

Deliverable D13.4: Initial report on bias correction activities

Work package no	WP13
Deliverable no.	D13.3
Lead beneficiary	ECMWF
Deliverable type	<input checked="" type="checkbox"/> R (Document, report) <input type="checkbox"/> DEC (Websites, patent filings, videos, etc.) <input type="checkbox"/> OTHER: please specify
Dissemination level	<input checked="" type="checkbox"/> PU (public) <input type="checkbox"/> CO (confidential, only for members of the Consortium, incl Commission)
Estimated delivery date	Month 23
Actual delivery date	18/07/2017 (Month 26)
Authors	Julie Letertre-Danczak, Angela Benedetti (ECMWF)
Version	Final
Comments	<p><i>The research scientist recruited for ACTRIS-2 could not take her post before September 2016 due to her previous contract obligations. The expected start of the position was June/July 2016 which explains part of the delay. Moreover, due to recent changes to the ECMWF's Integrated Forecast System (IFS) parts of the code which was needed to deliver Task 13.2 had been removed. Work to recover these missing pieces of software has progressed well, however, the timeliness of the delivery of the report has been affected.</i></p>

D13.3: INITIAL REPORT ON BIAS CORRECTION ACTIVITIES

The ECMWF forecast model is based on the assimilation of data in the 4D-var system. All data available over a time window of 12 hours are assimilated in a global minimization problem. To process the assimilation of different data (in-situ, remote sensing, radar, LIDAR, sounders, etc.), the data have to be consistent and they also need to be bias-corrected. As part of task Task 13.2 of WP13 (Joint Research Activity 3) which investigates the potential of ACTRIS-2 data for assimilation and satellite bias correction, technical and scientific developments have been undertaken at ECMWF to develop the 4D-Var assimilation of EARLINET data in order to correct the bias in satellite lidar backscatter, for example from the CALIPSO satellite. The assumption is that the ground-based data are more accurate and better calibrated than the satellite data, and could be used in the system to “anchor” the satellite data. This report covers the general principles of the variational bias correction (varBC) as it is implemented operationally at ECMWF and outlines the initial steps towards the fulfilment of Task 13.2.

1) varBC

The varBC process gives a correction of an estimated observational bias using available data and auxiliary information. Any data arriving in the code is first quality-checked to select good quality data. Filters for the quality check are usually fixed, for example land/ocean mask, range of observations, or any other indicators of data quality that can be used for the screening. The problem with this type of filters is they do not remove any biases deriving from inter-calibration or from the instrument itself. In 2005, the varBC was developed and implemented (Dee, 2005) to provide an online tool to adjust and remove observational bias. The strength of this tool is the consideration in parallel of global observations and model via the background but also the specific properties of the data set.

The principle of the var-BC is to use a simple bias correction, based on a regression model, which is pre-computed for all data and prior to the assimilation. The regression parameters depend on the dataset. For example, for some ozone observations the solar elevation angle is used as a regression parameter, while for Aerosol Optical Depth surface wind over ocean is used. The procedure starts with a first correction which is based on an estimation of the observational error. As for the minimization of the cost function, a minimization loop is executed on the ensemble of bias parameter (matrices format) to adjust the bias. This preliminary step in 4D-var provides a preconditioning for the joint minimization problem.

The datasets acquired at ECMWF are divided in two categories: active assimilation and passive monitoring. In both cases, the var-BC is applied. Since the var-BC procedure is cycling on a matrix, it means all data (active or passive) have an impact on the different parameters. In order to apply the var-BC to any data (instrument, channels, etc) it is important to have first a good quality screening.

The parameters of the var-BC are trained day after day and need a time of adjustment before being fully efficient. In most cases, at least two weeks are necessary for the var-BC to be adjusted during the monitoring phase before processing the data for the full assimilation.

The Figure 1 is an example of LIDAR data from the CALIOP instrument on board of CALIPSO. It can be seen on that case the bias in the prior (O-B) was larger than after analysis (O-A), which confirm the reduction of the bias by assimilation. In principle, however, for a “perfect” 4D-Var the background should be unbiased and the assimilation should only reduce the random error. This is where the varBC comes into play, in reducing the bias in the background due to the observations.

For atmospheric composition data such as Aerosol Optical Depth (AOD) or lidar backscatter, the bias correction is often based on the background itself, for lack of data that can be used as anchor.

The problem with this kind of bias adjustment is the consideration of the model as the truth. Often model biases are folded into observation biases. This makes the system converge to the model rapidly and there is no more balance between the model and the observation (Healy, 2008). To prevent this kind of drift in the system, it was demonstrated that varBC should not be applied to high quality data (as GPS-RO data) and they should be used as an anchor to oblige the system to keep a link with the true observed earth-atmosphere system. This anchoring system will be explained in the following part.

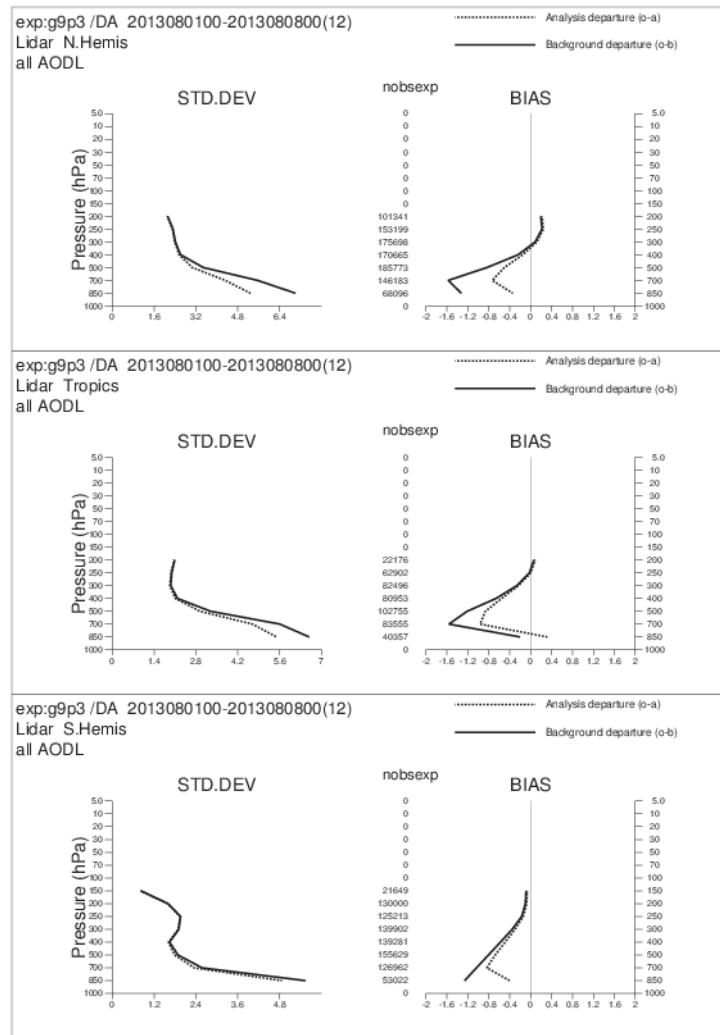


Figure 1. CALIOP observations minus analysis (dash), forecast (solid) standard deviation (left) and bias (right) for the north hemisphere (top), tropics (middle) and South hemisphere (bottom). The data are assimilated for 8 days (01st to 8th August 2013). Units of lidar backscatter are $10^{-7} (\text{sr.m})^{-1}$.

2) Anchor

The principle of anchor data is equivalent to consider those data as bias-free. This, however, does not mean that no screening is needed. It only means that the calibration errors and other sources of bias can be neglected. In IFS, the data which are considered without bias are GPS-Radio Occultation, AMVs, radio-sounders, and most of the time all in-situ measurement. In general, those are measurements from instruments than people can have direct access to perform the necessary calibrations or satellite products (level 2) which are by the nature of algorithm defined without bias.

In that case, the data are directly after the screening process assimilated without any correction and secure the system from converging too closely to the model. Eyre (2016) demonstrated that the more data are assimilated in the system with a bias correction, the more data assimilated with an anchor are needed to balance var-BC and the pre-conditioned minimization. This logical implication is often

forgotten in the complexity of the 4D-var system for several observations. For example, for the aerosol, all data routinely assimilated from the MODIS satellite are processed using var-BC without any anchor. This is due both to the fact that there is no redundancy in the observations and no single AOD product can be considered bias-free and used as anchor. In the experiments with the CALIOP data, it was observed that the introduction of the new dataset worsened the bias between the output of the assimilation and AERONET (Figure 2), which was not the case with MODIS only. This was most likely a consequence of the relative biases between MODIS and CALIOP as well as model bias in lidar backscatter which is higher than model bias in AOD.

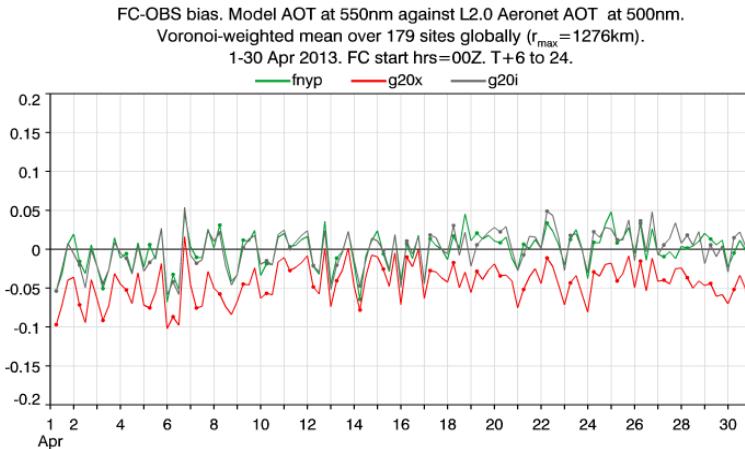


Figure 2. Bias in AOD with respect to AERONET data after assimilation of MODIS only (green: CY38R2, grey : CY40R1), MODIS and CALIOP (red) for April 2013.

The aim of this task is to take advantage of the EARLINET data as anchor for the CALIOP data. In order to achieve that the EARLINET data corresponding to the campaign between 09th and 12th July 2012 (Sicard et al 2015, D'Amico et al 2015) have been assimilated in the ECMWF's 4D-VAR. The data are located over Europe and are assimilated in the global system. Even with a weak global coverage, a positive impact can be observed after the assimilation of these data. Figure 3 shows the decrease of the standard deviation and bias after the data assimilation which indicates the correct functioning of the assimilation.

The next step towards using this data as an anchor is to assimilate them together with satellite lidar data, for example CALIOP data and apply the varBC only to the latter while taking the ground-based lidar data as bias-free. In the following section, some technical developments towards completing this task are described.

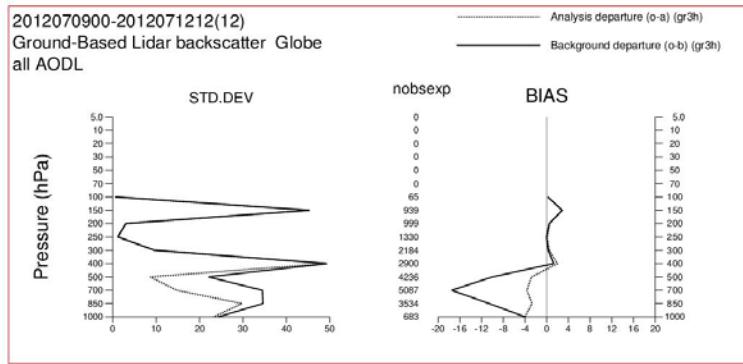


Figure 3. Ground-based LIDAR observations minus analysis (dash), forecast (solid) standard deviation (left) and bias (right) over all stations. The data are assimilated for 4 days (9th to 12th July 2012). Units for the horizontal axis are $10^{-7} (\text{sr.m})^{-1}$.

3) Technical developments

Several technical developments were needed to assimilate the ground-based lidar measurements. The development follows largely that which was implemented for the assimilation of CALIOP data in 2013. However, it has been realised that due to model changes and upgrades, part of that code had been removed. The code has now been re-written and adapted to the present configuration of IFS (CY43R1). Moreover, the data received from EARLINET are formatted in NETCDF4, whereas IFS is able to ingest only BUFR data or ODB data at the screening level. Because the ODB format is easiest to encode, it has been chosen to convert the NETCDF files into ODB files using python scripts.

Additionally, the original code was only able to assimilate one type of LIDAR (ground-based or satellite), one wavelength (to be chosen between 355, 532 or 1064) and one parameter (attenuated backscatter or aerosol backscatter). A lot of parameters and variables, such as wavelength and number of levels, were hardcoded. In order to be able to use different datasets such the ground-based lidar data in combination with the satellite lidar data, the system had to be further modified. At the moment, the system is flexible enough to ingest data in different configurations using information contained in the ODB files. However, these technical developments have still to be tested in order to be able to assimilate in parallel anchor data (EARLINET) and bias-corrected data (CALIOP).

4) Conclusions

This initial report highlights the importance to have different types of data (in-situ or satellite) with a different pre-treatment of the information. It also shows the difficulty to evaluate the cause of a bias in the data as this bias could be coming from the model or from the data themselves. Different techniques to remove observational bias have been implemented operationally for a number of years both to meteorological and composition data. Future work will involve the application of varBC with the anchor approach to the aerosol LIDAR case in a complex multi-instrumental setting.

5) Bibliography

D'Amico, G. et al.: EARLINET Single Calculus Chain - overview on methodology and strategy, *Atmos. Meas. Tech.*, 8, 4891-4916, doi:10.5194/amt-8-4891-2015, 2015.

Dee, D.: Bias and data assimilation, *Q.J.R.Meteorol.Soc.* 131: 3323 – 3343, 2005.

Eyre, J. R.: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties, *Q.J.R.Meteorol.Soc.* 142: 2284 – 2291, DOI:10.1002/qj.2819, July 2016.

Healy, S.: Assimilation of GPS radio occultation measurements at ECMWF, GRAS SAF workshop on Applications of GPSRO Measurements, 16-18 June 2008.

Sicard, M. et al.: EARLINET: potential operability of a research network, *Atmos. Meas. Tech.*, 8, 4587-4613, doi:10.5194/amt-8-4587-2015, 2015.