# Diagnosis of some error contributions in global and regional data assimilation cycling

Rachida EL OUARAINI[1], Loïk BERRE[2] and Claude FISCHER[2]
[1] *Maroc-Météo, CNRMSI, Casablanca, Morocco* (elouaraini@gmail.com)
[2] *Météo-France, CNRM/GMAP, Toulouse, France*

Ensemble data assimilation experiments are often run by perturbing observations and the model in a cycled way. The links between associated ensemble sensitivity experiments and a possible diagnosis of the amplitude of different error contributions are here made explicit and discussed, in the context of both global and regional systems.

**1. Formal evolution of ensemble perturbations and of errors in data assimilation (DA)**

As discussed in El Ouaraini and Berre (2011) and El Ouaraini et al (2015), during a given analysis/forecast step (whose index is denoted by $l$, associated to time $t_l$), while the full model state of each ensemble member evolves in a non linear way (due to non linearities of the forecast model in particular), ensemble forecast perturbations $\tilde{\boldsymbol{\epsilon}}_l^f$ can be formally conceived as resulting from linearized accumulation and propagation of three different contributions (equation (1)) :

$$\tilde{\boldsymbol{\epsilon}}_l^f = \mathbf{T}_{l+1}\ \tilde{\boldsymbol{\epsilon}}_0^b + \sum_{i=0}^{l} \mathbf{T}_{l-i}\ \mathbf{M}_i\ \mathbf{K}_i\ \tilde{\boldsymbol{\epsilon}}_i^o + \sum_{i=0}^{l} \mathbf{T}_{l-i}\ \tilde{\boldsymbol{\epsilon}}_i^m \tag{1}$$

where $\tilde{\boldsymbol{\epsilon}}_0^b$, $\tilde{\boldsymbol{\epsilon}}_i^o$ and $\tilde{\boldsymbol{\epsilon}}_i^m$ correspond respectively to initial background perturbations (introduced at a given initial time $t_0$) and to recent observation perturbations and model perturbations, both introduced at recent successive steps $t_i$ between $t_0$ and $t_l$. The $\mathbf{T}$ matrices represent the following cycling operators over several past successive analysis/forecast steps $j$ :

$$\begin{aligned} \mathbf{T}_{l+1} &= \prod_{j=0}^{l} \mathbf{M}_j\ (\mathbf{I} - \mathbf{K}_j \mathbf{H}_j) \\ \mathbf{T}_{l-i} &= \prod_{j=i}^{l-1} \mathbf{M}_j\ (\mathbf{I} - \mathbf{K}_j \mathbf{H}_j) \end{aligned}$$

where $\mathbf{M}_j$ is a linearized version of the model operator (around the non linear (unperturbed) deterministic state), $\mathbf{K}_j$ is the specified gain matrix, and $\mathbf{H}_j$ is the linearized version of the observation operator.

While this formalism is valid for the evolution of ensemble perturbations, it can also be considered to be representative of the actual error evolution. This means not only that this formalism can be used for investigating some sensitivity of ensemble spread to different perturbation sources (e.g. in order to show that ensemble spread is not much sensitive to old background perturbations, after some spinup period), but it also implies that this formal framework can also be applied to derive estimates of different contributions to the forecast error amplitude. This is illustrated and discussed in the next two sections, using some Figures from El Ouaraini and Berre (2011) in section 2, and from El Ouaraini et al (2015) in section 3. Note that these Figures were discussed in terms of sensitivity experiments in these two papers, while we discuss here their implications with respect to estimation of error contributions.

**2. Contribution of initial background errors in a global data assimilation system**

Figure 1 illustrates the temporal evolution of global spread of vorticity near 500 hPa for a cold-start configuration (solid line, associated to zero values for the initial background perturbations : $\tilde{\boldsymbol{\epsilon}}_0^b = 0$) and for a warm-start configuration (dashed line, corresponding to non-zero values for the initial background perturbations $\tilde{\boldsymbol{\epsilon}}_0^b$, thanks to ensemble DA cycling over 6 preceding days) of the ARPEGE global ensemble, in a perfect model framework (i.e. $\tilde{\boldsymbol{\epsilon}}_l^m = 0$ at every step $l$).

According to equation (1), the difference between the squared values of the two curves in Figure 1 can be interpreted as an estimate of the initial background error contribution (namely $\mathbf{T}_{l+1}\ \tilde{\boldsymbol{\epsilon}}_0^b$) to the forecast error variance $V(\tilde{\boldsymbol{\epsilon}}_l^f)$ at different steps $l$ of the cycling. The results indicate that the initial background error contribution has an amplitude which represents nearly half of the forecast error magnitude at initial time $t_0$ (while the other half corresponds to the contribution of observation errors introduced at time $t_0$). It also appears that this initial contribution tends to vanish after 3 days of data assimilation cycling, due to successive analysis damping effects during the cycling. This comparative analysis is thus also a way to study and show to which extent background errors (depending on their "age" with respect to the current cycling step $l$) contribute to uncertainties in the data assimilation cycling.

## 3. Contribution of Lateral Boundary Condition (LBC) errors in a regional data assimilation system

Figure 2 shows the temporal evolution of horizontally averaged analysis spread at 00 UTC for temperature near 850 hPa for the ALADIN-France regional ensemble, using either unperturbed Lateral Boundary Conditions (solid line) or perturbed Lateral Boundary Conditions (dashed line), in a perfect model framework (i.e. model perturbations are equal to zero at every step $l$, except for Lateral Boundary Condition perturbations obtained through the coupling to a global ARPEGE ensemble).

According to equation (1), the difference between the squared values of the two curves can be interpreted as the contribution of LBC errors (which are one part of the model errors $\tilde{\epsilon}_l^m$) to forecast error variance (on average over the considered ALADIN-France domain). The results indicate that LBC errors explain about one third of forecast error variance on average over the considered area. As illustrated in Figure 3, this contribution varies much over the domain, with values that tend to be largest near the boundaries and in associated downstream regions. Such sensitivity experiments can thus be explicitly used to investigate the influence of LBC errors on regional forecast error amplitudes.

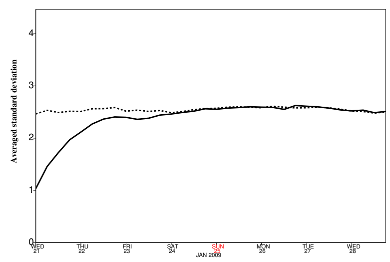


*Figure 1: Temporal evolution of global spread of vorticity near 500 hPa for cold-start (solid line) and warm-start (dashed line) configurations of the ARPEGE global ensemble.*

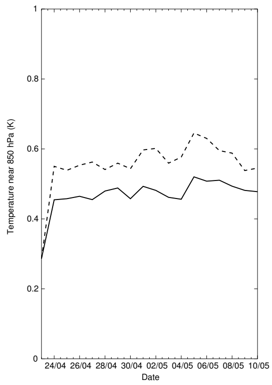


*Figure 2: Temporal evolution of horizontally averaged analysis spread at 00 UTC for temperature near 850 hPa for the ALADIN-France regional ensemble, using either unperturbed LBCs (solid line) or perturbed LBCs (dashed line).*

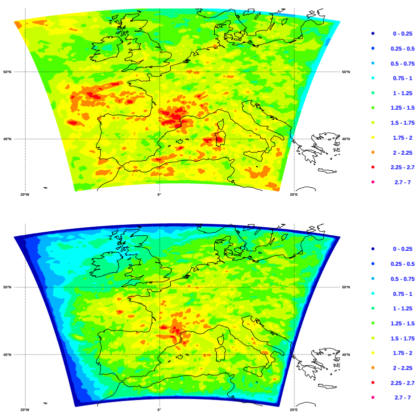

*Figure 3: Horizontal maps of time-averaged spread of 6h zonal wind forecasts at 06 UTC, using either unperturbed LBCs (bottom panel) or perturbed LBCs (top panel).*

## 4. Conclusions and future work

While sensitivity experiments can be relevant as such in ensemble data assimilation, we advocate their use for additionally deriving explicit estimations of error contributions, in both global and regional data assimilation systems. Such experiments and diagnostics may also be compared with innovation-based estimates in the future, in order to compare and derive estimates of model error contributions for instance.

## References

[1] El Ouaraini, R. and Berre, L. 2011. Sensitivity of ensemble-based variances to initial background perturbations. J. Geophys. Res. 116, D15106. DOI: http://dx.doi.org/10.1029/2010JD015075.
[2] El Ouaraini, R., C. Fischer, L. Berre and EH. Sayouty, 2015. Sensitivity of regional ensemble data assimilation spread to perturbations of lateral boundary conditions. Tellus A, 67. DOI: http://dx.doi.org/10.3402/tellusa.v67.28502.