A wavelet-based filtering of ensemble background-error variances

Olivier Pannekoucke[1], Laure Raynaud[1] and Marie Farge[2]
[1]CNRM/GAME, Météo-France/CNRS URA 1357, Toulouse, France
[2]LMD-IPSL-CNRS, Ecole Normale Supérieure, 24 rue Lhomond,75231 Paris Cedex 5, France
*Correspondence to: Olivier Pannekoucke, Météo-France CNRM/GMAP, 42 Av. G. Coriolis, 31057 Toulouse Cedex, France. E-mail : olivier.pannekoucke@meteo.fr

Background-error variances estimated from a small size ensemble of data assimilations need to be filtered out because of the associated sampling noise. Previous works showed that objective spectral filtering is efficient to reduce this noise, while preserving relevant features to a large extent. However, since such filters are homogeneous, they tend to smooth small-scale structures of interest.
In many applications, nonlinear thresholding of wavelet coefficients has proved to be an efficient technique for denoising signals. This algorithm proceeds by thresholding the wavelet coefficients of the noisy signal using an estimated threshold. This is equivalent to applying an adaptive local spatial filtering. A quasi optimal value for the threshold can be computed from the noise variance. We show that the statistical properties of the sampling noise associated with the estimation of background-error variances can be used to calculate the noise level and the appropriate threshold value.
This method is applied to 1D academic examples, for both Gaussian white and correlated noises. This approach is shown to outperform the commonly used homogeneous filters, since it automatically adapts to the local structure of the signal. We also show that this technique compares favourably to a heterogeneous diffusion-based filter, with the advantage of requiring less trial and error tuning.

Key Words: wavelet thresholding ; background-error variances ; ensemble data assimilation ; sampling noise ;

Received ; Revised ; Accepted

Citation: ...

1. Introduction

The background-error covariance matrix $\mathbf{B}$ plays a central role in data assimilation schemes by weighting the information from the observations and the background state in the analysis.

Recently there has been growing interest in estimating background-error covariances from ensemble data assimilation systems, either in the Kalman filter context (Evensen 2003) or in the variational framework (Kucukkaraca and Fisher 2006; Raynaud et al. 2009; Bonavita et al. 2011). However, the high computational cost of such ensembles in operational applications limits the ensemble size (namely $\mathcal{O}(100)$ members), leading to a significant sampling noise which has to be filtered out. The goal of the filtering step is to remove the noise while retaining as much as possible the important signal features. Traditionally, this is achieved by linear processing such as Wiener filtering.

Previous works on the filtering of ensemble-based variances (Raynaud et al. 2008, 2009) provided useful information with regard to the associated sampling noise, such as its statistical properties. Moreover, Wiener filtering of ensemble variances has been successfully implemented in large-scale applications at Météo-France (Raynaud et al. 2009) and ECMWF (European Centre for Medium-range

Weather Forecasts, Bonavita et al. (2011)).

The Wiener filter, however, optimizes the trade-off between an averaging of the signal discontinuities and the removal of the noise in the smooth regions in order to minimize the mean-square error. As a result, some noise is left in the smooth regions while the discontinuities are averaged a little. Discontinuities in background-error variance fields typically correspond to high forecast errors associated with severe weather events, e.g., midlatitude storms and tropical cyclones. The averaging of such error structures, which has for instance been observed by Bonavita et al. (2011), can then result in a smaller impact of relevant observations in these regions during the assimilation step.

In order to preserve such discontinuities in variance fields, the filter has to adapt to the local structure of the signal. This could be achieved by performing the filtering either in gridpoint space or wavelet space. A first contribution to heterogeneous variance filtering in gridpoint space has been proposed by Raynaud and Pannekoucke (2011) based on the integration of the heterogeneous diffusion equation. In the present paper, the application of nonlinear filtering in wavelet space is examined.

Wavelet transform, thanks to its excellent localization property, has rapidly become an essential signal and image processing tool for a variety of applications, including denoising. Denoising by wavelet coefficient thresholding is a commonly used method first proposed by Donoho and Johnstone (1994). The algorithm compares each wavelet coefficient of the noisy signal to a given threshold : if the coefficient is smaller than the threshold then it is set to zero, otherwise it is kept or modified (depending on the thresholding rule). The idea behind thresholding is to distinguish between the insignificant coefficients likely due to noise, and the significant coefficients consisting of important signal structures. The denoised signal is then reconstructed from the selected coefficients.

The paper is organized as follows. Section 2 introduces the technique of wavelet coefficients thresholding. Section 3 then details the application of this method to the filtering of ensemble background-error variances. Wavelet thresholding is applied in section 4 to 1D analytical signals corrupted by a Gaussian white noise. The extension to a correlated noise is examined in section 5. Finally, conclusions and perspectives are given in section 6.

2. Denoising by wavelet thresholding

In this section, we introduce the mathematical formalism associated with the denoising by wavelet thresholding, as initially proposed by Donoho and Johnstone (1994) to denoise signals affected by Gaussian white noise.

2.1. *Theoretical aspects*

We consider a discrete signal $\mathbf{S}$ of size $n = 2^J$, affected by a Gaussian white noise $\mathbf{W}$ of mean zero and variance σ_W^2, resulting in a noisy signal $\mathbf{X}$:

$$\mathbf{X} = \mathbf{S} + \mathbf{W}.$$

We decompose the noisy signal into an orthogonal wavelet series $\mathbf{X} = \sum_{j=0}^{J-1} \sum_{i=0}^{2^j-1} \tilde{X}_{i,j} \psi_{i,j}$, where $\tilde{X}_{i,j} = \langle \psi_{i,j} | \mathbf{X} \rangle$ is the wavelet coefficient at scale j and position index i, $\langle \cdot | \cdot \rangle$ denoting the inner product (Mallat 1999). The wavelet coefficients $\{\tilde{X}_{i,j}\}_{i=0,2^j-1}$ at scale j thus define an approximation of $\mathbf{X}$ on a grid whose resolution depends on j : the finer the scale the higher the resolution.

Denoising by thresholding wavelet coefficients consists in keeping only the coefficients whose modulus is above a given threshold value T :

$$\rho_T(\tilde{X}_{i,j}) = \begin{cases} \tilde{X}_{i,j} & \text{if } |\tilde{X}_{i,j}| > T \\ 0 & \text{if } |\tilde{X}_{i,j}| \leq T. \end{cases}$$

The denoised signal $\hat{\mathbf{X}}$ is finally reconstructed using an inverse wavelet transform $\hat{\mathbf{X}} = \sum_{i,j} \rho_T(\tilde{X}_{i,j}) \psi_{i,j}$.

Wavelet thresholding is first motivated by the fact that the decorrelating property of the wavelet transform reveals sparsity of the signal if any, i.e., most wavelet coefficients are close to zero (Mallat 1999). Moreover, since the noise is spread out equally over all coefficients, if the noise level is not too high it is then possible to discriminate between signal and noise coefficients.

The idea behind wavelet thresholding is to test each wavelet coefficient in order to check if it is compatible with a Gaussian white noise with standard deviation σ_W. This can be achieved by performing a statistical test, allowing us to verify that particular properties of the noise are consistent with this Gaussian distribution. A possible statistical property is the maximum magnitude that can be encountered when sampling a Gaussian random variable of standard deviation σ_W and size n. As detailed in Appendix A, this maximum magnitude should be lower than

$$T_D = \sigma_W \sqrt{2 \ln n}. \tag{1}$$

The sampling size n can be understood as the return time of the extreme event consisting in exceeding a magnitude strictly smaller than T_D (see appendix A for details). Since a Gaussian white noise in gridpoint space is transformed into an equivalent Gaussian white noise in an orthogonal wavelet representation, the test can be performed on the wavelet coefficients. The noise variance is then calculated as $\sigma_W^2 = \frac{1}{n} \sum_{l=1}^{n} (\mathbf{W}(l))^2 = \frac{1}{n} \sum_{i,j} |\tilde{W}_{i,j}|^2$, where $\mathbf{W}(l)$ is the noise at gridpoint l and $\tilde{W}_{i,j}$ is a wavelet coefficient of noise. Therefore, a wavelet coefficient whose modulus is larger than T_D is not compatible with the Gaussian assumption. In that case, the coefficient carries more signal information than noise and contributes to $\hat{\mathbf{X}}$.

It is worth noticing that T_D is equal to the *universal* threshold proposed by Donoho and Johnstone (1994), which results in an estimate asymptotically optimal (when $n \to \infty$) in the minimax sense (i.e. minimizing the maximum quadratic error, Mallat (1999)). This threshold is called universal since it depends on the sampling size n and on the variance of the noise σ_W^2, but not on the signal itself.

2.2. *Noise variance estimation*

When the statistical properties of the noise are known or can be calculated with an appropriate model, the determination of the optimal threshold is straightforward (Donoho and Johnstone 1994). However, the noise variance is unknown in many situations and has to be estimated.

Different methods have been proposed, such as the median absolute deviation (MAD), which estimates the level of noise by taking the median of the modulus of the smallest scale wavelet coefficients (Mallat 1999). An

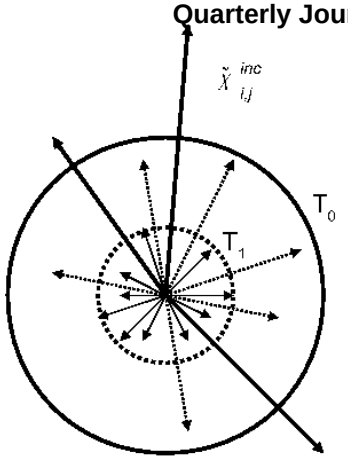


FIG. 1. Conceptual illustration of the recursive algorithm in $\mathbb{R}^n$ for estimating the noise variance and the threshold. The estimated thresholds T_0 and T_1 are represented by the bold and dashed spheres respectively. The arrows represent a selection of wavelet coefficients $\tilde{X}_{i,j}$.

alternative approach was introduced by Farge et al. (1999); Azzalini et al. (2004), based on a recursive estimation of the noise variance and the threshold.

The recursive approach of Azzalini et al. (2004) proceeds as follows. The wavelet signal $\tilde{\mathbf{X}}$ is split into a coherent (i.e. noise-free) part $\tilde{\mathbf{X}}^c$ and an incoherent (i.e. purely noisy) part $\tilde{\mathbf{X}}^{inc}$. The signal is first considered as incoherent (i.e. only due to noise) : $\tilde{\mathbf{X}}^{inc} = \tilde{\mathbf{X}}$, thus $\sigma^2_{W,0} = \frac{1}{n}\sum_{i,j}|\tilde{X}^{inc}_{i,j}|^2 = \frac{1}{n}\sum_{i,j}|\tilde{X}_{i,j}|^2$ and $T_0 = \sigma_{W,0}\sqrt{2\ln(n)}$. Wavelet coefficients above T_0 are then added to the coherent part

$$\tilde{X}^c_{i,j} = \rho_{T_0}(\tilde{X}^{inc}_{i,j}),$$

while wavelet coefficients below T_0 remain in the incoherent part

$$\tilde{X}^{inc}_{i,j} = (1 - \rho_{T_0})(\tilde{X}^{inc}_{i,j}).$$

The coherent and incoherent parts of the signal are thus recursively constructed, at loop $k+1$, based on the estimates $\sigma^2_{W,k+1} = \frac{1}{n}\sum_{i,j}|\tilde{X}^{inc}_{i,j}|^2$ and $T_{k+1} = \sigma_{W,k+1}\sqrt{2\ln(n)}$. This algorithm is repeated until the number N_W of non-zero coefficients in the incoherent part converges, i.e. $N_{W,k+1} = N_{W,k}$. At the end of this recursive algorithm, $\sigma_W = \sigma_{W,k}$, $T_D = T_k$, and the denoised signal is given by $\hat{\mathbf{X}} = \sum_{i,j}\tilde{X}^c_{i,j}\psi_{i,j}$. This algorithm is stable and converges with a finite number of iterations bounded from above by the number of samples n, although in practice very few iterations are needed (Azzalini et al. 2004).

This iterative process is illustrated in Figure 1. Since a white noise is isotropic it is spherical in an orthogonal basis, and the spheres correspond to the maximum noise magnitude (i.e the threshold) at iterations 0 and 1. Because the initial noise variance $\sigma_{W,0}$ is large, most of the wavelet coefficients are smaller than the calculated threshold T_0. Thus, only a few coefficients are larger than the threshold (they correspond to the bold arrows) and are added to the coherent part of the signal. After the first iteration, the estimated noise variance $\sigma_{W,1}$ is then smaller than $\sigma_{W,0}$, thus $T_1 < T_0$ and the wavelet coefficients such that $|\tilde{X}^{inc}_{i,j}| > T_1$ (dashed arrows) are added to the coherent signal.

2.3. *Adaptive local spatial filtering*

It is interesting to notice that the wavelet thresholding is equivalent to estimating the signal with a filtering that is locally adapted to the signal regularity. This property follows from the fact that the wavelet transform of a function f at scale j and position $x_j(i)$ locally measures the variation of f in a neighbourhood of $x_j(i)$ whose size is proportional to j (Mallat (1999), p165). Rapid transitions in a signal thus create large wavelet coefficients at fine scales.

Given that the wavelet thresholding selectively sets to zero all coefficients below a threshold T, it thus performs an adaptive filtering that depends on the amplitude of the wavelet coefficients. If $|\tilde{X}_{i,j}| > T$ then the coefficients are relatively large and thus are in the neighbourhood of sharp transitions of f at fine scales. Keeping them avoids smoothing sharp signal variations. In the regions where $|\tilde{X}_{i,j}| \leq T$, the coefficients are likely to be small, which means that f is smooth. The noise is then filtered out by setting the wavelet coefficients to zero.

3. Wavelet thresholding of background-error variances

In this section, we detail how the wavelet thresholding method can be applied to filter noisy background-error variances estimated from a finite-size ensemble of data assimilations.

3.1. *Ensemble variances and their sampling statistics*

Background-error variances $\tilde{\mathbf{v}}$ estimated from an ensemble of background errors are affected by a sampling noise, denoted $\mathbf{v}^e$, which directly arises from the finite size of the ensemble :

$$\mathbf{v}^e = \tilde{\mathbf{v}} - \mathbb{E}[\tilde{\mathbf{v}}], \tag{2}$$

where $\mathbb{E}$ stands for the expectation operator and $\tilde{\mathbf{v}}^\star = \mathbb{E}[\tilde{\mathbf{v}}]$ is the expectation of the ensemble-based variances that will be referred to as the noise-free estimated variances. It may be mentioned that this sampling error is non Gaussian, however the central limit theorem ensures that the sampling distribution of $\sum_{k=1}^{N}\mathbf{v}^e$ approaches the Gaussian distribution as the ensemble size $N \to \infty$.

The statistical properties of this sampling noise have been derived by Raynaud et al. (2009), under the assumption of Gaussian background errors. The spatial covariance of the sampling noise is given by

$$\mathbb{E}[\mathbf{v}^e{\mathbf{v}^e}^T] = \frac{2}{N-1}\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star, \tag{3}$$

where $\tilde{\mathbf{B}}$ is the ensemble-estimated $\mathbf{B}$ matrix, $\tilde{\mathbf{B}}^\star = \mathbb{E}[\tilde{\mathbf{B}}]$, N is the ensemble size and $\circ$ stands for the Hadamard product (i.e. an element-wise product). It follows that the noise variance is given by

$$\mathbb{E}[(\mathbf{v}^e)^2] = \frac{2}{N-1}\tilde{\mathbf{v}}^\star \circ \tilde{\mathbf{v}}^\star.$$

However, this formula cannot be used in practice to calculate local noise variances since we need to know in advance the noise-free signal $\tilde{\mathbf{v}}^\star$.

In the case of a white noise, the noise energy is equally distributed among all scales and the noise variance is then

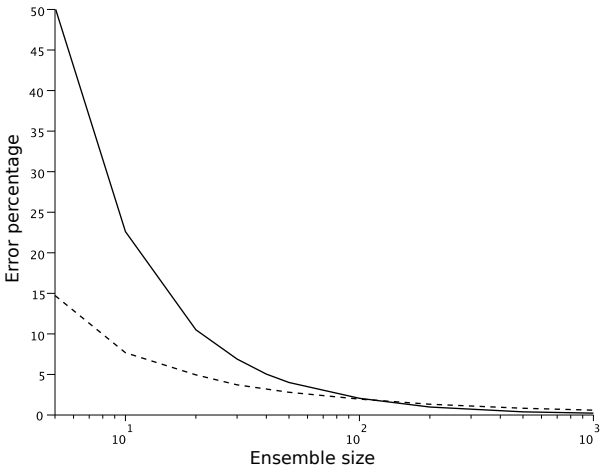


FIG. 2. Convergence of estimation error with the ensemble size : the bias (solid) and the standard deviation (dashed) are both normalized by $Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)$.

equal to the average noise variance :

$$\sigma_W^2 = Tr(\mathbf{E}[\mathbf{v}^e \mathbf{v}^{e^T}])/n.$$

It thus comes from equation (3) that

$$\sigma_W^2 = \frac{2}{N-1} Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)/n. \quad (4)$$

3.2. *Estimation of the noise variance*

It has been mentioned in section 2.2 that the noise variance σ_W^2 may be obtained through a recursive method. On the other hand, one can wonder if the noise variance equation (4) could be used instead. This point is detailed below.

Since $\tilde{\mathbf{B}}^\star$ is unknown in practice, a possible solution to estimate the white noise variance according to equation (4) is to replace $Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)$ by $Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})$. In order to better understand the influence of the finite ensemble size on the estimation of $Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})$, it is interesting to examine the sampling properties of this random variable, in particular its statistical expectation and its standard deviation.

It is shown in Appendix B that

$$\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})] = (1 + \frac{2}{N-1}) Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star). \quad (5)$$

This equation indicates the existence of a positive bias when $Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})$ is estimated from a finite-size ensemble. This bias decreases with the ensemble size at a rate $\mathcal{O}(1/N)$. Therefore, the relative error associated with the bias $\frac{\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})] - Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)}{Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)}$ is around 20% with a 10-member ensemble and decreases to 4% with a 50-member ensemble (Figure 2).

On the other hand, the quality of $Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})$ as an approximation of $Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)$ also depends on the impact of the standard deviation

$$\sigma_{Tr} = \sqrt{\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}}) - \mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})]]^2}$$

on the estimation error. The ratio $\frac{\sigma_{Tr}}{Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)}$, presented in Figure 2, decreases in $\mathcal{O}(1/\sqrt{N})$. Moreover, it can be seen that the standard deviation has a minor impact, compared to the bias, for small ensembles. With a 10 member-ensemble for instance, this ratio is around 7%. This thus shows that with current operational ensemble sizes (namely between 10 and $\mathcal{O}(100)$ members), $Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})$ is a quite accurate estimation of $Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)$. The white noise standard deviation can then be estimated with

$$\sigma_W^2 \approx \frac{2}{N-1} Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})/n. \quad (6)$$

In the context of the ensemble estimation of background-error variances, the statistical properties of the associated sampling noise thus allow us to calculate a relatively accurate estimation of the average noise level.

4. Denoising of 1D variance fields corrupted by a Gaussian white noise

The application of wavelet thresholding to ensemble-based variances is illustrated in this section by estimating variances in an idealized experimental one-dimensional set-up.

An ensemble of N random error realizations of size n is generated using a prescribed "true" background-error covariance matrix. Variances estimated from this ensemble are then decomposed into an orthogonal wavelet series, and the wavelet coefficients are thresholded using the threshold value calculated from equation (1). The filtered variances are finally obtained from an inverse wavelet transform of the selected wavelet coefficients. For the experiments presented in this paper, we use the Coiflet-5 wavelets.

The domain is a circle of radius $a = 6370$km, which corresponds to the Earth's great circle. This circle is divided into $n = 512 = 2^9$ equally spaced gridpoints.

4.1. *The prescribed covariance matrix*

Homogeneous and isotropic correlations are obtained from the Gaussian function

$$C^H(x, r) = \exp(-\frac{r^2}{2L_{\boldsymbol{\epsilon}^b}^2}), \quad (7)$$

where x is a point on the circle, r is a separation value and $L_{\boldsymbol{\epsilon}^b}$ is the correlation length-scale (Daley 1991; Pannekoucke et al. 2008).

Following Pannekoucke et al. (2007), heterogeneous correlations are then computed using a c-stretching Schmidt transformation (Courtier and Geleyn 1988), adapted to the circle and defined by

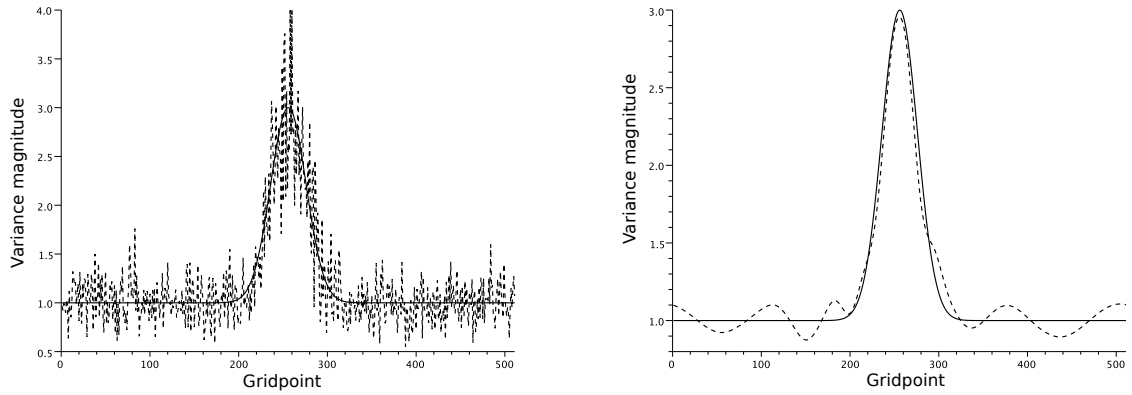
$$h(x) = a[\pi - 2A \tan\{\frac{1}{c} \tan(\frac{\pi}{2} - \frac{x}{2a})\}],$$

with c = 2.4. The resulting correlation function is

$$C(x, r) = C^H(h^{-1}(x), r)\{h^{-1}(x + r) - h^{-1}(x)\}.$$

The associated correlation length-scales are sharper in the center of the domain.

On the other hand, the prescribed variances $\mathbf{v}^\star$ are relatively smooth over a large part of the domain with a value of 1, and there is a sharp transition in the center of the domain with an increase of the variances by a factor 3 (Figure 3). This rapid increase of variances may simulate what is observed in the vicinity of low-predictability events (e.g., mid-latitude storms, tropical cyclones).



(a) (b)

FIG. 3. (a) Prescribed variances (solid line) and raw ensemble-estimated variances (dashed line). (b) Prescribed variances (solid line) and their estimation with a wavelet thresholding (dashed line), in the case of a 50-member ensemble.

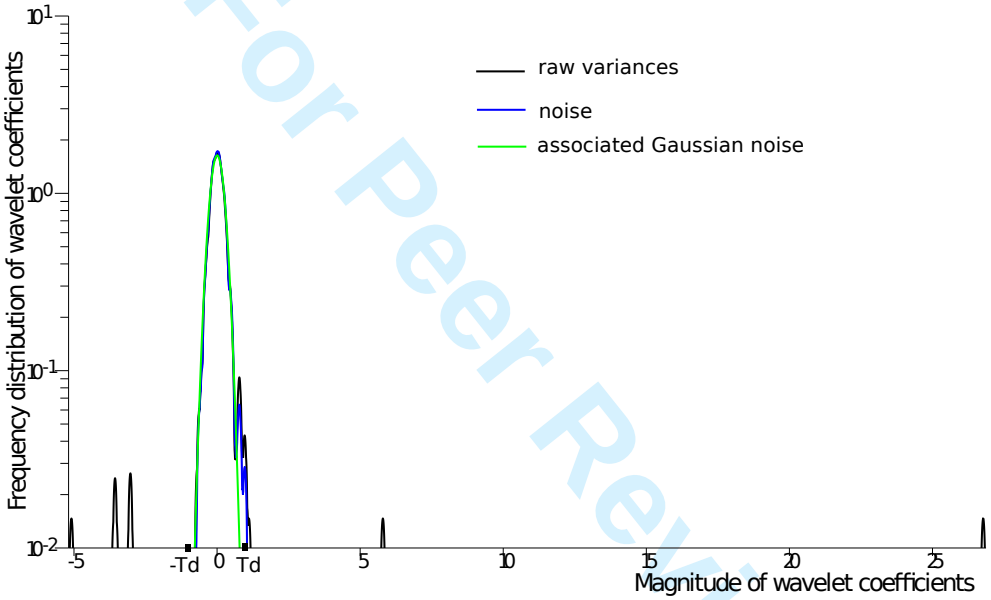


FIG. 4. Histograms of the wavelet coefficients for the raw variances (black) and for the noise (blue). The associated Gaussian distribution for the noise is represented in green. The threshold $T_D \approx 0.92$ is indicated by black squares.

4.2. *Filtering results*

In this section, the efficiency of a wavelet thresholding of estimated variances is examined in the context of a white noise. This is achieved by setting a very low value for the background-error length-scale in constructing the true covariance matrix (equation (7)). Moreover, the noise variance required for the calculation of the threshold value is calculated using both the recursive algorithm (section 2.2) and the theoretical formula (equation (6)).

Figure 3(a) shows the prescribed variances along with the raw variances estimated from a 50-member ensemble. The denoised variance field after a thresholding of wavelet coefficients is shown in Figure 3(b). The sampling noise is removed to a large extent while the spatial variations of the prescribed variances are preserved. The relative error of the estimated variances is reduced from 20% to 7% in average. It may be mentioned that threshold values calculated using the recursive algorithm and the theoretical formula lead to identical denoised variances.

The efficiency of the wavelet thresholding method is related to the separation between signal and noise wavelet coefficients. Figure 4 presents the histograms of the wavelet coefficients for the raw variances $\tilde{X}_{i,j}$ and for the noise $\tilde{W}_{i,j}$. It turns out that the noise coefficients are concentrated within the range $[-T_D, T_D]$ with $T_D \approx 0.92$. Moreover, the noise dominates the signal (i.e. $\tilde{X}_{i,j} \approx \tilde{W}_{i,j}$) within the range $[-T_D, T_D]$. As a result, the useful signal corresponds to the coefficients whose modulus is larger than T_D, and thus it can be efficiently retrieved through wavelet coefficient thresholding.

With smaller ensemble sizes (namely $\mathcal{O}(50)$), the larger amplitude of the noise makes the discrimination between signal and noise more difficult. Therefore, there is some residual noise after the wavelet thresholding, that

could be avoided by slightly increasing the threshold value for instance (not shown).

Under the assumption of a Gaussian white noise, wavelet thresholding is thus a straightforward and efficient method to extract the signal of interest. Moreover, one advantage of the wavelet thresholding is that it does not require any trial and error tuning.

5. Denoising of 1D variance fields corrupted by a Gaussian correlated noise

In this section, the efficiency of wavelet thresholding is examined in the presence of a correlated and heterogeneous noise.

5.1. *'Scale-dependent' threshold*

The theoretical basis of the wavelet thresholding described in section 2 relies on the assumption of a Gaussian white noise. However, in practical applications the sampling noise associated with the estimation of background-error variances is often correlated. This can be seen from equation (3), which implies the following relationship (Raynaud et al. 2009) between the spatial correlation length scales of sampling noise (denoted $L_{\mathbf{v}^e}$) and of background error (denoted $L_{\boldsymbol{\epsilon}^b}$) :

$$L_{\mathbf{v}^e} = \frac{L_{\boldsymbol{\epsilon}^b}}{\sqrt{2}}. \tag{8}$$

The assumption of a white noise is thus verified when background errors are not or slightly correlated, which may be the case in dynamical regions for instance (e.g., in the vicinity of lows and troughs). On the other hand, when background errors are correlated, the associated sampling noise is correlated as well. In addition, according to several studies (Thépaut et al. 1996; Ingleby 2001; Pannekoucke et al. 2007) background-error correlations in realistic NWP (Numerical Weather Prediction) applications are heterogeneous (i.e. $L_{\boldsymbol{\epsilon}^b}$ is not constant in space), which implies that the associated noise is also heterogeneous.

The method of wavelet thresholding has been generalized to correlated noise (Johnstone and Silverman 1997). In this case, one can apply a different threshold for each scale j :

$$T_D(j) = \sigma(j)\sqrt{2\ln(n_j)}, \tag{9}$$

where $\sigma(j)$ is the noise standard deviation associated with scale j and $n_j = 2^j$ is the number of wavelet coefficients at scale j. $\sigma(j)$ and $T_D(j)$ could be estimated with a scale-wise extension of the recursive algorithm presented in section 2.2 (Nguyen et al. 2011).

The difficulty of this approach lies in the estimation of the variance of the wavelet coefficients of the noise at each scale. Two problems can limit the quality of this estimation. First there is a statistical limitation, since the variance at scale j is estimated from 2^j realizations. Since the standard deviation of the relative error in the estimated standard deviation $\sigma(j)$ is equal to $\sqrt{\frac{2}{2^j-1}}$, a relative error smaller than 10% can be achieved only for scales $j \geq 8$. Secondly, at each scale the noise variance is estimated, it is necessary that only a few coefficients are due to the signal. In general, this is only the case at the smallest scales. It may then be expected that the larger scale the noise, the less efficient the denoising.

Finally, the 'scale-dependent' generalization of wavelet thresholding is particularly adapted to a homogeneous noise. In that case, the noise variance is constant within scales so that $\sigma(j)$ can be accurately calculated from the wavelet coefficients of the noise at scale j. If the noise is heterogeneous then the calculated $\sigma(j)$ corresponds to the average noise level at scale j and the 'scale-dependent' thresholding may then be sub-optimal. This problem could be treated with local adaptive thresholding using local noise levels estimated from a local window depending on the spatial statistics of the noise (Goosens et al. 2006), but this is beyond the scope of this paper.

5.2. *'Equivalent' white noise threshold*

Because of the limitations raised by the 'scale-dependent' formulation in the presence of a heterogeneous noise, we propose an alternative solution. The threshold value is calculated using the global *universal* threshold (equation (1)), under the assumption of a white noise with standard deviation $\alpha \times \sigma_W$:

$$T'_D = \alpha \times \sigma_W \sqrt{2\ln(n)}, \quad \alpha \geq 1, \tag{10}$$

where $\sigma_W = \sqrt{Tr(\mathbf{E}[\mathbf{v}^e\mathbf{v}^{e^T}])/n}$ is the average standard deviation of the correlated noise. It may be mentioned that, since the noise is correlated, the recursive algorithm described in section 2.2 is no more efficient to calculate σ_W. In that case, σ_W is estimated from the theoretical formulation (equation (6)), leading to

$$T'_D \approx \alpha \times \sqrt{\frac{4}{(N-1)n} Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})\ln(n)}.$$

A graphical interpretation of this choice is given in Figure 5. White noises are represented by spheres, while a correlated noise is represented by an ellipsoid (which reflects the variation of the noise level with the direction). The formulation of the threshold in equation (10) assumes that the correlated noise is replaced by an 'equivalent' white noise with standard deviation $\alpha \times \sigma_W$. Using a multiplicative factor $\alpha \geq 1$ helps reducing some residual noise arising from the scales where the noise level is above the average level (i.e. $\sigma(j) \geq \sigma_W$). A trivial upper bound for the parameter α is equal to $\max_j \sigma(j)/\sigma_W$. Using this upper bound would result in setting too much coefficients to zero. The choice of the parameter α is thus based on the optimization of the trade-off between the removal of the noise (where $\sigma(j) > \sigma_W$) and the averaging of the useful signal (where $\sigma(j) < \sigma_W$). A possible choice for α could be for instance $\alpha = median(\frac{\sigma(j)}{\sigma_W} \geq 1)$.

5.3. *Filtering results*

The experimental setup is as described in section 4, except that a non-zero correlation length-scale is now used. $L_{\boldsymbol{\epsilon}^b}$ is set to 250km in equation (7), which results in local length scales that vary between 100km in the center of the domain and 600km at the edges of the domain (Figure 7). The sampling noise associated with the estimated variances is then correlated and heterogeneous. In accordance with the prescribed local background-error length-scales, the noise

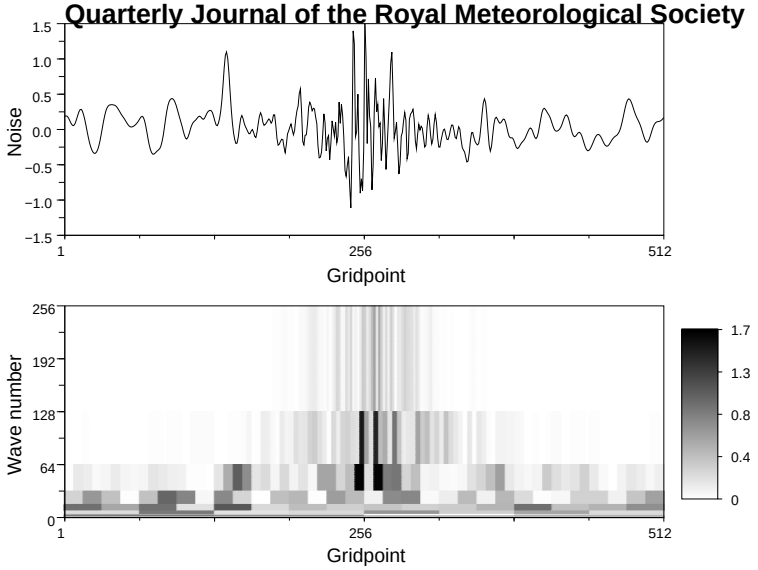


FIG. 6. (a) Geographical variations of the noise. (b) Absolute value of the noise wavelet coefficients $\tilde{W}_{i,j}$. Each coefficient is plotted as a filled rectangle whose color corresponds to the magnitude of the coefficient. The location and size of the rectangle are related to the space interval and the frequency range for this coefficient.

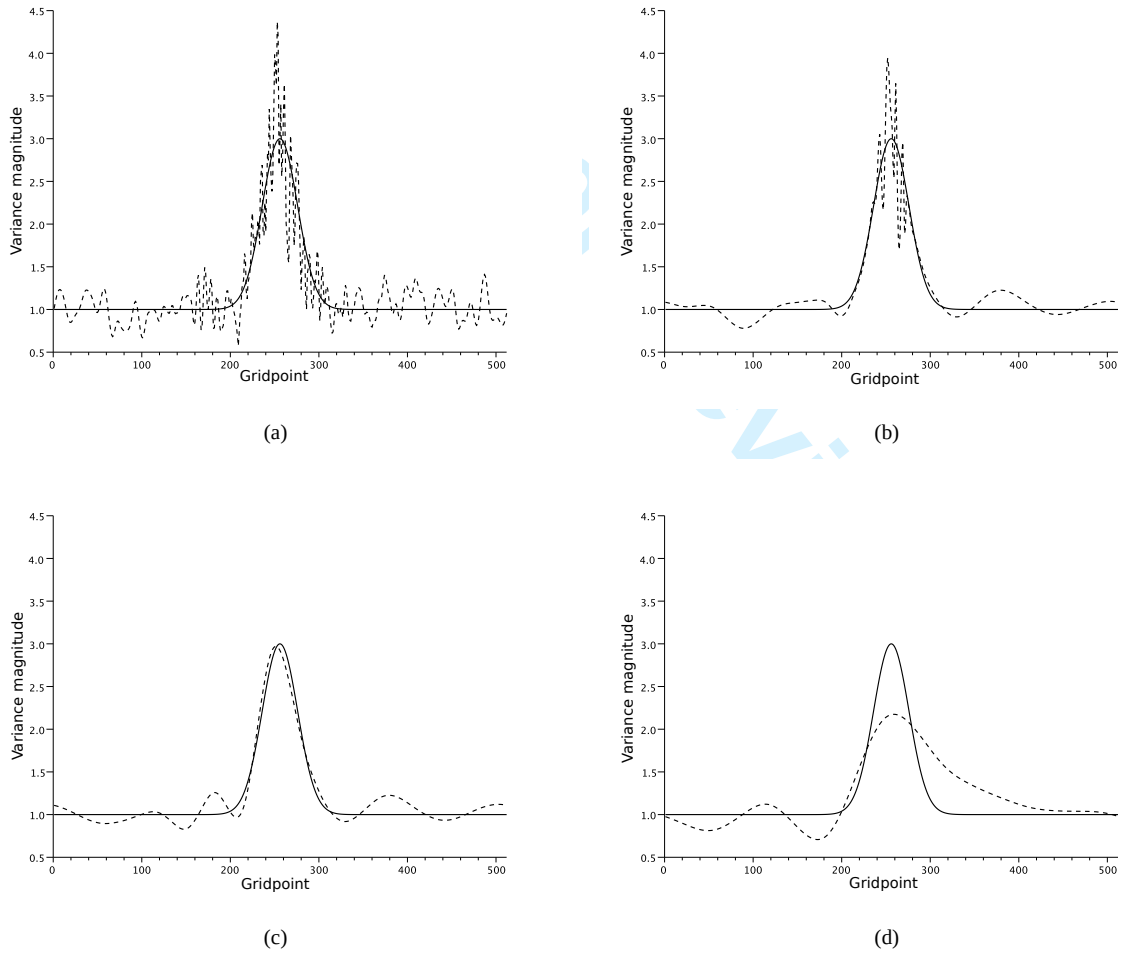


FIG. 9. (a) Raw estimated variances. (b)-(d) Estimated variances with wavelet thresholding using (b) $\alpha = 1$, (c) $\alpha = \alpha_{opt}$ and (d) $\alpha = 2 \times \alpha_{opt}$. The prescribed variances are represented in each panel by the solid line.

presents relatively short variations in the center of the domain while it is larger scale elsewhere (Figure 6(a)). This

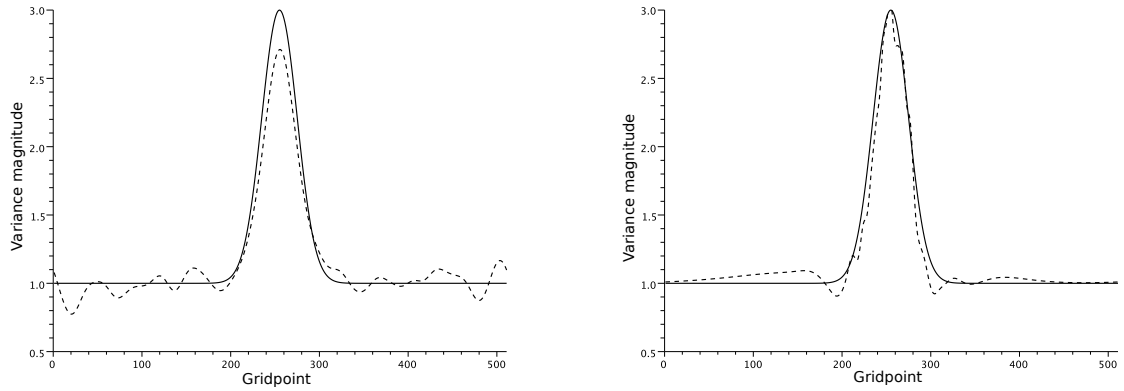


FIG. 10. (a) Estimated variances with an optimized homogeneous filter. (b) Estimated variances with an optimized heterogeneous diffusion-based filter. The prescribed variances are represented in each panel by the solid line.

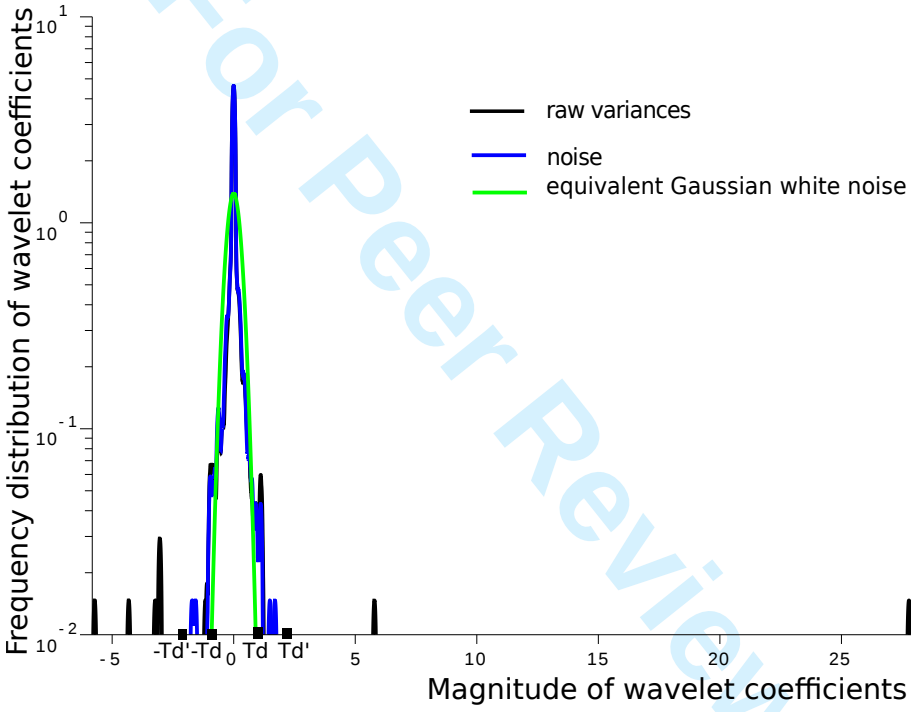


FIG. 11. Histograms of the wavelet coefficients for the raw variances (black) and for the noise (blue). The Gaussian distribution for the equivalent white noise of standard deviation σ_W is represented in green. The thresholds $T_D \approx 0.92$ and $T'_D \approx 2.12$ are indicated by black squares.

is also supported by the scalogram of the noise (Figure 6(b)). As expected, the amplitude of small-scale coefficients tends to be larger in the center of the domain.

The root mean square error of estimated variances as a function of the parameter α, defined by $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{v}}(\alpha)-\mathbf{v}^\star)^2}$, is shown in Figure 8 for a 50-member ensemble. The curve indicates that there is an optimal value that minimizes the error. In the present case $\alpha_{opt} = 2.3$ and the associated wavelet thresholding leads to estimated variances (Figure 9(c)) whose relative error is around 10% in average (compared to 20% for raw estimated variances, Figure 9(a)).

The impact of the choice of α is illustrated in Figures 9(b) and (d). With $\alpha = 1$ (Figure 9(b)), although the variances are much less noisy than the raw estimates, there remains some significant small-scale sampling noise in the vicinity of the variance peak. With $\alpha = 2 \times \alpha_{opt}$ (Figure 9(d)), the filtering is too strong and does not provide an accurate representation of the variance peak. With an appropriate choice of α, it thus turns out that the accuracy of the estimated variances is improved and is close to that obtained in the case of a Gaussian white noise.

The histograms of wavelet coefficients for signal and noise (Figure 11) indicate the presence of noise coefficients larger than the threshold $T_D = \sigma_W\sqrt{2\ln(n)} \approx 0.92$. This is consistent with the discussion of section 5.2 and this

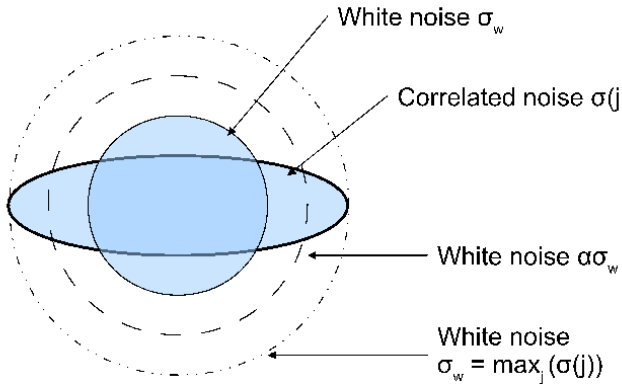


FIG. 5. Graphical representation of white and correlated noises in $\mathbb{R}^n$.

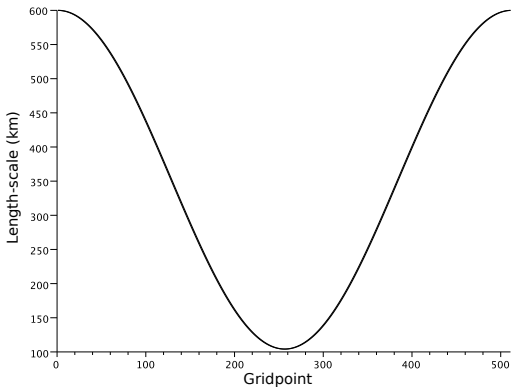


FIG. 7. Geographical variations of the background-error length scale in the analytical framework.

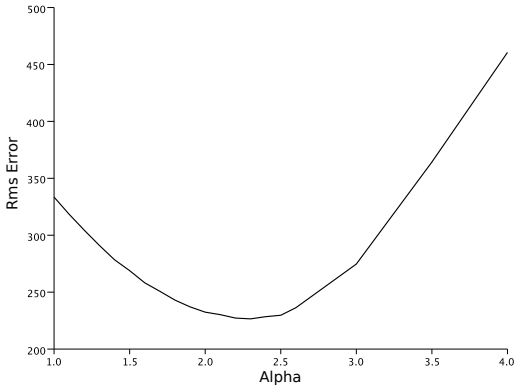


FIG. 8. Total root mean square error of the estimated variances as a function of the α value (equation (10)) in the case of a 50-member ensemble.

justifies the use of a factor $\alpha > 1$. With the optimal $\alpha = 2.3$, $T'_D \approx 2.12$ and one can see that the noise coefficients are concentrated within the range $[-T'_D, T'_D]$. The use of this threshold thus enables all the noise coefficients to be removed, while the useful signal coefficients are preserved. It may also be mentioned that the noise is less Gaussian than in the white noise case (compare blue and green curves).

For comparison purpose, Figure 10(a) presents the estimated variances after applying an optimized homogeneous filter. The signal is quite well-represented with this homogeneous filter, however, as detailed in Raynaud and Pannekoucke (2011), the amplitude of the variance peak is underestimated by around 10% on average. Raynaud and Pannekoucke (2011) also proposed a heterogeneous filter in gridpoint space, based on the integration of the diffusion equation. The comparison of Figures 9(c) and 10(b) indicates that the performance of the wavelet thresholding is comparable to that of an optimized heterogeneous diffusion-based filter.

6. Conclusions and perspectives

This paper introduces and tests a wavelet-based filter of ensemble background-error variances. The filtering is realized using a thresholding of the wavelet coefficients of the estimated variances. It is shown that this approach is equivalent to applying an adaptive local spatial filtering.

The most efficient application of wavelet thresholding is under the assumption of a Gaussian white noise. In that case, the threshold value is simply a function of the signal size and the noise standard deviation. Farge et al. (1999) proposed a recursive algorithm to estimate the noise level. On the other hand, an alternative solution to calculate the noise variance is based on the knowledge of the statistical properties of the associated sampling noise (Raynaud et al. 2009). We showed that the average noise variance can be accurately derived from the raw estimate of background-error variances, even with a small ensemble.

The method has been illustrated for a simple 1D framework. Under the assumption of a white noise, the wavelet thresholding is shown to work well without any trial and error tuning. Moreover, the filtering performance is similar whether the noise variance is calculated with the recursive algorithm or the theoretical formula.

In practical applications, however, the noise is correlated and heterogeneous. This makes the use of the *universal* global threshold and its 'scale-dependent' generalization sub-optimal. An alternative method has been proposed, based on the assumption that the correlated noise is replaced by a white noise with an appropriate variance to calculate a global threshold. The appropriate variance is larger than the average variance of the correlated noise, in order to remove some residual noise arising from the scales where the noise level is above the average level. The results indicate that this method provides variances whose accuracy is close to the white noise case. Moreover, this wavelet thresholding is shown to outperform the commonly used homogeneous filtering and it compares favourably to heterogeneous filtering in gridpoint space.

The encouraging results of this study thus suggest that wavelet thresholding is a feasible and efficient approach for heterogeneous filtering of ensemble variances. Future work will consider the application of the nonlinear thresholding in a more complex 2D framework.

The formalism presented in this paper is well adapted to the variance filtering in limited area models. Extension to global models on the sphere could be considered using biorthogonal wavelets and frames (especially tight frames) as suggested by Pannekoucke (2009).

Acknowledgments

The authors would like to thanks the French Agence Nationale de la Recherche (ANR), under grant Geo-FLUIDS (ANR- 09-SYSC-005- 01) "Analyse et simulation d'écoulements fluides à partir de séquences d'images : application à l'étude d'écoulements géophysiques".

A. Statistical interpretation of the universal threshold

While the mathematical proof of the optimality of the threshold can be found in e.g. Mallat (1999) for instance, we propose here a statistical interpretation of it.

The wavelet shrinkage employed in the paper can be considered as a statistical test on the magnitude of wavelet coefficients. It discriminates coefficients whose magnitude is compatible with a sampling of n Gaussian random variables of standard deviation σ.

For a given magnitude T, the probability that the magnitude of a centered Gaussian random variable X with standard deviation σ exceeds T is given by $P(|X| \geq T) = 2\int_T^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx$. The change of variable $x = T + u$ allows us to upper bound this probability as follows :

$$\begin{aligned} P(|X| \geq T) &= 2\int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(T+u)^2/2\sigma^2} du \\ &= 2e^{-T^2/2\sigma^2} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(2Tu+u^2)/2\sigma^2} du \end{aligned}$$

Since $e^{-2Tu} \leq 1$ for all T, it comes

$$P(|X| \geq T) < 2e^{-T^2/2\sigma^2} \int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} du.$$

Moreover, $2\int_0^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-u^2/2\sigma^2} du = 1$ then leading to $P(|X| \geq T) < p_T$ where $p_T = e^{-T^2/2\sigma^2}$. As an upper bound, p_T is the probability of an event that is more likely to happen than the event "$|X| \geq T$". In particular, p_T can be considered as the probability associated with the event "$|X| \geq T'$" for a certain T' such that $0 < T' < T$.

Let's $(B_k)_{k\in\mathbb{N}}$ be a sequence of identically independent Bernoulli variables associated with the event of probability p_T, so that $P(B_k = 0) = 1 - p_T$ and $P(B_k = 1) = p_T$. Then, the sum $\mathcal{B}_n = \sum_k B_k$ represents the number of occurrences of the event over a sampling of size n. The expected number of extreme events for a size n is thus $\mathbb{E}(\mathcal{B}_n) = np_T$. It is equal to one for a size

$$n_T \sim 1/p_T = e^{T^2/2\sigma^2}, \tag{11}$$

which corresponds to the return period associated with the extreme event "$|X| \geq T'$" of probability p_T (Wilks 2006). This implies that, on average, the event "$|X| \geq T$" should not happen for a sampling size n_T since its expectation of occurence $n_T P(|X| \geq T) = P(|X| \geq T)/p_T$ is lower than 1. Conversely, it follows that the extrem event "$|X| \geq T$" with $T = \sigma\sqrt{2\ln n}$ (resulting from equation (11)) should not occur on average for a sampling size n. Therefore, considering as components of noise wavelet coefficients of magnitude lower than T, while the maximum magnitude of the sampling of n Gaussian noises is T', ensures that the noise is effectively removed with the risk of losing part of the signal. $T = \sigma\sqrt{2\ln n}$ corresponds to the optimal *universal* threshold proposed by Donoho and Johnstone (1994).

Note that for a single sample of a normal law, it is common to assume that $P(|X| < 3\sigma) \approx 1$. Hence, that the threshold T increases with n seems counter-intuitive. However, the Gaussian function does not have a compact support, thus one should find larger values of T as n increases, otherwise the resulting distribution would not be Gaussian.

B. Derivation of $\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})]$

Using the decomposition of $\tilde{\mathbf{v}}$ in equation (2),

$$\begin{aligned} Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}}) &= \sum_l \tilde{\mathbf{v}}^2 = \sum_l (\tilde{\mathbf{v}}^\star + \mathbf{v}^e)^2 \\ &= \sum_l (\tilde{\mathbf{v}}^\star)^2 + \sum_l (\mathbf{v}^e)^2 + 2\sum_l \tilde{\mathbf{v}}^\star \mathbf{v}^e. \\ &= Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star) + \sum_l (\mathbf{v}^e)^2 + 2\sum_l \tilde{\mathbf{v}}^\star \mathbf{v}^e. \end{aligned}$$

Using the linearity of the expectation operator, it then comes

$$\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})] = Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star) + \sum_l \mathbb{E}[(\mathbf{v}^e)^2] + 2\sum_l \mathbb{E}[\tilde{\mathbf{v}}^\star \mathbf{v}^e].$$

Since signal and noise are not correlated $E[\tilde{\mathbf{v}}^\star \mathbf{v}^e] = 0$. Moreover, $\sum_l E[(\mathbf{v}^e)^2] = Tr(\mathbb{E}[\mathbf{v}^e \mathbf{v}^{e^T}]) = \frac{2}{N-1} Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star)$ using equation (3). Thus,

$$\mathbb{E}[Tr(\tilde{\mathbf{B}} \circ \tilde{\mathbf{B}})] = (1 + \frac{2}{N-1}) Tr(\tilde{\mathbf{B}}^\star \circ \tilde{\mathbf{B}}^\star).$$

References

Azzalini, A., M. Farge, and K. Schneider, 2004 : Nonlinear wavelet thresholding : A recursive method to determine the optimal denoising threshold. Appl. Comput. Harmon. Anal., **18**, 177–185.

Bonavita, M., L. Raynaud, and L. Isaksen, 2011 : Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system : some effects of ensemble size and day-to-day variability. Quart. J. Roy. Meteor. Soc., **137**, 423–434.

Courtier, P., and J.-F. Geleyn, 1988 : A global numerical weather model with variable resolution : application to the shallow-water equations. Quart. J. Roy. Meteor. Soc., **114**, 1321–1346.

Daley, R. Atmospheric data analysis. 1991.

Donoho, D. L., and I. M. Johnstone, 1994 : Ideal spatial adaptation by wavelet shrinkage. Biometrika, **81**, 425–455.

Evensen, G., 2003 : The Ensemble Kalman Filter : theoretical formulation and practical implementation. Ocean Dynamics, **53**, 343–367.

Farge, M., K. Schneider, and N. Kevlahan, 1999 : Non-gaussianity and coherent vortex simulation for two-dimensional turbulence using an adaptive orthogonal wavelet basis. Phys. Fluids, **11**, 2187–2201.

Goosens, B., A. Pizurica, and W. Philips, 2006 : Wavelet domain image denoising for non-stationary noise and signal-dependent noise. Proc. 2006 IEEE Int. Conf. Image Processing, Atlanta, USA, pages 1425–1428.

Ingleby, N. B., 2001 : The statistical structure of forecast errors and its representation in The Met. Office Global 3-D Variational data assimilation scheme. Quart. J. Roy. Meteor. Soc., **127**, 209–231.

Johnstone, I. M., and B. Silverman, 1997 : Wavelet threshold estimator for data with correlated noise. J. Roy. Stat. Soc., **59**, 319–351.

Kucukkaraca, E., and M. Fisher, 2006 : Use of analysis ensembles in estimating flow-dependent background error variances. ECMWF Technical Memorandum, **492**.

Mallat, S. A wavelet tour of signal processing. 1999.

Nguyen, R., M. Farge, and K. Schneider, 2011 : Scale-wise coherent vorticity extraction for conditional statistical modelling of homogeneous isotropic two-dimensional turbulence. Physica D.

Pannekoucke, O., 2009 : Heterogeneous correlation modelling based on the wavelet diagonal assumption and on the diffusion operator.. Monthly Weather Review, **137**, 2995–3012.

Pannekoucke, O., L. Berre, and G. Desroziers, 2007 : Filtering properties of wavelets for the local background error correlations. Quart. J. Roy. Meteor. Soc., **133**, 363–379.

Pannekoucke, O., L. Berre, and G. Desroziers, 2008 : Background-error correlation length-scale estimates and their sampling statistics. Quart. J. Roy. Meteor. Soc., **134**, 497–508.

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

Raynaud, L., L. Berre, and G. Desroziers, 2009 : Objective filtering of ensemble-based background-error variances. Quart. J. Roy. Meteor. Soc., **135**, 1177–1199.

Raynaud, L., and O. Pannekoucke, 2011 : Heterogeneous filtering of ensemble-based background-error variances. Quart. J. Roy. Meteor. Soc.

Thépaut, J., P. Courtier, G. Belaud, and G. Lemaître, 1996 : Dynamical structure functions in a four-dimensional variational assimilation : A case study. Quart. J. Roy. Meteor. Soc., **122**, 535–561.

Wilks, D. Statistical Methods in Atmospheric Science. 2006.