Bias correction using data assimilation: Application on the Lorenz '95 and NEMO-LIM models.

M. Canter[a,b] (martin.canter@ulg.ac.be), A. Barth[a,c]

[a] GeoHydrodynamics and Environment Research, MARE, University of Liège, BELGIUM
[b] Doctoral Student, BELGIUM
[c] Research Associate, F.R.S.-FNRS, BELGIUM

1. Introduction

Bias correction approaches can be separated into offline and online methods [1]. In offline methods, the bias is estimated from the model mean and the climatology (based on observations), using a preliminary model run. It is a rather basic estimation, but it has a small computational cost. In online methods, bias is updated during the data assimilation step, resulting in an analyzed bias. Several online bias correction approaches use a two-stage estimation technique [2,3,4]. They augment the model statevector with an estimate of the bias, and assume that the bias can be isolated from the other state-vector variables. This allows the successive yet separated estimation of the bias estimation and model estimation. It is assumed that the covariance distribution of the biased and unbiased error are identical except for a proportionality factor [2,5,6]. Literature shows that the performance of the assimilation system can be greatly enhanced when different covariance models are used for the unbiased and systematic errors [3,7]. However, the existing methods only allow to correct the model output. They do not help correcting the source of the bias, which generally originates from unresolved processes or bias in the surface forcing fields.

The main objective of this work is to develop an innovative and general method of bias correction using data assimilation. First developed with a twin experiment on a Lorenz '95 model [8], this new method is currently being applied and tested on the sea-ice ocean NEMO-LIM model, which is used in the PredAntar project.

2. Method Principle

This method aims at correcting the source of bias directly into the model's equations. Therefore, a good knowledge of the model is necessary, as well as a clear idea of the origin of the bias. The entire procedure can be summarized with the following steps:

- Estimate the model's bias and its source in the model's equations.
- Create an ensemble of stochastic forcing directly added into the model's equations.
- Run the model for each forcing field separately.
- Consider this stochastic forcing as a control variable for data assimilation.
- Estimate bias and correct the forcing field with data assimilation.
- Correct the source of the bias with the stochastic forcing.
- Interpret bias in terms of unresolved physical processes, bad parametrization, ...
- Validate the bias correction with external and independent data.

Data assimilation is thus used here as a tool in order to estimate and find the best forcing term to add into the model's equations. Previous bias correction methods with data assimilation only account for bias during the assimilation procedure. However, after the assimilation, the model tends towards its biased state again (Fig. 1). Here, with this new method, we aim at a continuous correction of the bias while the models is running.

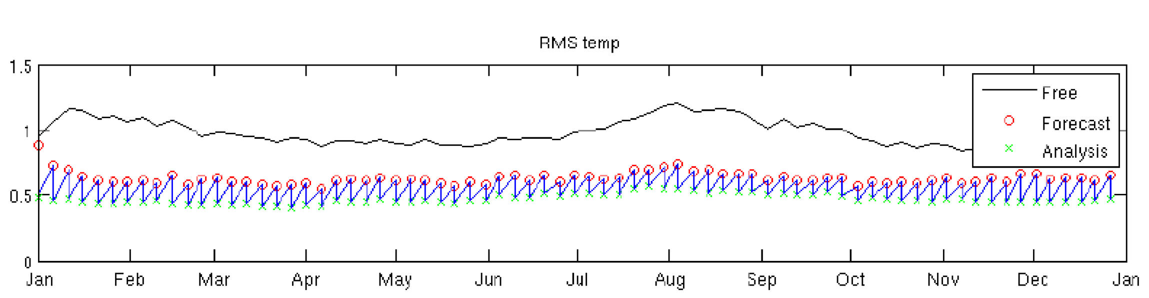

Figure 1: *Example of bias effect on the RMS of temperature around the Antartic(PredAntar Scientific report, 2012)*

3.1 Application on the Lorenz '95 model

The Lorenz '95 model is a chaotic model which we used here with $k = 40$ variables based on the following equation [8]. In order to have a realistic bias, we complicated the model, by using a different but spatially correlated forcing for each variable: $\mathbf{F_k}$, thus inducing a bias on the model.

$$\frac{dX_k}{dt} = -X_{k-2}X_{k-1} + X_{k-1}X_{k+1} - X_k + \mathbf{F_k} \quad (1)$$

We used the general procedure presented in paragraph 1, with a twin experiment. We only considered the mean over time of each variable k of the model, since there is a linear relationship between this mean, and $\mathbf{F_k}$. For each run (ensemble and reference), we used 15 different random initial conditions.

An initial run was considered as the reference state, with a random but spatially correlated forcing. Noise was added to create pseudo-observations: $\mathbf{F_{kref}} = \mathbf{F_k} + noise$.

We then created an ensemble of 100 different $\mathbf{F_k}$ and ran the model of each member, with 15 initial conditions. Using an extended state vector containing the model's variable mean over time, and the forcing terms $\mathbf{F_k}$, we assimilated the observations and corrected the ensemble of $\mathbf{F_k}$. We were able to improve the ensemble's model's mean and its bias (Fig. 2):

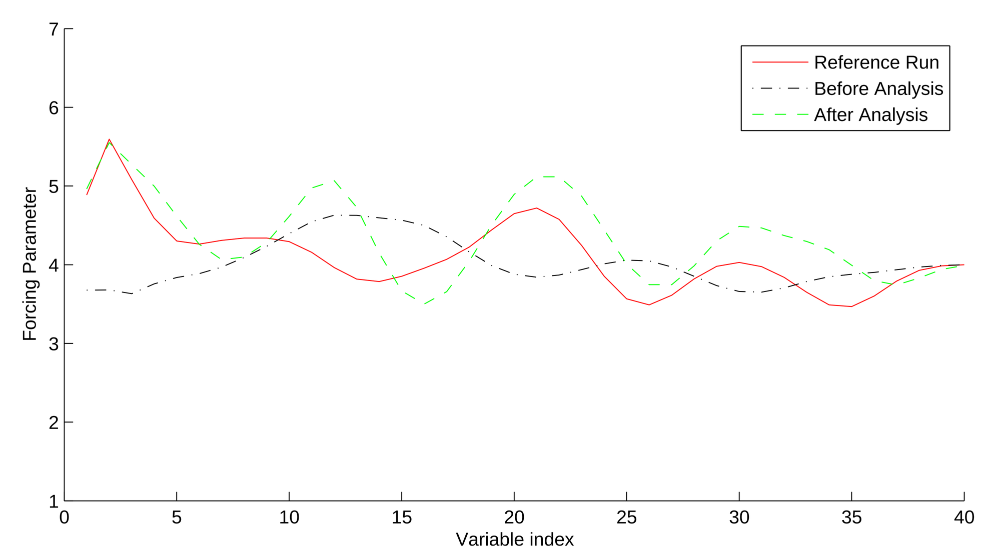

Figure 2: *Comparison of the model's variable mean between the biased ensemble, and the corrected ensemble through data assimilation.*

3.1 Multiple assimilation on the Lorenz '95 model

Since the Lorenz '95 model is non-linear, we tried to improve the data assimilation procedure by making smaller, but multiple corrections. Indeed, by changing the error covariance matrix of the observations accordingly [9], we can make multiple assimilation with smaller data batches:

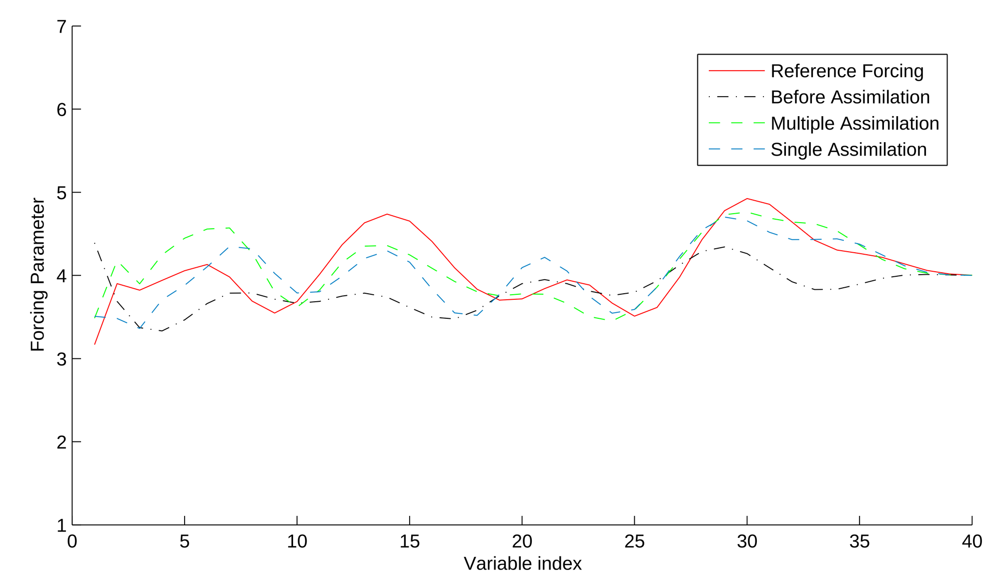

Figure 3: *Comparison of the model's variable mean between the biased enesmelbe, and the corrected ensemble through single and multiple data assimilation.*

The standard error deviation for the runs without assimilation, with a single assimilation and a multiple assimilation are respectively 0.2531, 0.0909 and 0.0626.

4.1 Application on the NEMO-LIM model

This method is currently being applied on a twin experiment, on the NEMO-LIM model from the PredAntar project (Belspo). NEMO-LIM is a global and low resolution (2 degrees) coupled model with long time steps allowing simulations over several decades. It is used in the PredAntar project (Belspo), which aims at understanding and predicting the Antartic sea ice variability at the decadal timescale. Because of this low resolution, ocean currents are badly represented and have been identified as a possible source of bias. They have a direct impact on heat transportation in the ocean, thus also on the sea surface temperature bias. Therefore, the forcing term (F_u and F_v) will be applied directly into the momentum equations of the ocean's dynamic equations of NEMO (Eq. 2,3).

$$\frac{du}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + fv + \frac{1}{\rho}\frac{\partial \tau_x}{\partial z} + F_u \quad (2)$$

$$\frac{dv}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - fu + \frac{1}{\rho}\frac{\partial \tau_y}{\partial z} + F_v \quad (3)$$

4.2 NEMO-LIM model: Creating the forcing term

The forcing in paragraph 2 has as only constraint that it is spatially correlated. However, since we are working with a realistic physical model, we need to add some conditions when forcing NEMO-LIM. Indeed, we do not want to add spurious gravitational waves, so we need the divergence of the velocity fields to be zero. We also want to dampen our forcing when going to higher depths, were currents are usually smaller. Finally, we can construct a forcing field by using higher resolution models and observations, instead of only using a random function. Here is how the forcing is built:

- Make a free run of the model.
- Compare this free run with a higher resolution model (Hycom) and subtract the difference between the two current fields.
- Generate a random, spatially correlated field, using Diva-nd [10], and use it as stream function.
- Derive zonal and meridional velocity fields.
- From the free run, extract a mean turbocline depth.
- Dampen the derived velocity fields with the mean turbocline depth.
- Finally, combine the field from the higher resolution comparison, and the randomly generated stream function.

$$Forcing = (Hycom - Nemo) - \exp^{\text{depth-turbocline}} * Random(Diva) \quad (4)$$

This way, we now have a forcing field which is based on a better resolution model, and completely random part, to create an ensemble of forcings. Different parameters have been tested concerning the correlation length of the random forcing ($2000km$), the amplitude ratio between the random forcing and the difference with the higher resolution model, ...

4.2 Assimilation with NEMO-LIM model

Now that we have a way to build a forcing term with physical constraints, we are currently proceeding with a twin experiment. Using the relationship between sea surface height and ocean currents, the assimilation procedure uses perturbed observations of the sea surface height to correct the forcing on the ocean currents. The twin experiment procedure is similar to the Lorenz '95 case test described in paragraph 2.

A reference run with a random forcing with noise is used for the observations. An ensemble of runs with random forcings is created, perturbed observations are assimilated, and the corrected forcing is compared to the reference forcing. However, some parameters still need to be improved.

Main references

1. C.L. Keppenne, M.M. Rienecker, D.A. Adamec (2005), Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction, Nonlinear Processes in Geophysics, **12**, 491-503
2. D.P. Dee, R. Todling (2000),Data Assimilation in the Presence of forecast Bias: The GEOS Moisture Analysis, Monthly Weather Review, **128**, 2368-3282
3. J.A. Carton, G. Chepurin, X. Cao, B.S. Giese (2000), A Simple Ocean Data Assimilation Analysis of the Global Upper Ocean 1950-1995, J. Phys. Oceanography, **30**, 294-309
4. L. Nerger, W.W. Gregg (2008), Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter, Journal of Marine Systems, **73 (12)**, 87-102
5. D.P. Dee, A.M. Da Silva (1998), Data assimilation in the presence of forecast bias, Q.J.R. Meteorol. Soc., **124**, 269-295
6. M.J. Martin, M.J. Bell, N.K. Nichols (2002), Estimation of systematic error in an equatorial ocean model using data assimilation, Int. J. Numer. Meth. Fluids, **40**, 435-444
7. G.A. Chepurin, J.A. Carton, D.P. Dee (2005), Forecast Model Bias Correction in Ocean Data Assimilation, Monthly Weather Review, **133**, 1328-1342
8. E.N. Lorenz, (1996), Predictability: a problem partly solved, Proc. Seminar on Predictability, European Centre for Medium-Range Weather Forecast, **1**
9. J. D. Annan, D. J. Lunt, J. C. Hargreaves, P. J.Valdes (2005), Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter, Nonlinear Processes in Geophysics, **12**, 363-371
10. C. Troupin, A. Barth, D. Sirjacobs, M. Ouberdous, J.-M. Brankart, P. Brasseur, M. Rixen, A. Alvera-AzcÃ¡rate, M. Belounis, A. Capet, F. Lenartz, M.-E. Toussaint, J.-M. Beckers (2012), Generation of analysis and consistent error fields using the Data Interpolating Variational Analysis (DIVA), Ocean Modelling, **52-53**, 90-101