Sensitivity of regional ensemble data assimilation spread to perturbations of lateral boundary conditions

By Rachida El Ouaraini *, Loïk Berre 2, Claude Fischer 2 and El Hassan Sayouty 3, 1Direction de la Météorologie Nationale, BP 8102, Casablanca, Morocco; 2Météo-France/CNRS, CNRM/GAME, 42 Avenue Coriolis, FR-31057 Toulouse, France; 3Physique des Hautes Energies et de la Matière Condensée, Faculté des Sciences Ain Chock, Hassan II University of Casablanca, BP 5366 Maarif, Casablanca, Morocco

(Manuscript received 12 May 2015; in final form 2 November 2015)

ABSTRACT

The implementation of a regional ensemble data assimilation and forecasting system requires the specification of appropriate perturbations of lateral boundary conditions (LBCs), in order to simulate associated errors. The sensitivity of analysis and 6-h forecast ensemble spread to these perturbations is studied here formally and experimentally by comparing three different LBC configurations for the ensemble data assimilation system of the ALADIN-France limited-area model (LAM). While perturbed initial LBCs are provided by the perturbed LAM analyses in each ensemble, the three ensemble configurations differ with respect to LBCs used at 3- and 6-h forecast ranges, which respectively correspond to: (1) perturbed LBCs provided by the operational global ensemble data assimilation system (GLBC), which is considered as a reference configuration; (2) unperturbed LBCs (ULBC) obtained from the global deterministic model; (3) perturbed LBCs obtained by adding random draws of an error covariance model (PLBC) to the global deterministic system. A formal analysis of error and perturbation equations is first carried out, in order to provide an insight of the relative effects of observation perturbations and of LBC perturbations at different ranges, in the various ensemble configurations. Horizontal variations of time-averaged ensemble spread are then examined for 6-h forecasts. Despite the use of perturbed initial LBCs, the regional ensemble ULBC is underdispersive not only near the lateral boundaries, but also in approximately one-third of the inner area, due to advection during the data assimilation cycle. This artefact is avoided in PLBC through the additional use of non-zero LBC perturbations at 3- and 6-h ranges, and the sensitivity to the amplitude scaling of the covariancemodel is illustrated for this configuration. Some aspects of the temporal variation of ensemble spread and associated sensitivities to LBC perturbations are also studied. These results confirm the importance of LBC perturbations for regional ensemble data assimilation. They also indicate that perturbing initial LBC is not sufficient to obtain realistic ensemble spread, whereas this can be achieved approximately by using random covariance draws for simulating LBC errors during the forecast and associated data assimilation cycling.

Keywords: lateral boundary conditions, regional ensemble, data assimilation, ensemble spread

1. Introduction

Lateral boundary conditions (LBCs) represent a significant source of analysis and forecast error in limited-area models (LAMs). In fact, several studies indicate that their simulations are likely to be more sensitive to LBCs than to initial condition errors, especially for small domains and long model integrations (Anthes et al., 1985; Anthes, 1986; Vukicevic and Errico, 1990; Du and Tracton, 1999).

LBC errors thus need to be accounted for and represented when designing and implementing a regional ensemble data assimilation and prediction system. A first possible approach is to use a set of perturbed LBCs provided by a global ensemble system (e.g. Hou et al., 2001; Storto and Randriamampianina, 2010). However, such a global ensemble may not be implemented and available locally in a given regional NWP centre. Moreover, fetching a large volume of global ensemble data from a global NWP centre is likely to be cumbersome and costly in terms of telecommunications, depending on data volume and telecommunication speed. Additionally, the regional
ensemble size may be larger than the global ensemble size, and there may also be timeliness issues. In order to reduce such dependency and cost issues, a second possible approach is to use data from a single deterministic global NWP system to provide the same LBC to all members of a given regional ensemble. This approach has been considered for instance in the context of ensemble Kalman filtering (see, e.g., Zhang et al., 2006) or variational ensemble assimilation (e.g. Storto and Randriamampianina, 2010). The drawback is that this amounts to assuming that LBC errors are negligible, which is likely to be unrealistic.

A third possible approach is to construct LBC perturbations as random draws from a specified error covariance model, which is assumed to be representative of LBC errors (e.g. Torn et al., 2006). This method is likely to be attractive, because it is relatively easy to implement and run locally, and it is also computationally cheap.

These different options have been partly studied experimentally in different regional NWP contexts. For instance, Torn et al. (2006) have focussed their diagnosis on the quality of the ensemble mean in an idealised context with simulated observations, by comparing the first and third approaches (i.e. either use of a larger-scale ensemble, or use of covariance draws). They have shown that the use of covariance draws leads to a quality of the ensemble mean that is similar to the use of a larger-scale ensemble, although some degradations are visible near the lateral boundaries. Gebhardt et al. (2011) have shown, in the context of the COSMO-DE ensemble, that LBC perturbations contribute significantly to the uncertainty simulation (compared to model perturbations based on a multi-parameter approach), when combining four different global deterministic forecasts and three different sets of physical parameters for intermediate-scale COSMO-SREPS coupling fields, in order to produce 12 different LBCs for the COSMO-DE ensemble. Moreover, in the context of a mesoscale ensemble system over Japan, Saito et al. (2012) have shown that lateral boundary perturbations increase the ensemble spread and improve the accuracy of the ensemble mean forecast. In the context of a 1.5-km ETKF over the southern U.K., coupled to a 24-km regional ensemble, Caron (2013) has found that a mismatch between the analysis perturbations and the LBC perturbations leads to the generation of spurious gravity waves, which confirms that a careful design of LBC perturbations should be sought.

While the influence of different LBC perturbation approaches has thus been partly studied experimentally in different regional NWP contexts, the purpose of the present paper is twofold. Firstly, while previous investigations have been mostly experimental, a theoretical development of underlying equations of errors and ensemble perturbations in the data assimilation cycle is presented here, in order to discuss the relative contributions of observation errors, LBC errors and model errors. Secondly, another purpose of the paper is to compare experimentally the spread associated to the three aforementioned options of LBC perturbations in the context of the ALADIN-France ensemble data assimilation system, for which the choice and effect of LBC perturbations is likely to vary at different steps of the ensemble data assimilation cycle.

On the one hand, in the ALADIN-France context, initial LBCs are perturbed because they are derived from the perturbed ALADIN-France 3D-Var analysis solution. On the other hand, LBCs and associated perturbations have still to be specified also at 3- and 6-h forecast ranges. Particular attention will thus be devoted to study the relative effects of observation perturbations and of LBC perturbations at initial and 3–6-h forecast ranges, by using the spread comparison between the three considered LBC approaches.

The paper is organised as follows: Section 2 describes the experimental framework. A formal analysis of error and perturbation equations is carried out in Section 3. Spatial variations of ensemble spread are investigated in Section 4, while some aspects of temporal variations are examined in Section 5. Conclusions are given in Section 6.

2. Experimental framework

2.1. The ALADIN-France LAM

The ensemble experiments which will be described in Section 2.3 have been performed using the ALADIN-France data assimilation and forecasting system. ALADIN-France is a high-resolution limited-area spectral model (e.g. Horanyi et al., 1996) with bi-periodic extension of the computation domain. It uses elliptical truncation of double-Fourier series, semi-implicit semi-Lagrangian integration scheme, fourth order horizontal diffusion and digital filter initialisation (DFI). It is run over a European domain centred over France, with a Lambert grid, a 7.5-km horizontal resolution (in both x and y directions) and 70 vertical levels from the surface up to 0.1 hPa.

It is built on the basis of the global model IFS/ARPEGE and keeps the same dynamics, physics and vertical coordinate (a terrain-following pressure hybrid coordinate). The primitive hydrostatic equations are solved for the horizontal wind components, temperature, specific humidity and surface pressure (e.g. Bubnova et al., 1995; Cordoneanu and Geleyn, 1998). ALADIN-France uses a 6-h 3D-Var data assimilation scheme (Fischer et al., 2005), which assimilates a wide range of conventional observations (e.g. SYNOP, BUOY, TEMP and PILOT) and of remote sensing data (such as AMSU-A and B, HIRES, MHS, AMV, SEVIRI, AIRS and IASI data). ALADIN-France was used at Météo-France operationally from 1996 to 2012, and other ALADIN versions are still used operationally in several NWP centres in Europe and North Africa.
2.2. LBCs for ALADIN-France

As any LAM, ALADIN needs information about the state of the atmosphere outside of its integration domain. Figure 1 shows the horizontal domain of ALADIN-France. The central zone represents the region of meteorological interest, where the forecast is fully adapted to resolved small-scale conditions. The width of this zone corresponds to 381 grid points. The lateral zone represents the coupling zone where a large-scale solution, generally obtained from ARPEGE, is mixed with the solution resulting from the LAM integration. LBCs are updated every 3 h and linearly interpolated in time between available updates in order to have LBC forcing field at each time step of ALADIN. These values on the boundaries are needed for solving the model system equations. The width of this zone corresponds to eight grid points. The one-way nesting is achieved by using a Davies relaxation scheme (Davies, 1976, 1983; Davies and Turner, 1977).

Initial LBCs (i.e. specified at 0-h forecast range) are determined by the ALADIN-France 3D-Var analysis itself, while the LBC specified at 3- and 6-h forecast ranges correspond to 3- and 6-h forecasts of the global ARPEGE system, which are interpolated onto the ALADIN-France grid. Using analysis fields as initial LBC data in all experiments avoids generating spurious spin-up behaviour in the early forecast ranges, as shown in Fischer and Auger (2011). This choice is also consistent with more recent experimental findings for the AROME assimilation at Météo-France (Pierre Brousseau, personal communication). An incremental DFI is additionally applied to ALADIN-France initial conditions, following the study of Fischer and Auger, 2011.

2.3. Ensembles description

Three regional ensemble data assimilation experiments are performed using the ALADIN-France model. Each ensemble consists of six ALADIN members, and it is run in the same way as described, for example, by Houtekamer et al. (1996), Fisher (2003) and Berre et al. (2006). For each member of each ensemble, and for each analysis step, observation perturbations $\epsilon_o$ are added to observation values, in order to simulate observation errors. These observation perturbations $\epsilon_o$ are provided by random draws of the observation error covariance matrix $R$: $\epsilon_o = R^{\frac{1}{2}} \zeta$, where $\zeta$ are random Gaussian draws which are spatially uncorrelated and which have zero mean and unit variance (i.e. $\zeta \sim N(0, I)$, $I$ being the identity matrix in observation space). In addition to these explicit perturbations of observations, the ensembles use implicit perturbations of the background, provided by the perturbation evolution during the 6-h data assimilation cycle. Apart from the use of perturbed LBCs, the ensemble simulations do not use explicit model perturbations (which amounts to using a perfect model assumption). Likely, this will be important to consider in future studies, because model errors are expected to limit the sensitivity of ensemble spread to perturbations of LBCs, to some extent, as will be indicated also in Section 3.1.

The three ensembles are run from April 23, through May 10, 2010. This 18-d period is relatively short, but it has been found to be sufficiently long in order to diagnose several

---

**Fig. 1.** Horizontal domain of ALADIN-France containing 389 grid points in $x$- and $y$- directions: the central zone (light grey) represents the region of meteorological interest. The coupling zone (dark grey), whose width corresponds to eight grid points, is the region where the coupling is added.
features which are either time-averaged (e.g. regarding spatial variations of spread and also its sensitivity to the amplitude of drawn perturbations) or time-varying (e.g. regarding the evolution of space-averaged spread but also some flow-dependent aspects during this period). The initial ALADIN-France background on April 23 at 00 UTC is taken from the operational deterministic system. This implies that initial background perturbations are equal to zero, and that the ensemble spread increases during the first data assimilation cycles before reaching stable values, as will be illustrated in Section 5.1.

Each of the three ensembles uses perturbed initial LBCs (i.e. at 0-h forecast range) which are determined by the corresponding ALADIN-France perturbed analyses of each member. The three considered ensembles differ with respect to the set of LBCs specified at 3- and 6-h forecast ranges:

1. The first ensemble, which will be called unperturbed LBC (ULBC), uses unperturbed 3- and 6-h LBCs which are the downscaled deterministic global forecasts of ARPEGE. Each ALADIN-France ensemble member thus receives the same 3- and 6-h LBC as the ALADIN-France deterministic configuration. The omission of such LBC perturbations amounts to assuming that there are no errors in these forecast LBCs which are used in the ALADIN-France data assimilation cycle.

2. The second ensemble has been run using 3- and 6-h LBCs taken from the global ensemble data assimilation AEARP, which has been operational at Météo-France since 2008 (Berre et al., 2007, 2009). It is made of six members with a global spectral truncation of T399 (i.e. a global uniform horizontal resolution of 50 km), 70 vertical levels and a 4D-Var assimilation scheme. This ensemble is considered as a reference for comparison purposes, and it will be referred to as GLBC.

3. The third ensemble uses perturbed 3- and 6-h LBCs, produced by adding LBC perturbations to the interpolated deterministic forecasts of ARPEGE, with LBC perturbations generated by using the same method as in El Ouaraini and Berre (2011) to create initial background perturbations: they are constructed from random draws of a specified error covariance matrix \( P_{LBC} \), which is assumed to be representative of LBC errors: \( \epsilon^{LBC} = P_{LBC}^{1/2} \eta \), where \( \eta \sim \mathcal{N}(0,1) \) being here the identity matrix in model space, and \( \epsilon^{LBC} \sim \mathcal{N}(0,P_{LBC}) \). In the present study, this specified covariance \( P_{LBC} \) has been simply constructed as a scaled version of the ALADIN-France background error covariance matrix \( B \); \( P_{LBC}^{d} = \alpha B \), where \( \alpha \) is a scaling factor whose value has been tuned to 0.3, in order to obtain roughly similar amplitudes of LBC perturbations as in the GLBC configuration. The sensitivity with respect to this amplitude factor \( \alpha \) will be illustrated in Section 4.3. This ensemble will be called PLBC in the remainder of the paper. Moreover, it can be noticed that ULBC can be seen as a limit case of PLBC corresponding to \( \alpha = 0 \), which is also a way to provide information on sensitivity with respect to \( \alpha \) in PLBC.

Note that the choice \( \alpha = 0.3 \) instead of a value close to 1 may reflect several aspects. Firstly, this may be related to effects of seasonal differences (as shown, e.g., in Monteiro and Berre, 2010), as the specified ALADIN-France matrix \( B \) has been computed over a winter period in January–February 2009, which is likely to be associated to larger background errors than in the spring period (April–May 2010) that is considered in this paper. Secondly, the relatively large values in the specified ALADIN-France matrix \( B \) may also suggest that these specified values were overestimated, which tends to be supported by the fact that such an overestimation has been diagnosed for the related ARPEGE system in 2009. Thirdly, \( P_{LBC} \) is supposed to represent ARPEGE analysis and short-range forecast errors, which are expected to have amplitudes that are partly different from ALADIN-France background errors, so that \( \alpha \) is not really expected to be equal to 1.

Note also that the specified matrix \( B \), and associated drawn lateral boundary perturbations, are partly flow-dependent. This corresponds to flow-dependent error standard deviations of vorticity provided by the global ARPEGE ensemble (Berre et al., 2007) and also to the use of flow-dependent non-linear and omega balances (Fisher, 2003).

In the next sections, associated error and perturbation equations are examined formally, and then the ensemble spread sensitivities with respect to LBC perturbations are studied experimentally for analyses and 6-h forecasts. Ensemble spread is computed using the following formula:

\[
\sigma (i,j,z) = \sqrt{\frac{1}{N - 1} \sum_{n=1}^{N} (\epsilon_n(i,j,z))^2}
\]

with \( \epsilon_n(i,j,z) = \bar{x}_n(i,j,z) - \bar{x}(z) \), where \( n \) is the ensemble index, \( N \) is the ensemble size, \( (i,j) \) are indices of horizontal position, \( z \) is the index of vertical position, \( \bar{x}_n \) is the model state of member \( n \), and \( \bar{x} \) is the model state of the ensemble mean \( \bar{x} = \frac{1}{N} \sum_{n=1}^{N} \bar{x}_n \).

3. Formal examination of the error evolution and of its ensemble simulation

In this section, the equations of states, errors and perturbations will be derived for the deterministic and ensemble configurations. This will be used to give a formal insight of
the implications of using observation and initial LBC perturbations, and either unperturbed or perturbed 3- and 6-h LBCs in the different ensemble experiments. The main derivations are partly analogous to those discussed, for example, in Berre et al. (2006) and El Ouaraini and Berre (2011), with adaptation to the considered LAM framework, by including LBC aspects. The model M and the observation operator H will be assumed to be linear for the sake of simplicity in the derivations and associated discussions. Model errors (i.e. errors in M) will be included in the error equations (Section 3.1), before being dropped in the perturbation equations [in accordance with the perfect model assumption used in the ensemble experiments (as discussed in Section 2.3)].

3.1. Evolution of errors during successive analysis/forecast steps

The evolution of errors during a given analysis step (denoted by the index i) is given by the following equation of analysis errors $e^i$:

$$e^i = (I - K_i H_i) e^i_0 + K_i e^m$$

(1)

where $e^i_0$ are background and observation errors, respectively, $K_i$ is the perfect model operator, that is, using not 3- and 6-h LBCs in the different ensemble experiments. The model M and the observation operator H will be assumed to be linear for the sake of simplicity in the derivations and associated discussions. Model errors (i.e. errors in M) will be included in the error equations (Section 3.1), before being dropped in the perturbation equations [in accordance with the perfect model assumption used in the ensemble experiments (as discussed in Section 2.3)].

To express model errors $e^m$ as the sum of LBC errors $e^{LBC}$ and of model errors $e^{PP}$, which are related to approximations in the dynamics and in the physics:

$$e^m = e^{LBC} + e^{PP}$$

(5)

where $e^{LBC}$ corresponds to the contribution of LBC errors during the considered 6-h forecast integration:

$$e^{LBC} = (M_i - M_{i0}) x_i$$

(6)

which is the difference between two forecast integrations starting from a perfect initial state $x_{i0}$, but using either imperfect or perfect LBCs (respectively), while $e^{PP}$ corresponds to accumulated model errors related to an imperfect representation of dynamical and physical processes:

$$e^{PP} = (M_i - M_{i0}) x_i$$

(7)

By considering the cycling of this analysis/forecast process in data assimilation, it is possible to expand forecast errors for the $i$th cycle as follows, as a function of previous observation errors, LBC errors and other model errors (note also that $e^0 = e^0_{i0}$, since the background for the analysis step $i$ is a forecast issued from the previous analysis step $i - 1$):

$$e^i = T_{i+1} e^i_0 + \sum_{j=0}^{i-1} T_{i,j} (M_i K_i e^j_0 + e^{LBC}_j + e^{PP}_j)$$

(8)

where $T_0 = I$ can be seen as a weight matrix associated to errors introduced during the current $i$th analysis/forecast step [which corresponds to $l=i$ in the sum above (eq. (8))], while for $l$ between 0 and $i-1$, $T_{i,l}$ is a weight matrix assigned to errors added during the $l$th analysis/forecast step:

$$T_{i,l} = \sum_{r=0}^{i-1} M_r \ (I - K_r H_r)$$

(9)

and $e^0$ is the initial background error (i.e. the error associated to the very first background from which the LAM data assimilation cycle is initiated), whose weight matrix in eq. (8) corresponds to

$$T_{i+1} = \sum_{r=0}^{i-1} M_r \ (I - K_r H_r)$$

(10)

Equations (8), (9) and (10) indicate that this operator $T$ can be seen as a temporal propagation operator of previous errors, which are accumulated (and partly damped) during the data assimilation cycle. As studied in El Ouaraini and Berre (2011), $T$ tends to act as a damping operator of previous error contributions, because eigenvalues of $I - K_r H_r$ are smaller than 1 (as discussed in Daley, 1991, p. 127). This implies for instance that the influence of the initial background error becomes negligible after 3-4 d of cycling, in the context of the study of El Ouaraini and Berre (2011).
Moreover, the forecast error equation [eq. (8)] indicates that the forecast error \( e' \) can be seen as the sum of contributions arising from observation errors \( e^o \), LBC errors \( e^{LBC} \) and other model errors \( e^{DP} \). Therefore, eq. (8) shows that the relative contribution of LBC errors to forecast errors depends on the relative amplitudes of observation errors, LBC errors and other model errors.

It also suggests that this relative LBC error contribution is likely to depend on the spatial coverage (and density) of the observation network. For instance, an observation network covering most of the LAM domain implies that the relative contribution of observation errors (compared to LBC errors) will be relatively large over the whole domain. Conversely, an extreme opposite configuration would be the case where the observation network would be reduced to a single observation (located in the centre of the domain for instance). In this extreme case, forecast errors \( e' \) (which are expected to be much larger than with a regular and dense observation network) would essentially arise from the accumulation of LBC errors \( e^{LBC} \) and of other model errors \( e^{DP} \) during the data assimilation cycle, except in the neighbourhood of the single observation.

The relative LBC error contribution is also expected to depend on the quality of the observation network. On the one hand, eq. (8) indicates that forecast errors \( e'_f \) are partly proportional to current observation errors \( e'_o \) in particular. On the other hand, this is expected to be partly mitigated by the fact that eigenvalues of the gain \( K_i \) are smaller than 1 (as discussed in Daley, 1991, p. 127), and also by the fact that these gain coefficients become smaller when specified observation error variances are increased. Examining the formula of the variance of \( K_i e'_o \) in a simple scalar case (see Appendix A) suggests that the contribution of observation errors tends to be maximum for observations whose quality is roughly similar to the background quality, and that this contribution decreases for accurate observations (because their noise becomes small) and also for inaccurate observations (because their weight in the analysis becomes very small). This suggests, in turn, that the LBC error contribution will be relatively large (compared with observation error contributions) in the case of accurate observations (because large LBC errors during the current forecast step will tend to predominate over small initial condition errors) and also in the case of inaccurate observations (because observation errors will be heavily damped by the analysis itself, while LBC errors will tend to accumulate during the data assimilation cycle; note also that background errors will be relatively large in such a configuration).

The effect of other model errors is more direct to some extent: eq. (8) indicates that, the larger the errors \( e^{DP} \) in the dynamical and physical schemes are, the smaller the relative contribution of LBC errors will be.

Last but not least, the appearance of the operator \( M \) in equations such as eqs. (6) and (9) reflects, for example, the influence of advective effects during each 6-h forecast integration: for instance, in the case of strong inflow, LBC errors will be advected towards a large part of the inner LAM domain.

3.2. Evolution of perturbations in the ensemble cycling

By analogy with the forecast error equation [eq. (8)], it is relatively easy to show that the equation of forecast perturbations \( \epsilon'_f \), for the different LAM ensembles which are conducted using a perfect model assumption, can be written as follows:

\[
\epsilon'_f = T_{i+1} \epsilon'_0 + \sum_{l=0}^{i-1} T_{i-l} \left( M_l K_l \epsilon'_i + \epsilon'^{LBC}_i \right)
\] (11)

where \( \epsilon'_0 \) is the initial background perturbation (which is equal to zero in the considered experiments, because all ensemble members start from the deterministic system), \( \epsilon'_f \) are observation perturbations which are drawn from the specified observation error covariances, and \( \epsilon'^{LBC}_i \) are LBC perturbations associated to the considered LAM ensemble configuration, whose equations will now be made explicit.

In order to derive equations of LBC perturbations, model states such as the LAM analysis and the LAM 6-h forecast will be denoted by \( x_{La,i} \) and \( x_{f,i} \), respectively: the \( L \) index refers to the LAM system, which can be distinguished from the \( G \) index referring to the global system that provides LBCs to the LAM forecasts; the \( a \) and \( f \) indices refer respectively to the analysis valid at initial time (akin to 0-h forecast range) and to the forecast valid at 6-h range.

In order to examine the perturbation evolution associated to the 3-h update of LBCs, the model operator \( M \) will be considered thereafter to correspond to either 3- or 6-h forecast integrations (as indicated by specific indices), and the LBCs that are used by the LAM model operator \( M \) will be indicated by two indices referring to the LBCs used at initial and final times: for instance, the 3-h forecast model operator to be applied to the analysis will be denoted by \( M_{La,a,Gf} \), in order to indicate that the initial LBC is the LAM analysis \( x_{La,a} \) and that the final LBC is the global 3-h forecast \( x_{f,Gf} \) (after interpolation onto the LAM domain). Similarly, the subsequent 3-h forecast model operator to be applied to the LAM 3-h forecast, and which provides the 6-h forecast state, will be denoted by \( M_{La,f,Gf} \), in order to express the use of the global 3-h forecast \( x_{f,Gf} \) as the initial LBC and the global 6-h forecast \( x_{f,Gf} \) as the final LBC. \( x_{La,a} \) is the LAM background state, which is used during the considered analysis state.

Using these notations, the evolution of the LAM state during a given analysis/forecast cycle can be expressed by
sensitivity of regional ensemble spread to LBCs

7

the following three equations for the deterministic LAM system:

\[ \mathbf{x}_{L0} = (I - KH) \mathbf{x}_{L0} + \mathbf{K} \mathbf{y}_0 \]
\[ \mathbf{x}_{L3} = M_{L0,G3} \mathbf{x}_{L3} \]
\[ \mathbf{x}_{L6} = M_{G3,G6} \mathbf{x}_{L3} \]

With respect to the GLBC ensemble for instance, the same kind of equations can be written for the perturbed LAM state, denoted by \( \tilde{\mathbf{x}}_L \), where the tilde symbol reflects the perturbed feature of the state:

\[ \tilde{\mathbf{x}}_{L0} = (I - KH) \tilde{\mathbf{x}}_{L0} + \mathbf{K} \tilde{\mathbf{y}}_0 \]
\[ \tilde{\mathbf{x}}_{L3} = M_{L0,G3} \tilde{\mathbf{x}}_{L3} \]
\[ \tilde{\mathbf{x}}_{L6} = M_{G3,G6} \tilde{\mathbf{x}}_{L3} \]

which allows equations of perturbations \( \epsilon = \tilde{x} - x \) to be derived:

\[ \epsilon_{L0} = (I - KH) \epsilon_{L0} + \mathbf{K} \epsilon_0 \]
\[ \epsilon_{L3} = M_{L0,G3} \epsilon_{L3} + \left( M_{L0,G3} - M_{L6,G3} \right) \tilde{\mathbf{x}}_{L0} \]
\[ \epsilon_{L6} = M_{G3,G6} \epsilon_{L3} + \left( M_{G3,G6} - M_{G6,G3} \right) \tilde{\mathbf{x}}_{L3} \]

Equations (12) and (13) can be combined in order to express the 6-h forecast perturbation as a function of analysis and LBC perturbations:

\[ \epsilon_{L6} = M_{L0,G6} \epsilon_{L0} + \epsilon_{LBC} \]

where \( M_{L0,G6} = M_{G3,G6} M_{L0,G3} \) is the unperturbed 6-h forecast operator, and \( \epsilon_{LBC} \) corresponds to the contribution of LBC perturbations during the 6-h forecast integration:

\[ \epsilon_{LBC}^{LBC} = M_{G3,G6} \epsilon_{LBC}^{L0,6} + \epsilon_{LBC}^{L6} = M_{G3,G6} \left( M_{L0,G3} - M_{L6,G3} \right) \tilde{\mathbf{x}}_{L0} \]

where \( \epsilon_{LBC}^{LBC} = (M_{L0,G3} - M_{L6,G3}) \tilde{\mathbf{x}}_{L0} \) is the 3-h forecast perturbation arising from the use of perturbed LBCs at both 0- and 3-h forecast ranges, and \( \epsilon_{LBC}^{L6} = (M_{G3,G6} - M_{G6,G3}) \tilde{\mathbf{x}}_{L3} \) is the subsequent 3-h forecast perturbation corresponding to the use of perturbed LBCs at 3- and 6-h forecast ranges. This equation [eq. (14)] of the LBC perturbation \( \epsilon_{LBC} \) in GLBC is similar to the LBC error equation [eq. (6)], which can be further expanded as follows (while omitting the index of the considered analysis/forecast step), as a function of LBC errors accumulated during the successive two 3-h forecast integrations (i.e. from 0 to 3 h and from 3 to 6 h, respectively):

\[ \epsilon_{LBC} = M_{G3,G6} \left( M_{L0,G3} - M_{L6,G3} \right) \mathbf{x}_0 + (M_{G3,G6} - M_{G6,G3}) \mathbf{x}_3 \]

(15)

where \( \mathbf{x}_0, \mathbf{x}_3 \) are the exact states at 0- and 3-h ranges, and \( M_{L0,G3}, M_{G3,G6} \) are the corresponding model operators based on perfect LBCs.

Comparing eqs. (14) and (15) indicates that LBC errors are simulated in GLBC through the difference between two integrations that use the same initial condition but different LBCs. Moreover, it can also be noticed that this common initial condition corresponds to \( \tilde{\mathbf{x}}_{L0} \) in GLBC (with respect to 3-h accumulated LBC perturbations), whereas it may seem preferable to use \( \mathbf{x}_0 \) ideally, according to eq. (15). This difference can be seen as corresponding to an additional term, for example, in the 3-h LBC perturbation equation:

\[ \epsilon_{LBC}^{L0} = (M_{L0,G3} - M_{L6,G3}) \tilde{\mathbf{x}}_{L0} = (M_{L0,G3} - M_{L6,G3}) \mathbf{x}_0 + (M_{L0,G3} - M_{L6,G3}) \tilde{\mathbf{x}}_{L3} \]

To the extent that the amplitude of the perturbation \( \tilde{\mathbf{x}}_{L0} - \mathbf{x}_0 \) tends to be smaller than the amplitude of \( \mathbf{x}_0 \), this additional term \( (M_{L0,G3} - M_{L6,G3}) \tilde{\mathbf{x}}_{L3} \) is expected to be a second order term in the LBC perturbation equation.

3.3. Evolution of analysis and forecast perturbations in the ULBC ensemble

The equations of state and perturbation evolution, such as (11), are similar in the ULBC ensemble, but the LBCs are partly different from those used in the GLBC ensemble, as expressed by the following equations of 3- and 6-h perturbed forecasts:

\[ \mathbf{x}_{L0} = M_{L0,G3} \tilde{\mathbf{x}}_{L0} \]
\[ \mathbf{x}_{L3} = M_{G3,G6} \tilde{\mathbf{x}}_{L3} \]

where the notation \( \tilde{\mathbf{x}}_{L0} \) refers to the fact that the 3-h forecast model in ULBC uses perturbed LBCs at initial time (provided by \( \tilde{\mathbf{x}}_{L0} \)), and ULBC at final time (corresponding to \( \mathbf{x}_{L0} \)) at 3-h range). By using similar derivations as in subsection 3.2, this allows the 6-h-accumulated contribution of LBC perturbations for the ULBC ensemble to be expressed as follows, which can be compared to eq. (14) for GLBC:

\[ \epsilon_{LBC}^{LBC} = M_{G3,G6} \epsilon_{LBC}^{L0,6} + \epsilon_{LBC}^{L6} = M_{G3,G6} \left( M_{L0,G3} - M_{L6,G3} \right) \tilde{\mathbf{x}}_{L0} \]

(16)

where \( \epsilon_{LBC}^{L0,6} = (M_{L0,G3} - M_{L6,G3}) \tilde{\mathbf{x}}_{L0} \) is the 3-h forecast perturbation arising from the use of perturbed LBCs at initial time, and \( \epsilon_{LBC}^{L6} = (M_{G3,G6} - M_{G6,G3}) \tilde{\mathbf{x}}_{L3} = 0 \) is the implicitly zeroed perturbation corresponding to the use of ULBC at 3- and 6-h forecast ranges.
3.4. Discussion on the effect of LBC perturbations in the different ensemble configurations

On the one hand, eq. (16) indicates that some significant contributions to forecast perturbations are expected from the use of perturbed initial LBCs in ULBC, in addition to the contributions of cycled observation perturbations, as indicated by eq. (11). On the other hand, the comparison between eqs. (14) and (16) also indicates that such contributions of LBC errors are obviously expected to be underestimated in ULBC, due to the use of ULBC at 3- and 6-h ranges. However, in accordance with the discussion in Section 3.1, the amplitude and spatial extent of this spread underestimation is likely to depend on the spatial coverage and quality of the observation network for instance.

As the ULBC configuration is the simplest one to implement for a NWP centre, which does not run a global ensemble system, one of the purposes of the considered experiments is thus to quantify the amount and spatial distribution of underdispersion associated to the use of LBC perturbations that are restricted to initial LBCs in ULBC.

Another purpose of interest is to examine the possibility to simulate LBC errors by using random draws of a specified error covariance model, as done in the PLBC ensemble. This will also include a sensitivity experiment with respect to the amplitude of such drawn perturbations: first, because it is likely to be a sensitive component of the PLBC ensemble, and second, because this is also a way to further document the effect of LBC perturbations in the LAM data assimilation ensemble.

4. Horizontal variations of time-averaged spread of 6-h forecasts

4.1. Horizontal maps of time-averaged spread

Horizontal maps of ensemble spread have been computed for each ensemble configuration, and their temporal average over the 18-d experimental period is plotted in Fig. 2 for 6-h forecasts of zonal wind near 500 hPa. Note that this 18-d average has been considered knowing that the ensemble spin-up has been found to be small (about 1 d), as will be shown in Section 5.1. The bottom panel (corresponding to ULBC) includes only observation perturbations and initial LBC perturbations, while the middle panel (associated to GLBC) additionally includes the effect of 3- and 6-h LBC perturbations. The middle and bottom panels of this Figure can thus be compared in order to diagnose the effect of 3- and 6-h LBC perturbations on ensemble spread, relative to the effect of observation and initial LBC perturbations.

As expected from the use of ULBC at 3- and 6-h ranges, the ensemble spread for ULBC (bottom panel) is artificially close to zero near the lateral boundaries. Moreover, the comparison with GLBC (middle panel) indicates that this underdispersion in ULBC is also pronounced in a relatively large part of the inner computation domain. This is particularly visible in the western part of the area, which covers approximately one-third of the LAM domain, and which corresponds to the Near Atlantic, where large amplitudes in GLBC appear to be largely underestimated in ULBC. This artefact is also well marked in the North-West corner of the domain, which corresponds to relatively large perturbations in GLBC, whereas this is associated to relatively small values in ULBC. This apparent contamination of the inner area by small LBC perturbation amplitudes in ULBC can be interpreted as corresponding to the advection of underestimated LBC perturbations towards the inner part of the domain, as discussed also in the previous section.

Conversely, it can be noticed that the spread underestimation in ULBC is much less visible in the central and North-East parts of the domain for instance. With respect to eqs. (11), (14) and (16), this indicates that observation perturbations and initial LBC perturbations contribute to the main part of the forecast perturbations in these areas, whereas the 3- and 6-h LBC perturbations (as represented in GLBC) contribute much more significantly in the western part of the LAM domain.

The underestimation of ensemble spread is avoided in PLBC (top panel), in accordance with the addition of 3- and 6-h LBC perturbations in this ALADIN-France ensemble configuration. As a consequence, the horizontal variations of ensemble spread are relatively similar in PLBC and in GLBC, with for instance large spread in the Near Atlantic, and smaller spread in data dense continental areas such as Germany and Spain. This finding can be explained by the fact that both GLBC and PLBC ensembles use identical observation perturbations, similar amplitudes of LBC perturbations (related to the tuning of factor $\alpha$ as explained in Section 2.3) and the same data assimilation system for cycling perturbations. It can also be noticed that the ensemble spread tends to be somewhat larger in PLBC than in GLBC, for instance in the Near Atlantic and in the Mediterranean Sea. This aspect will be illustrated further in Section 4.2.

4.2. Meridional and zonal cross-sections of spread maps

The meridional and zonal cross-sections of time-averaged spread of zonal wind near level 500 hPa, temperature near 300 and 850 hPa and specific humidity near 850 hPa are respectively illustrated in Figs. 3 and 4 (using meridional and zonal averages, respectively). At first, these various panels illustrate the fact that the spread underestimation
Fig. 2. Horizontal maps of time-averaged spread of 6-h zonal wind forecasts valid at 06 UTC, near level 500 hPa for: PLBC (top), GLBC (middle) and ULBC (bottom). The scaling factor $a$ in PLBC is 0.3.
in ULBC, and its propagation towards the inner area, is a general feature valid for different variables and for different vertical levels.

It can also be seen that the underdispersion is particularly pronounced in the western part of the zonal cross-sections (Fig. 4). This feature is consistent with the zonal cross-domain flow which predominates on average during the considered period, as shown in Fig. 5a. One can also notice that the time-averaged wind plotted in Fig. 5a is consistent with the horizontal variations of ULBC.

Fig. 3. Meridional cross-sections of time-averaged spread of 6-h forecasts valid at 06 UTC of: (a) temperature near level 850 hPa, (b) temperature near 300 hPa, (c) zonal wind near 500 hPa and, (d) specific humidity near 850 hPa.
underdispersion visible in the bottom panel of Figs. 2 and 5b: for instance, small perturbations near the North-West boundary tend to be advected by the inflow towards the inner area, whereas significant inner perturbations tend to be advected by the outflow towards the North-East boundary.

Fig. 4. Zonal cross-sections of time-averaged spread of 6-h forecasts valid at 06 UTC of: (a) temperature near 850 hPa, (b) temperature near 300 hPa, (c) zonal wind near 500 hPa and, (d) specific humidity near 850 hPa.
In contrast with ULBC, meridional and zonal cross-sections of PLBC ensemble spread are relatively similar to those of GLBC. This is particularly visible in the a-panels of Figs. 3 and 4, which correspond to temperature near 850 hPa. One can also notice that the spread tends to be somewhat larger in PLBC than in GLBC, as can be seen in the profiles for wind near 500 hPa. This is likely to reflect the fact that the ensemble spread in PLBC is sensitive to the tuning of the amplitude of the covariance model, from which LBC perturbations are drawn. The tuning of the LBC perturbation amplitude (which corresponds to the factor $a$ in Section 2.3) is relatively rough, leading here to somewhat overestimated ensemble spread in PLBC compared to GLBC. The sensitivity of ensemble spread to the tuning of $a$ is studied in the next section.

4.3. Sensitivity to LBC perturbation amplitudes

In order to examine the sensitivity of PLBC spread to the tuning of the specified covariance amplitude of LBC perturbations, a variant of the PLBC configuration has been run experimentally, based on the amplitude factor $a = 1$ instead of $a = 0.3$ in PLBC. The corresponding time-averaged map of ensemble spread is shown in Fig. 6 for zonal wind near 500 hPa, which can be compared to the top panel of Fig. 2.

It can be seen in Fig. 6 that the ensemble spread is largely increased not only near the lateral boundaries, but also in a large part of the inner area, which corresponds approximately to one-third of the domain size. This spread increase is also more visible in the western part of the domain, in accordance with aforementioned advection effects of the predominant western flow.

The deliberate choice of an excessive amplitude factor ($a = 1$), which is the opposite extreme scenario of taking ($a = 0$) in ULBC, allows the sensitivity of the PLBC ensemble spread to the amplitude of LBC perturbations to be confirmed. This indicates that an adequate tuning of this amplitude is an important component for obtaining realistic ensemble spread in the regional system.

More generally, this kind of sensitivity experiment also confirms the importance of LBCs for both ensemble and
deterministic configurations, for instance in the sense that it indicates that forecast error amplitudes in the LAM domain can be significantly influenced by the amplitude of LBC errors.

5. Temporal variations of ensemble spread and its sensitivity to LBC perturbations

5.1. Temporal evolution of analysis domain-averaged spread

Figure 7 represents the temporal evolution, over the 18-d considered period, of horizontally averaged analysis spread for temperature near 850 hPa (a-panel) and 300 hPa (b-panel), zonal wind near 500 hPa (c-panel) and specific humidity near 850 hPa (d-panel). Figure 7 shows features for the three ensembles that are consistent with previous results in terms of forecast spread, namely similar spread in PLBC and in GLBC (slightly larger in PLBC due to the roughly tuned amplitude factor) and a clear underestimation of ULBC dispersion (e.g. for temperature near 850 hPa).

While the ensemble spread is nearly stable beyond the first day for most variables, it can be noticed that humidity spread (d-panel) exhibits stronger temporal variations, with larger spread in April than in May. This is related to the fact that this April period is associated to a prevailing westerly moist flow from the Atlantic, while in this May period, a northern continental polar air mass (relatively cold and dry) predominates. The evolution of specific humidity spread thus reflects associated changes in atmospheric water content.

The larger effect of overestimation of spread in PLBC for specific humidity than for temperature and wind, further suggests that a parameter-dependent tuning of covariance amplitudes is likely to be needed. This tends to be supported by the examination of spread maps of surface pressure (not shown), which indicate that, in the vicinity of lateral boundaries, PLBC spread tends to be much larger (by a factor 2–3) than in GLBC (although spread is relatively similar in the innermost part of the domain).

In addition, Fig. 7 indicates that the ensemble spread needs about four analysis cycles to converge towards stable values for, for example, temperature and wind. This spread spin-up time (about 1 d) is relatively short for the regional ALADIN-France ensemble system. This is rather similar to the spin-up period found for the global ARPEGE ensemble system in Raynaud et al. (2012).

This short spin-up period may be interpreted as resulting at least partly from the influence of LBCs. For instance, in the case of GLBC and PLBC, LBC perturbations contribute to increase the spread from zero values in the initial background towards stable values after a 1-d period. In the case of ULBC, the null LBC perturbations are expected to contribute to counteract the spread increase observed during the first day period, but the observed evolution of ULBC spread suggests that observation and background perturbations remain large enough to maintain a stable ensemble spread in the innermost part of the domain, although spread remains globally underestimated throughout the period due to the use of ULBC.

On the whole, whatever LBC perturbations are, the spread spin-up period is about 1 d, suggesting that LBC perturbations affect mostly the asymptotic value to which each ensemble spread converges.

The evolution of ensemble spread has also been examined as a function of forecast range (not shown). It appears that the underdispersion of ULBC (compared to GLBC and PLBC) tends to be augmented during the 3- and 6-h forecast integrations, due to increasing LBC effects. For zonal wind near 500 hPa as an example, the ULBC space-averaged spread underestimation (compared to GLBC) increases from 20 % at analysis time to 25 % at 6-h forecast range. This indicates that the neglect of LBC errors in ULBC has a visible effect during the 6-h forecast (with a 5 % increase of underdispersion from 20 to 25 %) and that this effect tends to be accumulated (up to 20 % at analysis time) during the data assimilation cycle.

5.2. Temporal evolution of domain-averaged correlations between spread maps

While Section 5.1 shows the behaviour of horizontally averaged spread maps over time, this section focuses on a diagnosis of the spatial coherence between flow-dependent spread structures in the different ensembles. This is diagnosed by computing the domain-averaged correlation between the reference (i.e. GLBC) and experimental spread maps, for each date of the 18-day period. The a-panel of Fig. 8, which corresponds to domain-averaged correlations for temperature spread near level 850 hPa, indicates that PLBC produces flow-dependent spread structures which are more consistent with GLBC than those produced by ULBC. This is much less visible for zonal wind near 500 hPa (b-panel of Fig. 8), which is likely to be related to some predominant sampling noise effects, associated to the small ensemble size (namely six members).

In order to investigate this further, some local spatial averages of ensemble spread have been computed, in order to reduce sampling noise effects (as discussed, e.g., in Raynaud et al., 2008; Berre and Desroziers, 2010):

$$\hat{\sigma}(i,j) = \frac{1}{(2N_m^o + 1)^2} \sum_{k=-N_m^o}^{N_m^o} \sum_{l=-N_m^o}^{N_m^o} \hat{\sigma}(i + k, j + l)$$
where \((i, j)\) are indices of horizontal position (varying between 1 and \(N_i, N_j\), respectively, \(N_i\) and \(N_j\) being the numbers of grid points in the zonal and meridional directions of the LAM domain), and \((2N^\text{z}_m + 1)^2\) is the total number of grid points used in the local 2D spatial average; the averaging radius \(N^\text{z}_m\) has a constant value equal to \(N^\text{z}_m\) in the inner part of the domain, and it is progressively reduced when approaching the lateral boundaries: \(N^\text{z}_m = \min(N^\text{z}_m, i - 1, N_i - i, j - 1, N_j - j)\).
Domain-averaged correlations between such spatially filtered spread maps are plotted in Fig. 9 for $n_w = 12$. It can be seen that the increase of correlation for PLBC (dotted line), relative to ULBC (solid line), is more pronounced than in Fig. 8, for both temperature and wind. This supports the idea that a spatial filter is useful to highlight the spread increased realism of PLBC compared to ULBC. One can also note that the difference between ULBC and PLBC correlations for wind is relatively larger at the beginning of the experimental period, which is believed to be related to advection effects of the prevailing western winds for these dates (as suggested in Fig. 2 by the severe underestimation by ULBC of the large values of GLBC spread which are located west of the French Atlantic coast).

5.3. Flow dependency of ensemble spread sensitivities to LBC perturbations

In this section, the bottom panel of Fig. 2, which corresponds to the time-averaged ensemble spread map of ULBC, is compared to the corresponding map for a given date. Two special cases are examined, a case of a strong continental flow which crosses the area from the northwestern side on May 4 of 2010, and a case of a jet which crosses the domain from the West on April 25, 2010 (see Fig. 10).

The conspicuous rolling-up pattern in Fig. 10b, corresponding to the 6-h horizontal wind deterministic forecast near 500hPa on 4 May 2010, implies a specific and recognisable cyclonic-like structure in the ensemble spread of 6-h zonal wind forecast in Fig. 10d. In the vicinity of the central part of the North boundary, the advection of zero-valued LBC perturbations by a northerly jet towards a large part of the inner area is well marked. This feature is interesting to compare with Fig. 2, as an illustration that, depending on the meteorological situation, the underestimation of ensemble spread in ULBC can be more pronounced in terms of spatial coverage than what can be seen in a time-averaged sense (bottom panel of Fig. 2). Figure 10f which corresponds to the ratio of spread of 6-h zonal wind forecasts near level 500hPa of ULBC to that of GLBC on 4 May 2010 can also be compared to Fig. 5b, in order to notice the specific location of underestimated spread along the North-West and South-East boundaries, which correspond to inflow borders of the domain (Fig. 10b).

Figure 10a shows the 6-h horizontal wind deterministic forecast near 500hPa on 25 April, with an apparent jet that crosses the area from the west, and another one, less intensive, which crosses the domain from the extreme northeastern side. The effect of these jets is again well visible in the map of zonal wind spread (Fig. 10c), with the inward advection of zero-valued LBC perturbations by the jets in
these two regions. The difference between Fig. 10e and f, along with the difference with the time-averaged spread ratio (Fig. 5b), confirm the strong flow dependency of spread sensitivities to LBC perturbations.

6. Conclusions

Unlike global models which are naturally periodic, LAMs and associated ensembles require the specification of values along the lateral boundaries of the model domain. In this paper, the ensemble data assimilation spread of the regional ALADIN-France system has been studied with different choices of LBC perturbations for 3- and 6-h forecast ranges, while observations and initial LBCs are perturbed in all experiments. A first ensemble configuration (GLBC), considered as a reference, is based on the use of the global ensemble data assimilation system AEARP, in order to provide perturbed 3- and 6-h LBCs. A second ensemble configuration (ULBC) uses the global deterministic forecasts of ARPEGE, in order to provide 3- and 6-h LBCs to each member of the ensemble; this amounts to using zero-valued LBC perturbations for these forecast ranges. A third ensemble (PLBC) uses 3- and 6-h LBC perturbations, which are provided by random draws of an error covariance model which is assumed to be representative of LBC errors.

A formal analysis of error and perturbation equations has been carried out for these deterministic and ensemble configurations, in order to provide an insight of the relative effect of 3- and 6-h LBC perturbations, compared to observation and initial LBC perturbations. Due for instance to the contribution of observation perturbations to forecast perturbations, it has been noticed that the relative effect of LBC perturbations is likely to depend on the spatial coverage and quality of the observation network. This formal analysis has been complemented by experimental studies of ensemble spread sensitivities to LBC perturbations.

Time-averaged horizontal maps of spread indicate that the use of unperturbed 3- and 6-h LBCs leads to an underestimation of the ensemble spread not only near lateral boundaries, but also in a relatively large part of the inner domain (roughly one-third of the ALADIN-France area), due to advection. Conversely, this is much less pronounced in the central and North-East parts of the LAM domain, due to the predominant contribution of observation and initial LBC perturbations in these regions. The spread underestimation is avoided when using LBC perturbations which are randomly drawn from an error covariance model, leading to spread maps which are similar to those of the reference regional ensemble coupled to the global ensemble. An additional experiment indicates also that the spread maps are sensitive to the amplitude scaling of these drawn

Fig. 9. Temporal evolution of domain-averaged correlations between respective spatially filtered spread maps 6-h forecasts valid at 06 UTC of ULBC and of GLBC (solid line) and of PLBC and GLBC (dotted line) for: (a) temperature near 850 hPa and (b) zonal wind near 500 hPa.
LBC perturbations. This is consistent with the spread sensitivity to LBC perturbations which has been studied in Section 4.

Three aspects of the temporal variations of ensemble spread and of associated LBC sensitivities have also been studied. The time evolution of ensemble analysis spread over the period of study indicates that the spread spin-up period is relatively short (1 d), the spread increases towards stable values within a few data assimilation cycles. This can be seen as resulting partly from the influence of LBC perturbations, but also from the contribution of observation perturbations and of associated cycled background perturbations. Correlations of spread maps, computed over the whole domain and test period, indicate that the spread obtained with the PLBC method is better correlated with the reference case (GLBC) than the spread obtained with ULBC. Filtering spatially the spread maps further confirms this finding, by reducing sampling noise effects associated to the six-member ensemble. On the whole, this result increases the confidence for the PLBC method to provide realistic dispersions. Two case studies have been conducted also, in order to illustrate the strong flow dependency of LBC sensitivities. They indicate that the spread underestimation can be much pronounced in terms of amplitude and of spatial coverage in the case of strong inflow associated to jet streams in particular.

These different sensitivity experiments thus confirm the key role of LBC perturbations for regional ensemble data assimilation and provide an indirect reminder of the importance of good quality LBCs for regional deterministic systems too. The results also indicate that using LBC perturbations which are drawn from an error covariance model allows realistic ensemble spread to be obtained. Therefore, this provides a practical alternative approach, for example, for regional centres which do not run global ensemble data assimilation systems and which would like

Fig. 10. Case of a Westerly jet on April 25, 2010 (left panels) and of a strong northern flow on May 04, 2010 (right panels). (a) and (b) 6-h horizontal wind forecast near 500 hPa from ALADIN-France deterministic model, valid at 06 UTC. (c) and (d) spread in ULBC of 6-h zonal wind forecasts near 500 hPa valid at 06 UTC. (e) and (f) ratio of spread of 6-h zonal wind forecasts valid at 06 UTC near 500 hPa of ULBC to that of GLBC.
to avoid, for instance, remote communications of large ensemble data volumes from another centre running a global ensemble.

While the present study has focussed on analysis and short-range (6-h) forecast spread, it would be interesting to consider this kind of approach over longer forecast ranges, in addition to similar studies over different experimental periods. Moreover, while this work has been conducted with a perfect model assumption (apart from LBC uncertainties), future studies may be considered in a context where model error is accounted for in the ensemble simulations.

7. Acknowledgements

We would like to thank the two anonymous reviewers for their helpful remarks. Rachida El Ouaraini benefited from stays in Toulouse, which have been funded by Météo-France, La Direction de la Météorologie Nationale Marocaine as well as by the French Ministry of Foreign Affairs. She thanks all the GMAP team for their generous support.

8. Appendix

A.1. Contribution of current observation errors to analysis errors in a simple scalar case

According to eq. (8), the contribution of errors introduced specifically during the $i$th analysis/forecast step (which can be referred to as the ‘current’ analysis/forecast step) corresponds to $\mathbf{M} \mathbf{K} e_i^o + e_i^{LBC} + e_i^{PP}$. This indicates that the contribution of current observation errors $e_i^o$ to forecast errors $e_i^f$ is proportional to $\mathbf{K} e_i^o$, which means that observation errors are weighted by the analysis gain $\mathbf{K}$. Equation (1) also indicates that $\mathbf{K} e_i^o$ is the current observation error contribution to the analysis error $e_i^f$.

As will be shown below in a scalar case with a single observation located at a single grid point (for the sake of simplicity), this implies that the current observation error contribution $\mathbf{K} e_i^o$ to the analysis and forecast errors tends to be maximum when the observation quality is roughly similar to the background quality and that this contribution decreases towards zero in the cases of either very accurate or very inaccurate observations.

In the considered scalar case, the analysis error contribution of the single observation is equal to $k_i e_i^o$, where $k_i = \frac{\sigma_i^2}{\sigma_b^2 + \sigma_i^2}$ is the observation weight resulting from specified background and observation error standard deviations ($\sigma_b$ and $\sigma_i$, respectively).

The variance of $k_i e_i^o$ (assuming that the observation error is unbiased) thus corresponds to:

$$\text{var}(k_i e_i^o) = E[(k_i e_i^o)^2] = k_i^2 E[(e_i^o)^2] = \left(\frac{\sigma_i^2}{\sigma_b^2 + \sigma_i^2}\right)^2 \sigma_o^2$$

where $\sigma_o$ is the exact observation error standard deviation, which is equal to the specified value $\sigma_i$ only in the optimal case. One can then define $\gamma = \frac{\sigma_b}{\sigma_i}$ as the ratio between observation and background error standard deviations, in order to express $\text{var}(k_i e_i^o)$ as follows (after noting that $\frac{\sigma_b^2}{\sigma_i^2} = \frac{1}{\gamma^2 + 1} = \frac{1}{1 + \gamma^2}$):

$$\text{var}(k_i e_i^o) = \left(\frac{\gamma}{1 + \gamma^2}\right)^2 \sigma_o^2$$

These expressions (A2) and (A3) are convenient, because they allow the observation error variance contribution $\text{var}(k_i e_i^o)$ to be expressed as a function of the relative observation accuracy, as measured by $\gamma = \frac{\sigma_b}{\sigma_i}$.

Equations (A2) and (A3) indicate that the observation error contribution $\text{var}(k_i e_i^o)$ is proportional to $(\frac{1}{1+\gamma^2})^2$. As shown by the thick full line in Fig. 11a (which corresponds to the standard deviation $\sigma(k_i e_i^o) = \sqrt{\text{var}(k_i e_i^o)}$ in the optimal case and with $\sigma_b = 1$), this contribution is maximum when $\sigma_o = \sigma_b$ (i.e. $\gamma = 1$). Furthermore, it appears that $\sigma(k_i e_i^o)$ decreases towards zero when $\gamma$ approaches zero: this corresponds to $\frac{1}{1+\gamma^2} \approx \gamma \approx 0$; that is, the observation errors become so small that their contribution to the analysis errors becomes negligible. In addition, Fig. 11a indicates that $\sigma(k_i e_i^o)$ decreases towards zero also when $\gamma$ becomes very large: this corresponds to $\frac{1}{1+\gamma^2} \approx \frac{1}{\gamma} \approx 0$; that is, the observation errors become so large that the analysis tends to dampen observations and their errors very heavily, so that their contribution to analysis errors becomes very small. Finally, the thin dash-dotted line in Fig. 11a corresponds to the standard deviation of the contribution of background errors to analysis errors, namely $\sigma(1 - k_i) e_i^o$ [see eq. (1)]: it appears that the current observation error contribution tends to predominate when $\sigma_o < \sigma_b$, whereas it is the opposite when $\sigma_o > \sigma_b$.

Equations (A2) and (A3) indicate that the observation error contribution is determined not only by the ratio $(\frac{1}{1+\gamma^2})^2$, but also by $\sigma_o^2$. Furthermore, $\sigma_b$ is itself expected to be partly dependent on the observation quality, due to cycling effects (e.g. increased observation errors imply that the analysis and forecasts are degraded, leading to a lesser background quality too). The effects of this aspect on the observation error contribution are illustrated in Fig. 11b.
The thick full line in Fig. 11b corresponds to the standard deviation of the current observation error contribution to analysis errors (thick full line) and standard deviations of background error contributions to analysis errors (thin dash-dotted line), as a function of the ratio $\gamma$ between observation and background error standard deviations $\sigma_b \equiv\sigma_o / \sigma_b$. (a) case where $\sigma_b$ is the same whatever $\gamma$ is (i.e. $\sigma_b = 1$); (b) case where $\sigma_b$ increases when $\gamma$ is larger (i.e. by specifying $\sigma_b = \sqrt{\sigma_o / (1 + \sigma_o^2)}$) in this example.

Fig. 11. Specified background error standard deviations (thick dashed line), standard deviations of current observation error contributions to analysis errors (thick full line) and standard deviations of background error contributions to analysis errors (thin dash-dotted line), as a function of the ratio $\gamma$ between observation and background error standard deviations $\sigma_o / \sigma_b$. (a) case where $\sigma_b$ is the same whatever $\gamma$ is (i.e. $\sigma_b = 1$); (b) case where $\sigma_b$ increases when $\gamma$ is larger (i.e. by specifying $\sigma_b = \sqrt{\sigma_o / (1 + \sigma_o^2)}$) in this example.

The thick full line in Fig. 11b corresponds to the specified background error standard deviations (thick dashed line), standard deviations of current observation error contributions to analysis errors (thick full line) and standard deviations of background error contributions to analysis errors (thin dash-dotted line), as a function of the ratio $\gamma$ between observation and background error standard deviations $\sigma_o / \sigma_b$. (a) case where $\sigma_b$ is the same whatever $\gamma$ is (i.e. $\sigma_b = 1$); (b) case where $\sigma_b$ increases when $\gamma$ is larger (i.e. by specifying $\sigma_b = \sqrt{\sigma_o / (1 + \sigma_o^2)}$) in this example.

The thick full line in Fig. 11b corresponds to the standard deviation of the current observation error contribution for an optimal analysis in which $\sigma_b$ tends to increase when the ratio $\gamma = \sigma_o / \sigma_b$ gets larger. In this example, $\sigma_b$ has been specified to be equal to $\sigma_b = \sqrt{\sigma_o / (1 + \sigma_o^2)}$, in order to obtain significant variations of $\sigma_b$ as functions of $\gamma$, as shown by the thick dashed line. It is interesting to notice that, despite these variations of $\sigma_b$, the shape of the observation error contribution $\sigma(k, e'_o)$ (thick full line) remains nearly the same as in Fig. 11a: $\sigma(k, e'_o)$ remains maximum when the observation and background qualities are roughly similar, although the maximum is now
reached for $\gamma$ close to 1.4, that is, when observations are 40% less accurate than the background; $\alpha(k, \epsilon^o)$ still tends to zero both for very accurate and very inaccurate observations.

It is also possible to consider the case where the analysis is suboptimal, namely $\frac{\epsilon^o}{\gamma} \neq 1$. Comparing eqs. (A2) and (A3) indicates that the relative observation error contribution is almost the same as in the optimal case, except that this is further modulated by the ratio $\frac{\epsilon^o}{\gamma}$.

This means that, in the case where specified observation errors are underestimated (i.e. $\frac{\epsilon^o}{\gamma} > 1$), the relative observation error contribution will be amplified in proportion. Conversely, in the case where specified observation errors are overestimated (i.e. $\frac{\epsilon^o}{\gamma} < 1$), the relative observation error contribution will be attenuated in proportion.

References

Anthes, R. A. 1986. Chapter 27: The general question of predictability. In: Meso-scale Meteorology and Forecasting (ed. P. S. Ray). The American Meteorological Society, Boston, pp. 636–656.

Anthes, R. A., Kuo, Y.-H., Baumhefner, D. P., Errico, R. M. and Betgge, T. W. 1985. Predictability of mesoscale atmospheric motions. Adv. Geophys. 28, 159–202.

Berre, L. and Desroziers, G. 2010. Filtering of background error variances and correlations by local spatial averaging. Mon. Weather Rev. 138, 3693–3720.

Berre, L., Desroziers, G., Raynaud, L., Montrot, R. and Gibier, F. 2009. Consistent operational ensemble variational assimilation. In: Proceedings of the Fifth WMO International Symposium on Data Assimilation, Melbourne, Australia, 5–9 October 2009, Paper N.196.

Berre, L., Pannekoucke, O., Desroziers, G., Stefanescu, S. E., Chapnik, B. and co-authors. 2007. A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. In: Proceedings of the ECMWF Workshop on Flow-Dependent Aspects of Data Assimilation, 11–13 June 2007, pp. 151–168.

Berre, L., Stefanescu, S. E. and Belo Pereira, M. 2006. The representation of the analysis effect in three error simulation techniques. Tellus A 58, 196–209.

Bubnova, R., Hello, G., Bnard, P. and Geleyn, J.-F. 1995. Integration of the fully-elastic equations cast in the hydrostatic pressure terrain-following coordinate in the framework of the ARPEGE/ALADIN NWP system. Mon. Weather Rev. 123, 515–535.

Caron, J.-F. 2013. Mismatching perturbations at the lateral boundaries in limited-area ensemble forecasting: a case study. Mon. Weather Rev. 141(1), 356–374.

Cordoneanu, E. and Geleyn, J. F. 1998. Application to local circulations above the Carpathian-Black Sea area of a NWP-type meso-scale model. Contr. Atmos. Phys. 71, 191–212.

Daley, R. 1991. Atmospheric Data Analysis. Cambridge University Press, Cambridge, UK, 457 pp.

Davies, H. C. 1976. A lateral boundary formulation for multi-level prediction models. Q. J. Roy. Meteorol. Soc. 102, 405–418.

Davies, H. C. 1983. Limitations of some common lateral boundary schemes used in regional NWP models. Mon. Weather Rev. 111, 1002–1012.

Davies, H. C. and Turner, R. E. 1977. Updating prediction models by dynamical relaxation: an examination of the technique. Q. J. Roy. Meteorol. Soc. 103, 225–245.

Du, J. and Tracton, M. S. 1999. Impact of lateral boundary conditions on regional-model ensemble prediction. In: Research Activities in Atmospheric and Oceanic Modeling (ed. H. Ritchie). Report 28, CAS/JSC Working Group Numerical Experimentation (WGNE), WMO/TD-No. 942, Rep. 28, 6.7–68.

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

Fischer, C. and Auger, L. 2011. Some experimental lessons on digital filtering in the ALADIN-France 3DVAR based on near-ground examination. Mon. Weather Rev. 139, 774–785.

Fischer, C., Montmerle, T., Berre, L., Auger, L. and Stefanescu, S. 2005. An overview of the variational assimilation in the ALADIN/FRANCE NWP system. Q. J. Roy. Meteorol. Soc. 131, 3477–3492.

Fisher, M. 2003. Background error covariance modelling. In: Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, Shinfield Park, Reading, UK, 8–12 September 2003, pp. 45–64.

Gebhardt, C., Theis, S. E., Paulat, M. and Ben Bouallégue, Z. 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmos. Res. 100, 168–177.

Horanyi, A., Ihasz, I. and Radnoti, G. 1996. ARPEGE/ALADIN: a numerical weather prediction model for Central-Europe with the participation of the Hungarian Meteorological Service. Idojarasz. 100, 277–301.

Hou, D., Kalnay, E. and Drogemeier, K. K. 2001. Objective verification of the SAMEX ‘98 ensemble forecasts. Mon. Weather Rev. 129, 73–91.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. Mon. Weather Rev. 124, 1225–1242.

Monteiro, M. and Berre, L. 2010. A diagnostic study of time variations of regionally averaged background error covariances. J. Geophys. Res. 115, D23203. DOI: http://dx.doi.org/10.1029/2010JD014095

Raynaud, L., Berre, L. and Desroziers, G. 2008. Spatial averaging of ensemble-based background error variances. Q. J. Roy. Meteorol. Soc. 134, 1003–1014.

Raynaud, L., Berre, L. and Desroziers, G. 2012. Accounting for model error in the Meteo-France ensemble data assimilation system. Q. J. Roy. Meteorol. Soc. 138, 249–262.

Saito, K., Seko, H., Kunii, M. and Miyoshi, T. 2012. Effect of lateral boundary perturbations on the breeding method and the local ensemble transform Kalman filter for mesoscale ensemble prediction. Tellus A 64, 11594. DOI: http://dx.doi.org/10.3402/tellusa.v64i0.11594
Storto, A. and Randriamampianina, R. 2010. Ensemble variational assimilation for the representation of background error covariances in a high-latitude regional model. *J. Geophys. Res.* **115**, D17204. DOI: http://dx.doi.org/10.1029/2009JD013111

Torn, R. D., Hakim, G. J. and Snyder, C. 2006. Boundary conditions for limited-area ensemble Kalman filters. *Mon. Weather Rev.* **134**, 2490–2502.

Vukicevic, T. and Errico, R. M. 1990. The influence of artificial and physical factors upon predictability estimates using a complex limited area model. *Mon. Weather Rev.* **118**, 1460–1482.

Zhang, F., Meng, Z. and Aksoy, A. 2006. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part I: perfect model experiments. *Mon. Weather Rev.* **134**, 722–736.