Filtering properties of wavelets for local background-error correlations

Olivier Pannekoucke, Loïk Berre, Gérald Desroziers

QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **133**: 363–377 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.133

Filtering properties of wavelets for the local background error correlations

O. Pannekoucke*, L. Berre and G. Desroziers
GAME/CNRM (Météo-France, CNRS), Toulouse, France

Abstract:
Background error covariances can be estimated from an ensemble of forecast differences. The finite size of the ensemble induces a sampling noise in the calculated statistics. It is shown formally that a wavelet diagonal approach amounts to locally averaging the correlations, and its ability to spatially filter this sampling noise is thus investigated experimentally.
This is first studied in a simple analytical one dimensional framework. The capacity of a wavelet diagonal approach to model the scale variations over the domain is illustrated. Moreover, the sampling noise appears to be better filtered than when only using a Schur filter, in particular for small ensembles.
The filtering properties are then illustrated for an ensemble of Météo-France Arpège forecasts. This is done both for the "time-averaged correlations", and for the "correlations of the day". It is shown that the wavelets are able to extract some length scale variations that are related to the meteorological situation.

KEY WORDS Spherical wavelet; wavelet on the circle; background error covariance; assimilation ensemble; sampling noise; Ensemble Kalman Filter.

Received 20 December 2005; Revised 22 November 2006; Accepted 26 November 2006

1 Introduction

Most data assimilation schemes seek to provide an optimal combination of observations and of a background given by a short-term forecast. The optimal analysis is basically derived from statistical estimation theory. In such a theory, the two sets of information are associated with covariance matrices corresponding to their respective errors. The error covariance matrices determine the respective weights given to each piece of information in the analysis. However, the correct specification of those statistics remains a major challenge in data assimilation systems.

The estimation of background error covariances is a particularly difficult problem since in operational practice the background state is a vector of dimension $10^5 - 10^7$. In that case, it is not only intractable to handle such a huge corresponding error covariance matrix, but it is also impossible to specify it exactly, since there is a lack of available statistical information (Dee 1995).

To overcome these difficulties, a statistical model for the background error covariances has to be defined. Such a model often relies on the hypothesis that the background error correlations are homogeneous and isotropic (Gaspari and Cohn 1999). This assumption is equivalent to considering that the background error correlation matrix is diagonal in spectral space (Courtier et al 1998) and thus facilitates the representation of background error statistics.

One technique for specifying the background error covariance matrix is to use an ensemble of assimilations, obtained by a perturbation of observations and of the background (Houtekamer et al. 1996). This procedure has recently been applied at ECMWF and Météo-France for specifying the stationary component of background error covariances (Fisher 2003; Belo Pereira and Berre 2006). In that case, the covariances are computed over several weeks and the hypothesis of homogeneity is often assumed. Such an approach is also partly related to the Ensemble Kalman Filter (EnKF), originally proposed by Evensen (1994), where flow-dependent covariances are calculated from the ensemble.

The hypotheses of homogeneity and isotropy, are known to be rather crude representations for the "real" error structures. Lönnberg (1988) suggested that horizontal and vertical correlations vary geographically: horizontal scales tend to be broader in the tropics than at high latitudes because of atmospheric dynamics (Ingleby 2001). Bouttier (1993, 1994) also showed that correlation scales depend on the meteorological situation and on data density.

Using an ensemble of assimilations, Belo Pereira and Berre (2006) have shown such heterogeneities and

*Correspondence to: Météo-France CNRM/GMAP, 42 av. G. Coriolis, 31057 Toulouse Cedex France. e-mail: olivier.pannekoucke@meteo.fr

anisotropies in the background error correlations by using a new economical algorithm for estimating the correlation length scales.

The EnKF is one approach for obtaining heterogeneous and flow-dependent background error correlations. However, due to the relatively small sampling size, covariances are noisy and have to be filtered by an additional treatment. A Schur product (Houtekamer and Mitchell 2001) is generally applied to the raw statistics. The use of the Schur product in EnKF has been discussed by Lorenc (2003).

From a different point of view, Fisher (2003) has recently introduced the idea to use wavelets on the sphere to improve the representation of background error correlations, and in particular to allow some heterogeneity in the description of those errors. Such a formulation is now operationally applied at ECMWF to represent stationary but heterogeneous correlations, with for example larger correlations in the Tropics. A similar approach has been considered by Deckmyn and Berre (2005) for a limited area model.

The wavelet representation of background error covariances implemented at ECMWF has been obtained by using an ensemble of analyses over several weeks, which is expected to provide valid statistics. Moreover, there is scope to combine the use of wavelets and ensembles in order to obtain at the same time heterogeneous and flow-dependent background error correlations.

The aim of the paper is to show that wavelets provide an effective tool to allow some variability in the correlations, but also to filter the noise due to the small size of an ensemble. These properties of the wavelet formulation are in particular investigated by using the diagnostic of the local correlation length scale proposed by Belo Pereira and Berre (2006). It may be also mentioned that some analogous filtering effects are under investigation at the Meteorological Service of Canada (Buehner and Charron 2006), through spectral and spatial localization.

The structure of the paper is the following. In section 2, we explain that a wavelet diagonal approach amounts to applying a local spatial averaging of covariance functions. This allows the sample size to be increased and to preserve the representation of geographical variations. The ability of wavelets to extract useful information from a small ensemble is first discussed in section 3 for a toy analysis problem on a circle. Section 4 shows the application of the same wavelet representation in 2D on the sphere, with actual background errors provided by an ensemble of forecasts from the Météo-France Arpège system. Results are produced both over a long period, and on a single date with only a few members. Conclusions are given in section 5.

2 Wavelet spatial averaging of covariance functions

2.1 Local covariance functions

For the sake of simplicity, we will consider a 1D cyclic domain in this section. Derivations are also valid in a 2D cyclic domain, by considering horizontal positions as vectors with two components (x and y). They can be also generalized easily to 2D spherical domains and to 3D contexts (by including the appropriate metrical terms in the formulae).

The separation s between two positions x and x" is defined as the difference $s = x" - x$. Note that s can be positive or negative: the absolute value $|s|$ is the separation distance, while the sign of s corresponds to the orientation of the separation. (In a 2D cyclic context, the separation direction is given by the argument of s, when seeing s as a complex number: $s = s_x + is_y = |s|\, exp(i\, arg(s))$, where s_x and s_y are the x and y components of s).

The local error covariance function $f^x(s)$ at a reference point x is often calculated from an ensemble of N_e forecast differences ε, according to the following equation (in a horizontal context) :

$$f^x(s) = \overline{\varepsilon(x)\varepsilon(x+s)} = \frac{1}{N_e}\sum_k \varepsilon(x,k)\varepsilon(x+s,k),$$

where $\varepsilon(x,k)$ is a forecast error realization at position x, s is the separation value between the two considered points, k is the ensemble member index, and the overline is the ensemble average. This ensemble average is thus calculated over N_e members of the ensemble.

The sampling size is thus equal to N_e only. A first problem is that the finite size of the sample induces a noise, which often renders necessary the use of e.g. a Schur filter (Houtekamer and Mitchell 2001). A second important problem is that the sample covariance matrix will be rank deficient if the sample size is too small. This will limit the analysis increments to lie in a low-dimensional subspace in the analysis. In the remainder of the study, we will focus on the first problem (sampling noise).

2.2 Spectral diagonal approach: a global spatial averaging

It is possible to represent the background error covariance matrix $\boldsymbol{B}$ in spectral space (e.g. Courtier et al 1998). Under the assumption of horizontal homogeneity, the spectral covariance matrix is simply a diagonal matrix (or block-diagonal in a three-dimensional context, with vertical covariances in the diagonal blocks). The diagonal of $\boldsymbol{B}$ contains the variances of the error spectral coefficients.

Equivalently, it can be shown (see appendix A) that this spectral diagonal approach amounts to calculating $\frac{1}{N_g}\sum_{x'} f^{x'}(s)$, which is the global spatial average of the covariance functions (N_g is the number of gridpoints in the domain). The resulting local covariance functions, noted $f^x_{\boldsymbol{S}}(s)$, where the subscript $\boldsymbol{S}$ refers to the spectral diagonal approach, are all equal to this global spatial average:

$$f^x_{\boldsymbol{S}}(s) = \frac{1}{N_g}\sum_{x'} f^{x'}(s),$$

or equivalently

$$f^x_{\boldsymbol{S}}(s) = \frac{1}{N_g\, N_e}\sum_{x'}\sum_k \varepsilon(x',k)\varepsilon(x'+s,k).$$

DOI: 10.1002/qj

The globally averaged covariance function is thus estimated as an average over a number of pairs of error realizations, which is equal to $\mathcal{N} = N_g N_e$ (instead of $\mathcal{N} = N_e$ when estimating the local covariance functions).

Such a huge increase of sample size has an obvious counterpart: the *global spatial average* does not allow one to represent any geographical variations. Therefore, one may wonder if it is possible to consider a *local spatial average*, in order both to increase the sampling size and to represent some geographical variations.

2.3 Wavelet diagonal approach: a set of local spatial averages

The direct and inverse wavelet transforms, which are used by Fisher (2003), are defined as follows: $\hat{\boldsymbol{\varepsilon}}_j = \boldsymbol{\varepsilon} \otimes \boldsymbol{\psi}_j$ and $\boldsymbol{\varepsilon} = \sum_j \hat{\boldsymbol{\varepsilon}}_j \otimes \boldsymbol{\psi}_j$, where $\otimes$ is the convolution product in physical space for both expressions, and $\boldsymbol{\psi}_j$ are radial band-pass functions for different scales j (see Courtier et al 1998 for the properties of the convolution with radial functions on the sphere). It may be mentioned that such functions $\boldsymbol{\psi}_j$ are not orthogonal, and different from traditional wavelets, which instead are orthogonal with respect to integer dilation and translation on a regular lattice.

The functions $\hat{\boldsymbol{\varepsilon}}_j$ and $\boldsymbol{\psi}_j$ can be represented in grid-point space. Some examples of wavelet functions $\boldsymbol{\psi}_j$ are shown in a one dimensional framework (Fig. 1) and can be compared with spectral functions. Note that each wavelet function has both a specific position and a specific scale. The associated coefficients $\hat{\varepsilon}_{j,x_j(i)}$ of the transformed error field correspond to the error values at scale j and at position $x_j(i)$ on a grid whose resolution depends on j (with $i = 1, N_x(j)$).

In a horizontal framework, the wavelet diagonal approach for $\boldsymbol{B}$ consists in calculating variances of these wavelet coefficients $\hat{\varepsilon}_{j,x_j(i)}$. As each wavelet function contains information both on position and scale, the wavelet variances contain information on the local shape of the covariance functions.

Moreover, it can be shown (see appendix A) that this wavelet diagonal approach amounts to computing a set of weighted local spatial averages of the covariance functions. The resulting local covariance functions, noted $f^x_{\boldsymbol{W}}(s)$ with the subscript $\boldsymbol{W}$ refering to the wavelet diagonal approach, can be expressed as follows:

$$f^x_{\boldsymbol{W}}(s) = \sum_{x',s'} f^{x'}(s') \Phi^{x,s}(x',s'),$$

where $\Phi^{(x,s)}(x',s')$ is defined by

$$\sum_j \sum_{i=1}^{N_x(j)} \psi_j(x' - x_j(i))\, \psi_j(x' - x_j(i) + s')\, \psi_j(x - x_j(i))\, \psi_j(x - x_j(i) + s)$$

and can be seen as a weighting coefficient in the calculation of the spatial average.

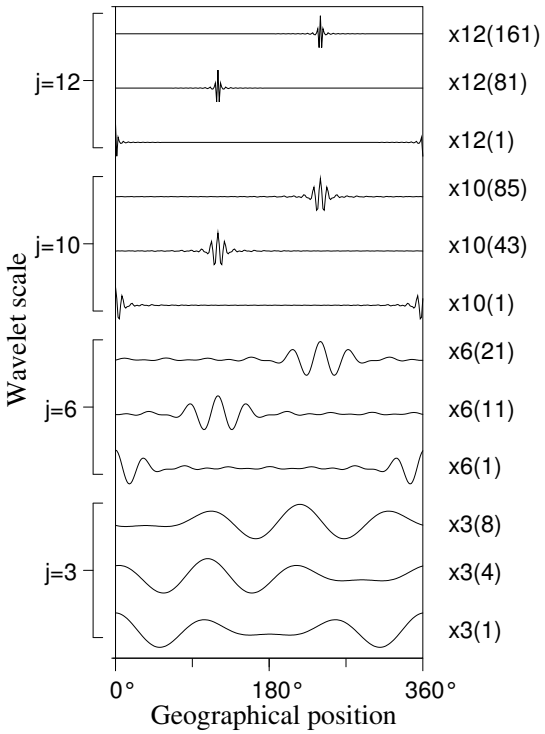

Figure 1. Some wavelet functions $\psi_j(x - x_j(i))$ for different scales j and different points $x_j(i)$.

As expected, $\Phi^{x,s}(x',s')$ will give more weight to positions x' in the neighbourhood of x, and to separation values s' that are close to s. Figure (2) represents two examples of function $\Phi^{x,s}(x',s')$, for $x = 121$ and $s = s' = 20$, and for two choices of wavelet bands. The larger weight given to position x' close to x is illustrated by the larger values of $\Phi^{x,s}(x',s')$ when x' is near x.

As in the spectral approach, the implied functions $f^x_{\boldsymbol{W}}(s)$ are the result of a spatial average. The spatial sample size is nevertheless likely to be smaller than in the spectral case, because the weighting functions $\Phi^{x,s}(x',s')$ have values close to zero except in the neighbourhood of x (as illustrated in fig. 2).

The two examples in fig. (2) differ by the choice of wavelet bandwidths. It may thus be noticed that the spatial average will tend to be more localized when the bandwidth is larger. This is related to the fact that the bandwidth determines the trade-off between spatial and spectral resolutions, as will be further discussed in section 2(e).

It may also be mentioned that the calculation, representation and spatial filtering of the matrix $\boldsymbol{B}$ remain efficient with a wavelet diagonal approach. It only requires the calculation of a diagonal matrix, which contains the wavelet variances. This is much cheaper than calculating a full gridpoint covariance matrix, and then applying a spatial averaging operator.

To summarize, the wavelet approach is similar to the spectral approach, in the sense that the local covariance functions are averaged spatially. This potentially allows one to reduce the level of sampling noise, compared to the

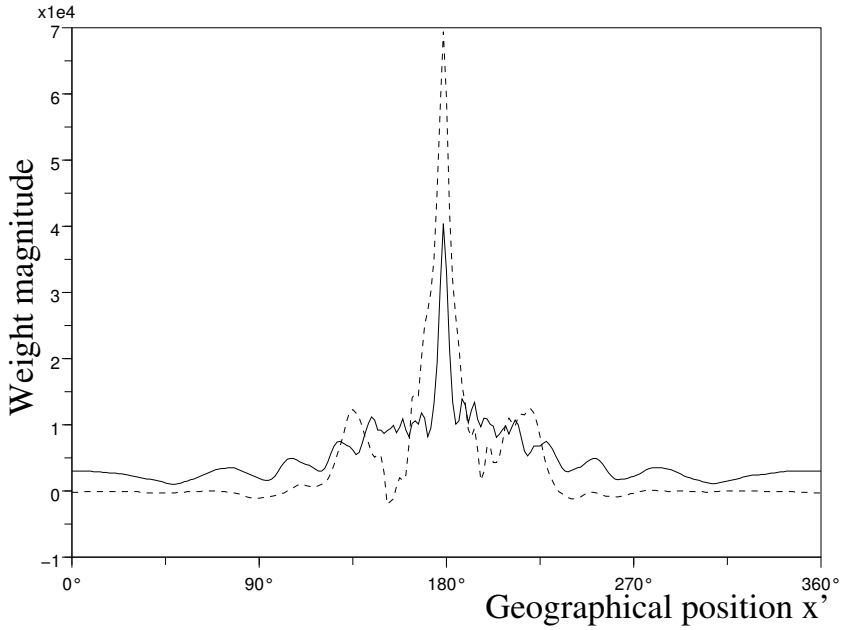

Figure 2. Representation of $\Phi^{x,s}(x',s')$ for $x = 121$ (corresponding to 180 °) and $s = s' = 20$ (corresponding to 30 °) and for two different choices of wavelet bands, as defined by the cutoff wave numbers $\{N_j\}$ (see section 2(e)) : a set of relatively tight bands $\{N_j\} = \{0, 1, 2, 3, 5, 7, 10, 15, 21, 30, 42, 63, 120\}$ (solid line) and a set of relatively wide bands $\{N_j\} = \{0, 4, 8, 12, 20, 28, 40, 60, 84, 120\}$ (dashed line).

estimation of the local covariance functions. Moreover, as the spatial averaging is local instead of global, it remains possible to represent some geographical variations of the covariance functions.

These two features (spatial filtering and geographical variations) will be studied experimentally in two different frameworks, in sections 3 and 4 respectively.

2.4 Isotropy of wavelets and of covariance averaging

As illustrated in fig. (1) on the circle and in Fisher (2004) on the sphere, the band-pass functions ψ_j are radial. This implies that the corresponding local averaging of covariances tends to be isotropic. In other words, covariances along different separation directions are averaged together, for a given separation distance.

On the one hand, this may be seen as a drawback, as it prohibits the representation of possible anisotropies. This limitation has been illustrated and discussed by Deckmyn and Berre (2005) with Meyer wavelets. On the other hand, averaging over several directions allows the sample size to be further increased, which can reduce the amplitude of sampling error. In other words, if the covariances are nearly isotropic, this isotropic averaging may rather be beneficial.

The balance between these pros and cons will thus depend on the degree of anisotropy of the actual covariances and on the available ensemble size. Another related point is that part of the covariance anisotropies arises from the geographical variations of the background error standard deviations. The latter can be represented in gridpoint space, while modelling correlations in wavelet space, as in Fisher (2003).

2.5 Details of the wavelet diagonal approach

The concept of local averaging of covariance functions has been introduced in the previous section. This concept can also be considered for correlation functions, by applying the equations to the background error normalized by the gridpoint standard deviations, namely $\varepsilon'(x,k) = \varepsilon(x,k)/\sigma_b(x)$, with $\sigma_b(x) = (\frac{1}{N_e}\sum_k \varepsilon(x,k)^2)^{1/2}$. This will be the approach used in the remainder of the paper, in order to have a similar covariance formulation as Fisher (2003) and Deckmyn and Berre (2005). The formulation of $\boldsymbol{B}$ is often determined by the design of $\boldsymbol{B}^{-1/2}$, which is detailed in appendix B. Thus, in a horizontal context, the square root of the wavelet-modelled gridpoint covariance matrix can be written

$$\boldsymbol{B}_w^{1/2} = \boldsymbol{\Sigma}_g \boldsymbol{\Sigma}_s \boldsymbol{W}^{-1} \boldsymbol{D}_w^{1/2}, \tag{1}$$

where $\boldsymbol{\Sigma}_g$ is a diagonal matrix of gridpoint standard deviations, $\boldsymbol{\Sigma}_s$ corresponds to a multiplication by spectral standard deviations, and $\boldsymbol{D}_w^{1/2}$ is a diagonal matrix of wavelet standard deviations. Matrices $\boldsymbol{W}$ and $\boldsymbol{W}^{-1}$ correspond respectively to the direct and inverse wavelet transforms (note that $\boldsymbol{W}$ is a rectangular matrix, and that $\boldsymbol{W}^{-1}$ is the left inverse of $\boldsymbol{W}$).

The particular wavelet functions ψ_j introduced by Fisher (2003) on the sphere are band-limited and defined in spectral space as follows. For $(N_j)_{j\in[0,J]}$, with $N_j < N_{j+1}$, the spectral coefficients of functions ψ_j are given by, for $j \neq 0$ (n being the total wave number in the spherical case): $\frac{1}{\sqrt{2n+1}}\check{\psi}_{j,n} = \sqrt{\frac{n-N_{j-1}}{N_j-N_{j-1}}}$ for $N_{j-1} \leq n < N_j$, $\sqrt{\frac{N_{j+1}-n}{N_{j+1}-N_j}}$ for $N_j \leq n < N_{j+1}$, and 0 otherwise. For $j = 0$, the definition is the same, except that the range

DOI: 10.1002/qj

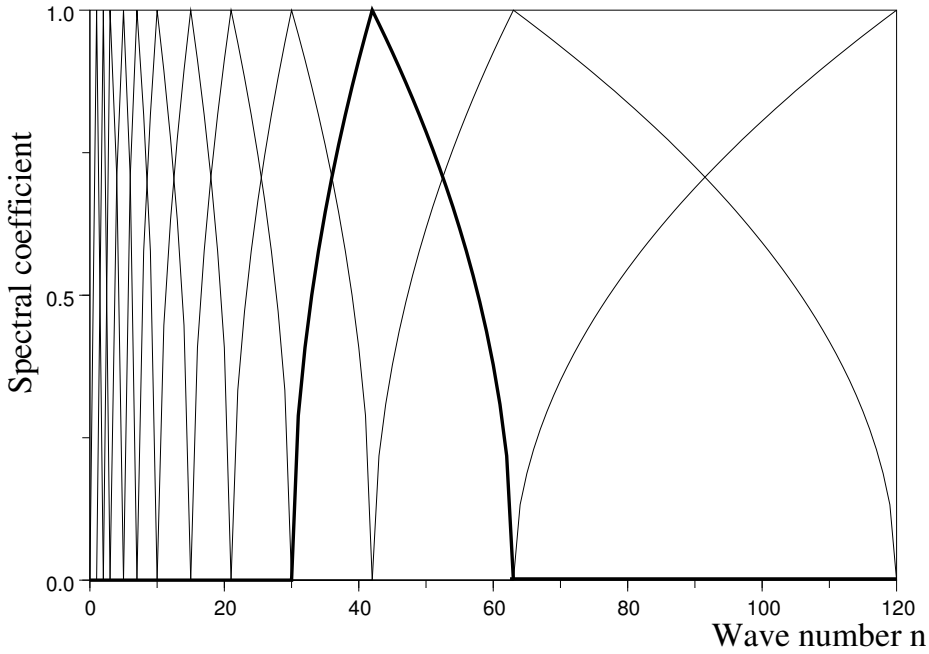

Figure 3. Spectral coefficients of wavelet functions $\psi_j(x)$ for different scales j (there is one curve for each function) and truncation $T = 120$. The spectrum associated with a particular $\psi_j(x)$ function ($j = 10$) is shown in bold.

$N_{j-1} \leq n < N_j$ is replaced by $0 \leq n < N_0$, for which $\frac{1}{\sqrt{2n+1}}\check{\psi}_{j,n} = 1$.

It is possible to define equivalent wavelets in a one dimensional Fourier space. The corresponding Fourier spectral coefficients $\check{\psi}_{j,n}$ are represented in Fig. 3 , for the following set $\{N_j\} = \{0, 1, 2, 3, 5, 7, 10, 15, 21, 30, 42, 63, 120\}$. This set has been chosen in order to be similar to Fisher (2003). Ways to define an optimized choice for this set may be explored in future studies. In particular, as discussed by Fisher and Andersson (2001), it is the choice of the bandwidths which determines the trade-off between spectral and spatial resolution. When the bands are broader, the spectral variations are smaller, but the geographical variations are allowed to be larger. This is connected to the discussion of Figure 2 in section 2(c). The spatial averaging is more localized when the bands are broader, which allows more geographical variations to be represented.

It may also be mentioned that each wavelet field $\hat{\boldsymbol{\varepsilon}}_j$ can be represented exactly on a low-resolution grid, which corresponds to a truncation equal to $T_j = min(N_{j+1}, T)$, T being the maximum truncation, associated to the original full resolution grid ($T = 120$ in the example above). The direct wavelet transform is thus applied as follows:

$$\hat{\boldsymbol{\varepsilon}} = \begin{pmatrix} . \\ \hat{\boldsymbol{\varepsilon}}_j \\ . \end{pmatrix} = \boldsymbol{W}\boldsymbol{\varepsilon} = \begin{pmatrix} . \\ \boldsymbol{G}_j\check{\boldsymbol{\Psi}}_j\boldsymbol{S}_j \\ . \end{pmatrix} \boldsymbol{\varepsilon},$$

where $\check{\boldsymbol{\Psi}}_j$ is the diagonal matrix containing the $\check{\psi}_{j,n}$ spectral coefficients, $\boldsymbol{S}_j$ is the spectral transform associated with truncation T_j, and $\boldsymbol{G}_j$ is the corresponding inverse transform onto a grid that corresponds to truncation T_j. The inverse wavelet transform is conversely defined by $\boldsymbol{\varepsilon} = \boldsymbol{W}^{-1}\hat{\boldsymbol{\varepsilon}} = \boldsymbol{G}_J \sum_{j=1}^{J} \check{\boldsymbol{\Psi}}_j\boldsymbol{S}_j\hat{\varepsilon}_j$, where $\boldsymbol{G}_J$ is the full resolution inverse spectral transform (since $T_J = T$). A representation of an example of wavelet field $\hat{\boldsymbol{\varepsilon}}$ will be shown in figure 4.

3 Wavelet filtering properties in a 1D analytical framework

3.1 A simple 1D analytical case with varying length scales

A simple 1D analytical framework has been considered to highlight the filtering properties of wavelets. In this framework, the geographical domain is supposed to be an earth great circle of radius a, and the coordinate x is the geographical position varying from $0\,^\circ$ to $360\,^\circ$ in terms of angle (or 0 to $2\pi a$ in terms of distance). On this circle, only one field is considered. A homogeneous covariance matrix is obtained from a radial correlation function $f_H^x(s) = e^{-\frac{s^2}{2L_H^2}}$, where x is a point on the circle, s is a separation value, and L_H is the length scale, which is here arbitrarily set equal to 250 km.

Then, a heterogeneous correlation is computed using a c-stretching Schmidt transformation (Courtier and Geleyn 1988), adapted to the circle and defined by $h(x) = a\left[\pi - 2Atan\left(\frac{1}{c}tan(\frac{\pi}{2} - \frac{1}{2}\frac{x}{a})\right)\right]$ with $c = 2.4$ (the Schmidt transformation is used for a different purpose in the Arpège global stretched model to obtain a variable resolution). The resulting correlation function is:

$$f^x(s) = f_H^{h^{-1}(x)}(h^{-1}(x+s) - h^{-1}(x)).$$

The associated matrix obtained, noted $\boldsymbol{B}_a$, is characterized by heterogeneous correlations which are relatively

sharp around 180 ° and broad around 0 °. This is illustrated in the top panel of Fig. 4, which represents the local correlation length scales $L(x)$ at point x, approximated by (Belo Pereira and Berre 2006)

$$L^2(x) = \frac{\sigma(\varepsilon)^2(x)}{\sigma(\partial_x \varepsilon)^2(x) - (\partial_x \sigma(\varepsilon))^2(x)}, \qquad (2)$$

where $\sigma(\varepsilon)(x)$ is the standard deviation of $\varepsilon(x)$, and ∂_x is the derivative along the coordinate.

In the 1D framework, evaluating the length scales that are implied by the formulation (1) involves application of the gradient operator and its adjoint to the covariance matrix $\boldsymbol{B}_w$. For instance, the covariance matrix $\boldsymbol{B}'_{w,xx}$ of $\partial_x \varepsilon$ corresponds to

$$\boldsymbol{B}'_{w,xx} = \overline{(\partial_x \varepsilon)(\partial_x \varepsilon)^*} = \partial_x \overline{\varepsilon \varepsilon^*} \partial_x^* = \partial_x \boldsymbol{B} \partial_x^* \qquad (3)$$

The diagonal of $\boldsymbol{B}'_{w,xx}$ then provides the variances of $\partial_x \varepsilon$ that are used in the length scale equation.

3.2 Randomization of $\mathbf{B}_a$ and Schur product

In order to examine the effects of the sampling noise, random perturbations $\boldsymbol{\varepsilon}$ have been generated from $\boldsymbol{B}_a$: $\boldsymbol{\varepsilon} = \boldsymbol{B}_a^{1/2} \boldsymbol{\zeta}$, where $\boldsymbol{\zeta}$ is a random sample of a normal distribution with zero mean and the identity as a covariance matrix (Fisher and Courtier 1995). This allows one to obtain an ensemble covariance matrix $\boldsymbol{B}_e = \frac{1}{N_e} \sum_k \boldsymbol{\varepsilon}(k) \boldsymbol{\varepsilon}(k)^T$, where N_e is the ensemble size.

The middle panel of Fig. 4 shows an example of random error realization. One may notice that this field presents shorter variations near 180 ° (Z1 region) than near 0 ° (Z2 region), in accordance with the local length scales of the specified covariance function (upper panel of Fig. 4). These varying length scales are expected to be captured by the wavelet diagonal approach, with a larger amplitude of small scale wavelet variances in Z1 than in Z2.

This is supported by the bottom panel of Figure 4, which shows the amplitudes of the wavelet coefficients of the error field example. As expected, the amplitudes of small scale coefficients (for $j \geq 10$) tend to larger in Z1 than in Z2.

The ensemble covariance matrix $\boldsymbol{B}_e$ can be compared to the exact covariance matrix $\boldsymbol{B}_a$ to examine the effects of the sampling noise, and also to the corresponding wavelet ensemble covariance matrix $\boldsymbol{B}_w$ defined by equation (1). Sampling noise is a particularly important issue in assimilation schemes such as the Ensemble Kalman filter, and makes the use of a Schur product appropriate (Houtekamer et al 2001). The Schur product corresponds to an element-wise product with a matrix $\boldsymbol{F}_{L_s}$. A filtered ensemble covariance matrix $\boldsymbol{B}_e^{L_s}$ can be obtained with

$$\boldsymbol{B}_e^{L_s} = \boldsymbol{F}_{L_s} \circ \boldsymbol{B}_e,$$

where $\boldsymbol{F}_{L_s}$ is the matrix which corresponds to a compactly-supported correlation function. In this paper, a fifth-order piecewise rational function (Gaspari and Cohn 1999, Eq. (4.10)) is used, as in Houtekamer and Mitchell (2001).

The Schur filter ensures that the long distance correlations are set to zero, at the expense of an artificial sharpening of the correlation functions in the intermediate distances.

This filter does not change the local variances themselves, while they are also affected by sampling noise. A normalization by the diagonal values of $\boldsymbol{B}_e^{L_s}$ is therefore applied in addition, to ensure that $\boldsymbol{B}_e^{L_s}$ becomes a valid correlation matrix. This procedure allows us to concentrate on the sampling noise effects on the correlation functions.

The impact of $\boldsymbol{F}_{L_s}$ on the final correlation is sensitive. If L_s is too short (left panels of Fig. 5, which correspond to $L_s = 200\ km$), the exact correlation functions are replaced by an excessively homogeneous and sharp correlation function. In this case, the correlation related to the short length scale area Z1 (grid-point 121) is well represented, but the correlation function in the large length scale area Z2 (grid-point 1) is shorter than it should be.

Therefore, L_s may be increased (here to 1000 km), in order to preserve the large length scale in the Z2 area. However, in the Z1 area, some spurious oscillations in the short distances are no longer filtered in this case (Fig. 5, right panels).

The scale L_s has thus to be chosen judiciously. This optimal value depends also on the ensemble size (Houtekamer and Mitchell 2001, Lorenc 2003).

A Schur filter is also applied to provide a filtered version of the wavelet ensemble covariance matrix:

$$\boldsymbol{B}_w^{L_s} = \boldsymbol{F}_{L_s} \circ \boldsymbol{B}_w.$$

The relative necessity of this Schur filter for the wavelet approach and its effect are illustrated in Figure 6. It is a comparison between correlations calculated from a 10 member ensemble, either directly (solid line) or based on a wavelet diagonal approach (dashed line). Zeroing spurious long distance correlations remains desirable for the wavelet approach in this example, although this is much less marked than in the direct estimation. The effect of a Schur filter with $L_s = 6000$ km on the wavelet-based correlations is illustrated (bold solid line). It sets large distance correlations to zero. The possibility to avoid applying this Schur filter may be explored in future studies, for instance by using bandwidths that are broad enough to make the correlations sufficiently local.

In the remainder of the paper, $\boldsymbol{B}_e^{L_s}$ and $\boldsymbol{B}_w^{L_s}$ will be simply refered to as the ensemble and wavelet correlation matrices respectively.

3.3 Wavelet filtering of the correlation functions and of their variations

The wavelet filtering properties can be examined by comparing the left and right panels of Figure 7, which have been produced with $L_s = 6000$ km and $N_e = 10$ elements. A large value of L_s is chosen to illustrate the typical amplitude of sampling noise. The left panels show the raw

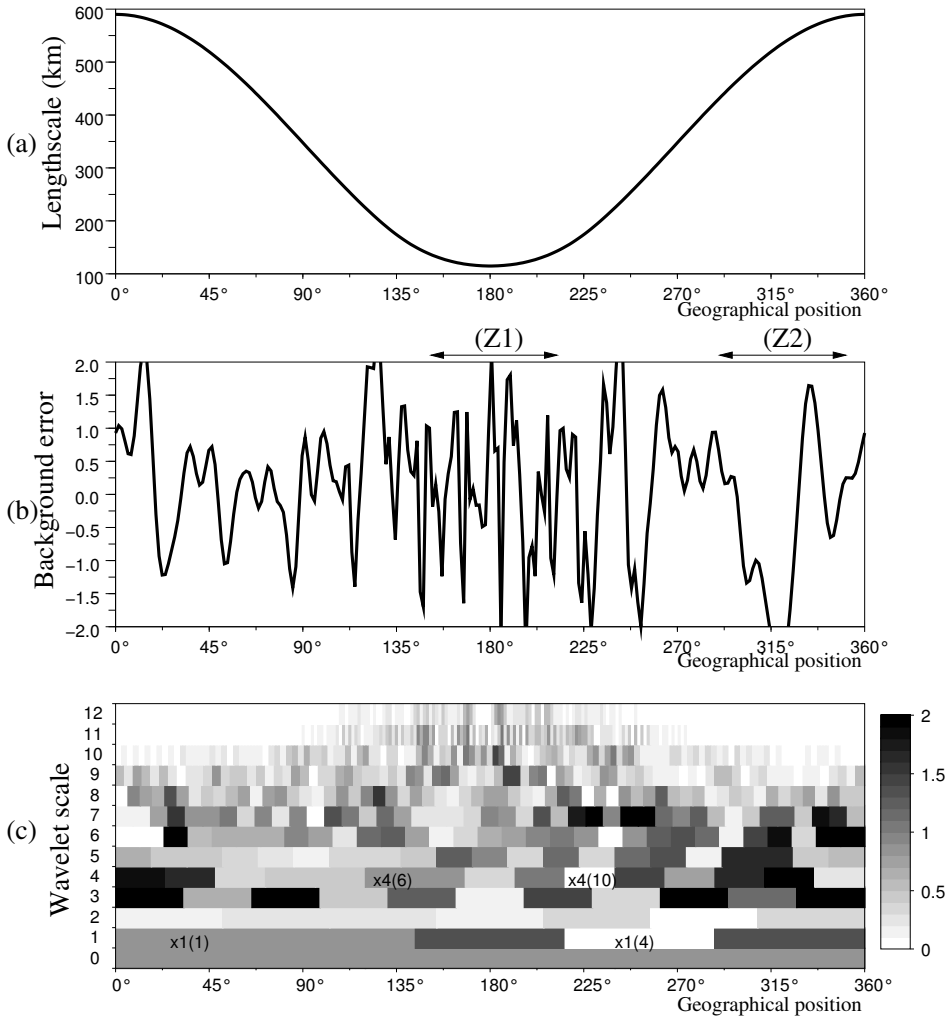

Figure 4. (a) Geographical variations of the length scale in the analytical framework. (b) Sample $\boldsymbol{\varepsilon}$ of a Gaussian background error associated with the analytical background covariance model on the circle and (c) the absolute value of its wavelet coefficients : each row is the absolute value of vector $\hat{\boldsymbol{\varepsilon}}_j = \boldsymbol{S}_j^{-1}\check{\boldsymbol{\Psi}}_j\boldsymbol{S}_j\boldsymbol{\varepsilon}$; for a given scale j there is one box per subgrid-point, such as $x_1(1), x_1(4)$ for $j = 1$ or $x_4(6), x_4(10)$ for $j = 4$.

ensemble correlations at gridpoints 1 (top panel) and 121 (bottom panel). The corresponding right panels are for the wavelet-based correlations at these gridpoints. The sampling noise is quite visible in the ensemble correlations, with many large and spurious oscillations. Such artefacts are much less marked in the wavelet correlations.

The wavelet formulation thus appears to partly filter the sampling noise. This is expected knowing that the wavelet diagonal assumption amounts to locally averaging the covariance functions, which implies an increase of the total size of the sample. A wavelet illustration of these filtering properties is also discussed in appendix C.

Fig. 8 illustrates the effects on the length scale variations of the sampling noise and of the wavelet filtering. The case $N_e = 10$ (top left panel) is the most spectacular. The raw length scales exhibit some large and spurious small scale oscillations, and the largest length scale values are often much exagerated (e.g. with 800 km raw values around 70 °, while the exact value is around 400 km). In contrast, the wavelet implied length scales have relatively smooth variations and accurate values. The wavelet approach thus appears to be able to capture and represent the main geographical variation of interest (increase of length scale towards 0 °) from a small ensemble with only 10 members.

Again the wavelet local averaging reduces the effects of sampling noise. These beneficial effects decrease when the size of the ensemble increases.

3.4 Data assimilation experiments

Assimilation experiments on the circle have also been performed. As the true correlations are known, the true solution can be computed, and can be compared with that of the correlation model.

In these experiments, the true state is taken as being zero everywhere. The background error is generated from the true correlation matrix (remember that the standard deviation is equal to 1) with $\boldsymbol{B}_a^{1/2}$, while observations are generated with $\boldsymbol{R}^{1/2}$ (which is the square-root of the observation error covariance matrix $\boldsymbol{R}$). $\boldsymbol{R}^{1/2}$ is assumed to be a diagonal matrix $\sigma_o\boldsymbol{I}$ where $\boldsymbol{I}$ denotes the identity matrix (thereafter $\sigma_o = 0.95$: the observations are

DOI: 10.1002/qj

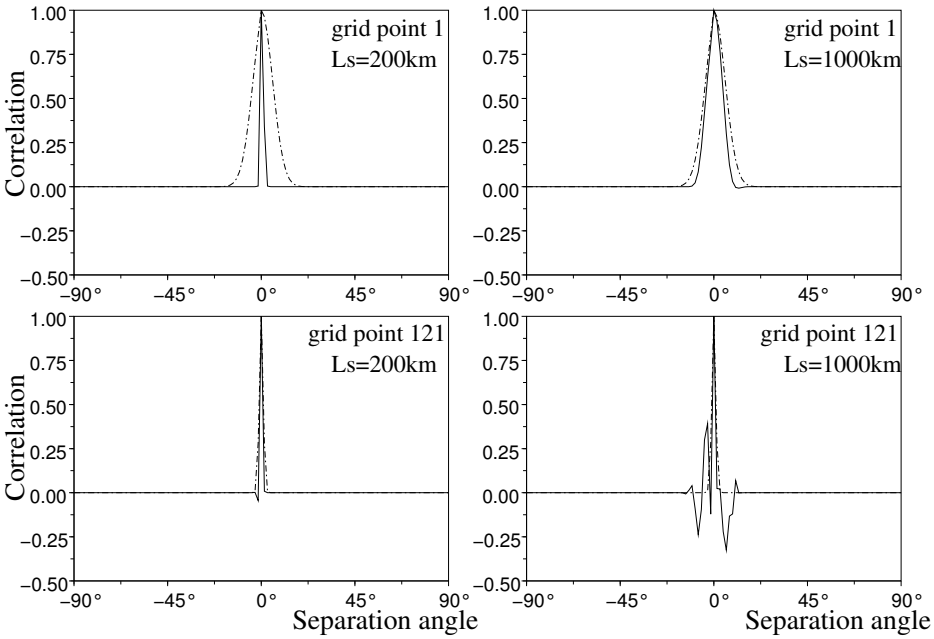

Figure 5. Ensemble correlations (solid line) related to grid points 1 (corresponding to the large length scale area) and 121 (corresponding to the short length scale area). Correlations are directly computed from a 10 element ensemble, and then filtered with a Schur filter $\boldsymbol{F}_{L_s}$ with $L_s = 200\ km$ (left panels) and $1000\ km$ (right panels). Exact correlations are also shown (chain-dotted line).

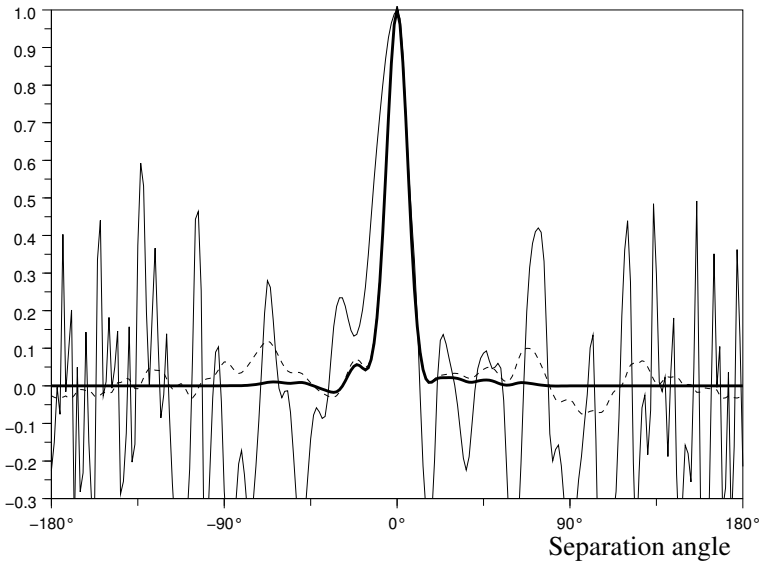

Figure 6. Ensemble correlations related to gridpoint 1, calculated from a 10 element ensemble: directly (solid line), with a wavelet diagonal approach and no Schur filter (dashed line), with a wavelet diagonal approach and a Schur filter such as $L_s = 6000km$ (bold solid line).

assumed to have a similar quality as the background). There is one observation every five gridpoints.

Root Mean Square (RMS) errors (averaged over the domain) of analysis can be calculated for three different covariance approximations: the direct ensemble estimation, the homogeneous formulation (i.e. the spectral diagonal approach), and the wavelet formulation. These RMS can then be compared with the RMS corresponding to the true covariances. Fig. 9 corresponds to the difference $RMS - RMS(true)$, for different Schur scales L_s and for different ensemble sizes N_e. In this figure, every curve has the same behaviour (similar to Lorenc 2003): the shape is convex with a minimum that determines an optimal L_s value.

For the direct ensemble estimation (dashed line), the smaller the ensemble is, the larger the RMS is, and the shorter the scale L_s has to be. Such dependencies of the RMS and optimal L_s on the ensemble size are much smaller for the wavelet approach (full line). Moreover, the wavelet results are relatively good even for a small ensemble size such as $N_e = 10$. This illustrates the beneficial impact of wavelet filtering.

Compared with the direct ensemble estimation, another attractive result of the wavelet approach is that the RMS slope is small beyond the optimal choice for L_s. This means that the analysis quality will be less affected

DOI: 10.1002/qj

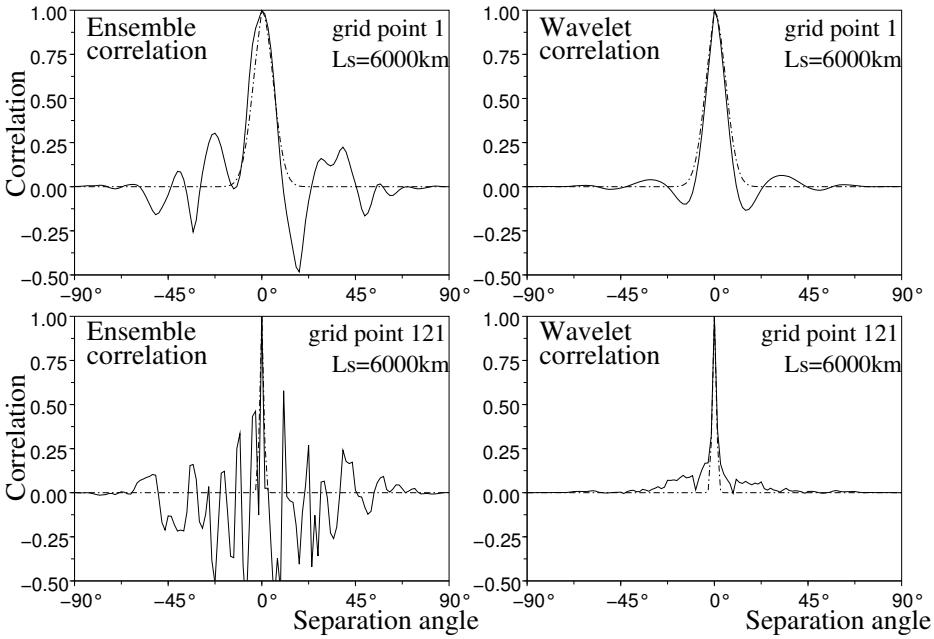

Figure 7. Ensemble and wavelet correlations (solid line) related to grid points 1 (corresponding to the large length scale area) and 121 (corresponding to the short length scale area). Correlations are computed from a 10 element ensemble and then filtered with a Schur filter $\boldsymbol{F}_{L_s}$ with $L_s = 6000\ km$. Exact correlations are also shown (chain-dotted line).

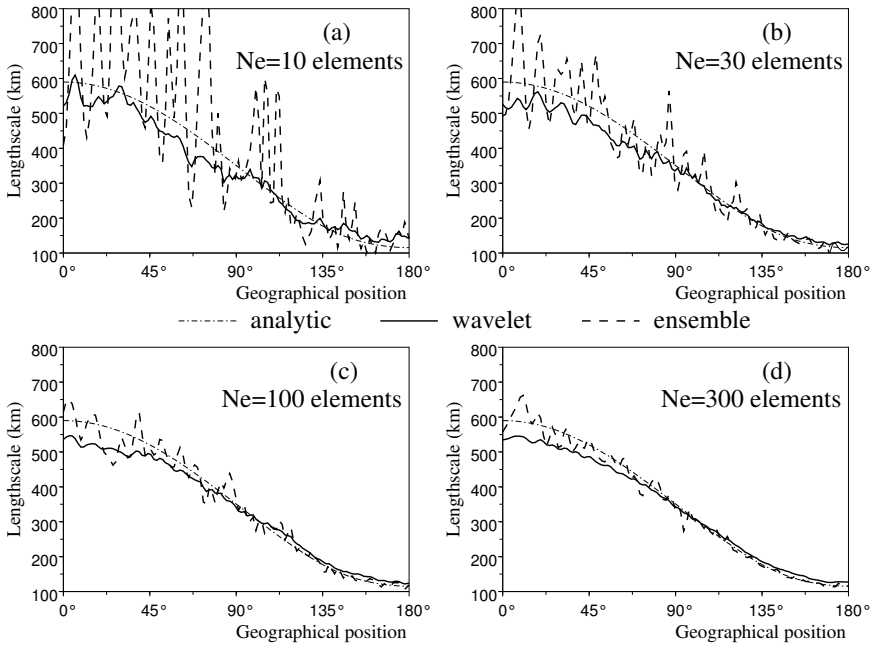

Figure 8. Geographical variation of the length scale: exact (chain-dotted line), wavelet estimated (solid line) and ensemble estimated (dashed line). Length scales are computed for different ensemble sizes, namely $N_e = 10, 30, 100$ and 300. For each case, the wavelet and ensemble correlation estimations have been filtered using a Schur filter $\boldsymbol{F}_{L_s}$ with $L_s = 6000\ km$.

by a suboptimal Schur scale than in the direct ensemble estimation.

As expected, analyses produced with the wavelet approach are closer to optimality than those produced with the homogeneous formulation. On the other hand, it is interesting to notice that for $N_e = 10$, the homogeneous approach can give better results than a direct ensemble estimation, if the Schur length is too large. This is another illustration of the importance of sampling noise in small ensembles, and of the potential benefit of spatially averaging the ensemble covariances in this case.

It is also possible to compare geographical variations the lengthscales, when the optimal Schur lengths are used. The results are illustrated in Fig. 10. The wavelet implied length scale values still appear to be more accurate, and their variations are smoother. This is particularly noticeable for small ensembles.

It may be mentioned also that, compared to Figure

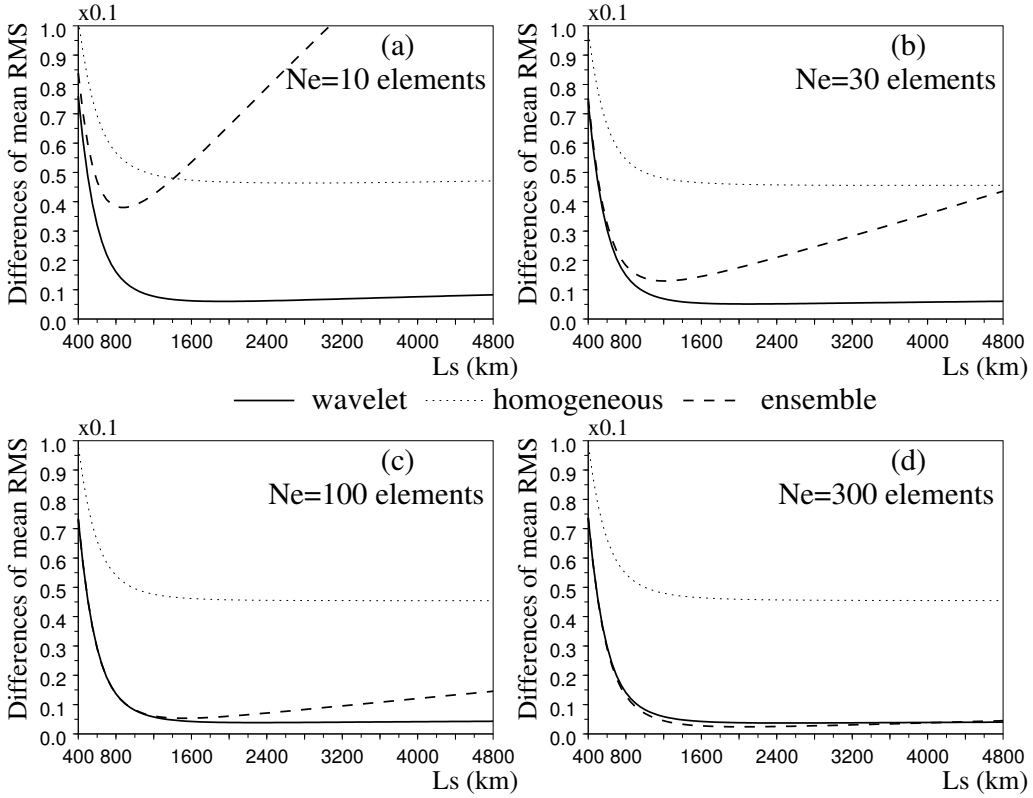

Figure 9. Differences between analysis RMS error and true analysis RMS error, as a function of the scale of the Schur filter L_s and of the size N_e of the ensemble (respectively 10, 30, 100 and 300 for the four different panels). Differences are shown in each panel for wavelet analysis (solid line), analysis with homogeneous formulation (dotted line) and analysis with raw ensemble covariances (dashed line). All analyses are performed with one observation every 5 gridpoints.

(8), the use of shorter Schur lengths L_s implies a shortening of the length scales. This is consistent with the fact that the smaller L_s is, the faster the filtered correlation functions will decrease (as a function of separation distance). With respect to Eq. (2), this is linked with an increase of $\sigma(\partial_x \varepsilon)^2$, because small scale contributions are emphasized both in $\partial_x \varepsilon$ (compared to ε) and when decreasing the correlation length scale.

4 Application to an ensemble of Arpège forecasts

4.1 The ensemble data set of global Arpège forecasts

The Météo-France operational NWP system is based on the Arpège model (Courtier and Geleyn 1988), and on a 4D-Var assimilation scheme (Rabier et al 2000, Veersé and Thépaut 1998). The background error covariance matrix is calculated by using an ensemble of perturbed assimilation runs (Houtekamer et al 1996, Fisher 2003). The detailed results for Arpège are described in Belo Pereira and Berre (2006).

We propose to illustrate some typical results of the wavelet covariance modelling on this kind of NWP ensemble data. The formalism and the cutoff wave numbers are the same as those mentioned in section 2 (e) (and Fisher 2003). The available ensemble consists in a set of 6 forecast differences for each day of the period from 9 February to 24 March 2002.

In this 2D high dimensional framework, local length scales are calculated by using equation (2), with a specific randomization technique for the wavelet-implied length scales. In the latter case, a set of 1000 random vectors ε has been generated from the square root of $\boldsymbol{B}_w$ (Fisher and Courtier 1995), and the gradient operator ∂_x has been applied to these vectors ε. The variances of $\partial_x \varepsilon$ are then used in the length scale equation.

A first possible option is to temporally average the spatial covariances, in order to examine the "climatology" of the error covariances. The total sample is made of $N_e = 264$ elements in this case.

A second option is to study the covariances for a particular date. The total sample is reduced to only $N_e = 6$ elements in this case, which correspond to differences between six perturbed forecasts, which are valid on 10 February at 12 UTC.

4.2 Local "climatological" length scales

As shown in Fig. 11, the ensemble method provides some interesting local information on the "climatological" correlations, which appear to be well represented by the wavelet formulation of equation (1). The length scales are smaller in the (data rich) Northern Hemisphere than in the (data poor) Southern Hemisphere. Some local length scale minima can be identified in the storm track region of the Northern Atlantic, and near the Inter Tropical

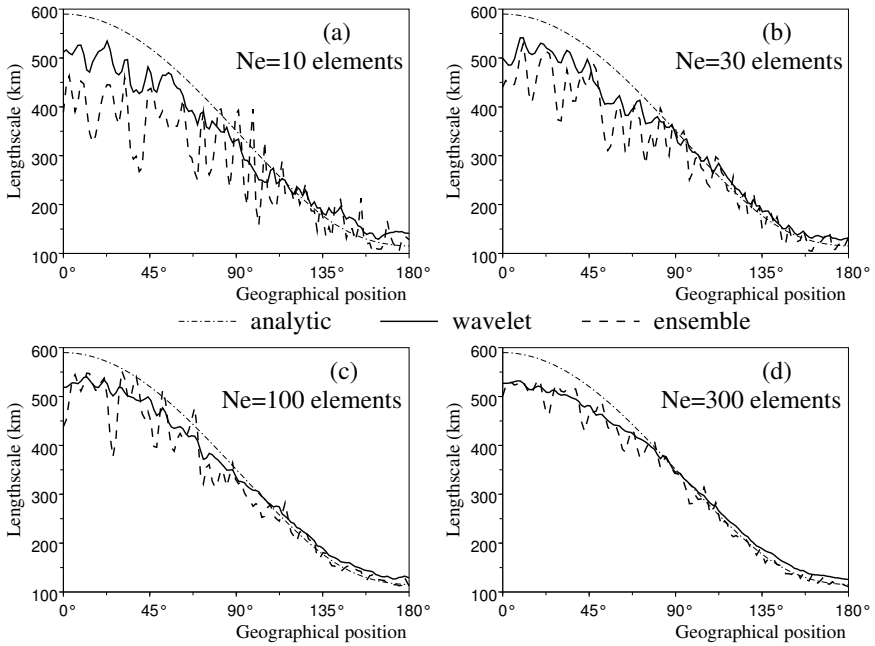

Figure 10. Same as figure 8, but with optimal Schur lengths L_s, chosen according to Fig. 9. For the wavelets, L_s is constant and equal to 4000 km. For the ensemble estimation, L_s depends on the size N_e : (a) $(N_e = 10, L_s = 900\ km)$, (b) $(N_e = 30, L_s = 1250\ km)$, (c) $(N_e = 100, L_s = 1600\ km)$ and (d) $(N_e = 300, L_s = 2050\ km)$.

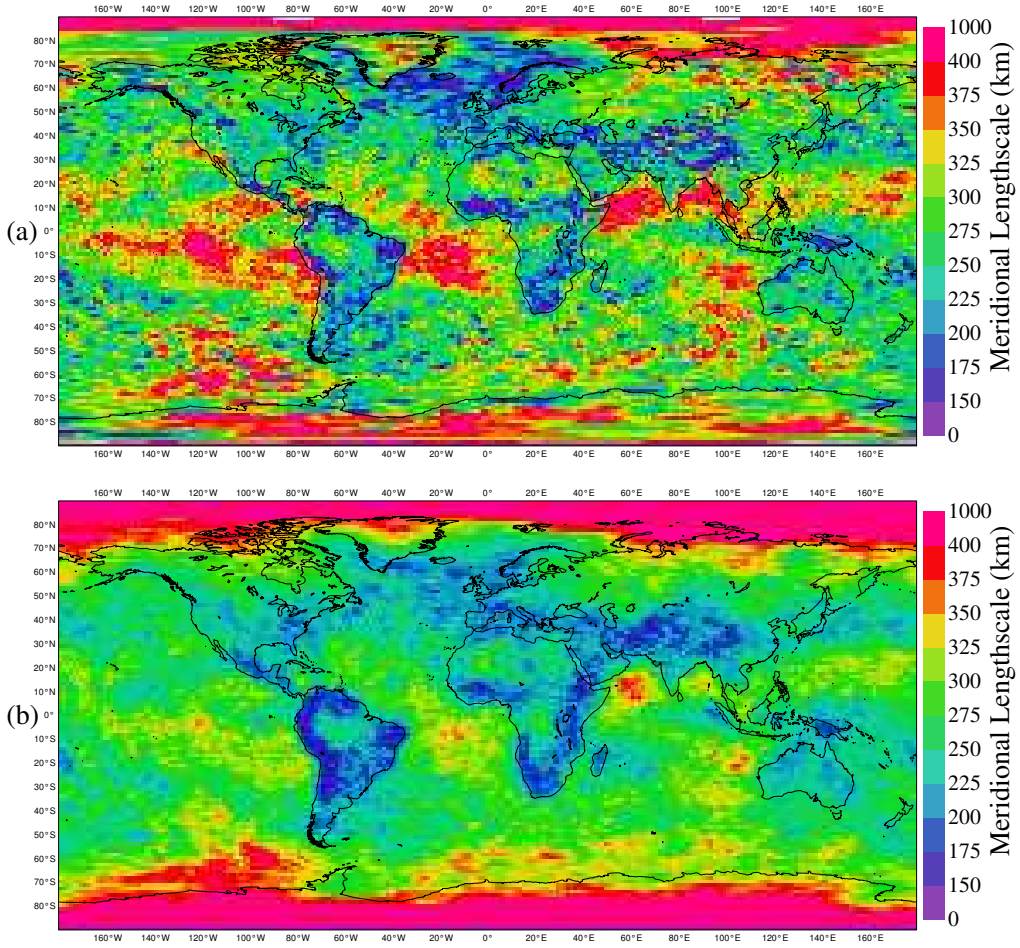

Figure 11. Meridional length scales (in km) for surface pressure, averaged over the 46 day period and the 6 ensemble members. (a): raw length scales. (b): wavelet implied length scales.

DOI: 10.1002/qj

Convergence Zone (ITCZ) area in Western Africa. Length scales are also smaller over orographic regions such as the Himalaya and the Andes, and in the Southern part of Africa. In contrast, some length scale maxima are visible over tropical oceans.

In accordance with the results in section 33.3, these behaviours on the whole globe confirm one of the main advantages of a wavelet covariance formulation (compared e.g. with a spectral formulation): it allows some correlation variations to be represented, such as those that are induced by atmospheric processes and by data density contrasts.

It can be also noticed that the extreme variations tend to be smoothed in the wavelet map. This is due to the filtering properties of the wavelet diagonal approach. These filtering properties will now be shown to be even more important when the ensemble size is reduced to $N_e = 6$.

4.3 Local length scales of a particular date

Fig. 12 shows the corresponding length scales for a particular date, the 10 February at 12 UTC. The raw length scales (top panel) appear to be very noisy, due to the small size of the ensemble. In contrast, the wavelet implied length scale map (bottom panel, superposed with the background field of sea level pressure) is relatively smooth and well structured. Large values appear clearly in the southern circumpolar ocean and over tropical oceans, as in the climatological case but in a more pronounced way. Small length scales are visible over land. Some other structures related to the local weather situation can also be identified, such as small values in the vicinity of some mid-latitude lows over sea (see e.g. the lows near Scandinavia and south of Argentina). This is consistent with results described by Thépaut et al (1995) for instance.

Such strong differences between the raw and wavelet implied length scales, when the sampling size is small ($N_e = 6$), are in agreement with the large length scale differences found in the 1D framework with $N_e = 10$ (top left panel of Fig. 8). They support the idea that the wavelet formulation is able to capture and to represent the main relevant length scale variations, from a small ensemble of forecasts, thanks to local spatial averaging.

5 Conclusions

In this paper, the ability of a wavelet diagonal approach to ensure a smooth representation of geographical variations of the correlation functions was studied. Representing the covariances by a diagonal matrix in wavelet space amounts to locally averaging the covariance functions. Due to this spatial averaging, the statistical estimate is more sampled than when estimating the purely local covariance functions. Moreover, as this spatial averaging is local, the representation of geographical variations remains possible. Such filtering properties look particularly attractive when estimating error covariances from an ensemble of forecasts. These aspects of wavelet covariance modelling were made explicit formally, and they were illustrated experimentally in two different frameworks.

The first experimental framework is a simple 1D context with varying length scales. The wavelet covariance model was shown to be able to represent the local correlation functions and their length scale variations, in a smoother and more realistic way than when only using a Schur filter. This is particularly noticeable when using small ensembles with e.g. 10 or 30 members. The wavelet formulation remains competitive with up to 100 members.

The second experimental framework consists in an ensemble of global NWP forecasts. The wavelet approach appears to represent well the "climatological" variations of the local length scales, which are induced by atmospheric processes and by data density contrasts. Moreover, a preliminary examination of the length scales of a particular day suggests that the wavelet approach allows some important length scale variations to be extracted, which are connected with the local weather situation.

These results are consistent with the expected filtering properties of wavelets in terms of local spatial averaging. They suggest that wavelets may be a promising tool to estimate and represent flow-dependent covariances from a small ensemble of forecasts.

Acknowledgement

The authors would like to thank Mike Fisher, Anthony Weaver and Andrew Tangborn for fruitful discussions. We also thank Bernard Chapnik, Claude Fischer and Florence Rabier, for their careful reading of the manuscript.

A Appendix A : Expressions of covariance spatial averaging

It is shown, in the general layout of frames, that the diagonal covariance assumption leads to a weighted spatial averaging of covariances.

A.1 Background error expansion in a frame

A frame (Daubechies 1992, Fisher 2004) is a family of functions $\{\boldsymbol{\phi}_m, m \in \mathcal{M}\}$, where $\mathcal{M}$ is a countable set. This family is associated to a dual frame $\left\{\tilde{\boldsymbol{\phi}}_m, m \in \mathcal{M}\right\}$, so that the error field $\boldsymbol{\varepsilon}$ can be analyzed as a set of frame coefficients $\hat{\varepsilon}_m$ with

$$\hat{\varepsilon}_m = \sum_x \varepsilon(x) \tilde{\phi}^*_m(x), \tag{4}$$

where the exponent $*$ stands for the conjugate transpose operator. The signal is recomposed according to

$$\varepsilon(x) = \sum_{m \in \mathcal{M}} \hat{\varepsilon}_m \phi_m(x),$$

or in vector form:

$$\boldsymbol{\varepsilon} = \sum_{m \in \mathcal{M}} \hat{\varepsilon}_m \boldsymbol{\phi}_m.$$

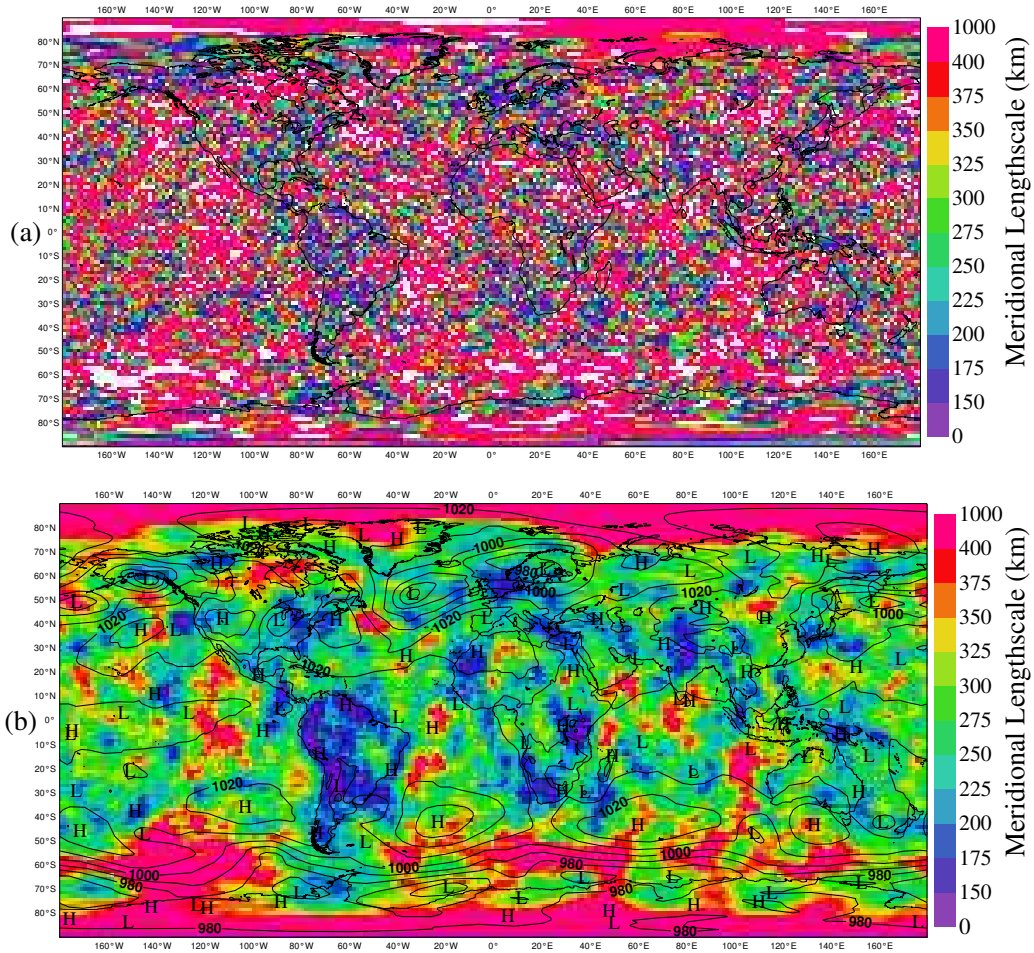

Figure 12. Meridional length scales (in km) for surface pressure on 10 February 2002 at 12 UTC, computed from 6 ensemble members. (a): raw length scales. (b): wavelet implied length scales, superposed with the background field of sea level pressure.

A.2 Expansion of covariances in a frame under the diagonal assumption

Using the previous frame decomposition, the covariance matrix $\boldsymbol{B} = \overline{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^*}$ can be expanded as $\boldsymbol{B} = \sum_{m,m'} \overline{\hat{\varepsilon}_m \hat{\varepsilon}^*_{m'}} \boldsymbol{\phi}_m \boldsymbol{\phi}^*_{m'}$. When using the diagonal assumption in the frame space, covariances $\overline{\hat{\varepsilon}_m \hat{\varepsilon}^*_{m'}}$ are zero except if $m = m'$. Then, the resulting covariance matrix $\boldsymbol{B}_d$ is $\boldsymbol{B}_d = \sum_m B_{mm} \boldsymbol{\phi}_m \boldsymbol{\phi}^*_m$, where $B_{mm} = \overline{\hat{\varepsilon}_m \hat{\varepsilon}^*_m}$ is the variance for the frame coefficient m. Thus, the associated covariance function f^x_d at position x has the following expression:

$$f^x_d(s) = \sum_m B_{mm} \phi_m(x) \phi^*_m(x+s),$$

where s is a separation value (in gridpoint space). From (4), coefficients B_{mm} can be rewritten

$$\begin{aligned} B_{mm} &= \sum_{x',s'} \overline{\varepsilon(x')\varepsilon(x'+s')^*}\ \tilde{\phi}^*_m(x') \tilde{\phi}_m(x'+s') \\ &= \sum_{x',s'} f^{x'}(s')\ \tilde{\phi}^*_m(x') \tilde{\phi}_m(x'+s'), \end{aligned}$$

where $f^{x'}(s') = \overline{\varepsilon(x')\varepsilon(x'+s')^*}$ is the local covariance function at position x' for the full covariance matrix $\boldsymbol{B}$, and then the resulting expression for $f^x_d(s)$ is

$$f^x_d(s) = \sum_{x',s'} f^{x'}(s') \left(\sum_m \tilde{\phi}^*_m(x') \tilde{\phi}_m(x'+s') \phi_m(x) \phi^*_m(x+s)\right),$$

which can be in turn rewritten as

$$f^x_d(s) = \sum_{x',s'} f^{x'}(s') \Phi^{x,s}(x',s'),$$

with

$$\Phi^{x,s}(x',s') = \sum_m \tilde{\phi}^*_m(x') \tilde{\phi}_m(x'+s') \phi_m(x) \phi^*_m(x+s).$$

The implied covariance function $f^x_d(s)$ may thus be seen as a weighted spatial average of the local covariance functions $f^{x'}(s')$, where the weights $\Phi^{x,s}(x',s')$, for a pair $\{x,s\}$, are functions of position x' and separation s'.

A.3 Diagonal assumption in Fourier space

Let N_g be the number of grid points on a circle, and $e_m(x) = exp(im\frac{2\pi x}{N_g})$. Then, the family

DOI: 10.1002/qj

$\left\{\phi_m = \frac{1}{\sqrt{N_g}} e_m, m \in [0, N_g - 1]\right\}$ is a frame whose dual is simply $\tilde{\phi}_m = \phi_m$. In this case, the weighting coefficients are given by

$$\begin{aligned}\Phi^{x,s}(x', s') &= \frac{1}{N_g^2} \sum_m e_m^*(x') e_m(x' + s') e_m(x) e_m^*(x + s) \\ &= \frac{1}{N_g^2} \sum_m e_m(s' - s) \\ &= \frac{1}{N_g} \delta(s' - s),\end{aligned}$$

where $\delta(s' - s) = 1$ when $s' = s$ and $\delta(s' - s) = 0$ elsewhere. This means that a zero weight is given to all values $f^{x'}(s')$ such as $s' \neq s$ and that a uniform weight $\frac{1}{N_g}$ is given to all values $f^{x'}(s')$ such as $s' = s$, whatever the position x' is. In other words, each implied covariance function f_d^x is simply a global spatial average of the local covariance functions $f^{x'}$.

A.4 Diagonal assumption in wavelet space

For the wavelet case, $\mathcal{M}$ is a set of $m = (j, \{x_j(i), i = 1, N_x(j)\})$, where j is a scale and $x_j(i)$ a position on a subgrid associated to j. This defines a frame such that, $\phi_m(x) = \tilde{\phi}_m(x) = \psi_j(x - x_j(i))$, where functions ψ_j are Fisher's wavelets. In this case, the weighting coefficients $\Phi^{(x,s)}(x', s')$ are given by

$$\sum_j \sum_{i=1}^{N_x(j)} \psi_j(x' - x_j(i))\, \psi_j(x' - x_j(i) + s') \\ \psi_j(x - x_j(i))\, \psi_j(x - x_j(i) + s).$$

In contrast with the spectral case, the resulting weights $\Phi^{(x,s)}(x', s')$ will vary with the position x'. As expected, numerical tests indicate that the weights $\Phi^{(x,s)}(x', s')$ tend to be maximum for positions x' close to x and for separation values s' that are also close to s. In other words, the implied covariance function f_d^x may be seen as a local spatial average of the covariance functions $f^{x'}$.

B Appendix B: Design of $\mathbf{B}_w^{-1/2}$ and of $\mathbf{B}_w^{1/2}$

The formulation of $\boldsymbol{B}_w^{1/2}$ is determined by the design of $\boldsymbol{B}_w^{-1/2}$. The latter matrix is conceived as an operator that transforms the background error variable $\boldsymbol{\varepsilon}$ into a transformed variable, whose covariance matrix is close to an identity matrix (see e.g. Deckmyn and Berre (2005) or Gustafsson et al (2001)):

$$\boldsymbol{B}_w^{-1/2} = \boldsymbol{D}_w^{-1/2} \boldsymbol{W} \boldsymbol{\Sigma}_s^{-1} \boldsymbol{\Sigma}_g^{-1},$$

where $\boldsymbol{\Sigma}_g$ is a diagonal matrix of gridpoint standard deviations of $\boldsymbol{\varepsilon}$, $\boldsymbol{\Sigma}_s^{-1} = \boldsymbol{S}^{-1} \boldsymbol{D}_s^{-1/2} \boldsymbol{S}$ corresponds to a normalisation by spectral standard deviations of $\check{\boldsymbol{\varepsilon}}' = \boldsymbol{S} \boldsymbol{\Sigma}_g^{-1} \boldsymbol{\varepsilon}$ ($\boldsymbol{S}$ being the spectral transform, and $\boldsymbol{D}_s$ is the corresponding diagonal matrix of variances in spectral space), and $\boldsymbol{D}_w^{1/2}$ is the diagonal matrix of wavelet standard deviations of $\hat{\boldsymbol{\varepsilon}}'' = \boldsymbol{W} \check{\boldsymbol{\varepsilon}}'$. Matrices $\boldsymbol{W}$ and $\boldsymbol{W}^{-1}$ correspond respectively to the direct and inverse wavelet transforms.

The associated expression of $\boldsymbol{B}_w^{1/2}$ is then $\boldsymbol{B}_w^{1/2} = (\boldsymbol{B}_w^{-1/2})^{-1} = \boldsymbol{\Sigma}_g \boldsymbol{\Sigma}_s \boldsymbol{W}^{-1} \boldsymbol{D}_w^{1/2}$.

C Appendix C: A wavelet illustration of the filtering properties

As mentioned in sections 2 and 3.3, the filtering properties of wavelets can be expected since the wavelet diagonal assumption amounts to locally averaging the covariance functions. This may be considered as a "gridpoint vision" of the filtering properties, in the sense that gridpoint covariance functions are seen as being averaged in gridpoint space. In this section, we will evoke two other complementary visions of the filtering properties (in wavelet and spectral spaces respectively).

Note that the full covariance matrix in wavelet space represents the covariances between different scales at different locations. A "wavelet vision" of the filtering properties at play is thus to note in particular that the wavelet diagonal assumption implies zeroing off-diagonal terms, which correspond to correlations between wavelet modes at different positions (e.g. for a given scale j). The formal examination of the associated equations (not detailed here for the sake of conciseness) suggests the following result: neglecting these wavelet off-diagonal correlations prevents small scale distant modes from (spuriously) contributing to the local covariance function (at a given reference position).

This interpretation is supported experimentally, as illustrated in Fig. 13. This figure shows the amplitude (absolute value) of the wavelet coefficients $\boldsymbol{W} f^0$ of the local covariance function at geographical position $0\,^\circ$, namely f^0. In the ensemble off-diagonal case (top right panel), the amplitude of the small scale distant wavelet modes (e.g. for $j \geq 10$ near $180\,^\circ$) can be spuriously large (due to sampling noise), while the exact values (top left panel) tend to be close to zero. In contrast, the wavelet diagonal approach produces realistic small values, even when a small ensemble is used ($N_e = 10$). This is consistent with the implicit zeroing of correlations between wavelet modes positioned around $0\,^\circ$ and those positioned around $180\,^\circ$ for instance.

Finally, it may be also mentioned that a "spectral vision" of the wavelet filtering properties can be considered as well. It will not be detailed here for the sake of conciseness. Briefly summarized, it consists in noticing that zeroing cross-correlations between different wavelet scales amounts to zeroing some small scale (noisy) contributions to the geographical variations of covariances. This is similar to the expected effect of a local spatial averaging in gridpoint space.

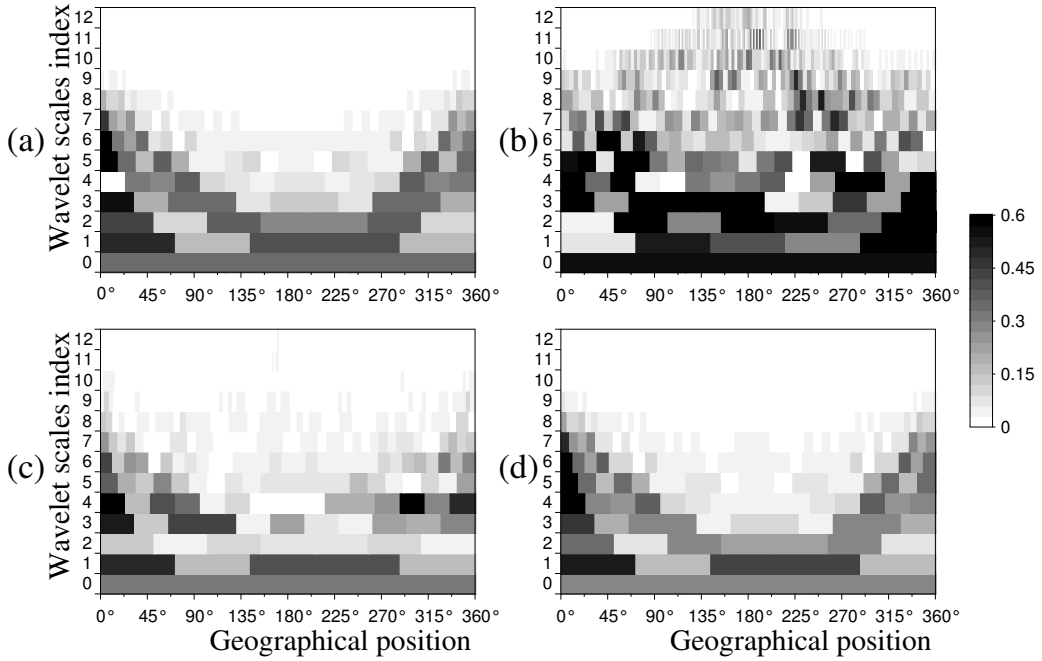

Figure 13. Amplitude of the wavelet coefficients of the local covariance function at geographical position 0 °. (a) exact covariance, (b) ensemble approach with 10 members and a full covariance matrix (in gridpoint or wavelet space), (c) ensemble wavelet diagonal approach with 10 members, and (d) analytical wavelet diagonal approach (equivalent to (c) but with an infinite number of members). The wavelet scale index is j: large j values correspond to small scale wavelets.

References

Belo Pereira M. and Berre L. 2006. *The use of an Ensemble approach to study the Background Error Covariances in a Global NWP model. Mon. Wea. Rev.*, **134**, 2466–2489.

Bouttier F. 1993. *The dynamics of error covariances in a barotropic model. Tellus*, **45A**, 408–423.

Bouttier F. 1994. *A dynamical estimation of error covariances in an assimilation system. Mon. Wea. Rev.*, **122**, 2376–2390.

Buehner M. and Charron M. 2007. *Spectral and spatial localization of background error correlations for data assimilation.* To appears in *Q.J.R. Meteorol. Soc.*

Courtier P. and Geleyn J.F. 1988. *A global numerical weather prediction model with variable resolution : Application to the shallow-water equations. Q.J.R. Meteorol. Soc.*, **114**, 1321–1346.

Courtier P., Andersson E., Heckley W., Pailleux J., Vasiljević D., Hamrud M., Hollingsworth A., Rabier F. and Fisher M. 1998. *The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1807.

Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press. p471.

Deckmyn A. and Berre L. 2005. *A wavelet approach to representing background error covariances in a LAM. Mon. Wea. Rev.*, **133**, 1279-1294.

Dee D. 1995. *On-line estimation of error covariance parameters for atmospheric data assimilation. Mon. Wea. Rev.*, **123**, 1128–1145.

Derber J. and Bouttier F. 1999. *A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus*, **51A**, 195–221.

Daubechies I. 1992. *Ten lectures on Wavelets.* CBMS-NSF regional conference series in applied mathematics, SIAM, 357pp.

Evensen G. 1994. *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error. J. Geophys. Res.*, **99** (C5) 10 143–10 162.

Fisher M. and Courtier P. 1995. *Estimating the covariance matrices of analysis and forecast error in variational data assimilation.* ECMWF Technical Memorandum, **220**, 29pp.

Fisher M. 2003. *Background error covariance modelling*. Processing of the ECMWF Seminar on "Recent developments in data assimilation for atmosphere and ocean", Reading, 8–12 September 2003, 45–63.

Fisher M. 2004. *Generalized frames on the sphere, with application to the background error covariance modelling.* Processing of the ECMWF Seminar on "Developments in Numerical Methods for Atmospheric and Ocean Modelling", Reading, 6–10 September 2004, 87–101.

Gaspari G. and Cohn S. 1999. *Construction of correlation functions in two and three dimensions. Q.J.R. Meteorol. Soc.*, **125**, 723–757.

Gustafsson N., Berre L., Hörnquist S., Huang X-Y., Lindskog M., Navasques B., Mogensen KS. and Thorsteinsson S. 2001. *Three-dimensional variational data assimilation for a limited area model. Tellus*, **53A**, 425–446.

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

Houtekamer P.L. and Mitchell H.L. 2001. *A sequential ensemble Kalman filter for Atmospheric Data Assimilation. Mon. Wea. Rev.*, **129**, 123–137.

Ingleby B. 2001. *The statistical structure of forecast errors and its representation in The Met. Office Global Model. Q.J.R. Meteorol. Soc.*, **124**, 1783–1807.

Lönnberg P. 1988. *Developpements in the ECMWF analysis scheme*. Proc. ECMWF Seminar on "Data assimilation and the use of satellite data", Reading, 5–9 September 1988, 75–120.

Lorenc A. 2003. *The potential of ensemble Kalman filter for NWP – a comparison with 4D-Var. Q.J.R. Meteorol. Soc.*, **129**, 3183–3203.

Rabier F., Jarvinen H., Klinker E., Mahfouf J.F. and Simmons A. 2000. *The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Q.J.R. Meteorol. Soc.*, **126**, 1148–1170.

Thépaut J-N., Courtier P., Belaud G. and Lemaître G. 1995 *Dynamical structure functions in a four-dimensional variational assimilation : a case study. Q.J.R. Meteorol. Soc.*, **122**, 535–561.

Veersé F. and Thépaut J-N. 1998. *Multiple-truncation incremental approach for four-dimensional variational data assimilation. Q.J.R. Meteorol. Soc.*, **124**, 1889–1908.