

A 3D ensemble variational data assimilation scheme for the limited-area AROME model: Formulation and preliminary results

Thibaut Montmerle^{id} | Yann Michel | Etienne Arbogast | Benjamin Ménétrier | Pierre Brousseau

Centre National de Recherches Météorologiques (CNRM), Toulouse, France

Correspondence

Thibaut Montmerle, Météo-France/CNRM/GMAP,
42 av. G. Coriolis, 31057 Toulouse, France.
Email: thibaut.montmerle@meteo.fr

This paper presents the formulation and preliminary results of a 3D ensemble-variational data assimilation algorithm (3DEnVar) for the AROME-France model at 3.8 km horizontal resolution. This algorithm is a deterministic, variational data assimilation scheme that uses background-error covariances sampled from an ensemble. Our ensemble is an ensemble of data assimilation (EDA) at convective scale, based on the same system with the same spatial resolutions.

In ensemble schemes, localization of the covariances is necessary to filter sampling noise. Two different localization schemes have been implemented, one in spectral space and one in grid-point space. We also evaluate hybrid formulations, where the background-error covariances are a weighted linear combination of the sampled covariances with the climatological ones.

Cycled experiments are performed over a five-week time period with 3 h updates. The 3DEnVar scheme largely outperforms standard 3D-Var in terms of forecast scores. The best experiment is the one with the grid-point localization scheme. A diagnostic of objective localization can provide guidance about the horizontal and vertical localization lengths that give best performance.

The hybrid configuration with 80% of ensemble covariances and 20% of climatological ones performs also significantly better than the 3D-Var, but to a lesser extent than the best 3DEnVar configuration, although it has better balanced initial fields.

KEYWORDS

convective scale NWP, covariance localization, variational ensemble assimilation

1 | INTRODUCTION

Since 2008, the AROME-France Numerical Weather Prediction (NWP) system (Seity *et al.*, 2011) has provided forecasts at convective scale over a domain that encompasses France and a large part of its neighbouring countries. Its spatial resolution (starting at 2.5 km and now moving to 1.3 km) allows realistic representations of high-impact weather, which adds real value to what global models with parametrized physics can represent. As demonstrated by Ménétrier *et al.* (2014), forecast errors at those scales are strongly flow-dependent. For instance, the variances are characterized by strong spatial variations and the local horizontal correlations by strong anisotropies. For some variables, especially for vorticity, divergence and specific humidity in cloudy

conditions, these errors can also clearly exhibit non-Gaussian behaviours (Legrand *et al.*, 2015).

The three-dimensional variational (3D-Var) data assimilation (DA) scheme of AROME (Application of Research to Operations at Mesoscale) has been improved over time. The calibration of the background-error covariance matrix is now performed from an Ensemble Data Assimilation (EDA; Brousseau *et al.*, 2011). The cycling strategy has been reduced from 3 h to 1 h (Brousseau *et al.*, 2016). However, the modelled background-error covariance matrix $\bar{\mathbf{B}}$ is still climatological and is written in a diagonal-spectral formulation (Berre, 2000) which implies horizontally homogeneous and isotropic background-error covariances. Some time-dependencies can be obtained by frequently upgrading the different coefficients implied in the covariance

formulation, providing that enough background perturbations are available (e.g. daily upgrade in Brousseau *et al.* (2012)). The technique of geographical masks can also help to deal with dependencies on different meteorological phenomena such as rain (Montmerle and Berre, 2010) or fog (Ménétrier and Montmerle, 2011). However, the common conclusion of the latter papers is that this formulation should evolve to account for more flow dependency of forecast errors.

The EnVar approach that introduces ensembles in variational DA has received considerable attention in the last few years. As explained by (e.g.) Hamill and Snyder (2000), Lorenc (2003), and Buehner (2005), the three-dimensional (3D) version consists of replacing $\bar{\mathbf{B}}$ by background-error covariances approximated from an ensemble of forecasts obtained by (e.g.) initial condition perturbations retrieved from background and observation perturbations. Such flow-dependent covariances are then used in a variational context to produce a deterministic analysis. As in Ensemble Kalman Filters (EnKF; Evensen, 1994) and its variants, these ensemble covariances must however be localized in order to reduce spurious noisy correlations induced by sampling noise. However EnVar applies localization exclusively in model space, while sequential EnKF requires both localization in observation space and localization between observation and model spaces (Houtekamer and Mitchell, 2001), and other square-root filters such as ETKF perform localization exclusively in observation space (review by Houtekamer and Zhang, 2016). No adaptive localization scheme, depending either on the observation types or on the variable that is observed (e.g. satellite radiances as in Zhen and Zhang, 2014) is thus needed.

The hybrid EnVar uses a weighted combination of the ensemble and climatological background-error covariances. This can be done directly as in the original study of Hamill and Snyder (2000), or through extension of the control variable by a part associated with the localized ensemble covariances (Lorenc, 2003; Buehner, 2005). As demonstrated by Wang *et al.* (2007), the presented methodologies are mathematically equivalent in 3D. For limited-area models (LAMs), several studies using hybrids have been published, but the vast majority have very coarse horizontal resolutions for the analysis increments (e.g. 90 km in Zhang and Zhang, 2012, 40 km in Poterjoy and Zhang, 2014). In the framework of the HIRLAM NWP system, Gustafsson *et al.* (2014) show clear improvements of 3D hybrids against 3D-Var with a 11 km resolution. The hybrid 3DEnVar used by the NOAA operational Rapid Refresh system (Benjamin *et al.*, 2016) still makes use of the GFS ensemble prediction system, which is based on an EnKF at global scale (Wu *et al.*, 2017).

There are various four-dimensional extensions to the 3DEnVar scheme. For instance, Ito *et al.* (2016) compare a hybrid EnKF/4D-Var with the operational 4D-Var at the Japan Meteorological Agency with a 5 km resolution. They show that such an approach significantly improves tropical cyclone and local heavy rainfall prediction. The hybrid EnKF-4D-Var still

makes use of the tangent linear and adjoint of the model to propagate the climatological error covariances as in Lorenc *et al.* (2014). This is different from the 4D extension of EnVar, where a localized linear combination of the ensemble trajectories can be used to model the time evolution of errors within each assimilation window (Liu *et al.*, 2008). This approach seems particularly interesting for applications at convective scale, as it allows avoidance of the linearization of strongly nonlinear processes such as those of the microphysical parametrizations, which is the main reason why running a 4D-Var still is a difficult task at those resolutions. Such a 4DEnVar has been successfully tested in a LAM context by Gustafsson and Bojarova (2014) and is applied operationally at Environment Canada (Caron *et al.* (2015)), however both systems produce analysis increments with rather coarse spatial resolutions over 30 km.

As a first step towards future developments in 4D at convective scale, this paper deals with 3D aspects, with focus on background-error covariance modelling and covariance filtering. A square-root free approach is taken in the minimization algorithm (e.g. Gürol *et al.*, 2014; Desroziers *et al.*, 2014). This approach allows us to express hybrid background-error covariance matrices in a straightforward way. Some specific developments are special to the limited-area geometry. In particular, two different localization schemes have been developed: the first one relies on Fourier expansion and shares many similarities with the correlation model within the climatological background-error covariance matrix (Berre, 2000). The second one implements recursive filters (Purser *et al.*, 2003; Michel and Auligné, 2010), together with spatial deformation in the vertical (Michel, 2013a). The localization lengths in the horizontal and in the vertical have been objectively diagnosed directly from the ensemble, following the methodology and algorithms detailed by Ménétrier *et al.* (2015a; 2015b). Finally, the ensemble perturbations are drawn from a variational EDA at the same spatial resolutions as the deterministic analysis, while other applications at regional scale mostly make use of ensembles with coarser resolutions, as evoked above. Inconsistencies between vertical grids, which can imply specific tuning coefficients of the hybrid depending on the vertical level as in e.g. Wu *et al.* (2017), and potential misrepresentation of the forecast errors of the deterministic model by the ensemble, are thus avoided. In addition, the EDA formalism is preferred as being more consistent with the deterministic DA system, in agreement with the results obtained at global scale by Bowler *et al.* (2017).

One limitation of this study is that, in order to find a compromise between computational cost and our wish to represent most of AROME's forecast errors, all experiments described in this paper consider a reduced horizontal resolution of 3.8 km, whereas the operational AROME-France suite runs nowadays with a 1.3 km horizontal mesh size. At that resolution, many convective systems still are realistically represented, even when using a hydrostatic dynamical model

core rather than a non-hydrostatic one. However, the same 90 levels of the hybrid pressure-based and terrain-following vertical coordinate are kept and the total domain encompasses the operational one. No forecast scores against the operational suite are performed at that point, since this paper rather aims at presenting the main algorithmic components (as well as some objective diagnostics) that are used in the AROME EnVar. However, at 3.8 km the dispersion of the ensemble has been found very close to different diagnostics that Ménétrier *et al.* (2014) have shown with an EDA AROME at 2.5 km, which was the former operational resolution for AROME-France. Even if the EDA is run with the hydrostatic approximation, the forecast errors which are estimated from it have been found at least sufficiently representative of the deterministic model at 2.5 km.

The structure of this paper is as follows: section 2 presents the formulations of the different DA methods, as well as the EDA that has been specifically set up to provide forecast perturbations. Section 3 firstly discusses some diagnostics deduced from the EDA, then the impacts on analyses and on forecast scores over a five-week time period are studied for configurations based on the use of the same homogeneous localization lengths, including hybrids. Section 4 focuses on extra EnVar experiments to evaluate the sensitivity of our results to the localization lengths. Conclusions and perspectives are finally given in section 5.

2 | DATA ASSIMILATION METHODS

This section describes the various DA methods considered in this study, as well as the localization operators used in the EnVar.

2.1 | Incremental 3D-Var

2.1.1 | General formulation

The incremental formulation of three-dimensional variational data assimilation schemes (3D-Var) is widely known but is presented here in order to set up the different notations that will be used throughout the paper.

The 3D-Var aims at retrieving the optimal analysis $\mathbf{x}^a \in \mathbb{R}^n$ by minimizing a cost function \mathcal{J} , given the background state $\mathbf{x}^b \in \mathbb{R}^n$ and observations concatenated in the vector $\mathbf{y}^o \in \mathbb{R}^p$:

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}^b - \mathbf{x})^T \mathbf{B}^{-1} (\mathbf{x}^b - \mathbf{x}) + \frac{1}{2} \{\mathbf{y}^o - \mathcal{H}(\mathbf{x})\}^T \mathbf{R}^{-1} \{\mathbf{y}^o - \mathcal{H}(\mathbf{x})\}, \quad (1)$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ and $\mathbf{R} \in \mathbb{R}^{p \times p}$ represent respectively the background- and the observation-error covariance matrices, and where \mathcal{H} gathers all nonlinear observation operators that allow us to compute simulated observations from the model state \mathbf{x} . The nonlinearities in \mathcal{H} (and of the nonlinear model in a 4D-Var) implies the non-quadraticity of \mathcal{J} . Standard methods to solve this problem are used in the NWP community, such as the Gauss–Newton approach Courtier *et al.* (1994) that makes successive linearizations of the problem.

The first linearization of the observation operator \mathcal{H} is around the background \mathbf{x}^b :

$$\mathcal{H}(\mathbf{x}^b + \delta\mathbf{x}) \approx \mathcal{H}(\mathbf{x}^b) + \mathbf{H}\delta\mathbf{x},$$

where $\mathbf{H} \in \mathbb{R}^{n \times p}$ and $\delta\mathbf{x}$ respectively are the observation operator linearized around \mathbf{x}^b and the increment. In AROME, the nonlinear trajectory is not updated, resulting in the minimization of the quadratic cost function:

$$J = \frac{1}{2} \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} + \frac{1}{2} (\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}\delta\mathbf{x}), \quad (2)$$

where $\mathbf{d} = \mathbf{y}^o - \mathcal{H}[\mathbf{x}^b]$ is the innovation vector. At optimality, the gradient vanishes and thus the following equation has to be solved:

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta\mathbf{x} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}. \quad (3)$$

This equation is typically solved by iterative Krylov or quasi-Newton algorithms with preconditioning.

2.1.2 | DPCG with a B-preconditionning

In AROME-France as well as in several other NWP systems (Bannister (2008)), a change of control variable is performed in Equation 2 by setting

$$\delta\mathbf{x} = \mathbf{B}^{1/2} \chi. \quad (4)$$

This allows us to greatly simplify the background term of the cost function (since \mathbf{B}^{-1} vanishes from Equation 2 under certain conditions; Ménétrier and Auligné, 2015b) and to better precondition the Hessian of the cost function, hence improving the convergence of the minimization problem. Other options can be taken. In particular, the Double Preconditioned Conjugate Gradient (DPCG) algorithm with a B-preconditioning of Derber and Rosati (1989) imposes

$$\delta\mathbf{x} = \mathbf{B}\delta\bar{\mathbf{x}}. \quad (5)$$

This algorithm may be interesting because it manipulate vectors of the same size as the model state $\delta\mathbf{x}$. This means that, contrary to the square-root approach (Equation 4), the vectors stored in the minimization scheme will have a size and thus a memory footprint that is independent of the size of the ensemble, or of whether a pure or hybrid scheme is employed. Introducing Equation 5 into Equation 3, we get:

$$(\mathbf{I} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta\bar{\mathbf{x}} = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d}. \quad (6)$$

The DPCG algorithm needs in particular to compute the following vector for each i th iteration of the inner loop :

$$\mathbf{h}_i = \mathbf{B}\mathbf{g}_i, \quad (7)$$

where \mathbf{h}_i and \mathbf{g}_i are vectors $\in \mathbb{R}^n$. Note that $\mathbf{B}^{1/2}$ and $\mathbf{B}^{1/2}$ can be successively used instead of applying a full \mathbf{B} to the control vector, ensuring its positive semi-definiteness.

2.2 | An EDA at convective scale

The main objective of an EDA (Fisher, 2003) is to approximate the true error covariances following a Monte Carlo

approach (Belo-Pereira and Berre, 2006). For this purpose, observations, boundary conditions and the forecast model are perturbed, giving in return the estimated perturbations of the analysis and of the background. In our case, the EDA consists in 25 cycled forecast/3D-Var minimizations at a horizontal resolution of 3.8 km. Each 3D-Var, which makes use of a climatological \mathbf{B} that has been specifically calibrated for this resolution (Brousseau *et al.*, 2016), aims at assimilating explicitly perturbed observations that are constructed as observations plus random errors drawn from their specified error covariances. The perturbed lateral boundary conditions (LBCs) are brought by the global EDA that is run operationally at Météo-France (Raynaud *et al.*, 2011; Berre *et al.*, 2015). For the surface errors, the sea surface temperature is explicitly perturbed accordingly to the errors provided by the OSTIA¹ analyses performed by the National Centre for Ocean Forecasting. For model error, a simple multiplicative inflation scheme is used, following Raynaud and Bouttier (2015). The inflation factor is derived online from the skill over spread ratio, and is applied every 3 h.

Following what is done in the operational AROME-France NWP system, this EDA is first used to calibrate the climatological background-error matrix ($\bar{\mathbf{B}}$ in the following, as in e.g. Brousseau *et al.*, 2011). More precisely, $\bar{\mathbf{B}}^{1/2}$ is modelled using a sequence of linear operators modelling the variances, the correlations and the cross-covariances between the variables. The cross-covariances follow the formulation of Berre (2000), which is an adaptation of Derber and Bouttier (1999) for a LAM. They include an extended multivariate balance relationship for specific humidity. The control variables are the vorticity ζ , the unbalanced divergence η_u , the unbalanced temperature and pressure at surface level $(T, P_s)_u$ and the unbalanced specific humidity q_u . In $\bar{\mathbf{B}}$, the variances and the correlation remain static (they do not vary with time), homogeneous and isotropic (they do not vary in the horizontal). The cross-covariances between the variables are also static, homogeneous and isotropic, which may weaken the convective coupling frequently observed in precipitating areas (Montmerle and Berre, 2010). The 3D-Var scheme simply uses $\bar{\mathbf{B}}$ instead of \mathbf{B} throughout the equations listed in the previous section.

Secondly, the EDA can also be used to compute the ensemble background-error covariances of the day in the EnVar. As evoked above, a perturbation ϵ_p^b is defined in this case as the difference between one forecast and the ensemble mean. The differences between two forecasts are used instead in the calibration of $\bar{\mathbf{B}}^{1/2}$. However, as discussed in Fisher (2003), the distribution of differences between such pairs of background fields has comparable statistical characteristics to those obtained with the former approach, with comparable correlation structures but twice the variance. These

differences have an implication when choosing normalization factors for the background-error variances applied in the different experiments, as discussed in section 3.2.

2.3 | 3DEnVar

As presented for instance by Lorenc (2003) or Buehner (2005), the basic principle of the 3DEnVar is to replace the climatological modelled $\bar{\mathbf{B}}$ used in the 3D-Var by the background-error covariances $\tilde{\mathbf{B}}$ sampled from an ensemble of N_e members, with:

$$\tilde{\mathbf{B}} = \frac{1}{N_e - 1} \sum_{l=1}^{N_e} (\tilde{\mathbf{x}}_l^b - \langle \tilde{\mathbf{x}}^b \rangle) (\tilde{\mathbf{x}}_l^b - \langle \tilde{\mathbf{x}}^b \rangle)^T, \quad (8)$$

where $\tilde{\mathbf{x}}_l^b$ is the l th background state and $\langle \tilde{\mathbf{x}}^b \rangle$ the mean of the ensemble

$$\langle \tilde{\mathbf{x}}^b \rangle = \frac{1}{N_e} \sum_{l=1}^{N_e} \tilde{\mathbf{x}}_l^b.$$

However, for two main reasons, $\tilde{\mathbf{B}}$ cannot be directly used in the variational system: it has a low rank limited by the size of the ensemble and it is strongly affected by sampling noise. As proposed by Houtekamer and Mitchell (2001), a way to alleviate this problem is to perform an element-by-element product (usually known as a Hadamard–Schur product) between $\tilde{\mathbf{B}}$ and a correlation matrix of the same dimensions $\mathbf{C} \in \mathbb{R}^{n \times n}$. The 3D localized ensemble background-error covariance matrix $\tilde{\mathbf{B}}_e$ then is written as

$$\begin{aligned} \tilde{\mathbf{B}}_e &= \tilde{\mathbf{B}} \circ \mathbf{C} \\ &= \mathbf{X}^b \mathbf{X}^{b^T} \circ \mathbf{C}. \end{aligned} \quad (9)$$

Here \mathbf{X}^b gathers all perturbations $(\epsilon_1^b, \dots, \epsilon_{N_e}^b)$ defined as

$$\epsilon_l^b = \frac{1}{\sqrt{N_e - 1}} (\tilde{\mathbf{x}}_l^b - \langle \tilde{\mathbf{x}}^b \rangle).$$

The linearized cost function of Equation 2 still is used and the minimization is also based on the DPCG, but this time preconditioned by $\tilde{\mathbf{B}}_e$. Equation 7 then is written as

$$\mathbf{h}_i = \left(\sum_{l=1}^{N_e} \mathbf{D}_l \mathbf{C} \mathbf{D}_l \right) \mathbf{g}_i = \sum_{l=1}^{N_e} \epsilon_l^b \circ \{ \mathbf{C}(\epsilon_l^b \circ \mathbf{g}_i) \}. \quad (10)$$

Here \mathbf{D} is a diagonal matrix that contains the elements of the vector of the 3D perturbations ϵ_l^b along its diagonal, following Buehner (2005). The control variables gathered in \mathbf{g}_i are the two wind components U and V , the temperature T , the pressure at surface level P_s and the specific humidity q , which are also considered as prognostic variables in AROME. The choice of using (U, V) instead of (ζ, η_u) for the wind variable may be justified by the fact that these variables present much more Gaussian behaviour (Legrand *et al.*, 2015). Also, their diagnosed localization length-scales are comparable to those obtained for T and q , as discussed in section 3.2.2. Using comparable localization length-scales is indeed important because it allows us to save computational time, but also because it probably avoids some balance problems. These

¹Operational Sea Surface Temperature and Sea Ice Analysis; http://grsstpp.metoffice.com/pages/latest_analysis/ostia.html; accessed 21 September 2018

particular topics are dealt with in sections 3.2.2 and 4.2 respectively.

Replacing those dynamical variables by the velocity potential and by the stream function, as proposed by Kepert (2009), has been tried, as those variables are supposed to be more isotropic and as they display comparable background-error correlation lengths to T and q (and consequently comparable localization lengths, as demonstrated by Ménétrier *et al.*, 2015a). However this has led in our experimental framework to a clear degradation in forecast scores and no evidence of a more isotropic behaviour has been found. These variables may be more related to synoptic-scale flows and accordingly may not be adapted to convective-scale models, where strongly anisotropic dynamical structures often occur.

2.4 | Hybrid scheme

As first proposed by Hamill and Snyder (2000), hybridizing the ensemble covariance matrix $\tilde{\mathbf{B}}_e$ of Equation 9 with the modelled $\bar{\mathbf{B}}$ allows us to overcome some of their respective shortcomings. The resulting hybrid matrix is expressed as a linear combination of the two matrices, respectively weighted by two positive coefficients β_e and β_c :

$$\mathbf{B}_h = \beta_e \tilde{\mathbf{B}}_e + \beta_c \bar{\mathbf{B}}. \quad (11)$$

Usually, $\beta_e + \beta_c = 1$, but different configurations may be chosen in practice as discussed in section 3.5.3. One of the advantage of the B-preconditioning is that \mathbf{B}_h can be used as such in the DPCG algorithm, Equation 7 straightforwardly becoming

$$\mathbf{h}_i = \mathbf{B}_h \mathbf{g}_i = \beta_e \sum_{l=1}^{N_e} \epsilon_l^b \circ \{ \mathbf{C}(\epsilon_l^b \circ \mathbf{g}_i) \} + \beta_c \bar{\mathbf{B}}^{1/2} \bar{\mathbf{B}}^{T/2} \mathbf{g}_i. \quad (12)$$

Here, the size of the control variable always equals the size of the state vector. In contrast, the square-root formulation of Lorenc (2003) and Buehner (2005) increases the size of the control variable with vectors of the size of N_e times the resolution of the localization correlations, possibly exceeding the size of the state vector. The use of the B-preconditioning easily allows to keep U , V , T , P_s and q as control variables for the ensemble part while using the same modelled $\bar{\mathbf{B}}$ as for the 3D-Var.

2.5 | Localization

Two variants of the localization \mathbf{C} have been developed, based either on a spectral or a grid-point formulation. In order to preserve values of the sampled background-error variances, it has been carefully checked that the resulting localization matrices are true correlation matrices, i.e. that their diagonal coefficients are strictly equal to one.

2.5.1 | Spatial localization with spectral formulation

For a LAM, and similarly to Gustafsson *et al.* (2014) and Gustafsson and Bojarova (2014), we use the inverse

bi-Fourier transform \mathbf{S}^{-1} , such as:

$$\mathbf{C} = \mathbf{S}^{-1} \mathbf{C}_s \mathbf{S}^{-T}. \quad (13)$$

The grid-point increments are converted to spectral space, then multiplied by the localization matrix in spectral space \mathbf{C}_s , before being converted back to grid point. \mathbf{C}_s is expressed as the sequence of operators

$$\mathbf{C}_s = \mathbf{L}_{h,s}^{1/2} \mathbf{L}_{v,s} \mathbf{L}_{h,s}^{T/2}, \quad (14)$$

where $\mathbf{L}_{h,s}$ is the horizontal part, consisting in a multiplication with the square-root of the power spectrum of a localization function, and $\mathbf{L}_{v,s}$ is the vertical correlation matrix.

The horizontal localization is based on the fifth-order piecewise rational function from Gaspari and Cohn (1999). This localization has a scale parameter c , and is compactly supported at the $2c$ separation distance. Considering a positive-definite and twice differentiable correlation ρ , the Daley length-scale is derived as (Daley, 1991)

$$\mathcal{L} = 1/\sqrt{-\rho''(0)}, \quad (15)$$

where $\rho''(0)$ is the local second derivative of ρ . The link between the length-scale \mathcal{L} and the scale parameter c is given by $c = \sqrt{10/3}\mathcal{L}$. Note that the rest of the paper uses \mathcal{L} as the definition of the length-scale.

As in Buehner (2005), the vertical localization function is defined with respect to a pressure scale height and is also based on the Gaspari and Cohn (1999) function. $\mathbf{L}_{v,s}$ is expressed with a truncated decomposition of empirical orthogonal functions (EOFs) that consider only a third of its eigenpairs. For the homogeneous vertical localization applied in section 3, it allows us to keep more than 96% of the total variance while saving computational time.

To diagnose the spatial structure of \mathbf{C} , we extract one of its columns by multiplying \mathbf{C} with a vector filled with zeros except for one grid point. Figure 1a shows a vertical cross-section of the resulting increment. Note that using the bi-Fourier decomposition in the horizontal implies the occurrence of residual small negative lobes at long separation distance. When the localization function is applied near a boundary of the domain, bi-periodicity of the fields yields also to unwanted positive values on the opposite side. To avoid the appearance of any residual increments, the extension zone can be expanded (as in Gustafsson *et al.*, 2014) or observations located within a $2c$ distance from the boundaries can be simply excluded from the assimilation process. However, in order to perform fair comparisons with other experiments, all observations have been kept in this study.

2.5.2 | Spatial localization

An alternative approach, based on the application of the localization exclusively in grid point space, has also been developed. In this case, the square root of \mathbf{C} is written as

$$\mathbf{C}^{1/2} = \mathbf{L}_{h,g} \mathbf{D}_{v,g} \mathbf{L}_{v,g}. \quad (16)$$

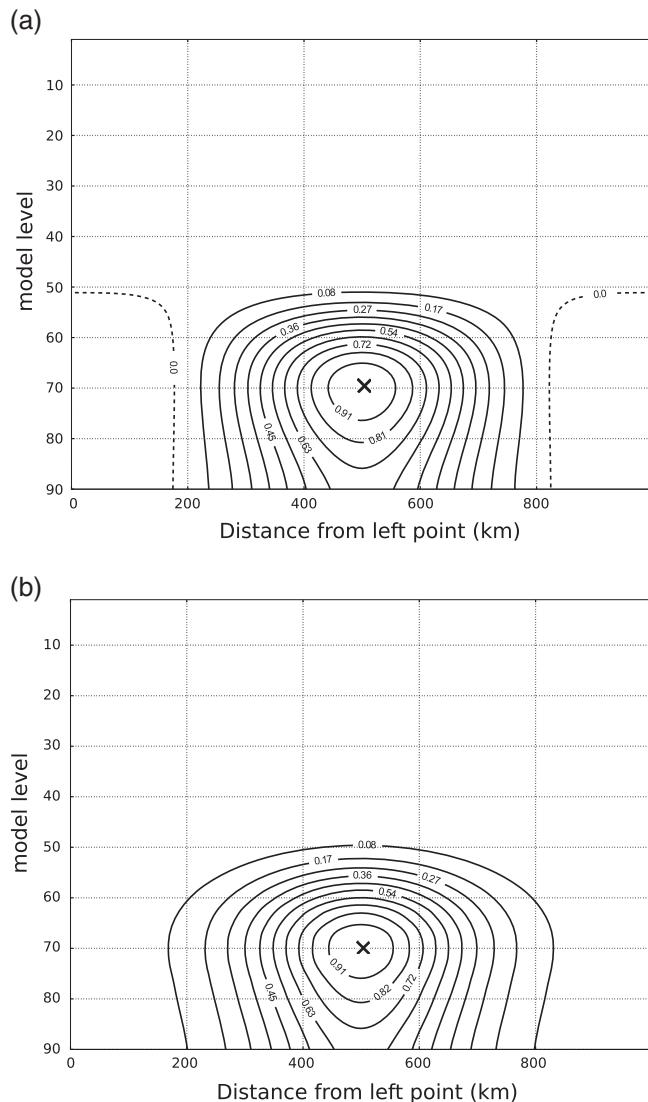


FIGURE 1 Vertical cross-sections of a localization matrix \mathbf{C} column, computed by multiplying \mathbf{C} by a vector which is null except at the grid point indicated by a cross. Localization functions are defined (a) in spectral and (b) in spatial spaces (see text for details), with localization lengths of 170 km in the horizontal, and of $0.3 \log(P)$ in the vertical

The operators $\mathbf{L}_{v,g}$ and $\mathbf{L}_{h,g}$ are the homogeneous recursive filters from Purser *et al.* (2003), which have been proven to be an efficient technique in grid-point correlation modelling. The operator $\mathbf{D}_{v,g}$ is a vertical grid deformation to make the correlation inhomogeneous on the AROME-France vertical levels, following Michel (2013b). The final correlation functions are quasi-Gaussians with localization length-scales defined by Equation 15.

As displayed in Figure 1, similar results are obtained when applying both localization methods. However, the recursive filters allow us to avoid the spurious negative lobes at long separation distances. They also avoid the increments wrapping around through periodicity, as when using the bi-Fourier decomposition.

3 | NWP TRIALS

In this section, the impacts of the different DA methods described in the previous section are studied in a NWP context. Only configurations that use homogeneous localizations with objective length-scales are tested here. Additional experiments are discussed in section 4, where other values of localization lengths are addressed.

3.1 | Meteorological conditions

All trials have been performed over a 33-day time period, from 6 February to 10 March 2016. This time period was firstly characterized by a succession of storms over Western Europe, where typical deep lows, formed over the North Atlantic, passed within a westerly flow over the UK and Northern Europe. On 6, 8 and 9 February, three successive storms produced very strong winds in different parts of France, often blowing over 120 km h^{-1} in several locations and accompanied by intense rainfall and coastal water surges. Between 9 and 10 February, a small but intense low developed over the Gulf of Genoa in the Mediterranean Sea, producing wind bursts over 160 km h^{-1} in northern Corsica. On 13 February, another storm passed over France, once again accompanied by strong wind and heavy precipitation. The following days were marked by very unstable and moist weather conditions, causing some snowfalls over Brittany around 17 February. After a week of more calm weather, snowfall occurred again over the plain in northeastern France after 24 February. Until the end of the period, very unstable conditions followed with strong winds around 1 March followed by intense convective activity leading to snowfall and showers. Finally, an active low passed over Brittany to conclude this very active end to winter 2016.

Thus this period seems to be particularly challenging for testing the impact of new DA configurations. Radar data may have a large impact as they are the only data type that directly sample the 3D structure of precipitating areas.

3.2 | Diagnostics from the EDA AROME

An EDA AROME of 25 members, with horizontal resolution of 3.8 km, has been set up and launched from 5 February 2016 and run to the end of the period of interest. It is preferable to start the day before the retrieval of the first deterministic analyses in order to stabilize the averaged background-error variances retrieved from the background perturbations, which need some assimilation cycles before reaching stable values (Ménétrier *et al.*, 2014).

3.2.1 | Background-error variances

As displayed in Figure 2 by the shaded areas, and analogously to what has been already noticed by former studies such as Brousseau *et al.* (2012), there are variations

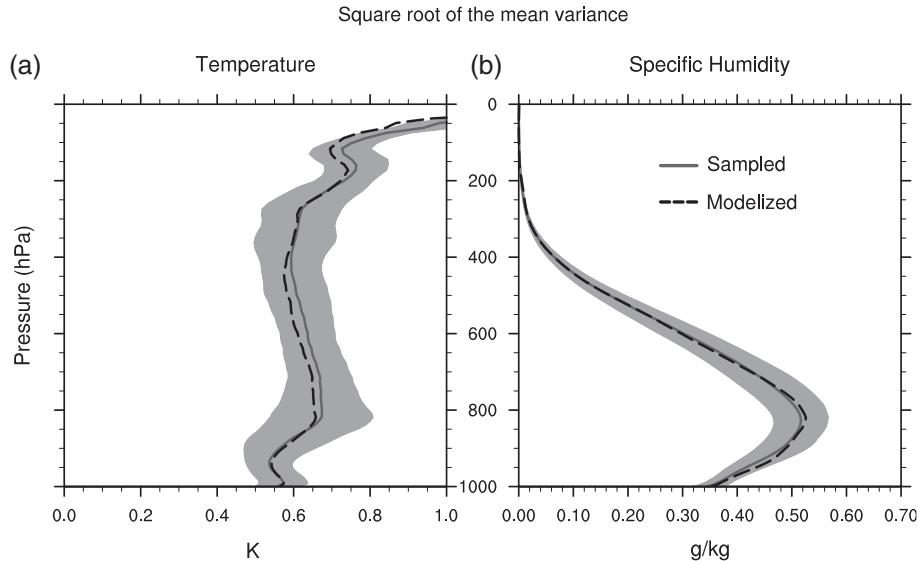


FIGURE 2 Profiles of the square root of the variances for (a) temperature (K) and (b) specific humidity (g kg^{-1}) diagnosed from 25 background perturbations provided every 6 h by the EDA AROME after 3 h of integration (solid lines), averaged from 6 to 21 February 2016. Shaded areas indicate the spread of the profiles and the dotted lines show the climatological values used in 3D-Var, which are computed considering six perturbations at 0000 and 1200 UTC over the same time period

in background-error variances retrieved by the EDA after 3 h of integration throughout the time period, reflecting the variations of weather regimes that have been sampled. For temperature, the variations are displayed for all tropospheric levels, the square root of the mean variances taking values approximately between 0.5 and 0.7 K. For specific humidity, the variations are mainly visible in the low troposphere with maximum values between 0.45 and 0.6 g kg^{-1} .

Furthermore, the EDA is used to model climatological background-error covariances $\tilde{\mathbf{B}}$ used in the 3D-Var experiment. Figure 2 also displays the total climatological variances that have been retrieved from six perturbations computed at 0000 and 1200 UTC assimilation times over the first two weeks of the time period, following the methodology described in section 2.2. Since in this particular case the later perturbations are computed by performing differences between pair of forecasts, a normalization factor of $\sqrt{2}$ is applied in order to get comparable standard deviations to those used in EnVars (discussion at the end of section 2.2). The resulting averaged values are very close to the averaged sampled ones. Similar weights will thus be given to observations in the different DA strategies discussed in the following.

3.2.2 | Localization lengths

Recently, Ménétrier *et al.* (2015a) have presented the general formulation of a methodology that allows us to objectively diagnose the localization, using only information from the ensemble. This methodology is based on both theories of the centred moments sampling and of the optimal linear filtering. As a matter of fact, Equation 9, where the correlation matrix \mathbf{C} is applied to $\tilde{\mathbf{B}}$ using the Schur product, can be seen as a linear filtering of sampled covariances.

In their companion paper, Ménétrier *et al.* (2015b) show how to derive practical algorithms that solve the problem of the determination of the optimal localization, under spatial ergodicity assumptions and considering only sampling error. In the simplest version of the algorithm, the spatial ergodicity assumption amounts to averaging over the horizontal. We obtain localizations that may vary with the variable and in the vertical, and possibly over time (although this is not considered in this study).

Figure 3 shows vertical profiles of horizontal \mathcal{L} estimated from 25 perturbations of the EDA for the first two weeks of the time period considered in this study and for the different model variables. The displayed averaged values are coherent with those shown by Ménétrier *et al.* (2015b): \mathcal{L} regularly increases with height from the ground up to the tropopause by a factor of 4, with similar values among U , V , T and q . The range of variation of \mathcal{L} over time is displayed by the shaded areas; there is more variation for T at the top of the boundary layer and at the tropopause and generally for q in the free troposphere. Above 400 hPa where very dry conditions occur for all members, the localization length-scales for q are much larger (as are the correlation length-scales) than for the other variables. We choose therefore to exclude q from the common averaged profile displayed in Figure 3e.

Analogously, a common profile of vertical localization lengths has been retrieved and is displayed in Figure 4. In contrast to what is found for the horizontal, a minimum around 0.2 units of the natural logarithm of pressure (hereafter referred as $\log(P)$ units) is visible above the boundary layer between 800 and 900 hPa, mainly because of the values retrieved for the wind components (not shown). It reflects the shorter vertical mixing at those levels, presumably linked to wind shear linked to diabatic processes. The presence of stable layers at low levels could also be an explanation.

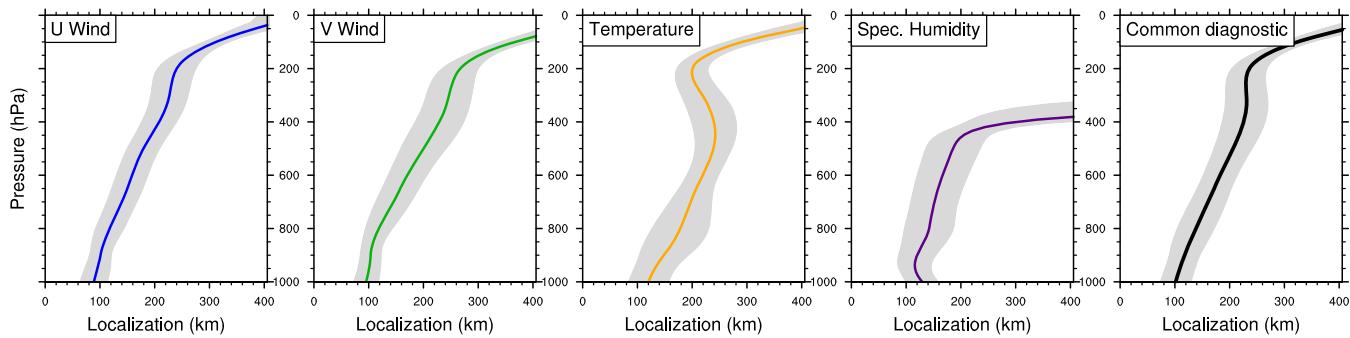


FIGURE 3 Profiles of averaged horizontal localization lengths for the different control variables (a) zonal wind, (b) meridional wind, (c) temperature and (d) specific humidity, deduced from the EDA AROME between 6 and 21 February 2016. (e) shows the common values which are further used in the EnVar (see text for details). The shaded areas illustrate the dispersion of the different values that have been diagnosed at each assimilation time from 25 perturbations [Colour figure can be viewed at wileyonlinelibrary.com]

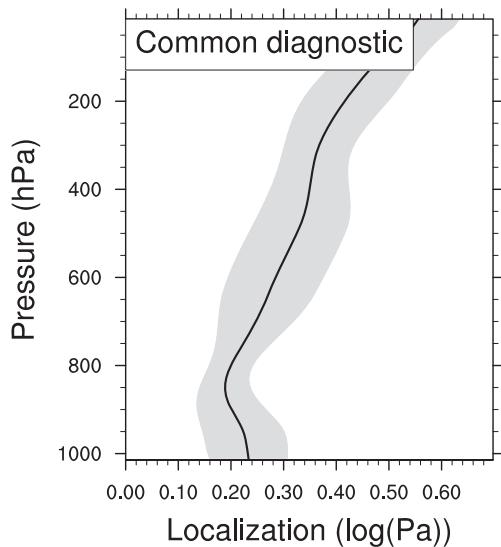


FIGURE 4 As Figure 3e, but for vertical localization lengths (in log(P))

Above 800 hPa, the common profile displays length-scales that linearly increase with height, as for the horizontal localization.

In this section, the experiments use a single pair of common horizontal and vertical localization lengths, derived from vertical averages over the tropospheric levels (i.e. below 200 hPa) of the two common diagnostics. This yields values of 170 km and 0.3 log(P) respectively. In section 4, other values have been empirically chosen in both directions in order to evaluate, in terms of forecast scores, the impact of using these objectively diagnosed length-scales.

3.3 | Experimental set-ups

Five different DA configurations, whose main features are summarized in the first four lines of Table 1, are evaluated at this stage. They gather the 3D configurations presented in section 2, and they all consider the same comprehensive set of different observation types: conventional and aircraft data, clear infrared radiances from geostationary and polar orbiting satellites (MSG/SEVIRI, NOAA and MetOp-A/B HIRS),

clear microwave radiances from AMSUA/AMSUB/MHS, Doppler winds and reflectivities from the French ARAMIS radar network.² Compared to the operational AROME suite, whose main features are described in Brousseau *et al.* (2016), only data from the ground-based Global Positioning System (GPS) stations, geowinds and infrared (IR) interferometers were not considered, mainly because the variational bias correction is not yet implemented in the EnVar. Finally, the thinning distance has been increased from 8 to 30 km for radar data, in order to avoid spurious observation-error correlations in our 3.8 km resolution configurations. Pseudo-profiles of relative humidity retrieved by the 1D Bayesian inversion of Caumont *et al.* (2010) may indeed be prone to correlations of their observation errors, as this methodology makes use of informations from the background in the vicinity of the observation locations (Wattrelot *et al.*, 2014 give more details). An example of active data coverage is given in Figure 5.

A 3 h update cycling strategy is chosen, as in the former AROME operational configuration described in Seity *et al.* (2011), and 30 h forecasts are launched for the 0000 and 1200 UTC assimilation times. The basic 3D-Var (BCLIM hereafter), which is considered as the reference in this section, uses a unique \bar{B} which has been calibrated from the EDA, as described in section 2.2. BENS-SP and BENS-GP match 3DEnVar experiments that respectively use spectral and spatial localization, following the approaches presented earlier in section 2.5, with constant localization length-scale for all variables and all vertical levels. As discussed above in section 3.2.2, the values of 170 km in the horizontal and of 0.3 hPa in the vertical have been retained.

Finally, HYB0.5 and HYB0.8 indicate Hybrid 3DEnVar with either equal 0.5/0.5 or 0.2/0.8 weights between climatological and ensemble background-error covariances. Both experiments share the same spatial localization functions as BENS-GP.

²MSG: Meteosat Second Generation. SEVIRI: Spinning Enhanced Visible and Infrared Imager. HIRS: High-resolution Infrared Radiation Sounder. AMSU: Advanced Microwave Sounding Unit. MHS: Microwave Humidity Sounder. ARAMIS: Application Radar à la Météorologie Infrasynoptique.

TABLE 1 List of the different experiments discussed in this paper, showing the DA method, the type of localization, the horizontal and the vertical localization lengths used to define the corresponding localization functions (km and $\log(P)$ respectively), and the respective weights given to the climatological and to the ensemble background-error covariances in the hybrid formulation given by Equation 11. The first set of experiments is discussed in section 3, and the second in section 4

Experiment	DA method	Localization type/	Climat/
		horizontal/ vertical lengths	ensemble weights
BCLIM	3D-Var		1/0
BENS-SP	3DEnVar	Spectral/170/0.3	0/1
BENS-GP	3DEnVar	Spatial/170/0.3	0/1
HYB0.5	Hybrid	Spatial/170/0.3	0.5/0.5
HYB0.8	Hybrid	Spatial/170/0.3	0.2/0.8
BENS-GP-0.6	3DEnVar	Spatial/170/0.6	0/1
BENS-GP-100	3DEnVar	Spatial/100/0.3	0/1
BENS-GP-350	3DEnVar	Spatial/350/0.3	0/1

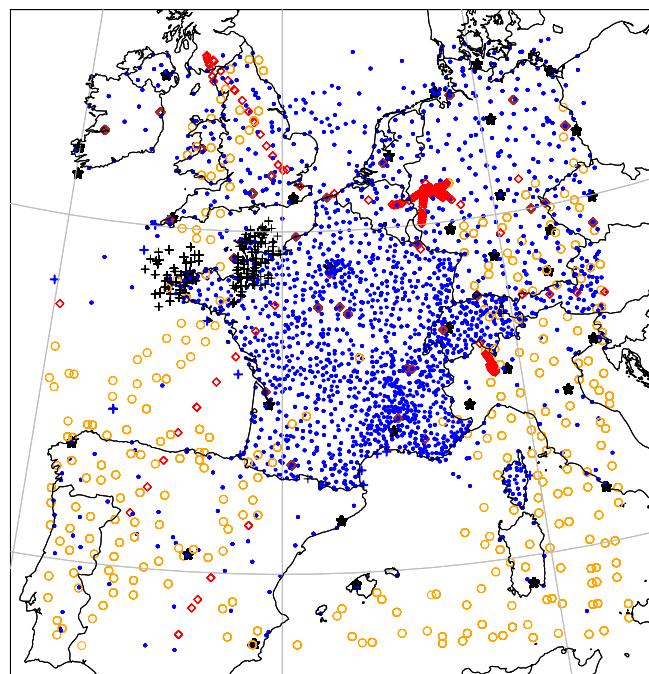


FIGURE 5 Position of the different types of observation that are effectively assimilated on 6 February 2016 at 0000 UTC in all DA experiments (dots: ground stations and buoys, diamonds: aircrafts, stars: radiosondes, circles: satellite radiances, crosses: radar data) [Colour figure can be viewed at wileyonlinelibrary.com]

3.4 | Impact on analysis increments

Horizontal cross-sections of analysis increments obtained using the same guess and the same observations but different algorithms are plotted for the first assimilation step, on 6 February 2016, for temperature around 900 hPa in Figure 6a–c. Clearly, using exclusively background covariances sampled from the EDA in BENS-SP (or very similarly in BENS-GP, not shown) allows us to get much more small-scale increments around assimilated observations than BCLIM, especially over the mainland where much more data are considered, as shown in Figure 5. While averaged

background-error variance values are close in \bar{B} and in the differently sampled \tilde{B} , as displayed in Figure 2, local values of increments can be much larger in BENS-SP. This is particularly the case in areas characterized by local variance structures of high intensities that reflect the strong uncertainties of predicting some local meteorological phenomena (not shown). Strong negative increments are for instance displayed over the Mediterranean Sea, partly because of the assimilation of low-peaking infrared SEVIRI channels, and also over Western Germany and Belgium where aircraft data are numerous. In such areas, the impact of observations is clearly increased. In addition, shorter length-scales of sampled covariances imply in most cases increment structures that are much more localized and elongated along the flow in BENS-SP than in BCLIM, e.g. for the rawinsonde launched over La Coruña in northwestern Spain. The increment structures displayed by HYB0.5 lies logically between the two latter experiments: the large-scale structures are mostly kept and some small-scale details are added where observations at low level are assimilated with high spatial densities.

However, at higher levels, satellite radiances have greater impact and increments are larger over the sea for BCLIM (Figure 6d). Dense mid-tropospheric data (e.g. radar over western France, aircraft above the UK or soundings) still produce noticeable increments for BENS-SP, but at smaller scales (Figure 6e). The smaller area of influence of some isolated observations, such as radiosondes, confirms that the horizontal background-error correlations supplied by the localized perturbations from the EDA are at much smaller scale than those modelled in \bar{B} and that the smaller scales of increments displayed by BENS-SP are not only explained by data density. When hybridizing these ensemble covariances, which are characterized by few structures and by a lack of spread at these levels (Figure 6f), the large-scale covariances brought by \bar{B} seem to dominate and the increment structures of HYB0.5 are close to those displayed by BCLIM.

In order to study which scales are analyzed by the different approaches, the spectra of the temperature increments displayed in Figure 6 are plotted in Figure 7. The visual impressions given by Figure 6 discussed above are clearly confirmed: compared to BCLIM, the localized and sampled background-error covariances used in BENS-SP (and also for BENS-GP which displays very similar results) imply increments that are characterized by larger contributions of the smallest scales. The energy of scales below 500 (200) km is indeed larger for the 900 (400) hPa vertical level. Furthermore, the gap between the two spectra regularly increases as scales reduce, especially near the surface. Conversely, scales larger than around 1,000 km contribute more in the BCLIM case, with values roughly one order of magnitude larger than in the BENS-SP case. In the HYB0.5 case, an intermediate spectra is shown, with values getting closer to that of BCLIM when height increases. At 900 (400) hPa, comparable values to BCLIM are displayed above 100 (50) km and contributions

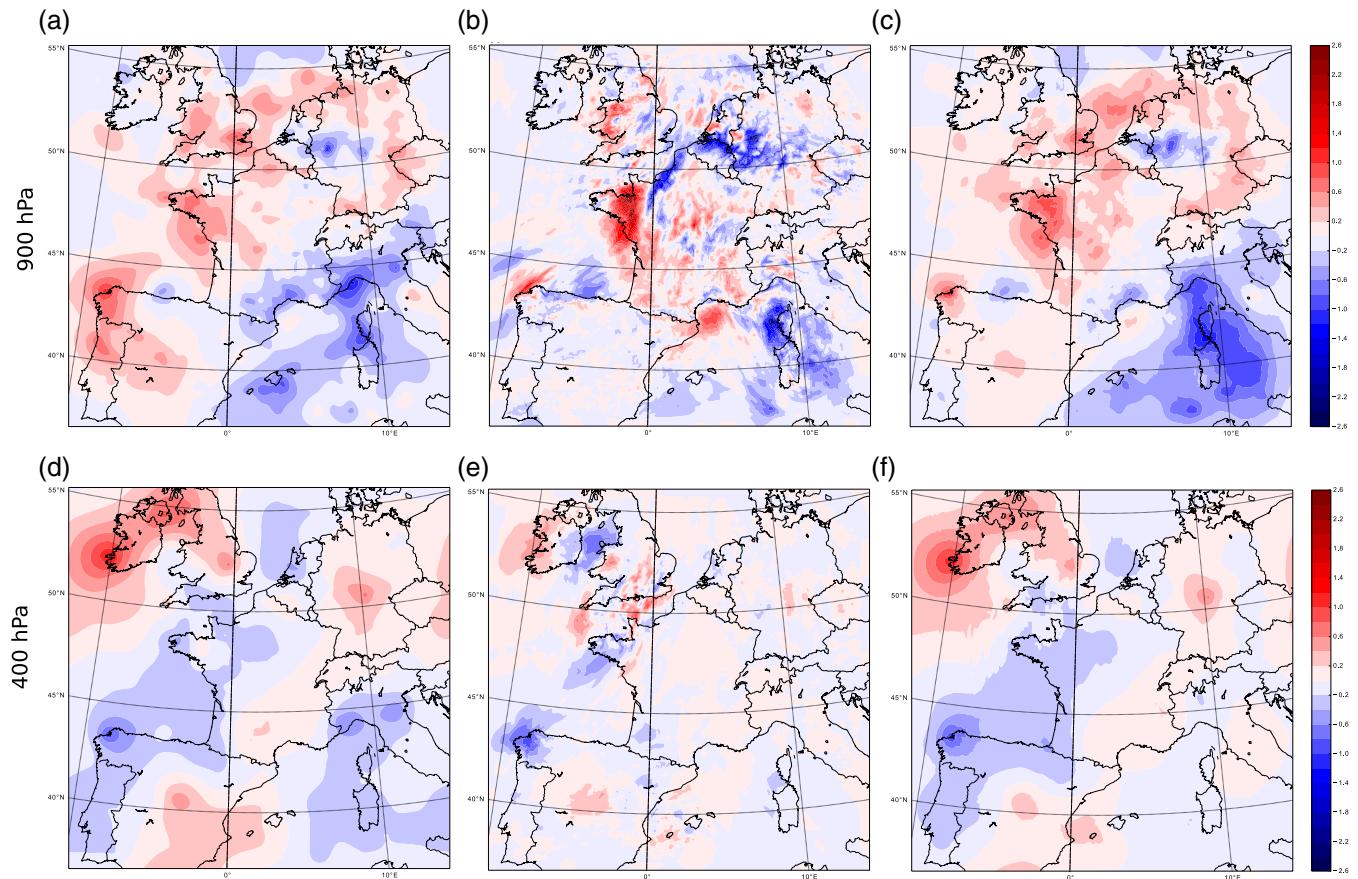


FIGURE 6 Horizontal cross-sections around (a–c) 900 hPa and (d–f) 400 hPa of temperature increments (K), computed on 06 February 2016 at 0000 UTC, for (a,d) BCLIM, (b,e) BENS-SP, (c,f) HYB0.5, as defined in Table 1. The same guess has been used for the three experiments [Colour figure can be viewed at wileyonlinelibrary.com]

of smaller scales increase below, but to a lesser extent than for BENS-SP. For the wind increments, the same tendencies are displayed for both altitudes, but the spectra of the three experiments are more comparable above 120 km (not shown). Similar spectra have been obtained for different dates, which suggests that similar shape characteristics of analysis increments are obtained by the three approaches throughout the considered time period (not shown).

3.5 | Impact on forecasts

Forecasts are evaluated using different types of scores computed against different types of observations. To ease the interpretation of the results, the scores for a particular experiment exp are most of the time given relative to the scores obtained by a reference experiment ref , such that

$$\Delta\text{Score} = (\text{Score}_{\text{exp}} - \text{Score}_{\text{ref}})/\text{Score}_{\text{ref}}. \quad (17)$$

The difference of the absolute values of the bias computed for each experiment is also preferred:

$$\Delta|\text{Bias}| = |\text{Bias}_{\text{exp}}| - |\text{Bias}_{\text{ref}}|.$$

Thus, in both cases, negative values show an improvement of the experiment compared to the reference as the better scores tend toward zero for the scores used here.

3.5.1 | Scores against conventional data

Figure 8 displays such scores against French ground-based measurements considering the BCLIM experiment as reference. This network provides about 900 observations for temperature and humidity at 2 m and 750 for the wind at 10 m every hour. The two EnVar experiments with basic localization definition BENS-SP and BENS-GP show overall close ΔRMSE results for most variables at all ranges. Significant improvement for sea level pressure particularly can be noticed in both cases, with score differences around 10% for all forecast ranges. Positive scores and improvement of the biases are also visible at all forecast ranges for wind speed and mainly for humidity. The larger gap of the initial analyzed wind against wind measurements, which is improved for the hybrids, may be due to the low rank of the ensemble covariances that implies a poorer initial fit to some observations. However, negative score differences are visible for temperature at longer forecast ranges, somewhat attenuated in the BENS-GP case. A slight increase with ranges of the $\Delta|\text{Bias}|$ is also displayed for the latter variable, but with values below 0.2 K. In contrast, results obtained by HYB0.5 are mostly neutral compared to those of BCLIM, but better than the 3DEnVar for temperature, in terms of biases and RMSE. Considering more ensemble covariances in HYB0.8 allows us to improve the scores of the latter experiment, but not

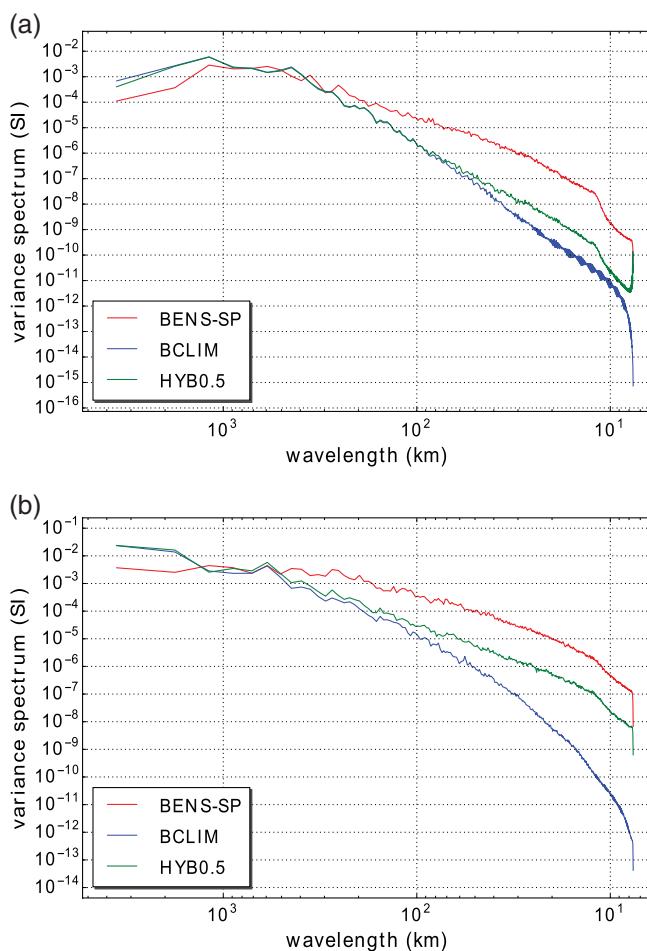


FIGURE 7 Spectrum of temperature increments at (a) 400 hPa and (b) 900 hPa, computed on 06 February 2016 at 0000 UTC, for BCLIM (blue), BENS-SP (red), HYB0.5 (green) [Colour figure can be viewed at wileyonlinelibrary.com]

to the same extent as BENS-SP and BENS-GP, except for temperature.

Scores against radiosondes, plotted in Figure 9, tend to confirm these main tendencies. For both EnVar experiments, improvements are visible up to 24 h forecasts for all variables, except for geopotential at 12 h above 500 hPa, where HYB0.8 performs better. For wind speed, improvements up to 3% are displayed regularly up to 200 hPa after 12 h. For temperature, relative improvements above 6% are mostly visible around 700 hPa, whereas for relative humidity the strong improvements up to 8% that are shown in the boundary layer decrease with height. Slightly better results are consistently displayed while applying localization in grid-point space. It can be noted that these improvements for EnVar experiments occur even if the initial fit to the observed values is degraded in the upper troposphere, especially for wind and temperature. The lack of coherent structures at those levels, in addition to the spectral relaxation applied in the top levels towards the forecasts at global scale brought by the ARPEGE EDA, may explain this degradation, as it clearly reduces the spread of the ensemble used to estimate the background covariances. The misleading averaged values of such spread in the upper

troposphere, displayed for temperature in Figure 2, are in fact mostly due to the perturbed LBCs drawn from the EDA ARPEGE (not shown). Uniformly adding some of the static climatological background-error variances in HYB0.5 logically allows improvement of this initial fit.

However HYB0.5 shows again rather neutral scores at longer ranges, which is quite disappointing given what has been obtained at global scale by e.g Kleist and Ide (2015), or for other LAMs such as the hybrid ETKF/3D-Var of Wang *et al.* (2008) or the E4DVar of Zhang and Zhang (2012). However, in the latter experiments, apart from the fact that the horizontal resolutions are much broader, more weight is always given to the ensemble part. In our case also, these first results seem to indicate that the weight of the ensemble background covariances should indeed be increased, which is confirmed by the scores of HYB0.8. Against ground-based measurements, the latter experiment indeed displays significant positive scores against BCLIM, but they are still below those obtained by both BENS experiments. A more general discussion about weight assignment in the hybrid formulation is presented later in section 3.5.3.

3.5.2 | Focus on BENS-GP

Here focus is on experiment BENS-GP which, at this stage, shows the overall best performances. Scores for accumulated precipitation and a scorecard that synthesizes the main tendencies are discussed.

Scores against rain gauges. Precipitation scores against the 1,800 rain gauges of the French real-time meteorological observation network have been computed for the whole period for all experiments. In order to evaluate the time evolution of the forecast precipitations accumulated every 6 h up to 30 h of the forecast, we made use of the Brier Skill Score (BSS) considering a neighbourhood of 52.8 km around each observation. The BSS measures the accuracy of the forecast precipitation by giving the probability of predicting accumulated rain above different thresholds (Stein, 2011). As for the conventional data, we consider here normalized relative scores by considering the BSS in Equation 17. However, since the closer the BSS is to unity the better is the score, positive values show an improvement compared to the reference. Figure 10 displays such values for BENS-GP considering BCLIM as reference, for four different precipitation thresholds and for different forecast lead times. The improvements against conventional data up to 30 h are confirmed with statistically significant scores, especially for the 0.5 and the 2 mm h^{-1} thresholds between 12 and 24 h of the forecast. A significant positive score is also visible for the afternoon forecasts up to 10 mm h^{-1} . However, for thresholds above 5 mm h^{-1} , caution should be applied to the interpretation of these results since the sample size is small.

Score card. In order to get an overview of the main forecast scores of one experiment compared to the reference at a glance, a score card has been recently developed for

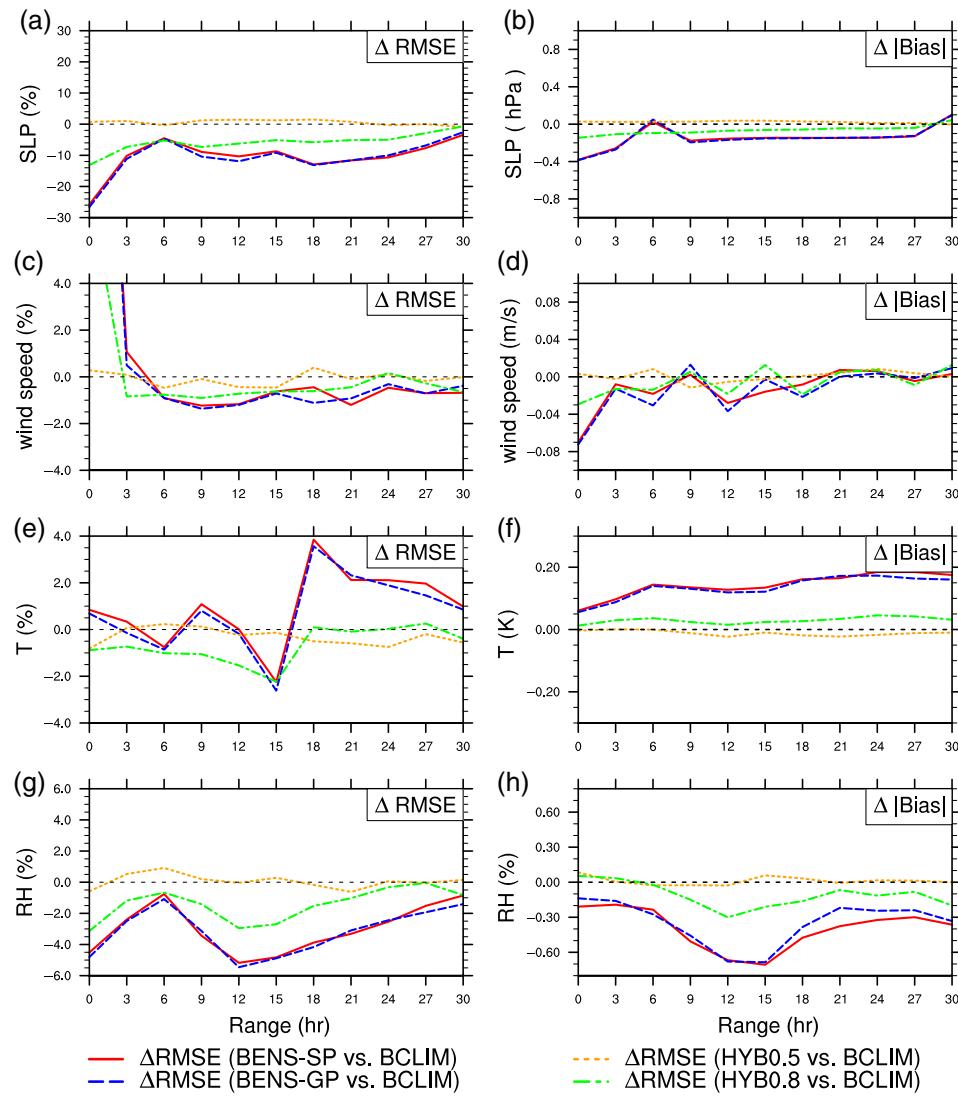


FIGURE 8 Scores versus forecast range against ground-based measurements from the French network. (a, c, e, g) represent the ΔRMSE , as defined by Equation 17 with BCLIM as reference, computed for the BENS-SP (plain lines), BENS-GP (dashed lines), HYB0.5 (dotted lines) and HYB0.8 (dash-dotted lines) experiments (as defined in the first part of Table 1) for (a) sea-level pressure, (c) wind speed at 10 m, (e) temperature and (g) relative humidity at 2 m (all in %). (b, d, f, h) display the corresponding $\Delta|\text{Bias}|$ in (b) hPa, (d) m s⁻¹, (f) K and (h) %. Values represent averages for forecasts launched at 0000 UTC from 06 February to 10 March 2016 [Colour figure can be viewed at wileyonlinelibrary.com]

AROME. Scores against a comprehensive set of upper-air and surface observations are computed for two experiments and for different forecast ranges. For the surface, all data from ground-based stations available over the domain covered by the model are considered. For the upper-air, (a) aircraft data are used to compute scores of temperature, (b) SEVIRI water vapour channels and ground-based GPS data are used to compute scores of humidity, and (c) aircraft and wind profilers are finally used to compute scores of horizontal wind. ΔRMSE are then calculated and the score card can be drawn, displaying its main variations in a visual way. As for the scores against conventional data previously discussed and contrarily to BSS against rain gauges, a positive value shows an improvement compared to the reference.

Such a score card is displayed in Figure 11 for BENS-GP against BCLIM. It clearly confirms the previous results, that a 3DEnVar with constant localization lengths and localization

functions defined in grid-point space gives significantly better forecast scores than a 3D-Var based on climatological background-error covariances that are calibrated using an EDA, which is run over the period of evaluation. The most striking improvements can be seen for the mean sea level pressure (MSLP), where values larger than 5% occur between 6 and 15 h of the forecast, which proves that the mass field integrated all over the troposphere is better predicted. Improvements are also clear for humidity at the ground and at low to mid-troposphere, where values of ΔRMSE between 2 and 5% are displayed for all forecast ranges. Overall, the total NWP index changes, which are the global averages for all variables and for all ranges, are 1.4 and 1.9% for upper-air and for the surface, respectively.

For most convective-scale NWP systems, the impact of the inclusion of a DA scheme compared to a simple dynamical adaptation of the coupling model is typically visible until

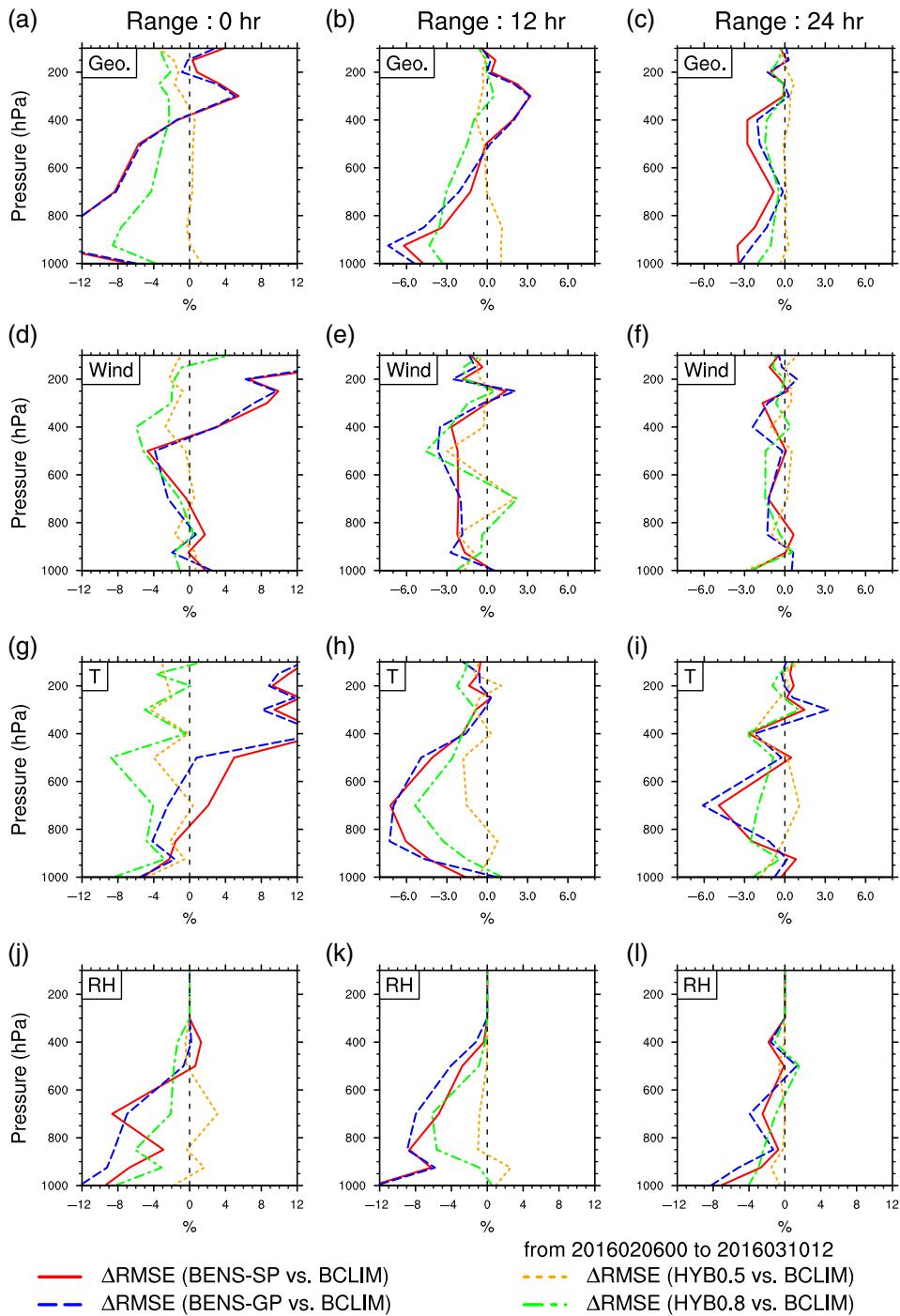


FIGURE 9 ΔRMSE , as defined by Equation 17 with BCLIM as reference and the BENS-SP (plain lines), BENS-GP (dashed lines), HYB0.5 (dotted lines) and HYB0.8 (dash-dotted lines) experiments, against radiosondes for (a, d, g, j) the analyses, (b, e, h, k) the 12 h and (c, f, i, l) the 24 h forecasts, computed for (a, b, c) the geopotential height, (d, e, f) the wind speed, (g, h, i) the temperature and (j, k, l) the relative humidity (all in %). These scores have been computed for forecasts launched at 0000 UTC from 06 February to 10 March 2016 [Colour figure can be viewed at wileyonlinelibrary.com]

12–18 h of integration, depending of the size of the domain and of the weather types (review by Gustafsson *et al.*, 2017). The significantly positive scores that have been shown here, which last at least until 30 h into the forecast, are thus very encouraging.

The impact of applying localization in grid-point space instead of spectral space is summarized in the score card shown in Figure 12. Global improvement of 0.38 and 0.26% is obtained for upper-air and the surface, respectively.

These positive values are mostly due to better forecasts of tropospheric wind components and of surface pressure for the first forecast ranges. However, BENS-SP still outperforms BCLIM significantly, with NWP indexes of 1.02 and 1.64%.

3.5.3 | Focus on hybrids

We have seen that imposing climatological and sampled background covariances with equal weights in Equation 11 in the HYB0.5 experiment leads to rather neutral forecast

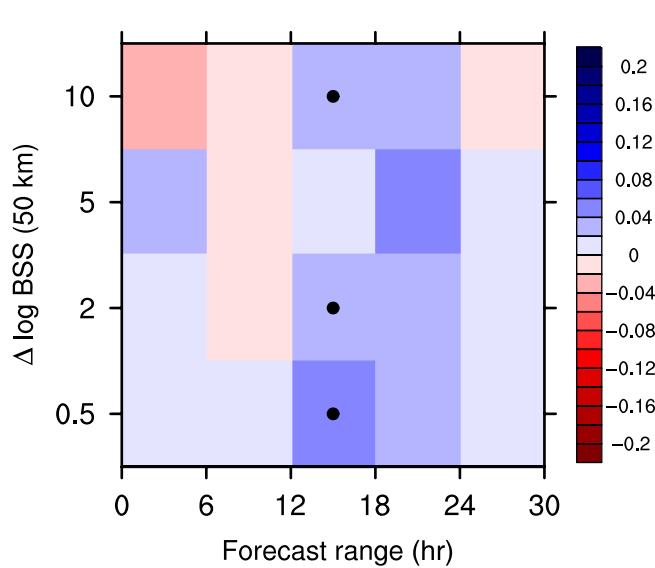


FIGURE 10 Relative variations (%) of the Brier Skill Score (BSS; see text for definition) for BENS-GP considering BCLIM as reference, computed considering a neighbourhood of 52.8 km, for precipitations forecast from the 0000 UTC runs and cumulated every 6 h. Results display averages over the whole time period from 6 February to 10 March 2016. Black dots denote scores that are statistically significant, as deduced from a boot-strap test [Colour figure can be viewed at wileyonlinelibrary.com]

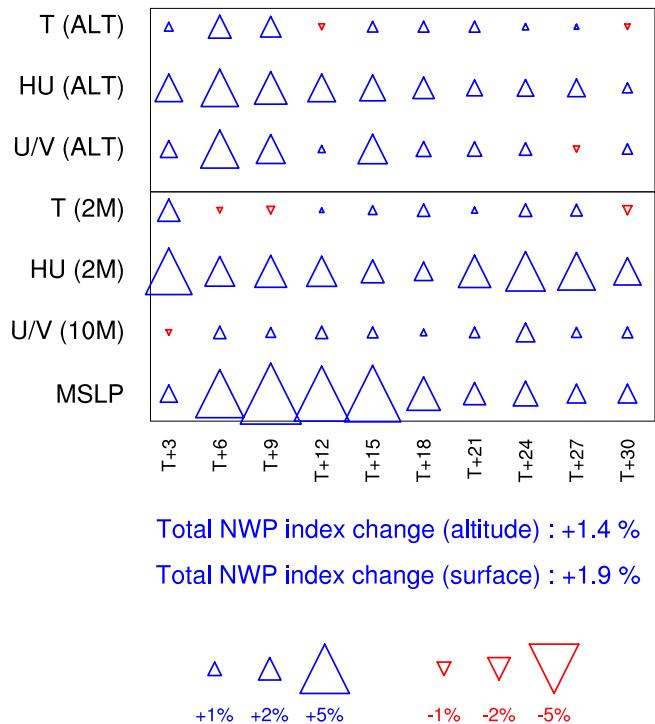


FIGURE 11 Score card summarizing the changes in forecast scores with forecast ranges brought by BENS-GP against BCLIM, for the main upper-air (ALT) and near-surface model variables. Only forecasts starting at 1200 UTC and performed over the whole time period from 6 February to 10 March 2016 are considered. A triangle pointing upwards (a triangle pointing downwards) shows an improvement (a degradation) in RMSE. The wider the area, the larger the $\Delta RMSE$ [Colour figure can be viewed at wileyonlinelibrary.com]

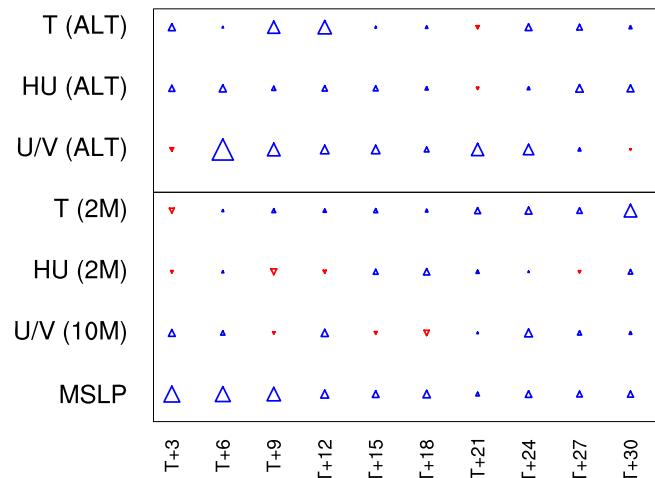
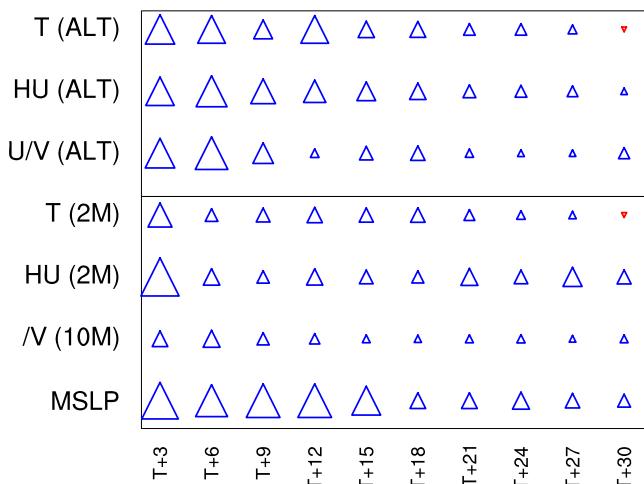


FIGURE 12 As Figure 11, but for BENS-GP against BENS-SP [Colour figure can be viewed at wileyonlinelibrary.com]

scores against conventional observations when compared to BCLIM. When 80% weighting is given to the sampled covariances in the HYB0.8 experiment, much better scores have been shown. This tendency is confirmed when comparing against other observation types, as summarized in the score card displayed in Figure 13. A clear improvement is now obtained compared to BCLIM, with positive signals for both surface and upper-air variables which last until the last forecast range. Improvements around 3% are particularly noticeable for the surface-level pressure and for the tropospheric temperature and humidity for the first 12 h of the forecast. Overall, NWP indices of +1.37% and +1.26% are obtained compared to BCLIM, which indicate a significant improvement.

Nevertheless, and as is shown in the score card of Figure 14, HYB0.8 shows a worse performances than BENS-GP especially against surface observations, with a degradation of the NWP indexes of -0.05% and -0.67% for upper-air and surface variables, respectively. However, a slight improvement is visible for most forecast ranges for both surface and tropospheric temperature.

Better results could probably be obtained with a proper combination of weights, however such tuning is outside the scope of this paper. The choice of weighting coefficients in previous studies is largely empirical and likely depends on different factors linked to the respective quality of static and flow-dependent covariances. Increasing the size of the ensemble (e.g. by using time-lagged ensembles), combined with an increase of the weight of the ensemble part, seems particularly beneficial as is shown at global (Lorenc, 2017), regional (Gustafsson *et al.*, 2014), and convective (Kong *et al.*, 2018)



Total NWP index change (altitude) : +1.37 %

Total NWP index change (surface) : +1.26 %



FIGURE 13 As Figure 11, but for experiment HYB0.8 against BCLIM [Colour figure can be viewed at wileyonlinelibrary.com]

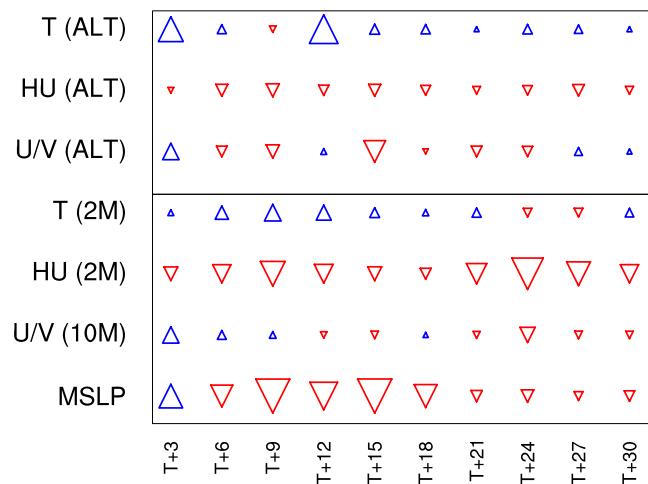
scales. More weight could also be given to the climatology for the higher levels (e.g. as in Buehner *et al.*, 2013 or Caron *et al.*, 2015). The theory of Ménétrier and Auligné (2015a), which aims at retrieving simultaneously optimal localization lengths and weighting coefficients in the hybrid formulation, could be used to apply more objective values in the future.

4 | SENSITIVITIES TO LOCALIZATION LENGTHS

The EnVar scheme may be sensitive to the localization lengths, which is the topic of this section. The main characteristics of additional experiments are gathered in the second part of Table 1. In addition to forecast scores, the impacts of all experiments on spin-up will also be detailed.

4.1 | Applying empirical length-scales

Three additional experiments using various localization lengths have been launched for comparisons with BENS-GP. The goal is to quantify the overall performance of using in the latter experiment the averaged diagnosed localization lengths of 170 km / 0.3 log(P) deduced from the EDA. Experimental horizontal localization lengths of 100 and 350 km are used respectively in BENS-GP-100 and BENS-GP-350, along with a 0.3 value for the vertical, while BENS-GP-0.6 uses a 170 km in the horizontal and 0.6 for the vertical.



Total NWP index change (altitude) : -0.05 %

Total NWP index change (surface) : -0.67 %

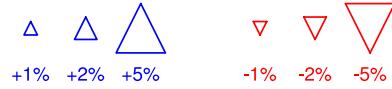


FIGURE 14 As Figure 11, but for experiment HYB0.8 against BENS-GP [Colour figure can be viewed at wileyonlinelibrary.com]

Forecast scores against radiosondes and ground-based data are respectively shown in Figures 15 and 16. These scores have been computed relative to BENS-GP, which is so far the experiment that gives the best overall results compared to BCLIM.

Increasing the vertical localization length to 0.6 generally degrades scores below 600 hPa for all variables, except for temperature at low levels and at 2 m, where slight improvements in $\Delta RMSE$ and bias can be noticed. Degradation of the MSLP bias can also be noticed. Using a large horizontal localization length in BENS-GP-350 clearly has also a detrimental effect on scores, especially before 18 h of the forecast for the 10 m wind, and for the temperature and the humidity at 2 m. Such degradation is also visible when comparing to radiosondes, although a slight improvement around 1% for geopotential height can be noticed after 12 h of the forecast. Scores are more neutral for BENS-GP-100 when compared to ground-based data, except for temperature after 15 h with degradations reaching 1% in $\Delta RMSE$ and with a degradation in bias at almost all forecast ranges. Larger $\Delta RMSE$ are nevertheless visible after 12 h for all tropospheric variables, except for relative humidity where improvement up to 2% in $\Delta RMSE$ is visible around 700 hPa. However 12 h later, this trend is reversed for this latter variable, while becoming more neutral for the others. Overall, as summarized in Table 2, all experiments display negative total NWP indices for upper-air and surface, the worst being obtained for BENS-GP-350. This confirms that the use of the constant values which have been obtained by averaging the objectively diagnosed localization lengths in both directions performs best in terms of forecast scores.

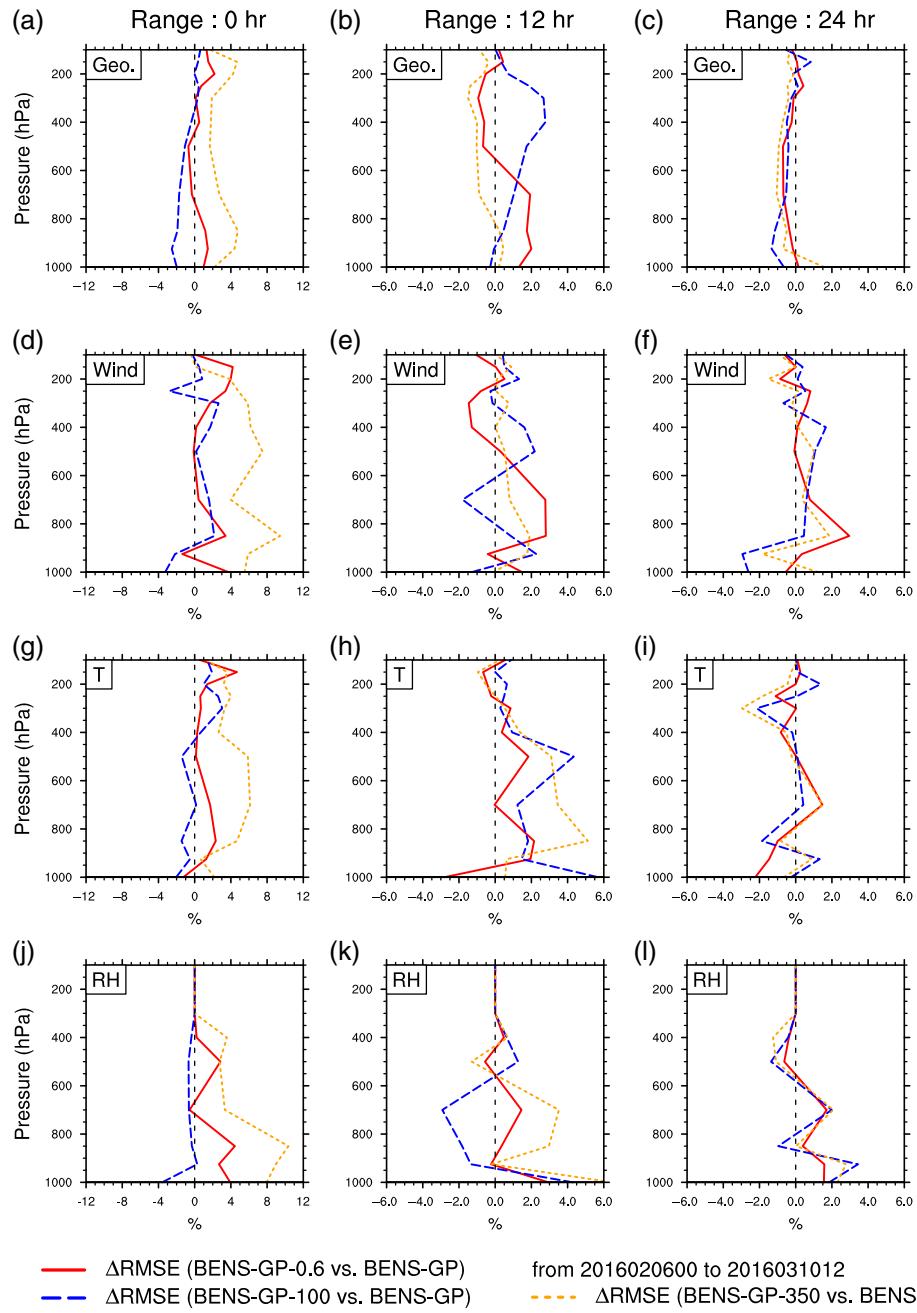


FIGURE 15 As Figure 9, but for experiments BENS-GP-0.6 (solid lines), BENS-GP-100 (dashed lines), BENS-GP-350 (dotted lines), relatively to BENS-GP [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Spin-up effects

As is demonstrated in different studies focusing either on En-Var (Lorenc, 2003) or EnKF (Kepert, 2009), the Schur product between the perturbations and the localization functions clearly induces imbalances in the analyzed fields. These imbalances appear by construction because of the presence of spatial gradients in the formulation of the different balances. For instance, for the horizontal, sub-geostrophic wind increments may appear, leading to a reduction of the degree of geostrophic balance. However at convective scale and as has been studied by Caron and Fillion (2010) and Montmerle and Berre (2010), the degree of geostrophy can be much smaller, especially in precipitating areas where balances

linked to diabatic processes prevail. In addition, in convective areas, the vertical balance between mass convergence and divergence into a model column may also be upset by vertical localization. Finally, and as discussed for applications at global scale by Clayton *et al.* (2013), such vertical localization may also have a detrimental effect on the hydrostatic balance in areas where the air mass is mostly stable.

To document these imbalances, Figure 17 shows the time evolution of the standard deviation of the surface pressure tendencies, computed during the first hour of integration for different experiments whose analyses are computed from the same background. For all EnVars, the spin-up is obviously

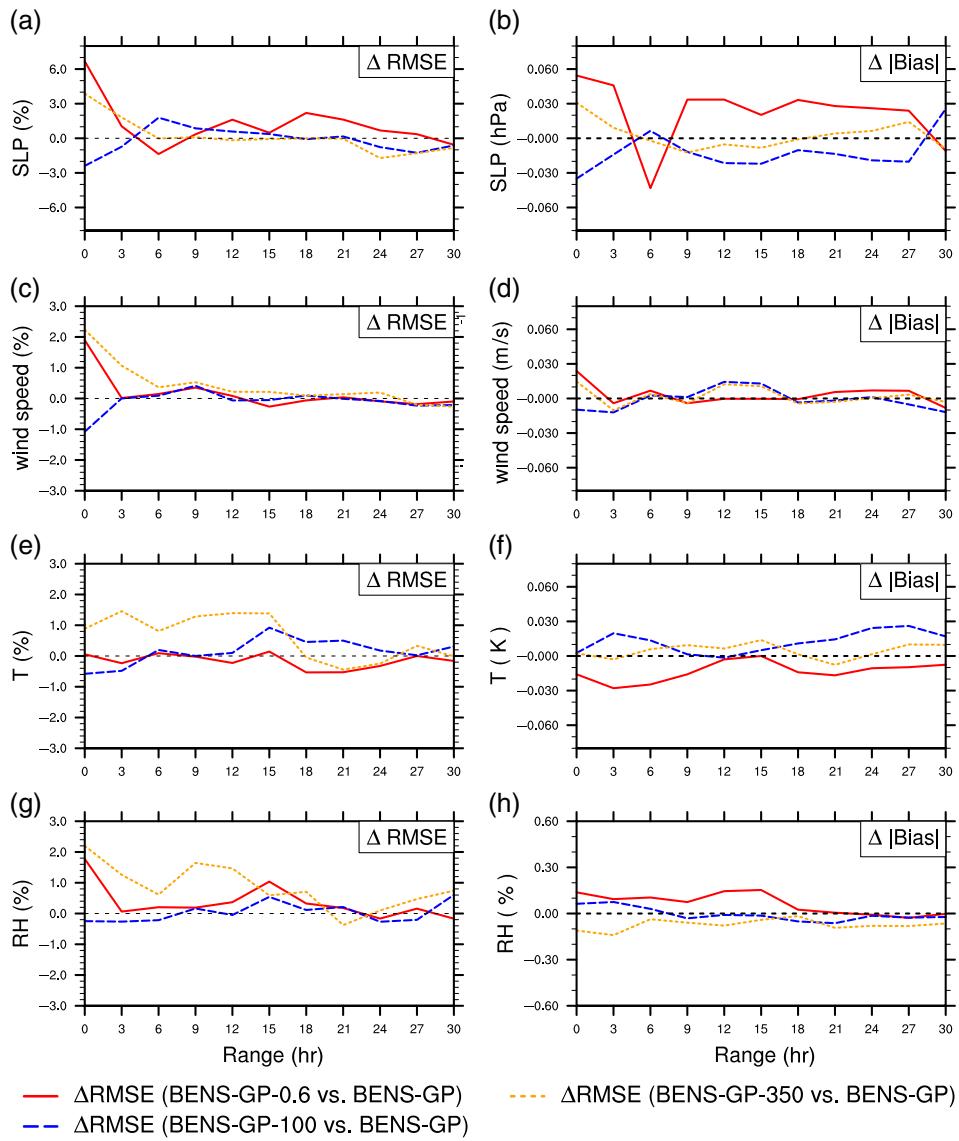


FIGURE 16 As Figure 8, but for experiments BENS-GP-0.6 (solid lines), BENS-GP-100 (dashed lines), BENS-GP-350 (dotted lines), BENS-GP-Vz (green dash-dotted lines) relatively to BENS [Colour figure can be viewed at wileyonlinelibrary.com]

increased compared to BCLIM with values that are at least larger by a factor of two, before reaching comparable values after 50 min. These values may depend on the ensemble size and on the inflation that can be applied to the perturbations in order to control the ensemble spread. Conversely, small values for BCLIM are expected as its modelled background-error covariances have been carefully calibrated precisely for the purpose of retrieving the most balanced fields to avoid the use of initialization techniques at the beginning of the forecasts, thus making frequent analysis updates feasible (Brousseau *et al.*, 2011). HYB0.8 displays a comparable spin-up to that of BCLIM, demonstrating that its initial analyzed fields are more balanced than those retrieved by all 3DEnVar configurations. This feature has been consistently observed for other dates with different meteorological conditions (i.e. with different assimilated observations, especially radar data). It is likely due to the use of a certain degree of climatological background-error covariances (and thus of its balance operators) and also probably to the increase of the

TABLE 2 Total NWP index changes computed for three experiments against BENS-GP for altitude and surface variables

Experiment	Altitude	Surface
BENS-GP-0.6	-0.14	-0.29
BENS-GP-100	-0.23	-0.09
BENS-GP-350	-0.56	-0.48

background-error covariance rank, as evoked by Gustafsson *et al.* (2014).

Applying localization in grid-point space in BENS-GP implies slightly better balanced fields than BENS-SP, for which the artifacts linked to the bi-Fourier decomposition, as discussed in section 2.5, seem to slightly enhance the amplitude of the 10 min spin-up wave that is visible up to 40 min. It has been found that such beneficial effect of localization in grid-point space is even more pronounced in rainy situations when more radar data are assimilated (not shown). Imposing shorter (longer) horizontal localization length than the optimal retrieved value in BENS-GP-100

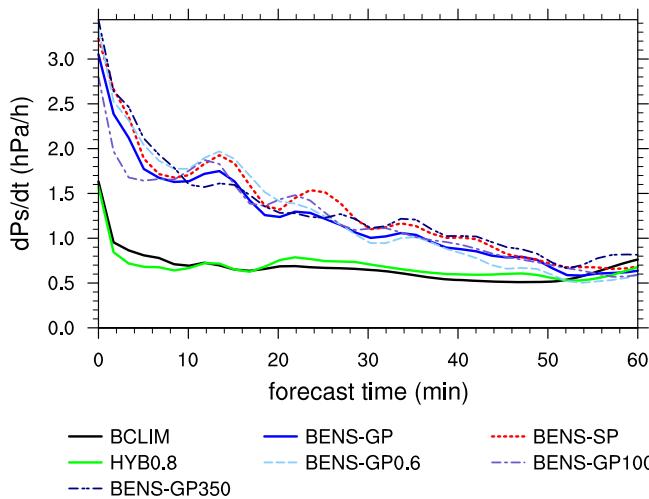


FIGURE 17 Time evolution of the standard deviation of surface pressure tendencies (hPa h^{-1}) computed for 6 February 2016, for different experiments: BCLIM (black solid line), BENS-GP (blue solid line), BENS-SP (red dotted line), HYB0.8 (green solid line), BENS-GP-0.6 (light blue dashed line), BENS-GP-100 (blue dash-dotted line), BENS-GP-350 (dark blue dash-double dotted line) [Colour figure can be viewed at wileyonlinelibrary.com]

(BENS-GP-350) implies more (less) balanced fields during the first 10 min of integration, before reaching values comparable to that of BENS-GP. Clearly, noisy covariances that are kept at longer distance in BENS-GP-350 induce more imbalances at the very beginning of the forecast.

Empirically enhancing the vertical localization length in BENS-GP-0.6 seems also to have a detrimental effect on spin-up, at least for the first 30 min. As discussed above, this may be due to the upset by the vertical localization of the hydrostatic and of the diabatic balances, roughly in clear air and in precipitating areas respectively.

The localization scheme developed for AROME is able to keep the vertical cross-covariances between surface pressure and the other variables, at least for the levels near the surface, before being damped with height. Yet, it does not allow for a larger horizontal localization length for surface pressure compared to 3D variables, which is probably a weakness that may affect the balance. Montmerle and Berre (2010) and Michel *et al.* (2011) showed strong coupling between humidity and divergence in precipitating areas that may also be harmed by the vertical localization. Some vertical adjustment, such as that of the divergence presented by Hamrud *et al.* (2015), may help in that context. Clearly, further investigation is needed for the interaction between balance and localization at convective scale.

5 | CONCLUSIONS

Ensemble variational assimilation is a promising path towards the improvement of NWP at convective scale. This paper presents the development and the first results of a 3DEnVar scheme for the AROME-France model. Our strategy has led

us to implement a variational EDA (based on the same system) of 25 members at a spatial resolution of 3.8 km in order to provide the background perturbations. The EnVar requires a localization scheme to filter out sampling noise in the covariances. We have developed two localization schemes, one in spectral space and one in grid-point space. The minimization algorithm in our EnVar does not require the square root of the background-error covariance matrix, which simplifies the implementation of hybrid schemes.

The 3DEnVar is evaluated against 3D-Var over a five-week period, at the same 3.8 km horizontal resolution. The best configuration of EnVar displays significant positive forecast scores for all variables and for precipitation that last up to 30 h of the forecast. The grid-point space localization scheme that uses recursive filters and spatial deformations performs significantly better than the one based on bi-Fourier expansion. This best configuration also uses constant horizontal and vertical localization lengths over the vertical, which are in agreement with the values deduced from the algorithm of Ménétrier *et al.* (2015a; 2015b). Making use of shorter or larger values clearly degrades the performances and adds some spin-up at the beginning of the forecasts.

In the present study, which makes use of AROME-France at 3.8 km resolution for both the deterministic model and the ensemble perturbations, the hybrid scheme using equal weights for the climatological and the ensemble covariances shows poorer performance than the best 3DEnVar configuration. Although the study of Li *et al.* (2012) leads to a similar conclusion, it is in contradiction to what has been shown in a LAM context by (e.g.) Gustafsson *et al.* (2014), Caron *et al.* (2015) or Wu *et al.* (2017). However, in the latter studies, the resolutions of the background-error perturbations are either much coarser than the final analysis increments or more compatible with regional-scale processes. Giving more weight to the ensemble covariances clearly improves the forecast scores, but the hybrid still is performing worse than the best 3D-Var configuration. In our experimental framework, using modelled background covariances in addition to the sampled ones may thus not be helpful in improving the smallest scales of forecast errors. However spin-up is lower and interestingly comparable to that of the 3D-Var, probably thanks to the increase of the background-error covariance rank and to the use of climatological balance operators.

Future research is needed before considering operational implementation. We have not yet fully considered the problem of ensemble size and of its spatial resolution. Future work will focus in implementing the dual-resolution EnVar, in which the analyses are computed at full resolution while the EDA is kept at lower resolution. We plan to apply interpolation operators ‘on the fly’ in the gradient computations. Further work will address advanced aspects of the localization, such as the scale-dependent localization from Buehner and Shlyaeva (2015). Another option is to introduce flow-dependency in the localization, using for instance spatial deformations (Michel,

2013b). The hybrid scheme has been found to be sensitive to the specification of its weights. A solution to the objective estimation of the weights has been proposed by Ménétrier and Auligné (2015a), and this requires more investigation. Some authors also consider weights that vary in the vertical, which may be of interest for the AROME EnVar.

Finally, work is also ongoing in order to set up a 4DEnVar for AROME (Liu *et al.*, 2008), whose main concept seems particularly interesting for DA at convective scale, as it avoids the linearization of strongly nonlinear processes such as those of the microphysical parametrizations. This opens new questions such as the time evolution of localization (Desroziers *et al.*, 2016).

ACKNOWLEDGEMENTS

This work was supported by the French national program LEFE-MANU of the CNRS's Institut National des Sciences de l'Univers and the AVENUE project funded by the RTRA-STAE foundation of Toulouse, France.

The authors thank Jean-François Caron from Environment Canada for the fruitful collaboration we had at CNRM during his recent stay. His work was in particular at the origin of the set-up of the new verification for AROME.

ORCID

Thibaut Montmerle  <https://orcid.org/0000-0003-3999-1633>

REFERENCES

- Bannister, R.N. (2008) A review of forecast-error covariance statistics in atmospheric variational data assimilation. II: modelling the forecast-error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134, 1971–1996.
- Belo-Pereira, M. and Berre, L. (2006) The use of an ensemble approach to study the background-error covariances in a global NWP. *Monthly Weather Review*, 134, 2466–2489.
- Benjamin, S., Weygandt, S., Brown, J., Hu, M., Alexander, C., Smirnova, T., Olson, J., James, E., Dowell, D., Grell, G., Lin, H., Peckham, S., Smith, T., Moninger, W., Kenyon, J. and Manikin, G. (2016) A North American hourly assimilation and model forecast cycle: the rapid refresh. *Monthly Weather Review*, 144, 1669–1694.
- Berre, L. (2000) Estimation of synoptic and mesoscale forecast-error covariances in a limited-area model. *Monthly Weather Review*, 138, 644–667.
- Berre, L., Varella, H. and Desroziers, G. (2015) Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quarterly Journal of the Royal Meteorological Society*, 141, 2803–2812. <https://doi.org/10.1002/qj.2565>.
- Bowler, N.E., Clayton, A.M., Jardak, M., Jermey, P., Lorenc, A.C., Wlasak, M.A., Barker, D., Inverarity, G. and Swinbank, R. (2017) The effect of improved ensemble covariances on hybrid variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143, 785–797. <https://doi.org/10.1002/qj.2964>.
- Brousseau, P., Berre, L., Bouttier, F. and Desroziers, G. (2011) Background-error covariances for a convective-scale data assimilation system: AROME-France 3D-Var. *Quarterly Journal of the Royal Meteorological Society*, 137, 409–422. <https://doi.org/1002/qj.750>.
- Brousseau, P., Berre, L., Bouttier, F. and Desroziers, G. (2012) Flow-dependent background-error covariances for a convective-scale data assimilation system.
- Quarterly Journal of the Royal Meteorological Society, 137, 4276–4292. <https://doi.org/1175/2009MWR3020.1>.
- Brousseau, P., Seity, Y., Ricard, D. and Léger, J. (2016) Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal Meteorological Society*, 142, 2231–2243. <https://doi.org/1002/qj.2822>.
- Buehner, M. (2005) Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operation NWP setting. *Quarterly Journal of the Royal Meteorological Society*, 131, 1013–1043.
- Buehner, M. and Shlyaea, A. (2015) Scale-dependent background-error covariance localization. *Tellus A*, 67. <https://doi.org/3402/tellusa.v67.28027>.
- Buehner, M., Morneau, J. and Charette, C. (2013) Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes in Geophysics*, 20, 669–682. <https://doi.org/5194/npg-20-669-2013>.
- Caron, J.F. and Fillion, L. (2010) An examination of background-error correlations between mass and rotational wind over precipitation regions. *Monthly Weather Review*, 138, 563–578.
- Caron, J.F., Milewski, T., Buehner, M., Fillion, L., Reszka, M., MacPherson, S. and StJames, J. (2015) Implementation of deterministic weather forecasting systems based on ensemble variational data assimilation at Environment Canada. Part II: the regional system. *Monthly Weather Review*, 143, 2560–2580.
- Caumont, O., Ducrocq, V., Wattrelot, E., Jaubert, G. and Pradier-Vabre, S. (2010) 1D+3DVar assimilation of radar reflectivity data: a proof of concept. *Tellus A*, 62, 173–187.
- Clayton, A.M., Lorenc, A.C. and Barker, D.M. (2013) Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 139, 1445–1461.
- Courtier, P., Thépaut, J.-N. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387.
- Daley, R. (1991) *Atmospheric Data Analysis*. Atmospheric and Space Science Series, Cambridge University Press, Cambridge, UK.
- Derber, J. and Bouttier, F. (1999) A reformulation of the background-error covariance in the ECMWF global data assimilation system. *Tellus A*, 51, 195–221.
- Derber, J. and Rosati, A. (1989) A global oceanic data assimilation system. *Journal of Physical Oceanography*, 19, 1333–1347.
- Desroziers, G., Camino, J.T. and Berre, L. (2014) 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140, 2097–2110.
- Desroziers, G., Arbogast, E. and Berre, L. (2016) Improving spatial localization in 4DEnVar. *Quarterly Journal of the Royal Meteorological Society*, 142, 3171–3185.
- Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162.
- Fisher, M. (2003) Background-error covariance modelling. In: *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, ECMWF, Reading, UK, pp. 45–63.
- Gaspari, G. and Cohn, S. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.
- Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G. and Gratton, S. (2014) B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 539–556.
- Gustafsson, N. and Bojarova (2014) Four-dimensional ensemble variational (4D-En-Var) data assimilation for the high-resolution limited-area model (HIRLAM). *Nonlinear Processes in Geophysics*, 21, 745–762. <https://doi.org/5194/npg-21-745-2014>.
- Gustafsson, N., Bojarova, J. and Vignes, O. (2014) A hybrid variational ensemble data assimilation for the high-resolution limited-area model (HIRLAM). *Nonlinear Processes in Geophysics*, 21, 303–323. <https://doi.org/5194/npg-21-303-2014>.
- Gustafsson, N., Janjić, T., Schraff, C., Leuenberger, D., Weissman, M., Reich, H., Brousseau, P., Montmerle, T., Wattrelot, E., Bučánek, A., Mile, M., Hamdi, R., Lindskog, M., Barkmeijer, J., Dahlbom, M., Macpherson, B., Ballard, S., Inverarity, G., Carley, J., Alexander, C., Dowell, D., Liu, S., Ikuta, Y. and Fujita, T. (2017) Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Quarterly Journal of the Royal Meteorological Society*, 144, 1218–1256.

- Hamill, T.M. and Snyder, C. (2000) A hybrid ensemble Kalman filter-3D Variational analysis scheme. *Monthly Weather Review*, 128, 2905–2919.
- Hamrud, M., Bonavita, M. and Isaksen, L. (2015) EnKF and hybrid gain ensemble data assimilation. Part I: EnKF implementation. *Monthly Weather Review*, 143, 4847–4882. <https://doi.org/1175/MWR-D-14-00333.1>.
- Houtekamer, P.L. and Mitchell, H.L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1), 123–137.
- Houtekamer, P.L. and Zhang, F. (2016) Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144, 4489–4532.
- Ito, K., Kunii, M., Kawabata, T., Saito, K. and Aonashi, K. (2016) Mesoscale hybrid data assimilation system based on JMA non-hydrostatic model. *Monthly Weather Review*, 144, 3417–3439.
- Kepert, J.D. (2009) Covariance localization and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135, 1157–1176.
- Kleist, D. and Ide, K. (2015) An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part I: system description and 3D-hybrid results. *Monthly Weather Review*, 143, 433–451.
- Kong, R., Xue, M. and Liu, C. (2018) Development of a hybrid En3DVar data assimilation system and comparisons with 3D-Var and EnKF for radar data assimilation with observing system simulation experiments. *Monthly Weather Review*, 146, 175–198.
- Legrand, R., Michel, Y. and Montmerle, T. (2015) Diagnosing non-Gaussianity of forecast and analysis errors in a convective-scale model. *Nonlinear Processes in Geophysics*, 2(4), 1061–1090. <https://doi.org/5194/npgd-2-1061-2015>.
- Li, Y., Wang, X. and Xue, M. (2012) Assimilation of radar radial velocity data with the WRF ensemble/3D-Var hybrid system for the prediction of hurricane *Ike* (2008). *Monthly Weather Review*, 140, 3507–3524. <https://doi.org/1175/MWR-D-12-00043.1>.
- Liu, C., Xiao, Q. and Wang, B. (2008) An ensemble-based four-dimensional variational data assimilation scheme. Part I: technical formulation and preliminary test. *Monthly Weather Review*, 136, 3363–3373. <https://doi.org/1175/2008MWR2312.1>.
- Lorenc, A.C. (2003) The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129, 3183–3203.
- Lorenc, A.C. (2017) Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quarterly Journal of the Royal Meteorological Society*, 143, 1062–1072. <https://doi.org/1002/qj.2990>.
- Lorenc, A.C., Bowler, N.E., Clayton, A.M. and Pring, S.R. (2014) Comparison of hybrid-4DEnVar and hybrid 4D-Var data assimilation methods for global NWP. *Monthly Weather Review*, 143, 212–229.
- Ménétrier, B. and Auligné, T. (2015a) Optimized localization and hybridization to filter ensemble-based covariances. *Monthly Weather Review*, 143, 3931–3947.
- Ménétrier, B. and Auligné, T. (2015b) An overlooked issue of variational data assimilation. *Monthly Weather Review*, 143, 3925–3930.
- Ménétrier, B. and Montmerle, T. (2011) Heterogeneous background-error covariances for the analysis and forecast of fog events. *Quarterly Journal of the Royal Meteorological Society*, 137, 2004–2013.
- Ménétrier, B., Montmerle, T., Berre, L. and Michel, Y. (2014) Estimation and diagnosis of heterogeneous flow-dependent background-error covariances at the convective scale using either large or small ensembles. *Quarterly Journal of the Royal Meteorological Society*, 140, 2050–2061. <https://doi.org/1002/qj.2267>.
- Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. (2015a) Linear filtering of sample Covariances for ensemble-based data assimilation. Part I: optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review*, 143(5), 1622–1643. <https://doi.org/1175/MWR-D-14-00157.1>.
- Ménétrier, B., Montmerle, T., Michel, Y. and Berre, L. (2015b) Linear filtering of sample Covariances for ensemble-based data assimilation. Part II: application to a convective-scale NWP model. *Monthly Weather Review*, 143(5), 1644–1664. <https://doi.org/1175/MWR-D-14-00156.1>.
- Michel, Y. (2013a) Estimating deformations of random processes for correlation modelling in a limited-area model. *Quarterly Journal of the Royal Meteorological Society*, 139(671), 534–547. <https://doi.org/1002/qj.1978>.
- Michel, Y. (2013b) Estimating deformations of random processes for correlation modelling: methodology and the one-dimensional case. *Quarterly Journal of the Royal Meteorological Society*, 139, 771–783. <https://doi.org/1002/qj.2007>.
- Michel, Y. and Auligné, T. (2010) Inhomogeneous background-error modeling and estimation over Antarctica. *Monthly Weather Review*, 138, 2229–2252.
- Michel, Y., Auligné, T. and Montmerle, T. (2011) Heterogeneous convective-scale background-error covariances with the inclusion of hydrometeor variables. *Monthly Weather Review*, 139, 2994–3015.
- Montmerle, T. and Berre, L. (2010) Diagnosis and formulation of heterogeneous background-error covariances at mesoscale. *Quarterly Journal of the Royal Meteorological Society*, 136, 1408–1420.
- Poterjoy, J. and Zhang, F. (2014) Intercomparison and coupling of ensemble and four-dimensional variational data assimilation methods for the analysis and forecasting of hurricane *Karl* (2010). *Monthly Weather Review*, 142, 3347–3364. <https://doi.org/1175/MWR-D-13-00394.1>.
- Pusher, R., Wu, W., Parrish, D. and Roberts, N. (2003) Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. *Monthly Weather Review*, 131, 1524–1535.
- Raynaud, L. and Bouttier, F. (2015) Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142, 854–866.
- Raynaud, L., Berre, L. and Desroziers, G. (2011) An extended specification of flow-dependent background-error variances in the Météo-France global 4D-Var system. *Quarterly Journal of the Royal Meteorological Society*, 137, 607–619. <https://doi.org/1002/qj.795>.
- Seity, Y., Brousseau, P., Malardel, S., Hello, G., Benard, P., Bouttier, F., Lac, C. and Masson, V. (2011) The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3), 976–991. <https://doi.org/1175/2010MWR3425.1>.
- Stein, J. (2011) Another look at the contingency tables: scores based on Manhattan distances in the phase space. *Meteorological Applications*, 18, 28–39.
- Wang, X., Snyder, C. and Hamill, T.M. (2007) On the theoretical equivalence of differently proposed ensemble/3D-Var hybrid analysis schemes. *Monthly Weather Review*, 135, 222–227.
- Wang, X., Barker, D.M., Snyder, C. and Hamill, T.M. (2008) A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Monthly Weather Review*, 136, 5132–5147. <https://doi.org/1175/2008MWR2445.1>.
- Wattrelot, E., Caumont, O. and Mahfouf, J.-F. (2014) Operational implementation of the 1D+ 3D-Var assimilation method of radar reflectivity data in the AROME model. *Monthly Weather Review*, 142(5), 1852–1873.
- Wu, W., Parrish, D., Rogers, E. and Lin, Y. (2017) Regional ensemble-variational data assimilation using global ensemble forecasts. *Weather and Forecasting*, 32, 83–96.
- Zhang, M. and Zhang, F. (2012) E4DVar: coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Monthly Weather Review*, 140, 587–600. <https://doi.org/1175/MWR-D-11-00023.1>.
- Zhen, Y. and Zhang, F. (2014) A probabilistic approach to adaptive covariance localization for serial ensemble square root filters. *Monthly Weather Review*, 142, 4499–4525. <https://doi.org/1175/MWR-D-13-00390.1>.

How to cite this article: Montmerle T, Michel Y, Arbogast E, Ménétrier B, Brousseau P. A 3D ensemble variational data assimilation scheme for the limited-area AROME model: Formulation and preliminary results. *QJR Meteorol Soc*. 2018;144:2196–2215. <https://doi.org/10.1002/qj.3334>