time τ.

Comparison with Var BC

Var BC adds an additional term to the usual cost function of variational data analysis by writing

$$\begin{aligned} J\left(\boldsymbol{x}, \boldsymbol{\beta}\right) = {} & \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{x}^b\right)^T \mathbf{B}^{-1}\left(\boldsymbol{x} - \boldsymbol{x}^b\right) \\ & + \frac{1}{2}\left(\boldsymbol{y}^o - \boldsymbol{b}^o\left(\boldsymbol{x}, \boldsymbol{\beta}\right) - H\left(\boldsymbol{x}\right)\right)^T \mathbf{R}^{-1}\left(\boldsymbol{y}^o - \boldsymbol{b}^o\left(\boldsymbol{x}, \boldsymbol{\beta}\right) - H\left(\boldsymbol{x}\right)\right) \\ & + \frac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^b\right)^T \mathbf{B}_\beta^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^b\right) \end{aligned}$$

where the bias correction vector $\boldsymbol{b^o}(\boldsymbol{x},\boldsymbol{\beta})$ was introduced in Eq.2. The covariance matrix $\mathbf{B}_\beta$ is generally block diagonal with respect to the subspaces related to the individual satellite channels so that the resulting equations for corresponding coefficients β_i (i.e., the components of $\boldsymbol{\beta}$ related to the respective satellite channels) largely decouple between these subspaces. Below for notational simplicity, we are restricting the discussion to the assimilation of a single satellite channel whose observation error is given by R.

Differentiation of J with respect to the model state $\boldsymbol{x}$, leads to the following equation for the value at the cost function minimum $\boldsymbol{x}^a$

$$\boldsymbol{x}^a - \boldsymbol{x}^b \quad = \quad \widetilde{\boldsymbol{K}}\left(\boldsymbol{y}^o - \boldsymbol{b}^o\left(\boldsymbol{x}^b,\boldsymbol{\beta}^a\right) - H\left(\boldsymbol{x}^b\right)\right) \tag{9}$$

where $\widetilde{\boldsymbol{K}}$ can be obtained from the usual Kalman gain matrix $\boldsymbol{K}$ by replacing the linearized observation operator (the linearized version of $H(\boldsymbol{x})$) by the corresponding linearization of $\widetilde{H}(\boldsymbol{x}) = H(\boldsymbol{x}) + \boldsymbol{b}^o(\boldsymbol{x},\boldsymbol{\beta})$. Including the gradient of $\boldsymbol{b}^o(\boldsymbol{x},\boldsymbol{\beta})$ with respect to $\boldsymbol{x}$ in this equation, physically amounts to computing the predictors at the analysis point (i.e., using $p_{k,i}(\boldsymbol{x}^a)$) and not with the background state as it is done for the online scheme (and also for many Var BC versions where this dependence is neglected[1]). While this may lead to noticeable impacts under some circumstances (particularly in the extra tropics if a cold or warm front is shifted by the analysis increments), in most regions one has $p_{k,i}(\boldsymbol{x}^a) \approx p_{k,i}\left(\boldsymbol{x}^b\right)$ and therefore $\widetilde{\boldsymbol{K}} \approx \boldsymbol{K}$. We therefore expect that the difference of the Kalman gain matrices has little influence on the global bias of the analysis increments.

Before discussing the corresponding equation for the analysis coefficients $\boldsymbol{\beta}^a$, we would like to note that, following the analysis of Eyre (2016), the bias of the asymptotic quasi-stationary equation towards which the coupled system tends does not depend on the particular form of the equation for $\boldsymbol{\beta}^a$ (as long as the system has a stationary attractor where $\boldsymbol{\beta}^a = \boldsymbol{\beta}^b = const$). In this limit one has $\boldsymbol{\beta}^a = const$ so that the equation for $\boldsymbol{\beta}^a$ and that for the analysis increments decouple and the remaining bias (i.e., the actual value of $\boldsymbol{\beta}^a$ in this limit) is fully constrained by the competition of the analysis weights on one side and the rate at which the forecast model pulls the analysis state towards its bias on the other (compare Eyre (2016)) . Similar constraints (though quantitatively different) apply also when correcting the bias with respect to the background state.

The equations for the analysis coefficients $\boldsymbol{\beta}^a$ is obtained from differentiating J by these coefficients and can be written as

$$\boldsymbol{\beta}^a \quad = \quad \left[\mathbf{B}_\beta^{-1} + \frac{\boldsymbol{p}^T\boldsymbol{p}}{R}\right]^{-1}\left(\mathbf{B}_\beta^{-1}\boldsymbol{\beta}^b + \frac{\boldsymbol{p}^T\left(\boldsymbol{y}^o - H\left(\boldsymbol{x}^a\right)\right)}{R}\right) \tag{10}$$

while the corresponding analysis error covariance matrix $\mathbf{A}_\beta$ for $\boldsymbol{\beta}^a$ can be written as

$$\mathbf{A}_\beta \quad = \quad \left[\mathbf{B}_\beta^{-1} + \frac{\boldsymbol{p}^T\boldsymbol{p}}{R}\right]^{-1}$$

(where we used the well known relation $\mathbf{A} = \left[\mathbf{B}^{-1} + \boldsymbol{H}^T\mathbf{R}^{-1}\boldsymbol{H}\right]^{-1}$ but replacing $\mathbf{B} \to \mathbf{B}_\beta$ and $\boldsymbol{H} \to \boldsymbol{p}$ as adequate for the $\boldsymbol{\beta}$-dependency of the cost function $J(\boldsymbol{x},\boldsymbol{\beta})$ above).

Choosing now

$$\mathbf{B}_\beta^{-1} = \frac{f_\tau C_p^b}{R} \tag{11}$$

one finds that the formulations to derive the bias correction coefficients from the Var BC (10) and from the online bias correction (8) are exactly the same when we set $x = x^a$ (i.e. perform online bias correction versus analysis). Further, the corresponding updating of the coefficients C_p^b in Eq.5a corresponds to a cycling of the B-matrix

$$\mathbf{B}_\beta(t + \Delta t) \quad = \quad f_\tau^{-1}\mathbf{A}_\beta(t)$$

where the new $\mathbf{B}_\beta$ is obtained by inflating the analysis covariance matrix from the last cycle with a factor f_τ^{-1}.

It is not uncommon to choose $\mathbf{B}_\beta$ to be diagonal, in which case the last term of the cost function $J(\boldsymbol{x},\boldsymbol{\beta})$ penalizes changes of the individual coefficients β_i. In contrast to this, using Eq.11 leads to a strongly nondiagonal matrix $\mathbf{B}_\beta$ which penalizes the resulting changes to the actual bias reduction. To see this we note that $\hat{C}_p^b =$

[1]E.g., in the DWD Var BC implementation, predictors are currently not based on x, but evaluated once for the background value x^b only.

C_p^b/m_n^b can be considered as an estimator for $\left(\boldsymbol{p}^T\boldsymbol{p}\right)/m$ so that the last term of the cost function $J\left(\boldsymbol{x},\boldsymbol{\beta}\right)$ can be written as

$$\begin{aligned}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^b\right)^T \mathbf{B}_\beta^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^b\right) &\simeq \frac{\alpha}{R}\sum_{i,j}\left(\beta_i-\beta_i^b\right)\left(\beta_j-\beta_j^b\right)\left[\boldsymbol{p}^T\boldsymbol{p}\right]_{ij} \\ &= \frac{\alpha}{R}\left(\boldsymbol{b}^o\left(\boldsymbol{x},\boldsymbol{\beta}\right)-\boldsymbol{b}^o\left(\boldsymbol{x},\boldsymbol{\beta}^b\right)\right)^T\left(\boldsymbol{b}^o\left(\boldsymbol{x},\boldsymbol{\beta}\right)-\boldsymbol{b}^o\left(\boldsymbol{x},\boldsymbol{\beta}^b\right)\right)\end{aligned}$$

where

$$\alpha = f_\tau\frac{m_n^b}{m}.$$

In practice we do not need to compute the constant α since C_p^b is directly computed off line. However, to get some impression of the relative magnitude of the $\mathbf{B}_\beta$ term in the Var BC cost function we give the expression for α obtained for a saturated system (which has been collecting data for a time t sufficiently larger than τ):

$$\alpha = \frac{f_\tau}{(1-f_\tau)} = \left(\exp\left(\frac{\Delta t}{\tau}\right)-1\right)^{-1}$$

which for large (or moderately large) $\tau/\Delta t$ can be approximated by

$$\alpha \approx \frac{\tau}{\Delta t}-\frac{1}{2}$$

(already for $\tau/\Delta t = 2$ the error of this approximation is less than 3 percent) yielding an approximately linear relation ship if the magnitude of τ is larger than Δt.

Some remarks may be appropriate:

- The correspondence of the variational and online schemes is not surprising since both methods are based on the minimization of quite similar cost functions.
- In practice, the choice of B_β in (11) may occasionally lead to large variations in the time series of background bias correction coefficients β^b while the bias correction $b(\beta^b)$ itself remains stable. This means that the bias correction gives meaningful results even if bias correction coefficients are highly correlated and certain linear combinations are not well constrained.
- The correspondence of equations (10) and (8) shows that online bias correction using analyses can be configured to behave almost exactly in the same way as the variational scheme. The only difference is that variational bias correction is able to use data of the current cycle to estimate bias correction coefficients, while the online bias correction is limited to statistics from the past. This makes the variational scheme more suitable for running with a shorter memory decay time and thus adapting faster to changes in observational bias (through using a larger B_β corresponding to a small τ).
- It should, however, be noted that also for the variational scheme, very short memory times may be problematic since for small α (or more precisely, small eigenvalues of $\boldsymbol{B}_\beta$) the inversion of the matrix in the first bracket on the right hand side of Eq.9 may require regularization measures which effectively increase the memory time.
- The correspondence between the online and variational scheme offers the opportunity of evaluting input quantities to the variational scheme on a broader statistical basis when using the computational setup of the online scheme. More precisely, including in Eqs.(5a,b) more data than have been actively assimilated at the previous cycles (e.g. by using a less restrictive thinning) allows a broader evaluation of β^b and B_β. The possibility to evaluate such quantities outside the assimilation scheme also offers opportunities for the preparation of the future assimilation of currently inactive data types.

The bias correction scheme at DWD is implemented as described above having options for online bias correction using background or analysis and variational bias correction as well. While, as mentioned above, theoretical considerations favor bias correction against the analysis, at DWD we currently obtain better forecast impact results when using the background instead. The reasons for this need to be investigated further and may include imbalances (spin up/down) or residual biases (possibly induced by the bias correction increments themselves) in the analysis state. Work is ongoing to use a bias correction versus analysis starting from analyses based on conventional data, atmospheric motion vectors and radio occultation data and complemented by radiances from carefully chosen instruments/channels only. Also, other aspects of observation bias correction are being looked at, e.g. the choice of predictors, including the use of non-linear terms, the bias correction for cloud radiances and bias correction for conventional observations.

Variational bias correction differs from online bias correction only in case of sudden changes of biases. They can be captured only if the contribution of the statistics of the current cycle is large. Currently the specified time constant τ between 7 and 30 days is large compared to the data assimilation cycle time $\Delta t = 3\,h$ so that the variational formulation is not expected to have any impact compared to the online scheme using analysis data.

References

Auligné, T., McNally, A. and Dee, D. (2007) Adaptive bias correction for satellite data in a numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, **133**, 631–642.

Eyre, J. (2016) Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. *Quarterly Journal of the Royal Meteorological Society*, **142**, 2284–2291.