Improving Background Error Covariances in a 3D Ensemble–Variational Data Assimilation System for Regional NWP

JEAN-FRANÇOIS CARON

Data Assimilation and Satellite Meteorology Research Section, Environment and Climate Change Canada, Dorval, Québec, Canada

YANN MICHEL, THIBAUT MONTMERLE, AND ÉTIENNE ARBOGAST

Centre National de Recherches Météorologiques/Météo-France, Toulouse, France

ABSTRACT

Following the recent development of a three-dimensional ensemble–variational (3DEnVar) data assimilation algorithm for the AROME-France NWP system, this paper examines different approaches to reduce the sampling noise in the ensemble-derived background error covariances in this new scheme without modifying the background ensemble generation strategy. We first examine two variants of scale-dependent localization: one method consists of applying different amounts of localization to different ranges of background error covariance spatial scales, while simultaneously assimilating all of the available observations. Another separate approach uses time-lagged forecasts in order to increase the effective ensemble size, up to a factor of 3 here. This approach of time-lagged forecasts is considered both on its own and together with scale-dependent localization. When the background error covariances are derived from the most recent 25-member ensemble forecasts, the results from data assimilation cycles over a 33-day winter period show that avoiding cross covariances between scales in the scale-dependent localization formulation first proposed by Buehner performs better than the more recent formulation of Buehner and Shlyaeva. However, when increasing the effective ensemble size to 75 members with time-lagged forecasts, the two scale-dependent formulations provide similar forecast improvements overall. It is also found that the lagged-members approach outperforms scale-dependent localization on its own. The largest forecast improvements are obtained when combining the two approaches.

1. Introduction

A three-dimensional ensemble–variational (3DEnVar; e.g., Lorenc 2003; Buehner 2005) data assimilation scheme for the AROME-France NWP system has been presented by Montmerle et al. (2018, hereafter M18). The scheme was evaluated in the context of a lower spatial resolution (i.e., 3.8 vs 1.3 km) and temporal updates (i.e., 3 h vs hourly), compared to the operational configuration (Brousseau et al. 2016). M18 showed that the new algorithm largely outperformed standard 3DVar in terms of forecast scores when the flow-dependent covariances are derived from a 25-member ensemble data

assimilation (EDA; e.g., Houtekamer et al. 1996; Fisher 2003; Belo Pereira and Berre 2006) at the same spatial resolution and considering the same observation types in the assimilation process. Similar to the recent work of Lorenc (2017) in a global EnVar system, the aim of this work is to examine different approaches to improve the ensemble-derived background error covariances in this new data assimilation system without modifying the EDA strategy.

Ensemble-derived background error covariances from limited ensemble sizes are strongly affected by sampling noise. Increasing the ensemble size is desirable but limited by the computational and time constraints for operational applications, especially for regional systems. As in most ensemble-based data assimilation schemes [e.g., see section 3e of Houtekamer and Zhang (2016)], M18 relied on spatial localization of the covariances. The localization consists of using a prescribed homogeneous and isotropic function to gradually damp

---

ⓞ Denotes content that is immediately available upon publication as open access.

Corresponding author: Jean-François Caron, jean-francois.caron@canada.ca

distant covariances to zero at a given distance. Such a distance can be objectively estimated from the EDA using the diagnostic of Ménétrier et al. (2015a).

Here, we examine the potential benefits of applying different localization length scales to different ranges of background error covariance spatial scales from two scale-dependent localization formulations: 1) the original approach of Buehner (2012) in combination with spectral localization, which assumes that the covariance between the scales are zero, and 2) the more recent formulation of Buehner and Shlyaeva (2015) that avoids such complete removal of the between-scale co-variances. Buehner (2012) and Lorenc (2017) showed that scale-dependent localization in combination with spectral localization has a positive impact on the accuracy of global NWP forecasts. Caron and Buehner (2018, hereafter CB18) recently revealed similar positive impacts from the formulation of Buehner and Shlyaeva (2015), also in a global context. This work represents the first application of scale-dependent localization in a high-resolution limited-area model (LAM) context. It is also the first time, to our knowledge, that the two approaches are directly compared. Previous scale-dependent localization approaches were implemented in an EnVar formulation that uses a preconditioning based on the square root of background error covariance matrix (denoted **B**). We demonstrate here that scale-dependent localization is also straightforward to implement in the Météo-France EnVar scheme, currently developed with a **B** preconditioning (Derber and Rosati 1989; Gürol et al. 2014). Although scale-dependent localization can be applied, in theory, to both horizontal and vertical scales, this paper only focuses on the horizontal, as in previous studies.

In addition, we test the time-lagged ensemble methodology (Hoffman and Kalnay 1983), which is based on the use of ensemble forecasts with different ranges but valid at the correct time. This simple and pragmatic approach allows increasing the rank of the ensemble-derived background error covariances without increasing the ensemble size in the EDA, but at the cost of the time extension of each ensemble forecast. In our LAM case, it enables information to be gathered from observations performed at different times, from forecasts at different ranges, and, in some cases, from different lateral boundary conditions (LBCs). For high-resolution NWP, it has been mostly used in the past for improving precipitation forecast in a nowcasting context, either in a probabilistic way or by using the ensemble mean (e.g., Lu et al. 2007; Yuan et al. 2009). In an EnVar context, the interest of adding lagged forecast to compute the ensemble-derived **B** has been recently shown at the global scale by Lorenc (2017) and at the regional scale

by Gustafsson et al. (2014) and Wang et al. (2017). In the latter studies, improvements in forecast scores have been clearly shown. In this paper and through this technique, 3DEnVar experiments with larger effective ensemble size were conducted both in the context of the traditional spatial localization and in combination with scale-dependent localizations.

The spatial localization approach implemented for AROME-France is first described in section 2. The three techniques tested to improve the ensemble-derived background error covariances are then presented in section 3. Section 4 presents the adopted configurations for the two scale-dependent localization methods and illustrates their impacts on the horizontal distribution of the analysis increments through the assimilation of a pseudo-single observation in two different flow regimes. Section 5 presents the impact on the AROME-France forecast scores of cycled experiments that make use of the above methods both individually and combined. Finally, the outcomes are summarized and further discussed in section 6.

2. Standard spatial localization

We first summarize how the same spatial localization can be implemented either when the preconditioning is based on **B**, as in M18, or on the square root of **B**, as is the case at Environment and Climate Change Canada (ECCC). Using the notation of CB18, the ensemble-derived **B** with standard spatial localization that does not depend on scale can be written as

$$\mathbf{B}_L = \left(\sum_{k=1}^{N_{\mathrm{ens}}} \mathbf{e}_k \mathbf{e}_k^{\mathrm{T}} \right) \circ \mathbf{L}, \qquad (1)$$

where $\mathbf{e}_k$ is the kth normalized ensemble member perturbation (i.e., the difference of the kth forecast and the ensemble mean divided by $\sqrt{N_{\mathrm{ens}} - 1}$, where N_{ens} is the ensemble size), **L** is the spatial (horizontal and vertical) localization matrix, and the symbol "∘" denotes a Schur product, which is an element-by-element product.

When using a preconditioning based on the square root of **B**, the analysis increment $\Delta\mathbf{x}$ of the EnVar scheme is obtained by computing (Lorenc 2003; Buehner 2005)

$$\Delta\mathbf{x} = \sum_{k=1}^{N_{\mathrm{ens}}} \mathbf{e}_k \circ (\mathbf{L}^{1/2} \mathbf{v}_k), \qquad (2)$$

where $\mathbf{v}_k$ represents the portion of the complete control vector corresponding to the kth ensemble member. At Météo-France, an alternative formulation based on a **B** preconditioning is used instead, leading to (M18)

$$\Delta \mathbf{x} = \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_k \circ [\mathbf{L}(\mathbf{e}_k \circ \Delta \tilde{x})], \qquad (3)$$

where $\Delta \tilde{x}$ represents the control vector having the exact same dimension as the analysis increment vector. Interested readers are referred to Desroziers et al. (2014) for a detailed description of this alternative formulation.

As detailed by M18, two variants of spatial covariance localization methods within the **L** operator were implemented for AROME-France and are based either on a spectral or a gridpoint formulation. As in Gustafsson et al. (2014), the spectral-space approach is similar to the spectral correlation model used in the 3DVar (e.g., Derber and Bouttier 1999; Berre 2000). It consists of applying the same vertical correlation matrix for each horizontal total wavenumber of a bi-Fourier decomposition (i.e., using a bi-dimensional Fourier transform; e.g., Berre 2000). In this approach, the choice has been made by M18 to use the compactly supported fifth-order piecewise rational function of Gaspari and Cohn (1999) as a localization function both on the horizontal and on the vertical. The second alternative approach makes use of homogeneous recursive filters on the horizontal (Purser et al. 2003) to impose a quasi-Gaussian localization function in gridpoint space. In the vertical, the localization employs homogeneous recursive filters together with a vertical grid transform (Michel 2013). The gridpoint approach has been shown by M18 to provide significantly better forecast performance than the spectral approach, possibly because the bi-Fourier approach introduces artificial periodicity of the analysis increments induced by observations located close to the boundaries.

3. Methods examined for improving the ensemble-derived covariances

a. Scale-dependent localization

Over the last decade, more advanced localization methods have been proposed mostly for ensemble Kalman filter applications (Bishop and Hodyss 2009; Lei and Anderson 2014; Flowerdew 2015), where spatial localization is performed in observation space, although Bishop and Hodyss (2011) proposed a method suited for variational algorithms where the localization is performed in model space. In the former method, flow-dependent (i.e., nonhomogeneous and anisotropic) localization functions are used to localize the ensemble-derived covariances. Scale-dependent localization (Buehner 2012; Buehner and Shlyaeva 2015) is another advanced method to improve model-space spatial localization and consists of applying appropriate (i.e., different) localization

length scales to different ranges of background error covariance (spatial) scales while simultaneously assimilating all the available observations. The methods still rely on homogeneous and isotropic correlation modeling and avoid the need to perform the analysis using multiple-step strategies like in Zhang et al. (2009) or Miyoshi and Kondo (2013). In practice, more severe localization is applied to small scales and less localization to large scales within the same single analysis procedure, but at the expense of an increase in the computational cost. As shown by CB18, a horizontal-scale-dependent localization allows the horizontal localization to vary implicitly as a function of the vertical level, the variable, and the horizontal location.

1) BUEHNER AND SHLYAEVA (2015) FORMULATION

The following description in this paragraph is derived from CB18, with only minor modifications and new formulations suited to the **B** preconditioning. To modulate the spatial localization as a function of horizontal scale following Buehner and Shlyaeva (2015), the normalized ensemble member perturbations are first each decomposed into a set of N_{band} overlapping wave bands (for the limited-area application in this paper, wave bands are defined in terms of bi-Fourier total wavenumbers), with scale denoted by index j:

$$\mathbf{e}_{j,k} = \boldsymbol{\Psi}_j \mathbf{e}_k, \qquad (4)$$

where $\boldsymbol{\Psi}_j$ is the spectral filter that isolates the jth wave band, and $\mathbf{e}_{j,k}$ is the kth normalized ensemble perturbation containing only the jth wave band. By introducing this scale separation into (1) and imposing that the scale-separated ensemble member perturbations sum to equal the original (i.e., full) perturbation, it can be shown that (1) can be rewritten as

$$\mathbf{B}_{\text{SDL}} = \sum_{j1=1}^{N_{\text{band}}} \sum_{j2=1}^{N_{\text{band}}} \left(\sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j1,k} \mathbf{e}_{j2,k}^{\text{T}} \right) \circ \mathbf{L}_{j1,j2}, \qquad (5)$$

where the spatial localization matrix applied to the cross covariances between wave bands j1 and j2 is expressed in terms of the specified within-scale localization matrix as

$$\mathbf{L}_{j1,j2} = \mathbf{L}_{j1,j1}^{1/2} \mathbf{L}_{j2,j2}^{\text{T}/2}. \qquad (6)$$

If the same localization matrix is applied to all within-scale and between-scale covariances, then the resulting matrix from (5) will be identical to the resulting matrix from (1). With this new formulation, the analysis increment resulting from this approach in an EnVar

scheme with a preconditioning based on the square root of **B** is obtained from

$$\Delta\mathbf{x} = \sum_{j=1}^{N_{\text{band}}} \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j,k} \circ (\mathbf{L}_{j,j}^{1/2}\mathbf{v}_k). \tag{7}$$

By comparing (7) with (2), it is obvious that the approach leads to an increase in the computational cost that is proportional to the (arbitrarily) chosen number of wave bands (N_{band}).

In the Météo-France **B**-preconditioning formulation used so far in EnVar, the analysis increment resulting from a direct application of this scale-dependent localization formulation is given by

$$\Delta\mathbf{x} = \sum_{j1=1}^{N_{\text{band}}} \sum_{j2=1}^{N_{\text{band}}} \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j1,k} \circ [\mathbf{L}_{j1,j2}(\mathbf{e}_{j2,k} \circ \Delta\tilde{\mathbf{x}})]. \tag{8}$$

By comparing (8) with (3), we remark that this formulation leads to an increase in the computational cost that is proportional to the square of N_{band}. However, using (6) allows rewriting (8) as

$$\Delta\mathbf{x} = \sum_{j1=1}^{N_{\text{band}}} \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j1,k} \circ \left\{ \mathbf{L}_{j1,j1}^{1/2} \left[\sum_{j2=1}^{N_{\text{band}}} \mathbf{L}_{j2,j2}^{T/2}(\mathbf{e}_{j2,k} \circ \Delta\tilde{\mathbf{x}}) \right] \right\}, \tag{9}$$

which is equivalent to (7) with the following variable substitution:

$$\mathbf{v}_k = \sum_{j=1}^{N_{\text{band}}} \mathbf{L}_{j,j}^{T/2}(\mathbf{e}_{j,k} \circ \Delta\tilde{\mathbf{x}}). \tag{10}$$

Because nothing in the sum over j2 depends on j1, (9) allows us to bring the increase in the computational cost close to a linear function of N_{band} compared to (3). This formulation was therefore implemented in the 3DEnVar for AROME-France. From now on, the scale-dependent localization formulation of Buehner and Shlyaeva (2015) will simply be referred to as SDL.

2) BUEHNER (2012) FORMULATION

The first SDL application was performed by Buehner (2012) in combination with the discarding of the covariances between the (horizontal) scales [so-called spectral localization (SL)[1]]. The normalized ensemble member perturbations are still decomposed into a set of N_{band} overlapping wave bands, but this time using

spectral filter coefficients whose squares sum to one (mandatory in this formulation to minimize the changes to the overall background error variance). The resulting scale-separated ensemble perturbations are then treated as independent, leading to an ensemble-derived **B** defined as

$$\mathbf{B}_{\text{SDLwSL}} = \sum_{j=1}^{N_{\text{band}}} \left(\sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j,k}\mathbf{e}_{j,k}^{T} \right) \circ \mathbf{L}_{j,j}, \tag{11}$$

which is different from (1), even if the same localization matrix is applied to all within-scale covariances due to the removal of the between-scale covariances. As shown by Buehner and Charron (2007), SL is equivalent to local spatial averaging of the covariances, which may be advantageous with a very small ensemble size [i.e., $O(10)$ members], as discussed by Berre and Desroziers (2010), given their large number of sampling errors. We remark that this formulation can be seen as a specific limit of the SDL method presented above, in which between-scale covariances become zero.

The analysis increments resulting from this formulation in an EnVar scheme with a preconditioning based on the square root of **B** are obtained from

$$\Delta\mathbf{x} = \sum_{j=1}^{N_{\text{band}}} \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j,k} \circ (\mathbf{L}_{j,j}^{1/2}\mathbf{v}_{j,k}), \tag{12}$$

where a different control vector is used for each wave band [compare (12) with (7)]. With a **B**-preconditioning approach, using again (6), it can easily be shown that the analysis increment is given by

$$\Delta\mathbf{x} = \sum_{j=1}^{N_{\text{band}}} \sum_{k=1}^{N_{\text{ens}}} \mathbf{e}_{j,k} \circ \left[\mathbf{L}_{j,j}^{1/2}\mathbf{L}_{j,j}^{T/2}(\mathbf{e}_{j,k} \circ \Delta\tilde{\mathbf{x}}) \right], \tag{13}$$

which is equivalent to (12) with the following variable substitution:

$$\mathbf{v}_{j,k} = \mathbf{L}_{j,j}^{T/2}(\mathbf{e}_{j,k} \circ \Delta\tilde{\mathbf{x}}). \tag{14}$$

Hereafter, the scale-dependent localization formulation with SL of Buehner (2012) will simply be referred to as SDLwSL.

The horizontal wave band decomposition used in both SDL and SDLwSL are described in detail in section 4, together with a series of single pseudo-observation assimilation experiments that illustrate the behavior of the two formulations.

b. Time-lagged ensemble members

The effective ensemble size can be easily increased by using ensemble forecasts of different ranges but valid at

---

[1] Not to be confused with the "spatial localization in spectral space" mentioned in section 2 (i.e., the technique of modeling isotropic, homogeneous spatial correlations by specifying statistics of spectral coefficients).

the correct time, using the so-called time-lagged approach (Hoffman and Kalnay 1983). For EnVar applications in an NWP context, this pragmatic approach has been successfully tested at regional and global scales by, for example, Gustafsson et al. (2014) and Lorenc (2017), respectively. Apart from potentially modifying the ensemble spread, it results in reduced noise and increased rank of the sampled background error covariances by a factor equal to the number of time-lagged perturbation batches.

Extending the EDA AROME forecasts up to 9 h enabled us to conduct 3DEnVar experiments with an increased ensemble size of 50 (with the addition of 25 6-h forecasts) and 75 (with the supplementary addition of 25 9-h forecasts) members. At $t = t_0$, the ensembles of 3-, 6-, and 9-h forecast ranges are, respectively, initialized from analyses retrieved at $t_0 - 3$, 6, and 9 h, considering perturbed observations performed at those different times. In our LAM framework, the perturbed LBCs that are applied during the forecasts can also differ between the different time-lagged ensemble batches, depending on the assimilation times. This is simply because the EDA ARPEGE at the global scale, which provides these LBCs, is based on a 6-h cycling strategy (Berre et al. 2015). For analyses performed at 0000, 0600, 1200, and 1800 UTC, the ensembles of 3- and 6-h forecasts consider LBCs drawn from forecasts performed 6 h before, while the ensemble of 9-h forecasts uses LBCs from forecasts initialized 12 h before. At 0300, 0900, 1500, and 2100 UTC, the ensemble of 3-h forecasts uses forecasts of the EDA ARPEGE performed 3 h before, while the ensembles of 6- and 9-h forecasts take their LBCs from forecasts initialized 9 h before.

4. Scale-dependent localizations: Setup and illustration

a. Horizontal wave band decomposition

For evaluating the SDL and the SDLwSL methods, we use a three-wave-band decomposition approach, as in CB18 [compared to six in Buehner (2012) and four in Lorenc (2017)]. This arbitrary choice was motivated primarily, as in CB18, to keep the total computational cost of the analysis step at a level that would not prevent a potential future operational implementation and to somewhat facilitate a comparison with the results of CB18 in a global context.

The spectral coefficients used to decompose the EDA-derived covariances are presented in Fig. 1. For SDL (solid lines), the response function used to isolate the large scale is a low-pass filter equal to 1 from wavenumber 0 (constant value over the domain) to wavenumber 2 (wavelength of ~1000 km) and then decays (following the
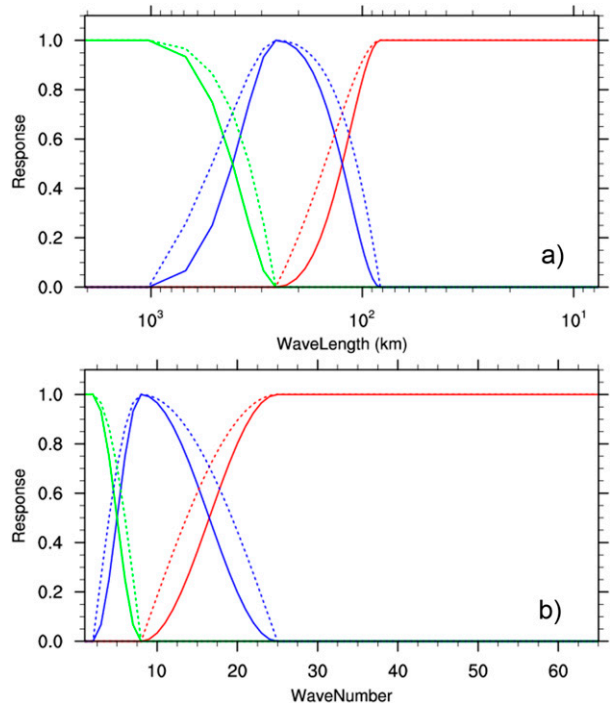

FIG. 1. Spectral filter coefficients used to separate the background error covariances into three horizontal wave bands (large scale in green, medium scale in blue, and small scale in red) expressed as a function of (a) wavelength and (b) total wavenumber. Solid lines (dashed lines) represent the coefficients used in the SDL (SDLwSL) formulation.

square of a cosine) to 0 at wavenumber 8 (wavelength of ~250 km). For the small scale, a high-pass filtering is applied by a response function imposing 0 up to wavenumber 8 (wavelength of ~250 km) and reaching a plateau of 1 starting from wavenumber 25 (wavelength of ~80 km). For the medium scale, the bandpass filtering is obtained by a third response function, which is simply equal to the differences between a value of 1 and the sum of the two previous ones. This choice ensures that the three overlapping functions sum to 1, which is essential for the SDL formulation to preserve as much as possible the overall background error variance. For the SDLwSL formulation (dashed lines), we simply used the square root of the above response functions. This way, the square of the response functions used for SDLwSL sum to one (as required by this approach), and the scale-separation functions used in the two approaches remain very similar.

The flow-dependent background error covariances used in this paper are provided by the 25-member EDA described by M18 (see their section 2.2) that uses a 3DVar scheme in combination with a perturbed observations approach on a domain that encompasses the operational AROME-France domain but with a grid spacing of 3.8 km. Figure 2 presents an example of (3 h)

a) **Full**

b) **Small scale**
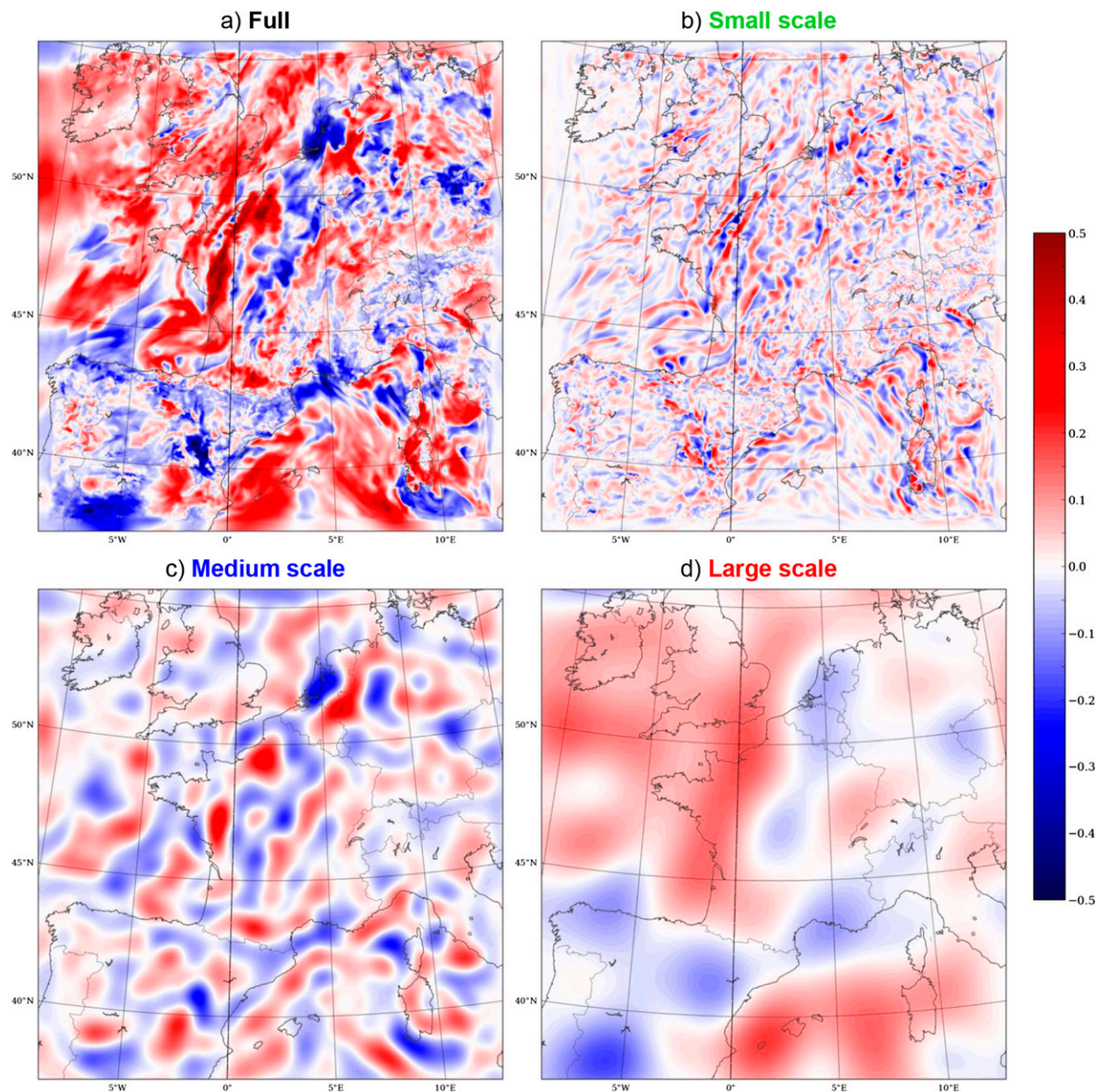
c) **Medium scale**

d) **Large scale**

FIG. 2. The 3-h ensemble forecast perturbations for temperature (K) on a model vertical level near 900 hPa from AROME EDA valid at 0000 UTC 6 Feb 2016. (a) The (full) ensemble perturbation (ensemble member 1 minus the ensemble mean). The other three figures show the scale-decomposed ensemble perturbation obtained after applying the three bandpass filters shown with solid lines in Fig. 1.

ensemble perturbations for temperature on a native model level near ~1 km AGL for a given ensemble member before (Fig. 2a) and after the scale decomposition (Figs. 2b–d) using the response functions for SDL. Therefore, in this case, the sum of the scale-decomposed perturbations equals the full ensemble perturbation. These three scale-decomposed perturbations were, as in the rest of this paper for both SDL and SDLwSL approaches, simply obtained by first transforming the

original ensemble member into spectral space (using a bi-Fourier technique[2]), then multiplying the resulting spectral coefficients by the filter coefficients shown in

---

[2] Since the forecast model used in AROME outputs horizontal fields that are biperiodic, no additional treatment of the gridded data was needed before applying the bi-Fourier transform from gridpoint space to spectral space.

Fig. 1, and finally transforming the results back into gridpoint space. Significant differences in scales in the three sets of scale-decomposed perturbations can be observed, and the amplitude of the perturbations is roughly similar in each of the wave bands (see Fig. 2).

b. Localization length scale diagnostics

To shed light on the scale differences between the scale-decomposed perturbations and to help assign realistic localization length scale values for the scale-dependent localization approaches, we applied the technique recently proposed by Ménétrier et al. (2015a,b) for objectively estimating the optimal localization amount based only on information from the ensemble. This technique, based on both the theories of centered moments sampling and of optimal linear filtering, was applied on the two sets of scale-decomposed ensemble perturbations (i.e., for SDL and SDLwSL), as well as the full ensemble perturbations for the EDA 3-h forecasts valid at 0000 UTC 6 February 2016.

The diagnosed horizontal length scales (Fig. 3) are very different for each wave band of the two sets of scale-separated ensemble perturbations, with vertical averages close to 500, 250, and 80 km for the large, medium, and small scales, respectively.[3] These results suggest that our bandpass filters design appears to be a realistic attempt for testing these scale-dependent localization approaches in the context of AROME-France.

c. Single pseudo-observation experiments

To illustrate the impact of the two scale-dependent localization approaches described in section 3a, we conducted, similar to CB18, a series of pseudo- (or synthetic) single temperature observation data assimilation experiments located here in the boundary layer (near ~1 km AGL) with the 3DEnVar configurations summarized in Table 1. Since the resolved scales here are totally different to those of CB18, this simple experimental framework aims to demonstrate the flexibility and the range of application of the two scale-dependent localization approaches that are considered here, as well as their main behaviors in terms of the resulting analysis increment structures.

The two chosen locations in the background for the AROME-France analysis valid at 0000 UTC 6 February 2016 exhibit very different flow regimes: 1) a frontal-like system in northwestern France (point A in Fig. 4), where large-scale background error covariances should be
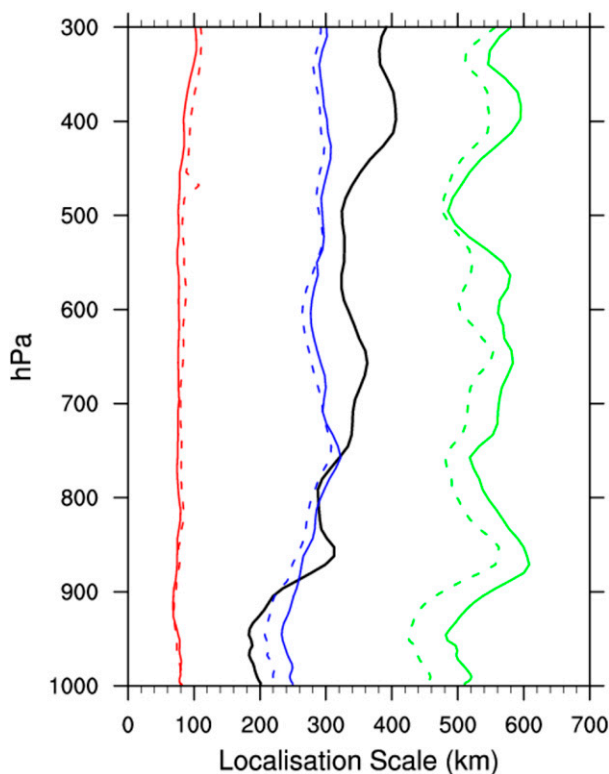

FIG. 3. Profiles of diagnosed horizontal localization scale (km) from the EDA AROME and averaged for the different control variables (temperature, zonal and meridional winds, and specific humidity) for the full ensemble perturbation (black line) and the scale-decomposed ensemble perturbations: small scale (red), medium scale (blue), and large scale (green). Solid lines (dashed lines) represent the scale-decomposed perturbations obtained with the filtering coefficients used in the SDL (SDLwSL) approach shown in Fig. 1. Data valid at 0000 UTC 6 Feb 2016.

predominant, and 2) a relatively cloud-free area in the highland of Galicia in northwestern Spain (point B in Fig. 4), where significant small-scale background error covariances were found (likely due to topographic effects; not shown). As in CB18, the results are presented in terms of analysis increments normalized by the value at the observation location, resulting in correlation-like patterns helping to distinguish changes in the horizontal distribution of the increment. In all the data assimilation experiments reported in this section, the localizations are applied in spectral space (note that the application of the localizations in gridpoint space would lead to the same conclusion since, as discussed by M18, both approaches give very close solutions far from the boundaries of the domain).

The differences in the scales of the background error covariances found in the two locations are obvious when comparing the resulting normalized analysis increments obtained with a single localization length scale of 250 km.

---

[3] In this paper, the localization length scales refer to one-half the distance over which the localization function reaches zero.

TABLE 1. Summary of single pseudo-observation experiments.

Name	N_{band}	Spectral localization	Total wavenumber peaks[a]	Localization length scales (km)
$CTL_{1\text{-obs}}$	1	—	—	250
$E\text{-}SDL_{1\text{-obs}}$	3	No	≤2, 8, ≥25	500, 250, 80
$E\text{-}SL_{1\text{-obs}}$	3	Yes	≤2, 8, ≥25	250, 250, 250
$E\text{-}SDLwSL_{1\text{-obs}}$	3	Yes	≤2, 8, ≥25	500, 250, 80

[a] Peaks refer to the total wavenumber where each response functions equal to one.

The results from this configuration named $CTL_{1\text{-obs}}$ (see Table 1) lead to a relatively large-scale and elongated (along the front axis) pattern in northwestern France (Fig. 5a), while a noisy signal with multiple maxima and a relatively strong negative normalized increment along the coast of the Bay of Biscay is depicted in northwestern Spain (Fig. 6a).

Based on the results previously presented in Fig. 3, we opted for length scale values of 500, 250, and 80 km for, respectively, the large-scale, medium-scale, and small-scale wave bands in both the $E\text{-}SDL_{1\text{-obs}}$ and $E\text{-}SDLwSL_{1\text{-obs}}$ configurations (see Table 1). Therefore, the ensemble-derived covariances in the medium-scale wave band will

be localized with the same amount as in the $CTL_{1\text{-obs}}$ configuration, but the small-scale (large scale) covariances will be localized significantly more (less). Identical sets of localization length scales were adopted here with the aim of facilitating the intercomparison of the effect of the two scale-dependent localization approaches.

For the frontal case, the impact of SDL (Fig. 5b) leads to a broadening of the normalized increments, together with a reduction of the noisy small-scale signal in the periphery of the observation location. With SDLwSL (Fig. 5d), the broadening is similar to with SDL, but the normalized increment is smoother both close and away
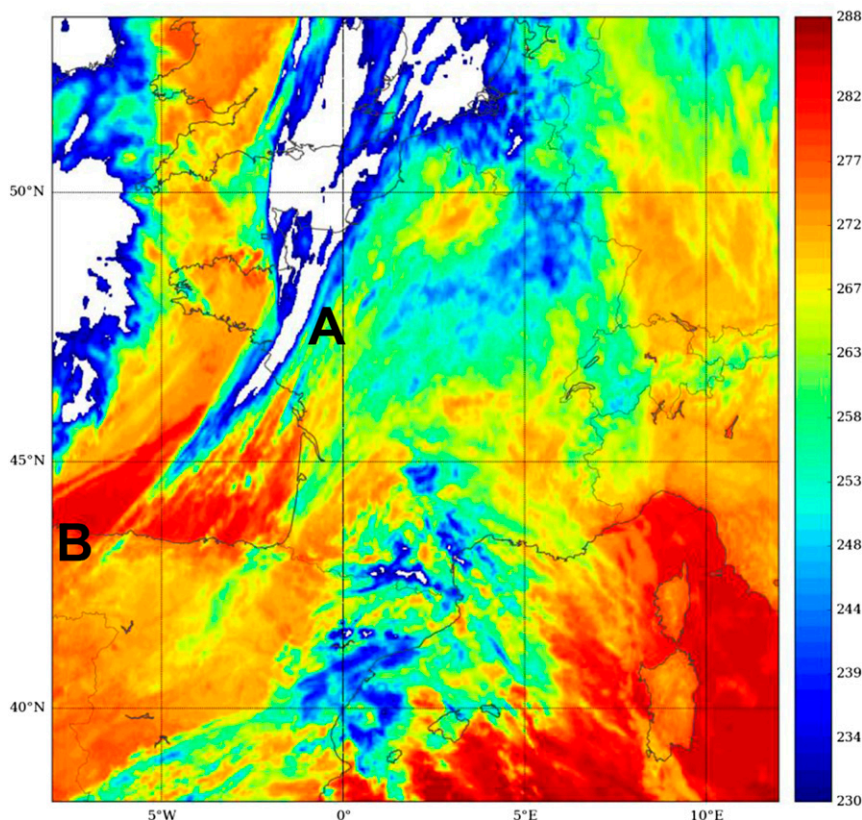

FIG. 4. Infrared (10.8 μm) brightness temperature observation (K) valid at 0000 UTC 6 Feb 2016 from SEVIRI on board Meteosat Second Generation (MSG) spacecraft. Points A and B indicate the location of the single pseudotemperature observations discussed in section 4c.
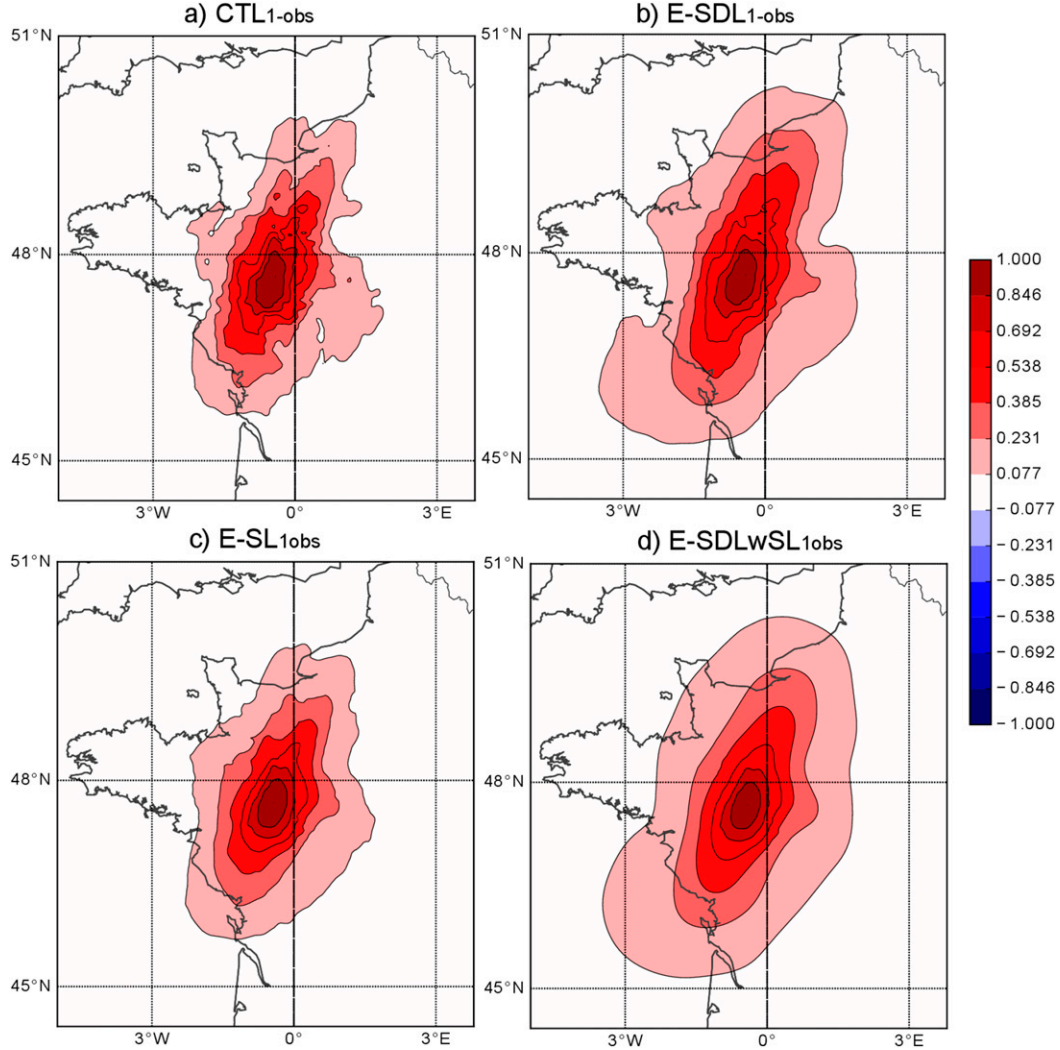
FIG. 5. Zooms of normalized temperature analysis increments near 900 hPa resulting from the assimilation of a single pseudotemperature observation located at point A in Fig. 4, obtained with four different configurations of spectral-space horizontal localization: (a) standard localization, (b) scale-dependent localization, (c) spectral localization, and (d) spectral and scale-dependent localization (see details in Table 1). The values are exactly equal to 1 at the grid point nearest to the pseudo-observation location. Data valid at 0000 UTC 6 Feb 2016.

from the observation, which is the signature of the impact of SL when the background error covariances project on multiple wave bands. To isolate the impact of SL, we computed the normalized increments with the SDLwSL approaches but using the same localization length scales of 250 km (as in CTL_{1-obs}) in the three wave bands (E-SL_{1-obs}; see Table 1). The results (Fig. 5c) show that the spatial extension of the normalized increments is similar than in CTL_{1-obs} (see Fig. 5a) but that the pattern is generally smoother, in agreement with the findings of Buehner and Charron (2007).

The impact of SDL (Fig. 6b) and SDLwSL (Fig. 6d) for the observation located in northwestern Spain leads, in both cases, to a compact core of small-scale normalized

increments surrounded by a small (less than 0.25) broader positive signal, where, again, SDLwSL presents the smoothest pattern. However, a particular feature is that the area of negative normalized increments located to the north of the observations that could be noticed with the localization that does not depend on the scale (CTL_{1-obs}; Fig. 6a) has almost completely vanished with both scale-dependent localization methods. This seems to be linked to changes in the between-scale covariances, as SL alone (Fig. 6c) also leads to the same behavior. Although SDL does not discard explicitly the between-scale covariances, in practice, the requirement that the complete localization matrix be positive–semidefinite to ensure that the resulting localized matrix is a valid
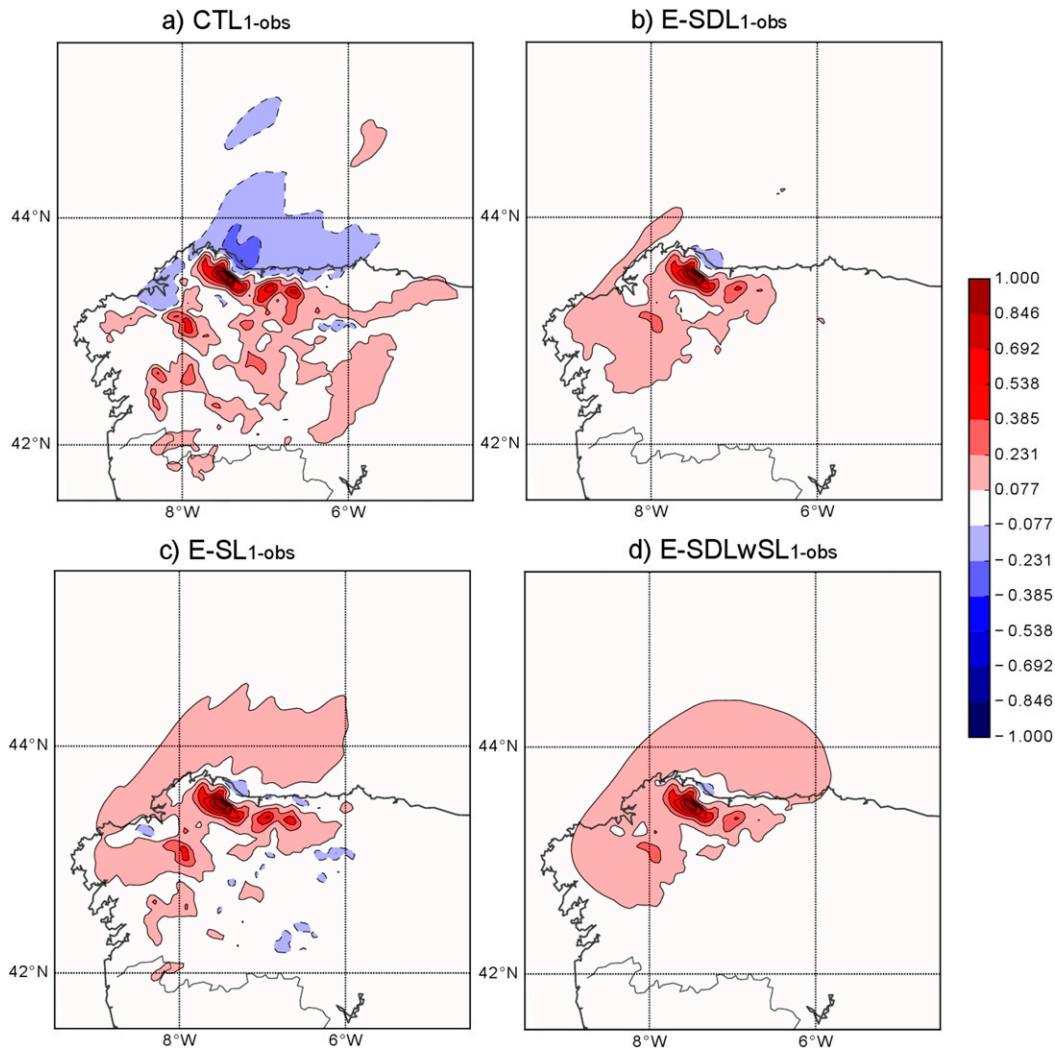
FIG. 6. As in Fig. 5, but for a pseudo-observation located at point B in Fig. 4.

covariance matrix inevitably reduces, to some extent, the between-scale covariances when different localization length scales are used for each of the wave bands (Buehner and Shlyaeva 2015). The fact that SDLwSL and SL, which completely remove the between-scale covariances, leads to a stronger positive signal on the northern side of the observation location compared to SDL supports the above hypothesis.

We remark that the changes to the between-scale covariances noted above have impacts on the implied background error statistics. It was observed that this leads to a reduction of the variances (greater in SDLwSL than in SDL due to the complete removal of the between-scale covariance in the former) that, on average, does not exceed a few points of percentage (however, locally, it can exceed 10%; not shown). In this study, no attempt was made to estimate the impact of

this effect on the changes to the forecast performances reported in the following section.

5. Impact on the forecasts

a. Experimental setup

The impact on the AROME-France forecast performances of scale-dependent localization, as well as time-lagged approaches described in section 3, was evaluated in data assimilation cycles that follow the strategy described by M18. All experiments span the 33-day winter period of M18 (from 0000 UTC 6 February to 0000 UTC 10 March 2016), with frequent stormy activity over western Europe (see section 3.1 in M18).

A 3-h update cycling strategy was also adopted, and 30-h forecasts were performed here four times per day

TABLE 2. Summary of NWP experiments.

Name	N_{band}	Spectral localization	Total wavenumber peaks	Localization length scales (km)	Time-lagged members
CTL	1	—	—	250	No
E-SDL	3	No	≤2, 8, ≥25	300, 150, 75	No
E-SL	3	Yes	≤2, 8, ≥25	250, 250, 250	No
E-SDLwSL	3	Yes	≤2, 8, ≥25	448, 224, 84	No
E-L3	1	—	—	250	3 h
E-L3&6	1	—	—	250	3 and 6 h
E-SDL-L3&6	3	No	≤2, 8, ≥25	430, 215, 100	3 and 6 h
E-SDLwSL-L3&6	3	Yes	≤2, 8, ≥25	448, 224, 84	3 and 6 h

(at 0000, 0600, 1200, and 1800 UTC) instead of twice per day (0000 and 1200 UTC) as in M18. In addition, for the time-lagged experiments, 9-h forecasts are performed from all retrieved EDA analyses every 3 h. As previously described, both the 25-member EDA and the AROME-France model adopted a grid spacing of 3.8 km. Interested readers are referred to M18 for a complete description of the assimilated observations, which include conventional and radar data, as well as radiances from geostationary and low-orbiting satellites.

For simplicity, only the experiments with the best-performing horizontal localization length scales found for each combination of localization approaches and ensemble sizes, in terms of forecast scores performed over the considered time periods, are reported (see Table 2). Scale-dependent localizations require the determination of different horizontal localization length scales for each wave band. Those values can be derived from the objective diagnostic of Ménétrier et al. (2015a,b) or through a more costly procedure of trials and errors. The application of the Ménétrier algorithm to the AROME EDA presented in section 4b suggests values of 500, 250, and 80 km for the localization length scales of the three horizontal bands. However, better results were obtained, with the refined values listed in Table 2 obtained by trials and forecast scores comparisons (using two to four attempts for each experiment, with some attempts covering only a shorter period of 2 weeks), a result shared by Lorenc (2017). The Ménétrier algorithm only considers sampling errors when determining the optimal localization, and better results can be obtained with fine tuning, but the values are useful as a first guess, especially when a large number of bands are considered.

In all the data assimilation experiments reported in this section, the localizations are applied in gridpoint space, since the latter was found to provide better forecast performance than the localization applied in spectral space by M18. The control experiment (CTL; see Table 2) uses a single localization length scale of 250 km, the best-performing value found by M18 in a one-size-fits-all context. Finally, all experiments consider a

vertical localization length scale of 0.3 unit of the natural logarithm of pressure, as in M18.

To get a global view of the impact on the forecasts, the verification observation database includes Doppler radial winds (only from the French network), ground-based weather stations, integrated water vapor from ground-based GPS receivers, wind profilers, and water vapor channels from the Spinning Enhanced Visible and Infrared Imager (SEVIRI). As in M18, these observations were used to compute relative changes in RMS error (RMSE) with respect to a given control experiment multiplied by −1 to simulate a change in a quality index and were represented with a "score card" graphical approach. Therefore, positive (negative) values represent improved (degraded) forecasts with respect to the observations. The observations in upper air were used to compute a change in quality index per 100-hPa vertical bin per variable, and the results were vertically averaged and gathered by observed variable type as detailed in Table 3. An F test was used to assess the significance of the differences for each variable and level (or vertical bin) and revealed that many differences exceed a confidence level of 90%, mostly in the first 15 h of the forecasts (not shown). Furthermore, the scores have been computed considering forecasts launched four times a day every 6 h. As a consequence, large samples of observations were available, avoiding, for example, the lack of aircraft data during nighttime.

b. SDL

The largest benefits of SDL were obtained when using localization length scales of 300, 150, and 75 km for the large-scale, medium-scale, and small-scale wave bands (experiment named E-SDL), respectively. The score card against the control experiment using a single localization length scale of 250 km (CTL; Fig. 7a) shows that the improvements both at the surface and in altitude are maxima at short lead time (close to 2%) and gradually decay to vanish beyond 15 h. The changes in this ad hoc quality index were also averaged over all variables and lead times in order to obtain a single value for each surface and altitude (see Table 4). This led to global forecast improvements of 0.27% and 0.26% for

TABLE 3. List of observations used for the altitude verification reported in the score cards.

Score card parameter	Observation type	Observation source
T (ALT)	Temperature	Aircraft
HU (ALT)	Humidity	Ground-based GPS and SEVIRI's water vapor channels
UV (ATL)	Zonal and meridional winds	Aircraft, Doppler radial wind, and wind profilers

altitude and surface, respectively. Using an hourly verification database (not shown) reveals further the significant improvement at very short lead time, with statistically significant error reduction reaching about 5% at some vertical levels in the first 3 h with respect to aircraft data.

c. SDLwSL

With SDLwSL, the best results were obtained when relaxing the localization in each wave band compared to

SDL (i.e., when using 448, 224, and 84 km instead of 300, 150, and 75 km for the large-scale, medium-scale, and small-scale wave bands, respectively; see experiment E-SDLwSL in Table 2). The improvements, compared to CTL, depicted in the associated score card (Fig. 7b) appear similar to the improvements found with SDL but are globally somewhat larger, as confirmed by the change in quality indices (see Table 4) at both the surface (0.34% vs 0.26%) and in altitude (0.46% vs 0.27%). An experiment with only SL (E-SL) was designed in an
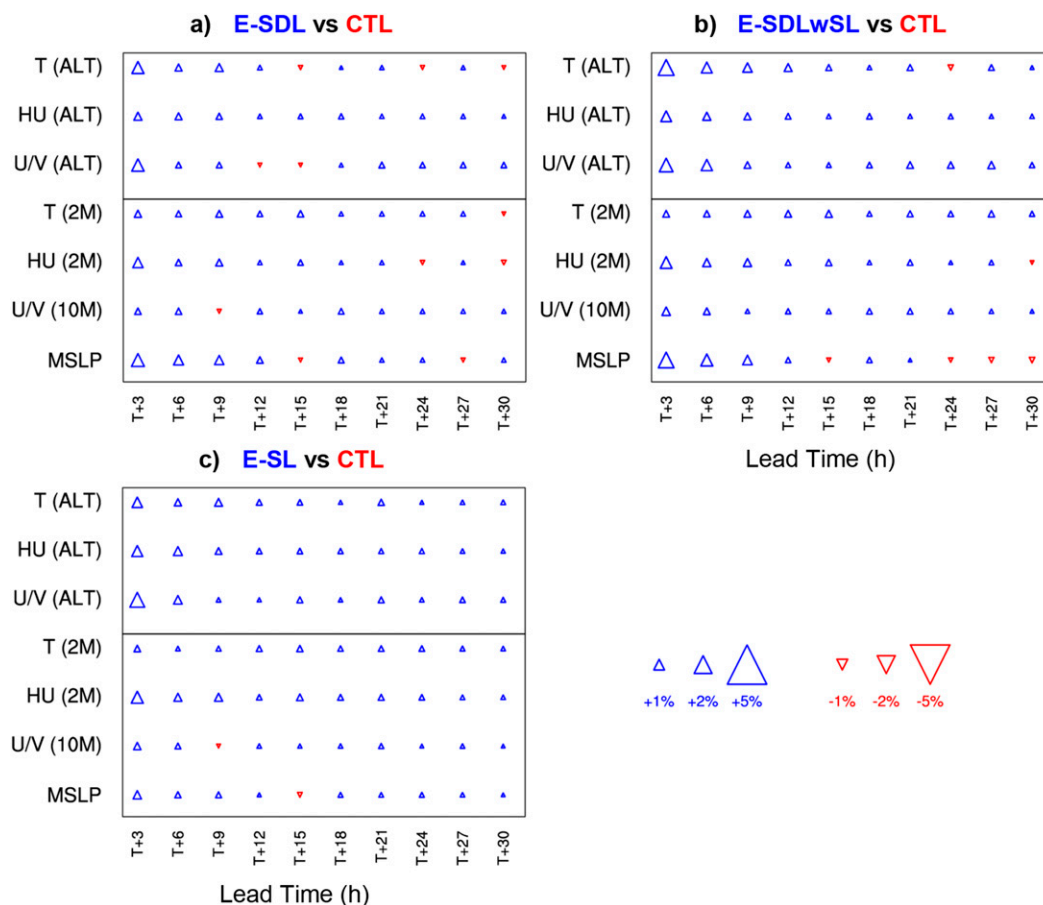

FIG. 7. Changes in the forecast quality index (defined as the relative changes in RMSE times −1) for paired configurations (experiment name in blue, reference name in red) listed in Table 2 and measured using various altitude and surface observations over a period of 33 days (from 0000 UTC 6 Feb to 0000 UTC 10 Mar 2016). Upward-pointing (downward pointing) blue (red) triangles indicate a reduction (increase) of the RMSE in the experiment (name in blue) with respect to the reference (name in red). The larger the area of the triangles, the greater the RMSE differences. See text for further details.

TABLE 4. Summary of changes in total forecast quality index (in %; see text for the details on the computation of this metric). The largest improvements are highlighted in bold.

Experiment	Altitude	Surface
CTL	Control	
E-SDL	0.27	0.26
E-SL	0.36	0.26
E-SDLwSL	0.46	0.34
E-L3	0.50	0.54
E-L3&6	0.70	0.68
E-SDL-L3&6	0.92	**0.90**
E-SDLwSL-L3&6	**1.01**	0.76

attempt to differentiate with SDLwSL in the case where the same amount of localization (length scale of 250 km; as in CTL) was enforced in the three wave bands. The results presented in Fig. 7c and Table 4 reveal that SL leads to less improvement than SDLwSL but that the former equals the improvements from the SDL at the surface and even outperform SDL in altitude, which tends to confirm the importance of covariance smoothing in very low-rank ensemble-derived background error covariances (Berre and Desroziers 2010).

d. Time-lagged ensemble members

The time-lagged ensembles are constituted of forecasts at different lead times that are initialized considering different perturbed observations and that, in some cases, use different LBCs (see section 3b). They may likely sample different probability density functions (PDFs), preventing the application of the theory of Ménétrier et al. (2015a). The horizontal localization length scale value of 250 km used by CTL has been used again with time-lagged ensembles since no significant forecast improvements were noticed when increasing the single horizontal localization length scale in this context, whereas one could expect the sampling noise to be reduced, which would imply less localization.

Increasing the ensemble sizes using ensemble forecasts with 3-h (i.e., 6-h forecasts) and 6-h (i.e., 9-h forecasts) initialization time lag (but still valid at the analysis time) is definitely beneficial in the context of this 3DEnVar prototype for AROME-France. With an additional 25 members from 6-h forecasts (E-L3), the improvements are similar (greater) at the surface (in altitude) than the best-performing scale-dependent localization experiments with only 25 ensemble members (E-SDLwSL; cf. Figs. 8a and 7b and see Table 4). Increasing the size of the ensemble to 75 members with the addition of 25 members with 9-h forecast ranges (E-L3&6) allows increasing further the benefits. The altitude scores now peak around 2% for the all the variables and reach as high as 4% for humidity at 2 m

and the mean sea level pressure (Fig. 8b) at $t + 3$ h. Overall, E-L3&6 outperformed all of the configurations tested so far (see Table 4). It is interesting to note that despite the fact that the improvements still decrease with lead time, some small benefits are now maintained up to 30 h when the ensemble size is increased. No attempt was made to further increase the ensemble size through the time-lagged members approach.

e. Combined approaches

The largest forecast improvements were obtained when combining scale-dependent localization with the increase of the size of the ensemble to 75 members through the time-lagged members approach (see experiments E-SDL-L3&6 and E-SDLwSL-L3&6 in Table 2). In this context, SDLwSL seemed to perform best when using the same localization length scale than with 25 members, but it was found beneficial to increase the length scales used in SDL to 430, 215, and 100 km for the large-scale, medium-scale, and small-scale wave bands, respectively (see Table 2). When looking at the score cards, both E-SDL-L3&6 and E-SDLwSL-L3&6 (Figs. 8c,d) appear to provide similar improvements, with changes in quality index reaching nearly 5% at $t + 3$ h for some of the variables both in altitude and at the surface. The overall measure (see Table 4) shows that E-SDLwSL-L3&6 performs best in altitude, with a total change in quality index of 1.01% (cf. to 0.92% for E-SDL-L3&6), but that E-SDL-L3&6 leads to the largest improvements for surface variables (cf. 0.90% to 0.76%).

The above results indicate that with a 75-member ensemble, SDLwSL still performs better than SDL in altitude but not at the surface. This is better illustrated by isolating the added value of the two scale-dependent localization variants in the context of 75-member ensembles (i.e., by computing the forecast performances of the two latter experiments with respect to E-L3&6). The associated score cards (Figs. 8e,f) and the overall changes in the quality index (Table 5) confirm that SDL outperforms SDLwSL for surface variables, and averaging the forecast improvements for altitude and surface variables indicates that the two formulations bring similar improvements with an increased ensemble size. This result tends to support the view of Buehner and Shlyaeva (2015) that preserving the heterogeneity of covariances by avoiding the removal of the between-scale covariances should be beneficial for larger ensemble sizes [i.e., $O(100)$ members].

6. Summary and discussion

Following the recent development of a 3DEnVar data assimilation algorithm for the AROME-France NWP
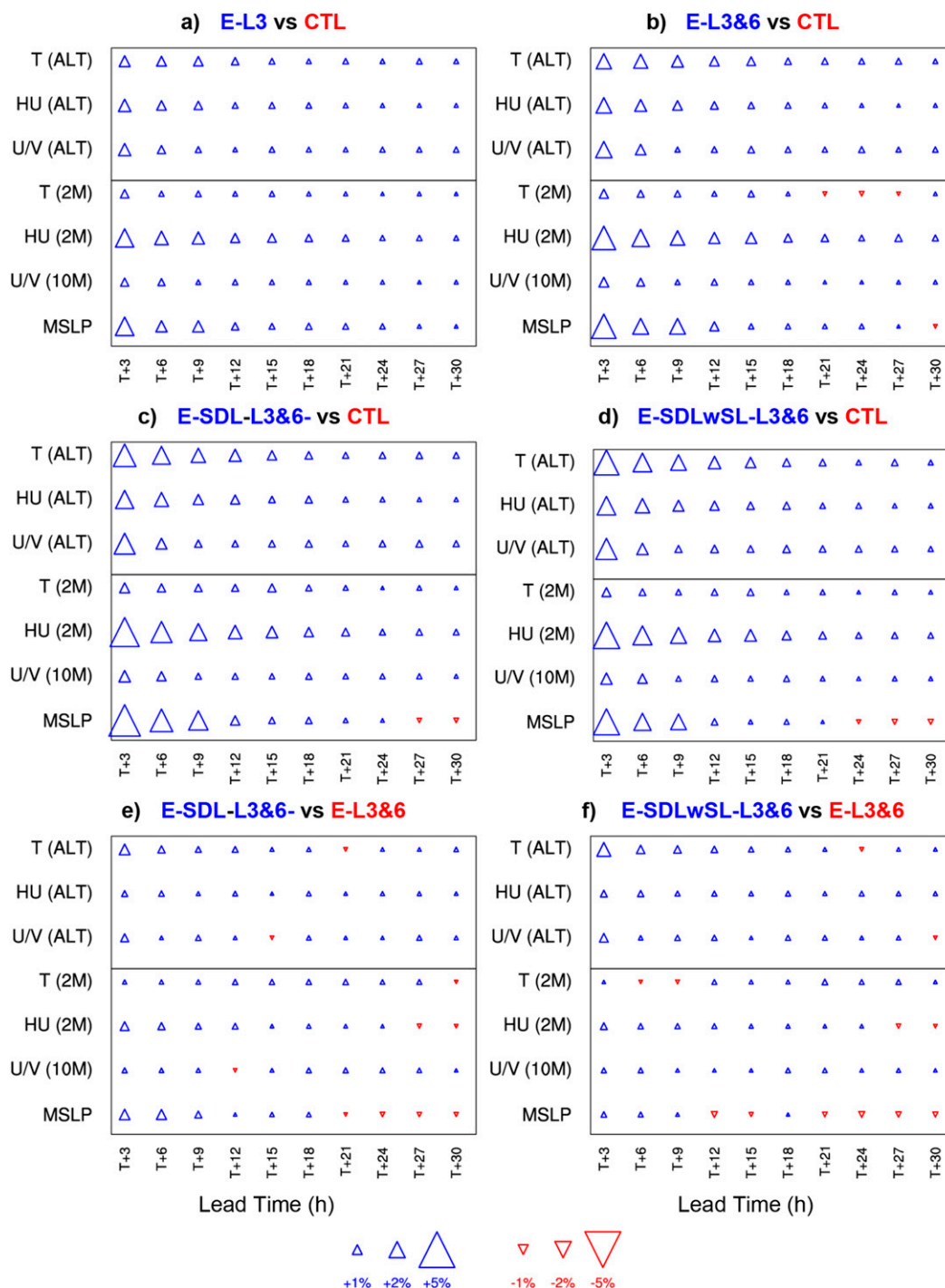
FIG. 8. As in Fig. 7, but for experiments using time-lagged ensemble members on its own or combined with scale-dependent localization. See Table 2 for the description of the experiments.

system at 3.8-km horizontal resolution, this paper examined different approaches to improve the ensemble-derived background error covariances that are exploited in this new data assimilation scheme without modifying the ensemble of background generation strategy. Two variants of the scale-dependent localization method that consist of applying appropriate (i.e., different) amounts of localization to different ranges of background error covariance spatial scales while simultaneously assimilating all of the available observations were examined

TABLE 5. As in Table 4, but with respect to E-L3&6.

Experiment	Altitude	Surface
E-L3&6	Control	
E-SDL-L3&6	0.22	**0.22**
E-SDLwSL-L3&6	**0.31**	0.08

and compared: 1) the original approach of Buehner (2012) in combination with spectral localization, which assumes that the covariance between the scales is zero, and 2) the more recent formulation of Buehner and Shlyaeva (2015) that avoids the complete removal of the between-scale covariances imposed in the former formulation. Increasing the effective ensemble size in the representation of the background error covariance from the use of time-lagged members was also considered both on its own and in combination with scale-dependent localization.

The results from data assimilation cycles over a 33-day winter period show the scale-dependent localization approach of Buehner (2012) performs better than the more recent formulation of Buehner and Shlyaeva (2015) when the background error covariances are derived from the most recent 25-member ensemble forecasts. However, when increasing the effective ensemble size to 75 members with time-lagged forecasts, the two scale-dependent formulations provide similar forecast improvements overall, albeit the approach of Buehner (2012) still performs better for variables in altitude. In contrast, the formulation of Buehner and Shlyaeva (2015) was found to provide the largest improvements for surface variables, where the heterogeneity of covariances could be greater due to, for example, land/sea contrasts, land-usage variability, and topographic effects. These results support the hypothesis of Buehner and Shlyaeva (2015) that local spatial averaging enforced by the elimination of the between-scale covariance in the method of Buehner (2012) may only be beneficial with very small ensemble sizes [i.e., $O(10)$ members] and that their alternative scale-dependent formulation should become advantageous with larger ensemble sizes [i.e., $O(100)$ members].

In every scenario, the positive impacts of each scale-dependent localization method gradually decrease with forecast lead times. This can be attributed in part to the intrinsic nature of limited-area data assimilation, where the impact of the lateral boundary conditions reduces the impact of any improvements to the initial conditions with increasing lead time, especially in cases of a strong synoptic forcing. Nevertheless, a similar dissipating positive signal can be seen in the results of Lorenc (2017) in a global context with SDLwSL (see his Fig. 14g), especially in the Northern Hemisphere, though the improvements

noted in the latter study lasted much longer (about 3 days), compared to our limited-area application (about 15 h). Using the SDL method in a global application, CB18 also reported a dissipating signal in the Northern Hemisphere (see their Fig. 9) that lasted for about 5 days, but, on the other hand, the beneficial impact in the Southern Hemisphere (tropics) was found to increase (stay constant) up to days 5 and 6 (up to day 7). These results could suggest that most of the time, scale-dependent localizations tend to improve mostly the smallest scales, which have a shorter predictability time.

Increasing the effective ensemble size in the ensemble-derived background error covariance with time-lagged ensemble members on its own improves the forecast performances as a function of the number of added ensemble members. Again, the largest benefits were found at short lead time but lasted somewhat longer than with scale-dependent localizations. These results are compatible with those of Wang et al. (2017), Gustafsson et al. (2014), and Lorenc (2017), though the latter employed time-lagged perturbations in combination with a time-shifting strategy. Here, the improvements from using time-lagged members outperform the impact of scale-dependent localizations on its own, whereas Lorenc (2017) reported about equal average benefits from the two approaches (but with a larger ensemble size). As in Lorenc (2017), the largest forecast improvements are obtained when combining the two approaches.

As stressed in CB18, finding the optimal scale-dependent localization configuration (i.e., the number of wave bands, the filtering response function design, and the amount of localization for each of the wave bands) is not trivial and still relies on ad hoc procedures. The methodology of Ménétrier et al. (2015a,b), however, allows localization length scales to be obtained that are close to the optimal values (i.e., those values found to result in the best forecast scores). Therefore, we cannot conclude that the results presented here with a 3DEnVar at 3.8-km horizontal resolution depict all of the benefits that could be obtained with this method in the AROME-France system that relies on a 1.3-km resolution.

The outcomes of this study will impact the future work in each of our organizations. At ECCC, it is planned to perform a similar investigation with the new prototype 4DEnVar scheme, developed for high-resolution application and presented by Bédard et al. (2018). This will shed light on the impact of the methods tested here in the context of a much larger ensemble (256 members) and a much larger domain (pan-Canadian; about 5 times the size of the AROME-France domain). At Météo-France, work is ongoing using a dual-resolution EnVar, with the goal of computing analyses from different flavors of EnVar (4DEnVar, hybrids) at the operational

resolution of 1.3 km, while keeping an EDA at a lower resolution that is comparable to the one used in this study. Given their clear positive contributions in the EnVar system, the scale-dependent localization, as well as the use of time-lagged ensembles, will be exploited in this context.

REFERENCES

Bédard, J., M. Buehner, J.-F. Caron, S.-J. Baek, and L. Fillion, 2018: Practical ensemble-based approaches to estimate atmospheric background error covariances for limited-area deterministic data assimilation. *Mon. Wea. Rev*, **146**, 3717–3733, https://doi.org/10.1175/MWR-D-18-0145.1.

Belo Pereira, M., and L. Berre, 2006: The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Wea. Rev.*, **134**, 2466–2489, https://doi.org/10.1175/MWR3189.1.

Berre, L., 2000: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model. *Mon. Wea. Rev.*, **128**, 644–667, https://doi.org/10.1175/1520-0493(2000)128<0644:EOSAMF>2.0.CO;2.

——, and G. Desroziers, 2010: Filtering of background error variances and correlations by local spatial averaging: A review. *Mon. Wea. Rev.*, **138**, 3693–3720, https://doi.org/10.1175/2010MWR3111.1.

——, H. Varella, and G. Desroziers, 2015: Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quart. J. Roy. Meteor. Soc.*, **141**, 2803–2812, https://doi.org/10.1002/qj.2565.

Bishop, C. H., and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111, https://doi.org/10.1111/j.1600-0870.2008.00372.x.

——, and ——, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.*, **139**, 1241–1255, https://doi.org/10.1175/2010MWR3403.1.

Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the AROME-France system. *Quart. J. Roy. Meteor. Soc.*, **142**, 2231–2243, https://doi.org/10.1002/qj.2822.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, https://doi.org/10.1256/qj.04.15.

——, 2012: Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Wea. Rev.*, **140**, 617–636, https://doi.org/10.1175/MWR-D-10-05052.1.

——, and M. Charron, 2007: Spectral and spatial localization of background-error correlation for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 615–630, https://doi.org/10.1002/qj.50.

——, and A. Shlyaeva, 2015: Scale-dependent background-error covariance localisation. *Tellus*, **67A**, 28027, https://doi.org/10.3402/tellusa.v67.28027.

Caron, J.-F., and M. Buehner, 2018: Scale-dependent background error covariance localization: Evaluation in a global deterministic weather forecasting system. *Mon. Wea. Rev.*, **146**, 1367–1381, https://doi.org/10.1175/MWR-D-17-0369.1.

Derber, J., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.*, **19**, 1333–1347, https://doi.org/10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2.

——, and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation. *system. Tellus*, **51A**, 195–221, https://doi.org/10.1034/j.1600-0870.1999.t01-2-00003.x.

Desroziers, G., J. T. Camino, and L. Berre, 2014: 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations. *Quart. J. Roy. Meteor. Soc.*, **140**, 2097–2110, https://doi.org/10.1002/qj.2325.

Fisher, M., 2003: Background error covariance modelling. *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, Shinfield Park, Reading, United Kingdom, ECMWF, 45–63, https://www.ecmwf.int/en/elibrary/9404-background-error-covariance-modelling.

Flowerdew, J., 2015: Towards a theory of optimal localisation. *Tellus*, **67A**, 25257, https://doi.org/10.3402/tellusa.v67.25257.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Gürol, S., A. T. Weaver, A. M. Moore, A. Piacentini, H. G. Arango, and S. Gratton, 2014: B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Quart. J. Roy. Meteor. Soc.*, **140**, 539–556, https://doi.org/10.1002/qj.2150.

Gustafsson, N., J. Bojarova, and O. Vignes, 2014: A hybrid variational ensemble data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Processes Geophys.*, **21**, 303–323, https://doi.org/10.5194/npg-21-303-2014.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118, https://doi.org/10.3402/tellusa.v35i2.11425.

Houtekamer, P. L., and F. Zhang, 2016: Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

——, L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242, https://doi.org/10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.

Lei, L., and J. L. Anderson, 2014: Empirical localization of observations for serial ensemble Kalman filter data assimilation in an atmospheric general circulation model. *Mon. Wea. Rev.*, **142**, 1835–1851, https://doi.org/10.1175/MWR-D-13-00288.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, https://doi.org/10.1256/qj.02.132.

——, 2017: Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quart. J. Roy. Meteor. Soc.*, **143**, 1062–1072, https://doi.org/10.1002/qj.2990.

Lu, C., H. Yuan, B. E. Schwartz, and S. G. Benjamin, 2007: Short-range numerical weather prediction using time-lagged ensembles. *Wea. Forecasting*, **22**, 580–595, https://doi.org/10.1175/WAF999.1.

Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre, 2015a: Linear filtering of sample covariances for ensemble-based

data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Wea. Rev.*, **143**, 1622–1643, https://doi.org/10.1175/MWR-D-14-00157.1.

——, ——, ——, and ——, 2015b: Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model. *Mon. Wea. Rev.*, **143**, 1644–1664, https://doi.org/10.1175/MWR-D-14-00156.1.

Michel, Y., 2013: Estimating deformations of random processes for correlation modelling: Methodology and the one-dimensional case. *Quart. J. Roy. Meteor. Soc.*, **139**, 771–783, https://doi.org/10.1002/qj.2007.

Miyoshi, T., and K. Kondo, 2013: A multi-scale localization approach to an ensemble Kalman filter. *SOLA*, **9**, 170–173, https://doi.org/10.2151/sola.2013-038.

Montmerle, T., Y. Michel, E. Arbogast, B. Ménétrier, and P. Brousseau, 2018: A 3D ensemble variational data assimilation scheme for the limited-area AROME model: Formulation and preliminary results. *Quart. J. Roy. Meteor. Soc.*, **144**, 2196–2215, https://doi.org/10.1002/qj.3334.

Purser, R., W. Wu, D. Parrish, and N. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.*, **131**, 1524–1535, https://doi.org/10.1175/1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.

Wang, Y., J. Min, Y. Chen, X.-Y. Huang, M. Zeng, and X. Li, 2017: Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event. *Atmos. Res.*, **183**, 1–16, https://doi.org/10.1016/j.atmosres.2016.07.026.

Yuan, H., C. Lu, J. A. McGinley, P. J. Schultz, B. D. Jamison, L. Wharton, and C. J. Anderson, 2009: Evaluation of short-range quantitative precipitation forecasts from a time-lagged multimodel ensemble. *Wea. Forecasting*, **24**, 18–38, https://doi.org/10.1175/2008WAF2007053.1.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, https://doi.org/10.1175/2009MWR2645.1.