REVIEW

Filtering of Background Error Variances and Correlations by Local Spatial Averaging: A Review

LOÏK BERRE AND GÉRALD DESROZIERS
CNRM/GAME, Météo-France, CNRS, Toulouse, France

(Manuscript received 24 June 2009, in final form 12 February 2010)

ABSTRACT

The use of local spatial averaging to estimate and validate background error covariances has received increasing attention recently, in particular in the context of variational data assimilation for global numerical weather prediction. First, theoretical and experimental results are presented to examine spatial structures of sampling noise and signal in ensemble-based variance fields in this context. They indicate that sampling noise tends to be relatively small scale, compared to the signal of interest. This difference in spatial structure motivates the use of spatial averaging techniques.

Based on the usual linear estimation theory, it is shown how this information can be taken into account in order to calculate and apply an objective spatial filter. This kind of approach can also be used to compare and validate ensemble-based variances with innovation-based variances. The use of spatial averaging is even more important for innovation-based variances because local innovations correspond to single error realizations.

Similar ideas can be considered for the estimation of correlation functions. The spatial structures of sampling noise and signal in correlation length scale fields suggest that space-averaging techniques could also be applied to correlation functions. The use of wavelets for this purpose is presented in particular. Connections with related approaches in different contexts such as ensemble Kalman filters and probabilistic forecasting are also discussed.

1. Introduction

Usual data assimilation systems for numerical weather prediction (NWP), using Kalman filter or variational techniques, are based on a statistical combination of observations and a background, which is usually a short-term forecast. This statistical estimation requires the specification of spatial covariances of errors in these two kinds of information. As presented in Hollingsworth (1987) and Daley (1991, p. 125), the role of background error covariances is to spatially filter and propagate the observed information. However, specifying these error covariances remains a difficult scientific and technical challenge, for several reasons. In particular, the true state is never exactly known in practice. This implies that such error covariances have to be estimated approximately.

A first type of technique to estimate these covariances is based on innovation vectors, which correspond to differences between observations and the background (e.g., Hollingsworth and Lønnberg 1986). Innovation covariances can be written as the sum of background and observation error covariances, and these two contributions can be distinguished, under some assumptions on the spatial structures of these two error components.

A second type of estimation technique relies on error simulation methods. These correspond in particular to Monte Carlo techniques, based on the use of an ensemble of perturbed assimilations (e.g., Houtekamer et al. 1996; Fisher 2003; Berre et al. 2006). In these techniques, observation and background perturbations are added (either explicitly or implicitly) to the unperturbed assimilation system in order to simulate associated error contributions and their effects on the error cycling of the assimilation system. Background
perturbations correspond partly to the forecast evolution of the previous analysis perturbations, and also to possible additional model perturbations representative of model errors.

In all cases, background error covariances are estimated from a finite sample of difference fields (typically, either background-minus-observation differences, or differences between ensemble members and the ensemble mean). The finite size of this sample implies that the covariance estimation is affected by a detrimental sampling noise. This noise should preferably be filtered out, in order to improve the accuracy of the covariance estimation.

The effects of sampling noise on the estimation of background error correlations have been studied in the context of ensemble Kalman filter (EnKF) systems in particular (e.g., Hamill et al. 2001). This has led to the use of Schur filter techniques (e.g., Houtekamer and Mitchell 2001), in which the correlation value is left unchanged at the origin of the correlation function, and then there is a gradual reduction of correlation values toward zero as separation distance increases.

There has been less research on possible filtering of the local variances, which are in fact another important component of the background error covariances. Such an issue can be raised both for ensemble-based and innovation-based variance estimates. Evaluating and improving the accuracy of these two kinds of variance estimates is all the more important as innovation-based estimates can be used to represent model error contributions, by calibrating additive model perturbations (e.g., Houtekamer et al. 2009) or ensemble inflation factors (e.g., Li et al. 2009). Moreover, other approaches than Schur filter techniques may be considered for filtering out sampling noise in correlation estimates.

In this paper, we will review the use of local spatial averaging techniques to filter variances (estimated either from an ensemble or from innovations) and ensemble-based correlation estimates. These issues have received increasing attention recently, especially in the context of variational data assimilation for global NWP. Therefore, recent approaches in the latter framework will be examined in the first instance. On the other hand, connections with related approaches in different contexts such as ensemble Kalman filters and probabilistic forecasting will also be discussed.

The paper is structured as follows. In section 2, the main context of ensemble variational data assimilation for global NWP is presented, together with motivations for the review. In section 3, the respective spatial structures of sampling noise and signal for ensemble-based variances will be examined in the context of interest. Accounting for these spatial structures leads to the calculation and application of objective spatial filters, as described in section 4. Applications to calculate spatially averaged innovation-based variances will be discussed in section 5. Spatial averaging of background error correlations will then be presented in section 6. Connections with related approaches in different contexts are discussed in section 7. Conclusions and perspectives are given in section 8.

2. Context and motivation

a. Variational data assimilation and covariance modeling

Data assimilation aims to determine an optimal combination of observations \( y \) and a background \( x^b \), which is usually a short-range forecast. Following classic linear estimation theory (Talagrand 1997), the best linear unbiased estimate of the model state, called the analysis \( x^a \), is given by

\[
x^a = x^b + K(y^o - Hx^b).
\]

Here \( H \) is the observation operator (assumed to be linear here for the sake of simplicity), which projects the model state to observation space. The difference between the observation vector and the background vector projected to observation space is the innovation vector, also denoted \( \delta y = y^o - Hx^b \). Here \( K \) is the gain matrix, corresponding to \( K = BH^T(HBH^T + R)^{-1} \), where \( B \) and \( R \) are the error covariance matrices of the background and of observations, respectively. In a three-dimensional framework, \( B \) is the spatial covariance matrix of background errors \( e^b = x^b - x^b \), where \( x^b \) is the true state in model space. In the case where background errors are unbiased, this covariance matrix is defined by \( B = E[e^b(e^b)^T] \), where \( E[\cdot] \) is the expectation operator.

The diagonal of \( B \) corresponds to background error variances, which reflect expected amplitudes of background errors. Normalizing the covariance matrix by standard deviations provides the correlation matrix \( C = \Sigma_b^{-1}B\Sigma_b^{-1} \) (where \( \Sigma_b \) is the diagonal matrix of background error standard deviations). These spatial correlations reflect the spatial structure of background errors: if the structure of background errors is predominantly large scale (small scale), correlation functions (which are columns of the correlation matrix) will tend to be broad (sharp).

As presented in Hollingsworth (1987) and Daley (1991, p. 125), the role of background error covariances is to spatially filter and propagate observed information. Typically, when background error variances are large (small), the analysis will tend to fit observations to a large (small) extent. Moreover, if spatial correlation functions of background errors are broad (sharp), small-scale observed
features will tend to be filtered out (preserved) by the analysis, and observed information will be propagated spatially to large distances (small distances).

It can be shown (Lorenc 1986, p. 1180) that an equivalent way to calculate the analysis \( x^a \) is to minimize a cost function, that measures quadratic departures from observations and from the background. This approach corresponds to variational data assimilation techniques (Le Dimet and Talagrand 1986), which can be either three-dimensional (i.e., 3D-Var), or, when the time dimension is taken into account, four-dimensional (i.e., 4D-Var).

In variational assimilation, it is usual to model the background error covariance matrix \( B \) with a sequence of sparse operators (e.g., Derber and Bouttier 1999):

\[
B = \mathbf{L}_B \mathbf{L}_u^T,
\]

where \( \mathbf{L} \) is a balance operator, describing multivariate relationships (such as geostrophy) between the different variables such as mass, wind, and humidity (in the context of NWP). Such multivariate relationships can be derived either analytically or statistically by calculating linear regressions between vorticity, divergence, temperature, surface pressure, and humidity. The residual terms in the balance equations, which correspond to the difference between full variables and their balanced components, are called the unbalanced variables. The matrix \( \mathbf{B}_u \) is the covariance matrix of vorticity and of the unbalanced components of the other variables.

In the simplest case, \( \mathbf{L} \) and \( \mathbf{B}_u \) are calculated under a horizontal homogeneity assumption, which enables them to be represented as sparse matrices in spectral space. In particular, \( \mathbf{B}_u \) can be represented as a block-diagonal covariance matrix in spectral space in this case (e.g., Courtier et al. 1998; Lorenc et al. 2000; Gustafsson et al. 2001).

Moreover, the horizontal homogeneity assumption also implies that background error covariances are calculated from global spatial averages. Covariance estimates are relatively robust in this case, even with a small sample of forecast differences, because a large spatial sample is used to calculate the statistics. On the other hand, the homogeneity assumption also means that geographical variations cannot be represented.

In contrast, ensemble Kalman filters (e.g., Evensen 1994; Houtekamer and Mitchell 1998) are often based on covariance functions that are implicitly calculated in a fully local way (i.e., independently for each grid point). This means that a lot of geographical variations can be represented potentially. On the other hand, this requires a relatively large ensemble to make the estimation robust. In practice, the small size of the ensemble implies that filtering techniques may be applied to reduce sampling noise effects.

In this context, an attractive compromise between these two extreme approaches may be to calculate local spatial averages of covariances. We will see that this technique has received increasing attention recently, in particular in the context of ensemble variational data assimilation for global NWP.

b. Error simulation techniques in variational data assimilation

The first versions of operational variational schemes often used background error covariances, which were estimated with the so-called NMC method (Parrish and Derber 1992; Rabier et al. 1998; Berre 2000; Ingleby 2001). This method is based on covariances of differences typically between 24- and 48-h forecasts valid at the same time. In this framework, covariances are usually calculated using a long time average over a dataset of several weeks.

However, as discussed in Berre et al. (2006), the theoretical justification for the NMC method is rather weak, in particular regarding the representation of data density effects on analysis and background error covariances. In this connection, previous studies (Rabier et al. 1998; Bouttier 1994; McNally 2000) indicate that geographical variations of error standard deviations provided by the NMC method are not expected to be realistic.

Therefore, the NMC method tends to be replaced by an alternative approach, which is based on an ensemble of perturbed assimilations (Houtekamer et al. 1996; Fisher 2003; Buehner 2005; Belo Pereira and Berre 2006). In this technique, observation and background perturbations are added (either explicitly or implicitly) to the unperturbed assimilation system, in order to simulate associated error contributions and their effects on the error cycling of the assimilation system. Background perturbations correspond partly to the forecast evolution of the previous analysis perturbations and also to possible additional model perturbations representative of model errors. When applied to a variational scheme, this system will be referred to as an ensemble variational assimilation, by analogy with the ensemble Kalman filter.

While this ensemble assimilation technique was used to specify globally averaged background error covariances in the first trials, it has also opened up new perspectives with respect to the estimation of space- and time-varying covariances.

c. From global to local space and time averages of covariances in variational data assimilation

After initial operational implementations, the continued development of variational assimilation schemes has relied on a gradual relaxation of the homogeneity
hypothesis, in order to represent space and time variations of background error covariances.

First, this was done by using space-dependent balance relationships (e.g., Derber and Bouttier 1999). In particular, it has been shown by Derber and Bouttier (1999) that the use of a linear balance equation (including an implicit representation of latitudinal variations of the Coriolis parameter) allows latitudinal variations of temperature vertical correlations to be represented (see their Figs. 12 and 13). Moreover, as shown later by Fisher (2003), using linearized versions of the flow-dependent nonlinear and omega balance equations allows for the representation of some effects of the jet dynamics on space and time variations of temperature covariances in particular.

Another way to relax the covariance homogeneity assumption is to represent geographical variations of variances while using the homogeneity assumption for correlations. This has been done in the European Centre for Medium-Range Weather Forecasts (ECMWF) variational assimilation system, as described in Derber and Bouttier (1999), by using a Hessian-based estimate of local analysis error standard deviations and a simple inflation algorithm to obtain local background error standard deviation estimates.

In such a context, another possible approach is to use an ensemble of perturbed variational assimilations to estimate geographical variations of background error standard deviations. This has been tried by Belo Pereira and Berre (2006) for the estimation of static but space-dependent standard deviations. A realistic representation of effects from data density variations and atmospheric spatial variabilities was noted, and positive impacts were observed on the forecast quality of the Météo-France NWP system. Deriving flow-dependent local variances from an ensemble variational assimilation was then attempted by Kucukkaraca and Fisher (2006), with encouraging results on cases of midlatitude storms and tropical cyclones.

Another possible step is to relax the application of the homogeneity assumption with respect to background error correlations. As will be further detailed in section 6, this can be achieved by using a wavelet approach (Fisher 2003; Deckmyn and Berre 2005; Rhodin and Anlauf 2007). Other possible techniques correspond to the use of gridpoint recursive filters (Wu et al. 2002) and diffusion operators (Weaver and Courtier 2001).

Calibrating all these techniques with an ensemble of perturbed variational assimilations gives the possibility of estimating and representing space and time variations of background error standard deviations and correlations. All these developments reflect a gradual change from covariance modeling based on global spatial averaging to a specification of local space- and time-varying variances and correlations. On the other hand, this change also means that the sampling noise issue becomes more important, because a smaller spatial sample is used in the local covariance estimation than in the global spatial averaging approach. It is in this context that local spatial averaging techniques have been developed. In particular, local spatial averaging will be shown to be an attractive approach to achieve a robust estimation of time and space variations of covariances.

d. Motivations for the paper

Another important aspect is that innovation-based covariance estimates remain important in this kind of ensemble assimilation context, for the estimation of model error contributions (e.g., Daley 1992; Houtekamer et al. 2009). However, the use of such innovation-based covariance estimates has been mostly restricted to global spatial averages of these estimates up to now. Moreover, while the sampling noise issue has received much attention for the estimation of correlations in the context of EnKFs, there has been less research on sampling noise for the estimation of local variances in such EnKF contexts. Therefore, one of the first goals of the present paper is to show that the sampling noise issue and the concept of local spatial averaging can be considered not only for ensemble-based correlations but also for ensemble-based variance fields and for innovation-based variance fields.

Another goal of the paper is to propose a synthesis of formal and experimental results of sampling noise studies and spatial averaging techniques that have been obtained by different NWP centers in the context of variational assimilation [in particular ECMWF, Météo-France, and the High-Resolution Limited-Area Model (HIRLAM)]. This gives an opportunity to discuss relevant results that are somewhat scattered among peer-reviewed journals, workshop proceedings, and Ph.D. manuscripts. The proposed synthesis also includes a new mathematical demonstration to show that spatial averaging techniques of the background error variance field can be optimized objectively, in order to minimize the estimation error variance of this field (section 4b). A new interpretation of this optimized filter, in terms of amplitude and phase information, is also presented in the appendix.

A third goal is to show that there are connections with related approaches that are emerging in other contexts, such as EnKF and ensemble prediction. While the approaches and contexts may be somewhat different, it is interesting to note that they tend to be related to each other.

Throughout the paper, in addition to 1D idealized small-dimension systems, results from operational NWP systems will be discussed. These NWP systems are high dimensional, with a typical size of model space around $10^9$. 

Unauthenticated | Downloaded 12/30/23 06:21 PM UTC
3. Spatial structure of sampling noise and signal in ensemble-based variance fields

The idea of spatial averaging is closely connected to the notion of spatial structure in the variance field of interest. This notion of spatial structure refers to the fact that a given field can be either relatively large scale (e.g., a forecast field of geopotential at 500 hPa), or relatively small scale (e.g., a forecast field of humidity).

Following previous studies, it will be shown here that spatial structures tend to be different for the expectation of ensemble-based variance fields on the one hand, and for the associated sampling noise on the other hand. Therefore, before discussing the concept of spatial averaging, in this section, we will summarize theoretical and experimental results regarding the respective spatial structures of sampling noise and of the signal of interest.

a. Ensemble estimation of variance fields and notations

We will consider the ensemble estimation of background error variance fields for a given date and for a given parameter (typically, in a global NWP context, temperature, surface pressure, vorticity, divergence, or specific humidity). Such variance fields are three dimensional (except for surface pressure, which is 2D), with varying values for different vertical levels and different grid points in the horizontal direction.

The true variance field will be denoted \( \mathbf{v}^* \). The associated ensemble variance estimate is denoted \( \hat{\mathbf{v}} \) as

\[
\hat{\mathbf{v}} = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{x}_k^b - \overline{\mathbf{x}}^b) \circ (\mathbf{x}_k^b - \overline{\mathbf{x}}^b),
\]

where \( N \) is the ensemble size, and \( \mathbf{x}_k^b \) is the perturbed background for member \( k \) of the ensemble, which can be seen as resulting from the implicit addition of background perturbations \( \mathbf{e}_k^b \) to an unperturbed background \( \mathbf{x}^b \) (corresponding to an unperturbed assimilation cycle):

\[
\mathbf{x}_k^b = \mathbf{x}^b + \mathbf{e}_k^b.
\]

Here \( \overline{\mathbf{x}}^b \) is the associated ensemble average:

\[
\overline{\mathbf{x}}^b = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i^b.
\]

The symbol \( \circ \) corresponds to the Schur product (element by element product, also called Hadamard product) of the two vectors.

By definition, the estimation error corresponds to \( \hat{\mathbf{v}} - \mathbf{v}^* \). Moreover, as for any estimation error, it can be expanded as the sum of two contributions, namely a systematic error component (usually called bias), corresponding to the expectation of the estimation error \( \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*] \), and a residual term, \( (\hat{\mathbf{v}} - \mathbf{v}^*) - \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*] \), which is the random error component (by definition):

\[
(\hat{\mathbf{v}} - \mathbf{v}^*) = \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*] + (\hat{\mathbf{v}} - \mathbf{v}^*) - \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*].
\]

Moreover, after noting that \( \mathbf{E}[\mathbf{v}^*] = \mathbf{v}^* \), the random error can be equivalently (and more simply) written as

\[
\hat{\mathbf{v}} - \mathbf{E}[\hat{\mathbf{v}}] = \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*] + (\hat{\mathbf{v}} - \mathbf{v}^*) - \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*].
\]

From a general point of view, the expectation \( \mathbf{E}[\hat{\mathbf{v}}] \) of the ensemble estimate of the variance field can be different from the true value \( \mathbf{v}^* \), and this corresponds precisely to the systematic error component: \( \mathbf{E}[\hat{\mathbf{v}}] - \mathbf{v}^* = \mathbf{E}[\hat{\mathbf{v}} - \mathbf{v}^*] \). In the context of an ensemble of perturbed assimilations, this systematic error can arise, for instance, from approximations in the observation covariance matrix from which observation perturbations are derived or from approximations in inflation factors which can be applied, either to represent model error contributions (e.g., Buehner 2005) or to compensate for possible in-breeding effects (e.g., Sacher and Bartello 2008). They can also correspond to approximations in the noisy way the reference analysis/forecast steps are simulated in the ensemble (e.g., the ensemble may have a lower resolution than the reference model whose errors are to be simulated by the ensemble).

The other contribution, namely, \( \hat{\mathbf{v}} - \mathbf{E}[\hat{\mathbf{v}}] \), corresponds to the random error component in the variance estimation. It will be referred to as the sampling noise, denoted \( \mathbf{v}^s \):

\[
\mathbf{v}^s = \hat{\mathbf{v}} - \mathbf{E}[\hat{\mathbf{v}}].
\]

If the number of members was infinite, this sampling noise would be equal to zero. Thus, this error component corresponds directly to the fact that the ensemble size is finite. Moreover, it may be observed that this random error arises from the fact that the ensemble-based variance estimate varies randomly around its expected value \( \mathbf{E}[\hat{\mathbf{v}}] \) (as will be illustrated later in Fig. 1), depending on the particular values of the random background perturbations.

Finally, it is also classic to note that the mean square of the estimation error is the sum of the squared systematic error and of the mean square of the random error:

\[
\mathbf{E}[(\hat{\mathbf{v}} - \mathbf{v}^*)^2] = \mathbf{E}[(\hat{\mathbf{v}} - \mathbf{v}^*)^2] + \mathbf{E}[(\hat{\mathbf{v}} - \mathbf{E}[\hat{\mathbf{v}}])^2].
\]
filtering. Moreover, the expectation of \( \mathbf{v} \) will be denoted \( \mathbf{v}^* = \mathbf{E}[\mathbf{v}] \), and will be referred to as the noise-free signal of interest in our review and study.

b. Spatial structure of sampling noise in 1D idealized contexts

The spatial structure of sampling noise has been notably examined in 1D idealized contexts, as in Raynaud et al. (2008, 2009). In these idealized frameworks, the ensemble variance estimates are calculated from random perturbations \( \mathbf{e}^b \), directly drawn from a specified covariance matrix, which is taken to represent a true variance matrix.

Figure 1 is an example of variance fields in a one-dimensional framework, which corresponds to an equatorial circle (of 40 000-km circumference). The full line is a very simple example of (specified) true variance field, which corresponds to a wavenumber-1 sinusoid. Specified background error correlation functions have a length scale equal to 200 km (which corresponds to the grid spacing in this example). It may also be mentioned that this 200-km length scale is similar to temperature length scale values diagnosed in Rabier et al. (1998) and Belo Pereira and Berre (2006). The dashed line is an associated ensemble estimate of the variance field, from a 50-member ensemble (\( N = 50 \)). A first obvious feature is that the large-scale sinusoidal shape of the true variance field is well visible in the ensemble variance field.

But another striking feature is also that the ensemble-based variance field is affected by sampling noise, which appears to be relatively small scale (as indicated by high-frequency oscillations), compared to the true variance field. In other words, while the signal of interest is large scale here, sampling noise is relatively small scale in this simple example. This may look surprising at first sight, in the sense that, knowing that the true variance field is large scale, one could have expected that the associated sampling noise \( \mathbf{v}^e \) should also be large scale. The reason for this apparent paradox can be found in the analytical expression of the spatial covariance of sampling noise (Raynaud et al. 2009):

\[
\mathbf{E}[\mathbf{v}^e(\mathbf{v}^e)^T] = \frac{2}{N-1} \mathbf{B}^* \circ \mathbf{B}^*,
\]

where \( \mathbf{B}^* \) is the expectation of the ensemble-based covariance matrix, and \( \mathbf{B}^* \circ \mathbf{B}^* \) is the Schur product of \( \mathbf{B}^* \) with itself. This expression implies, in particular, that the correlation function of sampling noise is the square of the (expected) correlation function of background error:

\[
\text{cor}(v_x^e, v_{x'}^e) = \left( \text{cor}(e_x^b, e_{x'}^b) \right)^2,
\]

where \( x \) and \( x' \) are indices of two grid points. This indicates that the spatial structure of sampling noise is closely connected to the spatial structure of background error. This can also be seen through the relationship between the associated correlation length scales of sampling noise [denoted \( \mathbf{L}(\mathbf{v}^e) \)] and of background error [denoted \( \mathbf{L}(\mathbf{e}^b) \)]:

\[
\mathbf{L}(\mathbf{v}^e) = \mathbf{L}(\mathbf{e}^b)/\sqrt{2}.
\]

In other words, Eqs. (6) and (7) indicate that sampling noise is smaller scale than the background error field. And the resulting length scale of sampling noise is then to be compared to the specified length scale of the true variance field. In Fig. 1, the specified length scale of background error is equal to 200 km (leading to a sampling noise length scale around 140 km), while the true variance field corresponds to a sinusoid with a 40 000-km wavelength. This explains why, in this simple example, sampling noise is much smaller scale than the (specified) true variance field. A similar 1D example is also shown in Fig. 6 of Fisher and Courtier (1995).

This contrast in terms of spatial structure is of course, particularly strong in these examples, because the specified true variance field is very large scale in this idealized framework. The next section is devoted to the examination of spatial structures in NWP ensemble variational assimilation experiments.

c. Qualitative studies in high-dimensional NWP systems

Indications about the spatial structure of variance fields can be obtained in the literature, by examining examples of variance maps in the context of ensemble variational assimilation for NWP. Figure 2 is an example...
of such a map from Isaksen et al. (2007), derived from a 10-member ensemble of perturbed 4D-Var experiments. (Note that this is a map of ensemble-based estimates of background error standard deviations, denoted $\sigma_b$, which are simply the square root of the variances $\tilde{\Psi}$.) What is striking is that large scale features tend to predominate in this typical example. They correspond to data density contrasts and synoptic variations, related to the position and amplitude of troughs and ridges. This is a first experimental example that illustrates that the signal of interest can be relatively large scale. This is also consistent with Fig. 8 in Houtekamer and Mitchell (1998) in an EnKF context. On the other hand, this single figure does not give indications about the specific spatial structure of sampling noise.

Therefore, one way to obtain more information is to examine how spatial structures in variance maps change when the ensemble size increases. Typically, it is expected that, when the ensemble size increases, the sampling noise amplitude is reduced, whereas the noise-free signal contribution will remain. An example is shown in Fig. 3 (from Isaksen et al. 2007). A map of 10-member $\sigma_b$ estimates is shown in the top panel and a corresponding map of 50-member $\sigma_b$ estimates is presented in the bottom panel. As mentioned by Isaksen et al. (2007), the 50-member map is relatively close to the 10-member map, except that it is smoother. This corresponds to the fact that apparent small scale details in the 10-member map, such as local oscillations around 38°N, 170°E, tend to disappear in the 50-member ensemble estimate. Because of this attenuation when the ensemble size increases, these small-scale details can be interpreted as sampling noise. In contrast, large-scale features tend to remain, such as the large-scale area of strong standard deviation values east of Japan. Since they are preserved when the ensemble size increases, these large-scale structures can be interpreted as relatively noise-free contributions, arising from the signal of interest. This simple experimental example suggests that, in the context of this study, sampling noise tends to be relatively small scale, compared to the signal of interest.

FIG. 2. Raw ensemble-based standard deviation field ($\times 2$) of 850-hPa zonal wind background error over the Northern Hemisphere from a 10-member ensemble variational assimilation. Isocontour spacing: 1 m s$^{-1}$. Isolines correspond to the 500-hPa geopotential field (isoline interval: 8 dm). [©ECMWF 2007, from Isaksen et al. (2007). Reproduced by permission of the ECMWF.]
Another way to confirm this is to compare spatial structures of ensemble-based estimates, derived from two independent ensembles. As discussed in section 3.1 of Berre et al. (2007), qualitatively, it is expected that common structures in the two associated $\sigma_b$ maps correspond to signal contributions, while differences correspond to sampling noise effects. An example of this kind of comparative studies is shown in Fig. 4, for two independent three-member ensembles of perturbed assimilations. A visible common feature is the large-scale contrast between large $\sigma_b$ values in the Atlantic trough, and small $\sigma_b$ values in the neighboring ridge. Conversely, there are also many structures that differ between the two maps, in particular at the small scales. Again, this example supports the idea that the signal of interest tends to be larger scale than the associated sampling noise in the ensemble estimation.

Beyond these first simple qualitative studies, some quantitative investigations have also been conducted. They are described in the next two sections.

d. Quantitative estimation formalism for high-dimensional systems

To characterize the spatial structure of global meteorological fields, energy spectra are generally calculated in the space of spherical harmonics (e.g., Boer 1983). In this section, we will see how this approach has been used formally and experimentally by Berre et al. (2007) and Raynaud et al. (2009), in order to characterize the respective spatial structures of signal and noise in ensemble-based variance estimates.

A spectral transform $S$ can be applied to a geographical field of variances, to provide spectral coefficients of this gridpoint variance field (which is different from the variance spectrum of the background error field $e^b$). This corresponds to the following two equations, for the ensemble-based variance field and for its expectation, respectively, $\mathbf{s} = S(\mathbf{v})$ and $\mathbf{s}^* = S(\mathbf{v}^*) = \mathbf{E}[\mathbf{s}]$. The transform $S$ can also be applied to the gridpoint field of sampling noise $v^*_s$, to provide the spectral field of sampling noise, denoted $s^s = S(v^*_s)$, which can thus be written as the difference between $\mathbf{s}$ and $\mathbf{s}^*$:

$$s^s = S(v^*_s) = S(\mathbf{v} - \mathbf{v}^*) = S(\mathbf{v}) - S(\mathbf{v}^*) = \mathbf{s} - \mathbf{s}^*.$$  

Therefore, the ensemble-based spectral field $\mathbf{s}$ can be decomposed as the sum of its expectation $\mathbf{s}^*$ and of sampling noise $s^s = S(v^*_s)$:

$$\mathbf{s} = \mathbf{s}^* + s^s.$$  

Here $\mathbf{s}$ and $\mathbf{s}^*$ will be referred to as the raw and noise-free signals, respectively.

Under the assumption that sampling noise is a random process (arising from the randomness of observation and background perturbation values), $s^s$ is uncorrelated with the signal $\mathbf{s}^*$. The power variance spectrum of the ensemble-based variance field may thus be decomposed as follows:

$$p(\mathbf{s}) = p(\mathbf{s}^*) + p(s^s),$$  

Where, for example, $p(\mathbf{s})$ is the power variance of $\mathbf{s}$ defined, for a given total wavenumber $n$, by

$$p(\mathbf{s})(n) = \sum_{m=-n}^{+n} \bar{s}(n,m)s(n,m)^*.$$  

where $m$ is the zonal wavenumber and $\bar{s}(n,m)^*$ is the conjugate of $\bar{s}(n,m)$. Note that this power variance appears to be a “variance of the variance.” Therefore, it has the units of the original variable to the fourth power.

In practice, $p(\mathbf{s})$ can be calculated directly from the ensemble-based estimate $\mathbf{s}$. Moreover, the power variance
of sampling noise $p(s^e)$ can be estimated in two different ways.

1) METHOD 1

In Berre et al. (2007), $p(s^e)$ is estimated by comparing variance fields from two independent ensembles of equal size $N$. To demonstrate this, previous equations can be applied to these two ensembles, by using $i$ as an ensemble index (i.e., $i = 1$ for the first ensemble and $i = 2$ for the second). In particular, $\hat{s}_i = s^* + s^e$ leads to

$$\hat{s}_1 - \hat{s}_2 = (s^* + s^e_1) - (s^* + s^e_2) = s^e_1 - s^e_2,$$  \hspace{1cm} (12)

which implies for the power variances:

$$p(\hat{s}_1 - \hat{s}_2) = p(s^e_1) + p(s^e_2) - 2\text{cov}(s^e_1, s^e_2),$$  \hspace{1cm} (13)

where $\text{cov}(s^e_1, s^e_2)(n) = \sum_{m=-n}^{n} s^e_1(n,m)s^e_2(n,m)$. As the random realizations of the two noises $s^e_1$ and $s^e_2$ arise from two independent random sets of background perturbations, it can be assumed that $\text{cov}(s^e_1, s^e_2) = 0$. Moreover, as the two independent ensembles have the same size, it can be assumed that $p(s^e_1) = p(s^e_2)$, which can thus be denoted $p(s^e)$. This leads to $p(\hat{s}_1 - \hat{s}_2) = 2p(s^e)$, which means that the noise power variance can be estimated as half of the variance of $\hat{s}_1 - \hat{s}_2$:

$$p(s^e) = \frac{1}{2} p(\hat{s}_1 - \hat{s}_2).$$  \hspace{1cm} (14)

2) METHOD 2

Another approach to estimate $p(s^e)$ has been proposed by Raynaud et al. (2009). It is based on an analytical expression of the spatial covariance of sampling noise, corresponding to Eq. (5). This equation indicates that the sampling noise covariance can be calculated from an estimation of the expectation of the ensemble-based covariance matrix.

If one is interested in diagnosing geographical variations of the spatial structure of sampling noise, one can use Eq. (5) at different geographical locations. Conversely, if one is interested in diagnosing the average spatial structure of sampling noise over the globe (as a first step), it is sufficient to estimate spectral variances corresponding to $\mathbf{B}^*$. These spectral variances are associated with the global spatial average of the sampling noise covariance functions through a direct spectral
Legrendre transform (e.g., Courtier et al. 1998). This means that, in this case, it is sufficient to estimate the homogeneous component of \( B^* \).

Moreover, once this homogeneous framework is considered, it can be shown that the associated global covariances are relatively stationary in time. This is illustrated by Fisher (2004) in his Fig. 6 and also by Raynaud et al. (2009) in their Fig. 6. These two figures show that the global spatial average of ensemble-based variances is relatively static. These features suggest that a time-averaged estimate of the homogeneous covariance functions could be used to estimate the global covariances of sampling noise. This is the approach followed by Raynaud et al. (2009).

3) COMPARISON BETWEEN THE TWO ESTIMATION METHODS OF NOISE POWER VARIANCE

It may be mentioned that the two methods differ slightly from a theoretical point of view. The derivation of the second method is based on the assumption that ensemble-based variances (and associated noise fields) are random variables that follow a Gaussian probability distribution function (see the beginning of the appendix of Raynaud et al. 2009). In contrast, the first method is based on an assumption that sampling noise is a random field, without assuming that the associated distribution function is Gaussian. In this sense, the first method is slightly more general.

From a practical point of view, it may also be mentioned that, in the first method, the noise power variance can be calculated directly from the variance fields of the two independent ensembles. In contrast, in the second method, the noise power variance is to be calculated from an estimate of the whole background error covariance matrix \( B^* \) (i.e., not only variances, but also spatial correlations). However, such an estimate is usually required and available for the data assimilation scheme anyway. This means that using this estimate for calculating the noise variance remains relatively easy in the end.

Moreover, the main advantage of the second method is that it requires calculations from a single ensemble, whereas the approach presented in Berre et al. (2007) requires calculations from two ensembles. Although the two-ensemble experiments can be conducted offline, avoiding this factor of 2 in the computation cost may be advantageous, especially if the ensemble system is costly.

Another theoretical and practical aspect is the robustness of power variance estimates. Note that \( s^* \) denotes the sampling noise with respect to the background error variance estimation. There is potentially another kind of sampling noise, this time with respect to the estimation of power variances, such as \( p(s) \) and \( p(s^*) \). In practice, the amplitude of this other sampling noise tends to be reduced by the fact that power variances are calculated as the sum over different zonal wavenumbers \( m \), as shown in the Eq. (11) of \( p(s)(m) \). When applying the first method in Berre et al. (2007), this was further strengthened by the calculation of power variances as a time average over a 49-day period. The implied robustness will be illustrated experimentally in the next section. When applying the second method over a single date in Raynaud et al. (2009), some sampling noise effects were observed (see their Fig. 3), in particular for the first wavenumbers \( n \), for which power variances are calculated from a relatively small number (equal to \( 2n + 1 \)) of wave vectors \((m, n)\). This was treated by fitting an analytical function to the final spatial filter, which was derived from estimated power variances.

The results of these approaches to estimate the average spatial structures of sampling noise and signal are illustrated in the next section.

e. Quantitative estimation results for high-dimensional NWP systems

The top panel of Fig. 5 is an example of power variance spectra for the raw signal (full and dashed lines) and for the associated sampling noise (dotted line) shown in Berre et al. (2007). These power variance spectra were obtained from a time average over a 49-day period. They correspond to ensemble-based \( \sigma_b \) maps calculated from two independent 3-member ensembles, for vorticity near 500 hPa. The full line is the signal power variance for the first ensemble \([p(s_1)]\), and the dashed line is for the second ensemble \([p(s_2)]\). The fact that \( p(s_1) \approx p(s_2) \) is consistent with the expectation that these two quantities are equal to \( p(s^*) + p(s^*) \), according to Eq. (10). This experimental result supports the idea that such time-averaged power variance spectra can be estimated in a robust way.

Moreover, the dotted line is the power variance of sampling noise \( p(s^*) \), and according to Eq. (10), the difference between the top and bottom curves corresponds to the power variance of the noise-free signal: \( p(s^*) \).

On the one hand, it thus appears that the noise contribution to the raw signal is relatively large at small scales, for which \( p(s^*) \approx p(s) \) and thus \( p(s^*) \approx 0 \). In contrast, the noise contribution to the raw signal is relatively small at large scales, for which \( p(s^*) \approx 0 \) and thus \( p(s^*) \approx p(s) \). These contrasted changes in function of horizontal scale are a direct consequence of the fact that the spatial structure of sampling noise tends to be smaller scale than the noise-free signal of interest.

These scale dependencies can be confirmed by calculating \( p \), the relative amount (i.e., the proportion) of noise-free signal in the power variance of the raw signal \( p(s) \), that is,
This proportion $\rho$ of noise-free signal is plotted in the bottom panel of Fig. 5. As expected (from the top panel of Fig. 5), $\rho$ is relatively close to 1 for small wavenumbers, and then it decreases progressively toward 0 for the larger wavenumbers.

Similar results have been obtained by Raynaud et al. (2009), using a different approach for the estimation of $p(s^x)$, as discussed in section 3d. This is illustrated by the top panel of Fig. 6, which shows power variance spectra of the raw signal and of the associated sampling noise, for a six-member ensemble estimation of vorticity variances near the surface. The associated proportion $\rho$ of noise-free signal is shown in the bottom panel of Fig. 6.

4. Optimized spatial averaging of ensemble-based variances

To summarize the previous section and as illustrated by Figs. 5 and 6, it appears on the one hand that...
ensemble-based variance fields tend to be relatively noise free at large scales. On the other hand, the smallest-scale components of these variance fields tend to be dominated by sampling noise. This suggests that this information could be taken into account in order to make the ensemble-based variance estimates more accurate thanks to spatial averaging. As will be illustrated in this section, this is related to the idea that spatial averaging could allow the large-scale noise-free signal of interest to be extracted, while filtering out small-scale sampling noise. This section starts with direct spatial averaging techniques in idealized frameworks and then it is shown how this can be generalized and optimized with simple spatial filtering techniques.

a. Direct spatial averaging in idealized frameworks

Direct spatial averaging of ensemble-based variance fields is presented here as an introduction to more general methods. It has been examined in 1D and 2D idealized frameworks by Raynaud et al. (2008). It is illustrated by Fig. 7 in a 2D experiment, where Fig. 7c corresponds to direct spatial averages, over local circles of diameter $D = 1500 \text{ km}$, of the raw 6-member variance field shown in Fig. 7b. While the raw variance map is contaminated by small-scale sampling noise, the large-scale signal of interest can be extracted through spatial averaging. Moreover, it can also be seen that the filtered 6-member variance field is even slightly more accurate than the raw 220-member variance field (shown in Fig. 7d) in this case.

The direct spatial average corresponds to the following formula, for the resulting filtered variance $\hat{v}(x)$ at grid point $x$:

$$\hat{v}(x) = \frac{1}{N_D} \sum_{x'} \hat{v}(x') = \frac{1}{N_D(N-1)} \sum_x \sum_{k=1}^N \left[ \epsilon_k(x) \right]^2,$$

where indices $x'$ correspond to a set of $N_D$ grid points whose separation distance with respect to grid point $x$ is smaller than or equal to $D$.

The double sum in Eq. (16) illustrates that background error variances are estimated from an average over a total number of perturbation values equal to $N_{\text{tot}} = N_D \times N$. In other words, compared to the raw $N$-member variance estimate, there is a potential increase of sample size up to $N_{\text{tot}} = N_D \times N$. This vision offers a simple interpretation of the increased accuracy of the locally averaged variance field $\hat{v}$: the increased sample size allows sampling error to be reduced.

However, in practice, as discussed in sections 2.1 and 2.2 of Berre et al. (2007) and illustrated in Fig. 8, the effective increase of sample size is likely to depend on background error correlation length scales. Typically, if
the $N_D$ background error values are uncorrelated, the effective total sample size will be equal to its maximum potential value, namely, $N_{\text{tot}} = N_D \times N$. Conversely, if the $N_D$ background error values are fully correlated, they all correspond to the same single value and the effective total sample size will remain equal to $N$. In this case, for which the background error correlation function is relatively large scale (i.e., broad), one would like to average raw variances over a relatively larger domain, in order to obtain an effective increase in sample size.

Another limitation of direct spatial averaging represented in Eq. (16) is that all $N_D$ grid points are given the same uniform weight in the calculation of the spatial average, whatever their distance with respect to the reference point $x$.

These issues suggest the need to search for an objective and optimized filtering approach that takes spatial structures of sampling noise and signal into account. This is presented in the next section.

b. Formalism for optimized spatial averaging

A simple and general formalism for objectively filtering ensemble-based variance fields has been proposed by Berre et al. (2007). The idea is to design the spatial filter as a linear regression between the raw ensemble-based variance field $\mathbf{s}$ and the noise-free signal of interest $\mathbf{s}^*$. The associated regression equation can be obtained by searching for the vector of linear coefficients $\alpha$ such that the linear estimator $\mathbf{\hat{s}} = \alpha \cdot \mathbf{s}$ minimizes the variance $\mathbf{p}(\mathbf{s}')$ of the residual estimation error $\mathbf{s}' = \mathbf{s} - \mathbf{\hat{s}} = \alpha \cdot \mathbf{s} - \mathbf{s}^*$.

This residual error variance $\mathbf{p}(\mathbf{s}')$ can be expanded as follows:

$$\mathbf{p}(\mathbf{s}') = \mathbf{p}(\alpha \cdot \mathbf{s} - \mathbf{s}^*) = \alpha^2 \mathbf{p}(\mathbf{s}) - 2\alpha \cdot \mathbf{cov}(\mathbf{s}, \mathbf{s}^*) + \mathbf{p}(\mathbf{s}^*).$$

Then, minimizing $\mathbf{p}(\mathbf{s}')$ amounts to imposing the condition $\partial \mathbf{p}(\mathbf{s}')/\partial \alpha = 0$, which corresponds to

$$\frac{\partial \mathbf{p}(\mathbf{s}')}{\partial \alpha} = 2\alpha \cdot \mathbf{p}(\mathbf{s}) - 2\mathbf{cov}(\mathbf{s}, \mathbf{s}^*) = 0.$$

It can be verified that this condition leads to a minimum value for $\mathbf{p}(\mathbf{s}')$, since the second derivative $\partial^2 \mathbf{p}(\mathbf{s}')/\partial \alpha^2 = 2\mathbf{p}(\mathbf{s})$ is always positive. This leads to the following classic formula of $\alpha$ as linear regression coefficients:

$$\alpha = \frac{\mathbf{cov}(\mathbf{s}^*, \mathbf{s})}{\mathbf{p}(\mathbf{s})},$$

and, thus, the optimized filtering equation, which provides a filtered signal $\mathbf{\hat{s}}$, is

$$\mathbf{\hat{s}} = \alpha \cdot \mathbf{s} = \frac{\mathbf{cov}(\mathbf{s}^*, \mathbf{s})}{\mathbf{p}(\mathbf{s})} \mathbf{s}.$$  \hspace{1cm} (17)

Moreover, it can also be shown that the regression coefficient is simply equal to $\rho$, the proportion of noise-free signal in the raw variance field $\mathbf{s}$:

$$\alpha = \frac{\mathbf{cov}(\mathbf{s}^*, \mathbf{s})}{\mathbf{p}(\mathbf{s})} \frac{\mathbf{p}(\mathbf{s}^*)}{\mathbf{p}(\mathbf{s})} = \rho.$$ \hspace{1cm} (18)

This corresponds to Eq. (10) for the denominator, while the numerator corresponds to the following expansion:

$$\mathbf{cov}(\mathbf{s}^*, \mathbf{s}) = \mathbf{cov}(\mathbf{s}^*, \mathbf{s}^* + \mathbf{s}^e) = \mathbf{p}(\mathbf{s}^*).$$

under the assumption that sampling noise is uncorrelated with the signal. This is justified by considering sampling noise as a random process arising from the randomness of observation and background perturbations in the ensemble.

In addition, Eq. (18) shows that the regression coefficient can be calculated from estimates of sampling noise and raw signal variances. This can be achieved using techniques described in section 3d.

It may also be mentioned that, according to Eq. (2) in Berre et al. (2007), when using the approach based on the comparison of two independent ensembles, an equivalent way of calculating the regression coefficients is to directly compute the correlations $\text{cov}(\mathbf{\hat{s}}, \mathbf{\hat{s}})$ between the two sets of spectral coefficients $\mathbf{\hat{s}}_1$ and $\mathbf{\hat{s}}_2$. This corresponds to the fact that these correlations are equal to $\rho$, the proportion of noise-free signal power variance in the raw variance fields.
c. Interpretation of the spectral linear regression in terms of convolution

Because of the spectral shape of $\rho$ (see Figs. 5 and 6), the regression corresponds to a multiplication of the raw spectral field by a low-pass filter. In accordance with section 3 concerning the spatial structures in play, this reflects the idea of extracting the large-scale signal of interest while filtering out small-scale sampling noise.

Knowing that spectral coefficients such as $\hat{s}$ contain both amplitude and phase information, it is shown in the appendix that the raw variance field must be filtered for two reasons. First, because of sampling noise, amplitudes of spectral coefficients $\hat{s}$ are too large compared to $\hat{s}^*$. Second, spatial structures in the raw variance field are partly out of phase compared to spatial structures in the signal, due again to sampling noise. As shown in the appendix, a mathematical analysis of these two effects indicates that amplitude and phase errors are of equal importance in the raw variance field.

Moreover, a multiplication in spectral space is equivalent to a convolution in gridpoint space [e.g., Courtier et al. (1998), and theorem 2.9 in Gaspari and Cohn (1999)]. This means that Eq. (17) can also be seen as a convolution, which produces a filtered variance field $\hat{v}$:

$$\hat{v}(x) = \hat{v}(x) \otimes \hat{\rho}(r) dr,$$

where $\otimes$ is the convolution product, and $\hat{\rho}$ is the inverse spectral transform of $\rho$.

Expressed in the 1D continuous case, this allows filtered variances to be seen as resulting from a weighted spatial average:

$$\hat{v}(x) = \int \hat{v}(x - r) \hat{\rho}(r) dr.$$

The application of this low-pass filter can thus be seen as a weighted spatial averaging technique, with weights $\rho(r)$ computed in an optimized way. It can also be mentioned that, because this convolution is calculated through single products in spectral space [see Eq. (17)], it is less costly than direct spatial averaging in physical space, which involves multiple products for each reference position $x$.

As described in the previous section, after applying such a filter, there is a residual estimation error with respect to the signal $\hat{s}^*$, defined by $s^* = \hat{s} - \hat{s}^*$. As shown in appendix C of Raynaud et al. (2009), an estimate of the associated residual error variances is given by

$$p(s^*) = \rho \cdot p(s^*). \quad (19)$$

This expression implies that the error variance reduction, namely $p(s^*) - p(s^*) = (1 - \rho) \cdot p(s^*)$, is proportional to $1 - \rho$. Therefore, the relative error reduction is maximum at the small scales, for which $\rho$ is close to 0. Conversely, there is not much error reduction at the large scales, in accordance with the fact that these large scales are relatively accurate from the beginning, since $\rho$ is close to 1.

d. Experimental results of optimized spatial averaging

Experimental results of this kind of approach are illustrated in a 1D idealized context in Raynaud et al. (2008, 2009), and also in the framework of a real ensemble assimilation in Berre et al. (2007) and Raynaud et al. (2009).

As expected, large-scale coherent structures are extracted by the filter while small-scale details associated with sampling noise tend to be eliminated. An impact study has been carried out by Raynaud et al. (2009) to evaluate the effect of filtered variances compared to raw (unfiltered) variances. It indicates that the use of such objectively filtered variances (instead of raw variances) is beneficial to the forecast quality. Moreover, using Eq. (19), the variance of the residual relative error is estimated to be around 10% for filtered variances from a 6-member ensemble assimilation (whereas the relative error variance of raw variances is around 40%). This is also consistent with quantitative results obtained in the 1D idealized context of Raynaud et al. (2008, 2009). Such a spatial filtering of flow-dependent variances is now used in the ensemble variational assimilation at Météo-France, which has been running operationally in real time since July 2008 (Berre et al. 2007, 2009).

Case studies also indicate that filtered variances are connected with the meteorological situation, with a tendency for the error variances to be higher in the vicinity of troughs and cyclones. This is consistent with previous studies in the literature, based on other variance estimation techniques such as 4D-Var (Thépaut et al. 1996), extended Kalman filtering (Bouttier 1993), or Hessian-based methods (Andersson and Fisher 1999).

On the other hand, such filtered variances are likely to remain imperfect to some extent, for instance due to neglect or approximations of model error contributions. Therefore, an important task is also to compare ensemble-based variances with innovation-based variances. Moreover, as discussed in the next section, the sampling noise issue is even more important in this context of comparison.

5. Spatial averaging of innovation-based variances

a. Innovation-based covariance estimates

The covariance matrix of the innovation vector $\delta y$ can be classically written as the sum of background and
observation error covariance matrices, under the assumption that the two errors are unbiased and uncorrelated (e.g., Daley 1991, p. 113):

$$E[\delta y \delta y^T] = HBH^T + R.$$  \hspace{1cm} (20)

where $HBH^T$ is the projection of the background error covariance matrix $B$ in observation space (through the observation operator $H$, assumed to be linear here for the sake of simplicity), and $R$ is the observation error covariance matrix.

As shown by Drozdov and Shepelevskii (1946), Gandin (1963), Rutherford (1972), Hollingsworth and Lonnberg (1986), and Lonnberg and Hollingsworth (1986), these two contributions can be separated, under the assumption that the observation error is spatially uncorrelated. In this case, the background error covariance can be estimated by extracting the spatially correlated component of innovation covariances.

In a more general idealized context, the exact extraction could be based on the multiplication between the exact optimal gain matrix $HK = HBH^T(HBH^T + R)^{-1}$ and the innovation covariances:

$$HK\{E[\delta y \delta y^T]\} = HBH^T(HBH^T + R)^{-1}(HBH^T + R) = HBH^T.$$  \hspace{1cm} (21)

However, in practice, the exact optimal gain matrix is not available, and the extraction has to be approximated. On the other hand, as discussed e.g., in Daley (1991, p. 128), $HK$ can be seen as a low-pass filter, in the case where the observation error is spatially uncorrelated. This indicates the innovation-based estimation of $HBH^T$ may be made by applying a well-chosen low-pass filter to innovation covariances.

To some extent, this is the approach used in Hollingsworth and Lonnberg (1986). In this case, background error covariances are estimated by adjusting a truncated Bessel spectral expansion to innovation covariance functions, with a least squares fit criterion. In other words, this may be seen as searching for a well-chosen function $F$, applied to innovation covariance functions, which mimics the effect of the exact low-pass filter $HK$. This can be written as follows, to express the fact that the estimation is based on the extraction of $HBH^T$, as the spatially correlated component of interest in the innovation covariances:

$$F\{E[\delta y \delta y^T]\} \approx HBH^T.$$  \hspace{1cm} (22)

This is a first possible approach to the estimation of background error covariances in observation space. A second possible approach is based on covariances of analysis residuals, as proposed by Desroziers et al. (2005). In this case, the background error covariance is estimated by the covariance between the analysis increment $H\delta x = H(x^a - x^b)$ and the innovation $\delta y$:

$$E[H\delta x \delta y^T] \approx HBH^T.$$  \hspace{1cm} (23)

Knowing that $H\delta x = HK\delta y$, where $K$ is the specified (usually suboptimal) gain matrix, it can be noted that this second approach is related to the first one through the following equation, by reformulating Eq. (23) as

$$HK\{E[\delta y \delta y^T]\} = HBH^T,$$  \hspace{1cm} (24)

which is the same as (22) with $F = HK$. In other words, the specified gain matrix is applied as a filter that extracts, from the innovation covariance, a spatially correlated component that corresponds to the background error covariance contribution.

Moreover, Eq. (24) can also be seen as an approximation of the exact extraction represented by Eq. (21), in which the exact (i.e., truly optimal) gain matrix $K$ is approximated by the specified gain matrix $K$. As discussed and shown in Desroziers et al. (2005), this is a reliable technique for the global variance estimation in particular, when there is a scale separation between observation and background errors.

It is also important to note that, in practice, the exact local innovation covariances $E[\delta y \delta y^T]$ are never available as such. The two kinds of innovation-based techniques are thus based on the approximation of exact expectations by space and time averages, corresponding to an ergodic approach (e.g., Courtier et al. 1998; Bannister 2008).

This is related to the fact that a given innovation value, at a specific location and date, corresponds to a single background error realization. In contrast, a variance estimate derived from a $N$-member ensemble is calculated from $N$ random realizations of background perturbations. In other words, the sample size issue is even more critical for innovation-based variance estimates than for ensemble-based variances.

One way to take this into account is to calculate global spatial averages of these background error covariance estimates. This is the approach that is usually adopted in order to calibrate additive model error (e.g., Houtekamer et al. 2009) or ensemble inflation factors (e.g., Li et al. 2009). In the next two sections, we will describe two application examples aiming to estimate space and time variations from innovation-based variances, based on local spatial averaging techniques.
b. Estimation of a climatological variance map

In Lindskog et al. (2006), innovations were used in order to estimate time-averaged spatial variations of background error variances for the HIRLAM data assimilation scheme. The aim of this work was to represent expected effects of data density contrasts and possible climatological features related to the Atlantic storm track for instance. For this purpose, a 1-yr time series of departures between surface pressure observations and corresponding background values was used for the HIRLAM domain, which is represented in Fig. 9.

The top panel of Fig. 9 is a map of raw innovation standard deviations from this study, averaged over the whole year period, and normalized in such a way that the mean field is 1.0 over the continental European area. Relatively small values can be observed over data-dense regions such as Europe and North America. In contrast, larger values are observed over oceanic regions, where storm tracks are located, and where fewer observations are available. On the other hand, some isolated peaks can also be identified near the North Pole for example. These questionable features are likely to be related to the specific time-varying irregular locations of ship data. This kind of spurious small scale features makes the raw map potentially difficult to use as such in a data assimilation scheme.

Therefore, Lindskog et al. (2006) applied a spatial filter to this raw map. The result of this approach is shown in the bottom panel of Fig. 9. This was achieved by only retaining the first five total wavenumber components of a spectral expansion of the raw field (on the limited-area grid) of innovation standard deviations. As expected, the filter extracts the relevant large-scale features related to data density contrasts and storm track areas while filtering out the small-scale doubtful structures, near the Pole for instance.

This filtered map was then implemented in the HIRLAM data assimilation scheme in order to represent some horizontal variations in the background error variances. Extensive data assimilation experiments indicated a positive impact from this kind of geographical information, compared to a HIRLAM version with horizontally homogeneous variances.

It may be mentioned that, in this example, the observation error contribution was not fully taken into account, when studying and implementing a normalized version of the innovation standard deviation map, as an estimate of the background error standard deviation field. The next section presents some results where this was accounted for when using innovation-based variance estimates from a specific date (instead of a time average).

c. Estimation of space- and time-dependent variances

An example of results for the estimation of space- and time-dependent innovation-based variances is presented in section 3.4 of Berre et al. (2007). In this case, Eq. (23) was used in order to estimate background error variances.

The top panel of Fig. 10 is a map of raw innovation-based variance estimates, for channel 7 of the High Resolution Infrared Radiation Sounder (HIRS) at 0000 UTC 28 August 2006. Each plotted value at location \(i_x\) is the product of the local increment value, namely, \((\mathbf{H}^\prime \mathbf{x})(i_x)\), with the corresponding local innovation value, namely, \((\mathbf{\delta y})(i_x)\). This local product \((\mathbf{H}^\prime \mathbf{x})(i_x) \times (\mathbf{\delta y})(i_x)\) is an estimate of the background error variance at location \(i_x\), according to Eq. (23) (including, in particular, an assumption that \(\mathbf{H}^\prime \mathbf{x}\) and \(\mathbf{\delta y}\) are unbiased).

As expected from the very small sample size (equal to 1, as this corresponds to a single error realization), the raw innovation-based variance map has little spatial structure: geographical variations are difficult to identify and interpret because this raw map is likely to reflect random sampling noise effects to a large extent.

Conversely, one may consider calculating spatial averages of these innovation-based variances in a way similar to that used for ensemble-based variances. This is illustrated in the middle panel of Fig. 10, where values have been averaged over a radius of 500 km for each observation point. In this case, some large-scale contrasts can be identified more clearly, such as relatively large values over the central Pacific compared to relatively small values over the southern Atlantic.

This filtered innovation-based variance map was also compared to an ensemble-based variance map. For this purpose, the background field of each ensemble member was projected to observation space through the observation operator \(\mathbf{H}\), and the same spatial average within a 500-km radius was applied to ensemble-based variances at each observation location. The resulting ensemble-based variance map is shown in the bottom panel of Fig. 10. A striking feature is that similar spatial structures are visible in the middle and bottom panels, such as the aforementioned contrast between the central Pacific and the southern Atlantic. Conversely, there are also regions where the two estimates differ, which could potentially be related to some deficiencies in the ensemble (such as approximations in the model error representation).

Therefore, this simple illustrative example suggests that one could consider applying spatial averaging techniques in order to systematically compare innovation-based and ensemble-based variance maps. On the other hand, it remains to be extensively studied how an objective
Fig. 9. Estimated background error standard deviation fields, averaged over a 1-yr period from surface pressure innovation data, on a $0.5^\circ \times 0.5^\circ$ grid of latitudes and longitudes. (top) Raw field. (bottom) Field smoothed by application of a horizontal filter. These fields include a normalization so that the mean of the field is 1.0 over the continental European area. The isoline interval is 0.1 and values $<$1.0 are in white. [©Blackwell Publishing 2006, from Lindskog et al. (2006). Reproduced by permission of Blackwell Publishing.]
FIG. 10. Estimated background error standard deviation fields, for channel 7 of HIRS at 0000 UTC 28 Aug 2006. (top) Raw innovation-based field, (middle) smoothed innovation-based field, and (bottom) smoothed ensemble-based field from a six-member ensemble. The smoothing in the (middle) and (bottom) was obtained by locally averaging raw variance values over a 500-km circular radius. Unit: K. [©ECMWF 2007, from Berre et al. (2007). Reproduced by permission of the ECMWF.]
spatial filter could be designed and applied to innovation-based variances for different observation types, in a way similar to that used for the ensemble-based variances.

6. Spatial averaging of correlation functions

As for ensemble-based variance fields, a similar issue of sampling noise exists for the estimation of background error correlations from an ensemble assimilation. As mentioned in the introduction, this is sometimes tackled by using a Schur filter approach, which gradually reduces correlation values toward zero as a function of separation distance. In this section, we will see that an alternative approach is to use local spatial averaging techniques in order to filter out this sampling noise. The study will focus on horizontal correlation functions, although an extension to vertical correlation functions can be considered easily.

First, we will consider the diagnosis of correlation length scales, as a classical characteristic of local correlation functions. This allows the spatial structures of sampling noise and signal to be examined, with respect to geographical variations in the correlation length scale field. Then we will review works on the use of wavelets, as a way to calculate local spatial averages of correlation functions.

a. Diagnosis of correlation length scales

Local correlation functions are more challenging to calculate and study than variances. This is due to the much larger number of values to be considered; for example, for a given parameter and vertical level: if \( N_x \) represents the number of horizontal grid points over the globe, there are \( N_x^2 \) horizontal correlation values to be calculated instead of only \( N_x \) variance values.

For this reason, it is usual and convenient to calculate diagnostics such as correlation length scales, as described by Hollingsworth and Lönneberg (1986), Daley (1991, p. 110), and Rabier et al. (1998). The 2D length scale \( L \) of a background error correlation function \( e \) is given by (e.g., Daley 1991, p. 110):

\[
L = \sqrt{-2/(\nabla^2 c)_{r=0}},
\]

where \( r \) is the geographical separation vector and \( \nabla^2 \) is the Laplacian operator. As discussed in Pannekoucke et al. (2008), \( L \) can also be seen as the distance at which the tangential parabola \( c_r \), at the origin [given by \( c_r(r) = 1 - |r|^2/(2L^2) \)] is equal to 0.5. This means that, the larger \( L \), the broader the correlation function.

Vertical variations of \( L \) have been studied by Hollingsworth and Lönneberg (1986), Rabier et al. (1998), and Berre (2000). Horizontal variations of \( L \) have been examined by Belo Pereira and Berre (2006) and Deckmyn and Berre (2005). As shown in the two latter papers, length scales can be calculated for each grid point, and then plotted in the form of a horizontal field. This is convenient in the study of how correlation functions vary geographically over the globe for instance.

b. Spatial structures of sampling noise and signal in length scale fields

Spatial structures of sampling noise and signal in correlation length scale fields have been studied in idealized 1D contexts by Pannekoucke et al. (2007, 2008). This is illustrated in Fig. 11 for a half domain in a 1D circular framework, where the dash–dotted line is a specified length scale field. Length scales are large in the left part of the domain, and small in the right part. The dashed line is a length scale field estimated from a 10-member ensemble of random realizations \( e^b = B^{1/2} \eta \), where \( B \) is the true background error covariance matrix and \( \eta \) is a Gaussian random field whose covariance is the identity matrix.

On the one hand, the large-scale signal of interest can be identified in the ensemble-based length scale field, with a tendency to have larger values in the left part than in the right part. On the other hand, the raw length scale field also contains spurious small-scale oscillations, which degrade the accuracy of the estimation.

This tendency of sampling noise to be much smaller scale than the signal of interest is reminiscent of what has been found for ensemble-based variance fields (see Fig. 1). These contrasted spatial structures for sampling noise and signal can be diagnosed in spectral space, too.
as shown in Fig. 8 of Pannekoucke et al. (2008). As expected from Fig. 11, the power spectra of the length scale fields show that sampling noise has much more small-scale energy than the signal of interest.

In the context of the Météo-France ensemble variational assimilation for NWP, Pannekoucke (2008) has studied how power spectra of ensemble-based length scale fields evolve when the ensemble size increases. This is shown in Fig. 12, which corresponds to correlations estimated using a time-expanded sampling approach [in a similar way as Xu et al. (2008) in an EnKF context] with the Météo-France ensemble variational system: correlations are calculated from a sample of $N_{\text{tot}} = N \times N_t$ perturbed forecasts, where $N$ is the ensemble size (varying between 1, 3, 6, and 12), and $N_t = 5$ corresponds to the number of 6-h forecasts produced by 5 consecutive perturbed analysis networks (namely between at 0000 UTC 18 January 2005 and 0000 UTC 19 January 2005). This means that the total sample size $N_{\text{tot}}$ varies between 5, 15, 30, and 60 in Fig. 12.

As expected, these spectral powers tend to decrease when the ensemble size increases because of the decrease of sampling noise amplitude. Moreover, this decrease is much more pronounced in the small-scale part of the spectra than in the large-scale part. This is consistent with the idea that sampling noise tends to be smaller scale than the signal of interest in the correlation length scale field. This is also supported by the fact that this behavior as a function of sample size is similar to what has been found for ensemble-based variance fields, as shown by Fig. 10a in Raynaud et al. (2009) and also by Fig. 3 in section 3c.

Results of this kind support the idea of applying space averaging techniques for the estimation of correlation functions, in a way similar to that used for ensemble-based variances.

c. Spatial averaging of correlations through a wavelet diagonal approach

The use of a wavelet diagonal approach to represent spatially varying correlation functions has been proposed by Fisher (2003). Wavelets are functions that are localized in both gridpoint and spectral spaces, as illustrated in Figs. 9 and 10 of Fisher (2003). This means, in particular, that the variance of a given wavelet component of the background error contains information about both spatial scale and geographical position. For instance, if the correlation function is relatively broad (sharp) in a given region, variances of large-scale (small-scale) wavelet components will tend to be large in this area. It is this fundamental property that allows spatially varying correlation functions to be represented with a wavelet diagonal approach.

In practice, the wavelet correlation model is calibrated by a projection of the normalized gridpoint background error field $e^b = \frac{e^b}{\sigma_b}$ to wavelet space (where $\sigma_b$ is the background error standard deviation in gridpoint space), followed by the calculation of associated variances (in the case of a 1D or 2D context) or vertical covariances (in a 3D context) for each wavelet component. The associated background error correlation model $C_W$ thus corresponds to

$$C_W = \Sigma_b^{-1} W^{-1} V_W W^{-T} \Sigma_b^{-1},$$

where $\Sigma_b$ is the diagonal matrix containing background error standard deviations $\sigma_b$ in gridpoint space and $W^{-1}$ is the inverse wavelet transform. In a 1D or 2D context, $V_W$ is the diagonal matrix containing variances of each wavelet component of the field $We^b$. In a 3D context, the matrix $V_W$ is block diagonal, with each block containing covariances of $We^b$ between different vertical levels for a given wavelet component.

This approach is now operational at ECMWF to represent geographical variations of time-averaged correlations derived from ensemble assimilation experiments. The ability of such wavelet techniques to represent horizontal heterogeneities related to land–sea contrasts and orographic features has also been illustrated by Deckmyn and Berre (2005).

Moreover, an extension of this approach to represent flow-dependent correlations can be considered. As shown
by Pannekoucke et al. (2007), such a wavelet diagonal approach is interesting for another reason in the latter context. This corresponds to its filtering properties for the estimation and representation of background error correlations.

This is related to the fact that, implicitly, using a wavelet diagonal approach amounts to calculating local spatial averages of correlation functions. This is expressed by the following equation for the implied wavelet-based correlation function $c_W^x(r)$ at grid point $x$, which can be written as a weighted average of raw correlation functions $c^x$ at neighboring positions $x'$:

$$ c_W^x(r) = \sum_{x',r'} c^x(r') \Phi^{x'}(x', r'), $$

(26)

where $r$ and $r'$ are horizontal separation values, and $\Phi^{x'}(x', r')$ is a weighting function; as illustrated in Fig. 2 of Pannekoucke et al. (2007), $\Phi^{x'}(x', r')$ gives more weight to contributions of grid points $x'$ and separation values $r'$, which are close to $x$ and $r$, respectively.

The impact of this wavelet approach is shown in Fig. 11 for the aforementioned 1D idealized context, with a 10-member ensemble. Spurious small-scale oscillations in the raw length scale field (dashed line) tend to be much attenuated, through the wavelet local averaging of correlation functions (full line).

This effect of sampling noise filtering can also be seen when plotting implied correlation functions. This is shown in Fig. 13, which corresponds to correlation functions at a given position $x$. The full line in the top panel is the raw ensemble-based correlation function, while the full line in the bottom panel is the wavelet-filtered ensemble-based correlation function. The true correlation function corresponds to the dotted line in both panels. A Schur filter with a scale $L_s = 6000$ km (i.e., $54^\circ$) was applied for both ensemble-based correlation functions. It appears again that spurious small-scale oscillations in the raw correlation function tend to be largely attenuated through the wavelet approach.

Data assimilation experiments have also been conducted in this 1D idealized context and the impact of the wavelet approach is shown in Fig. 9 of Pannekoucke et al. (2007). It indicates that the wavelet approach has a positive impact on the analysis quality, and that there is less need of Schur filtering. Moreover, the analysis quality appears to be less sensitive to a misspecification of the Schur filter when a wavelet approach is used.

Diagnostic studies of wavelet filtering effects have also been conducted in ensemble variational assimilation for NWP. Figure 14 (from Pannekoucke et al. 2007) is an example of correlation length scale fields for a specific date from a global six-member ensemble assimilation system.

As expected from the small sample size, the raw length scale field (top panel) looks relatively noisy, although some structures may be identified, such as a tendency for length scales to be smaller over South America and Europe for instance, and larger in the circumpolar Antarctic ocean. This is reminiscent of what was found in the 1D idealized framework of Fig. 11. These spatial structures appear much more clearly in the wavelet-implied length scale field (bottom panel). This is consistent with the expected filtering properties of the wavelet approach.

These results tend to be supported by Fig. 15 (from Pannekoucke 2008), which represents the robustness of length scale fields, as a function of sample size, either in a raw way from the ensemble (full line) or with the wavelet approach (dashed line). In this figure, the robustness is measured by the correlation between length scale fields estimated from two independent ensembles.
This is analogous to the methodology used in Berre et al. [2007; see their Eq. (2)], and presented in section 3e, to estimate the relative amount of noise-free signal in ensemble-based variance fields. If sampling noise is small compared to the noise-free signal contribution, this correlation is expected to be close to 1. Conversely, if sampling noise is dominant, this correlation is expected to be close to 0.

This correlation is calculated for different total sample sizes $N_{tot} = N \times N_t$, which vary between 6, 18, and 30, by using $N = 6$ members and a number of successive networks $N_t$, varying between 1, 3, and 5.

As expected from the previous results, it appears that the robustness of the wavelet-based length scale fields is systematically much larger than for the raw length scale fields. For instance, for $N_{tot} = 18$, the correlation is around 80% for the wavelet-based length scale field, instead of 40% for the raw length scale field.

Moreover, case studies presented in Pannekoucke (2008) indicate that wavelet-based length scale fields are
closely connected to the weather situation, with a tendency to have smaller length scales in the vicinity of lows and troughs. This is consistent with results described in Thépault et al. (1996).

This suggests that a wavelet diagonal approach may be a promising way to extract flow-dependent correlation information from an ensemble assimilation. This tends to be supported by results presented in Fig. 16 (from Lindskog et al. 2007). The top panel of this figure illustrates the weather situation in terms of 500-hPa geopotential height. The bottom panel shows associated examples of wavelet-based flow-dependent correlation functions, estimated from a sample of 12 forecast differences. In particular, correlation functions in the middle of the domain appear to be elongated in the meridional direction, in accordance with the shape of the geopotential field in this region.

7. Connections with related approaches in different contexts

While sampling noise studies and spatial filtering techniques have received increasing attention in the context of ensemble variational assimilation for NWP, it is interesting to note that some related approaches have been used in other contexts such as EnKF and probabilistic forecasting.

a. Methodologies for sampling noise studies

As discussed for Fig. 4 in section 3c, and for method 1 in section 3d, one way to estimate sampling noise features in ensemble-based variance fields is to compare variance fields estimated from two independent ensembles. This may be seen as a Monte Carlo estimation technique, in which the spread between variance estimates from two independent ensembles is used to estimate the sampling noise power variance. It may be observed that this is analogous to the vision of ensemble assimilation as a Monte Carlo estimation technique, in which the spread between state estimates from several independently perturbed members is used to estimate the state error variance.

Moreover, such an approach has been used in other contexts. In Fig. 6 of Houtekamer and Mitchell (1998),
correlation fields estimated by two different ensembles are compared in order to obtain a qualitative indication of the accuracy of these correlation estimates. It is noteworthy that short-distance correlations tend to be more consistently estimated by the two ensembles than long-distance correlations. This is an indication that sampling noise effects are relatively larger for long-distance correlations than for short-distance correlations. This is confirmed by a comparison with correlations estimated from ensembles whose size is much larger. A similar methodology has also been used by Buehner and Charron (2007) to estimate sampling noise effects in an EnKF context.

Another analogous approach has been proposed by Anderson (2007) in the context of an ensemble filter technique. In Anderson (2007), “groups” of ensembles are used to evaluate sampling errors in ensemble-based estimates. As suggested in the first paragraph of this section, the use of such groups of ensembles is described in Anderson (2007) as a Monte Carlo technique to estimate sampling errors. This information about sampling errors is then used to estimate optimized localization functions, so that sampling noise effects are minimized.

b. Local time averaging of covariances

While the present review focuses mostly on spatial averaging techniques, an analogous approach is to consider temporal averaging techniques. This was briefly mentioned in section 6c when studying statistical features of correlation length scale fields. This temporal approach has been used in the context of ensemble variational assimilation in oceanography (Daget et al. 2009). In this case, background error variances were calculated by using a sliding window with a total sample of 81 forecasts made of 9 perturbed forecast fields (from a 9-member ensemble) coming from 9 successive analysis networks. In particular, it was observed that this temporal averaging allowed the robustness of variance estimates to be strengthened.

A similar technique has been used in the context of a mesoscale ensemble square root filter (Xu et al. 2008), to compute local time averages of background error covariances. Experiments were carried out by calculating a temporal average of 10-member ensemble covariance estimates over 3 time levels in the vicinity of the analysis time. It was shown that this time-expanded sampling approach could lead to better analyses and forecasts than the use of raw 10- or 30-member covariance estimates (without any temporal averaging). Moreover, it was also demonstrated that the improvement of the filter performance could be optimized by properly selecting the sampling time interval. This method of sample size increase was found to be computationally attractive, as it does not require the computation of additional perturbed forecasts.

These techniques can be seen as the temporal counterpart of direct spatial averaging techniques described in section 4a. Moreover, this analogy suggests that an objective calibration of these time averaging techniques may be considered as well, although this is not explored in the above references.

c. Schur filtering of correlations

In section 6, the use of a wavelet diagonal approach was presented as a technique to calculate spatial averages of correlation functions.

In Buehner and Charron (2007), it is shown that spatial averaging of correlation functions corresponding to nearby locations can be achieved equivalently through a Schur localization in spectral space. This Schur localization corresponds to a multiplication, in spectral space, of the raw ensemble-based correlation matrix by an analytical correlation matrix, which has a compact support in spectral space.

Buehner and Charron (2007) also mention the idea that spatial averaging of correlations may be achieved by including spatially shifted copies of the ensemble perturbations. This approach for filtering correlations is relatively similar to the direct spatial averaging techniques mentioned in section 4a for the estimation of variances, with the additional use of similar weighting functions as described in section 4c.

One potentially interesting aspect of these two techniques (Schur localization in spectral space and use of spatially shifted perturbations) is the possibility of using weighting functions that are smoother than what is shown in Fig. 2 of Pannekoucke et al. (2007) for their wavelet diagonal approach.

As mentioned in the introduction, another possible approach for filtering correlations is to apply a Schur localization in gridpoint space (Houtekamer and Mitchell 2001). As discussed in Hamill et al. (2001), this approach can be justified by the fact that correlation estimates generally have a higher ratio of noise to signal with increasing distance between grid points. One potential difficulty is that localization functions are usually isotropic and state independent. Recent works by Bishop and Hodyss (2007, 2009a,b) aim at defining an adaptive localization approach. The idea is to use the ensemble information in order to apply flow-dependent localization functions. This is achieved by raising smoothed versions of ensemble correlations to a power.

On the one hand, these Schur localization techniques in gridpoint space are different from the spatial averaging approaches described in the present review. On the other hand, it would be interesting to establish connections
between possible optimization techniques for these two different approaches. For instance, a signal-to-noise calculation is used as a diagnostic of sampling noise in correlations in Hamill et al. (2001), while it is also used as a way to objectively optimize spatial filtering of variances in Berre et al. (2007) and Raynaud et al. (2009). Therefore, one could explore further the possibility of using signal-to-noise information in order to optimize the Schur localization techniques in gridpoint space, possibly in connection with approaches also proposed by Anderson (2007) and Bishop and Hodyss (2007, 2009a,b).

Moreover, as proposed by Pannekoucke et al. (2007) and Buehner and Charron (2007), it seems possible to use an optimal combination of gridpoint Schur filtering and spatial averaging techniques in order to minimize sampling errors of correlation estimates. Note also that the gridpoint Schur filter does not treat the problem of sampling noise in the variance field. Spatial (and possibly temporal) averaging techniques can be considered for this aspect.

d. Combining spatial statistical and ensemble information in probabilistic forecasting

In another context, associated with probabilistic weather forecasting, Berrocal et al. (2007) have proposed combining spatial statistical and ensemble information in Bayesian model averaging (BMA). In this *spatial BMA technique*, the spatial structures of weather fields and of forecast error fields are taken into account in order to optimize the estimation of local probability distribution functions over a domain of interest.

This is reminiscent of the idea of taking the spatial structure of signal and noise into account in order to optimize the estimation of local background error variances, for instance. The spatial BMA ensemble is shown to outperform both raw and “standard local” BMA ensembles (which do not take spatial structures into account).

In this context of probabilistic forecasting, it is noted again by Berrocal et al. (2007) that statistical samples of any size can be generated at small computational cost. This is in line with the idea that spatial sampling and averaging is a cost-effective way to increase the sample size and to enhance the accuracy of ensemble estimates.

8. Conclusions and perspectives

Recent works on spatial filtering of background error covariances have been reviewed in this paper, in particular in the context of variational data assimilation for global NWP. Several theoretical and experimental results indicate that sampling noise, which affects variance fields and correlation length scale fields, tends to be smaller scale than the signal of interest in this context. This suggests that local space averaging techniques may be considered in the aim of improving the accuracy of the ensemble-based covariance estimation.

Results from simple direct spatial averages were illustrated first. The resulting increased estimation accuracy can be interpreted as arising from an increased sample size. Then it was shown that the spatial filtering of ensemble-based variance fields could be formulated as a linear estimation problem, which is based on the minimization of the estimation error variance. This allows the optimized estimation of variance fields to be designed in the form of a spectral linear regression between the raw ensemble-based variance field on the one hand and the signal of interest on the other hand. The associated regression coefficients can then be calculated from either analytical or experimental estimations of spectral variances of sampling noise and of the noise-free signal, respectively. Experimental results in both 1D idealized and high-dimensional contexts indicate that, typically, the regression tends to act as a low-pass filter; it extracts the robust large-scale signal of interest while filtering out sampling noise, which predominates at the smallest scales. It may also be mentioned that this technique is used in the ensemble variational assimilation system at Météo-France, which has been running operationally in real time since July 2008, to provide flow-dependent background error variances to 4D-Var.

While this kind of approach has been explored for ensemble-based variance fields, it may be considered for innovation-based variance fields, too. This has been illustrated here for the estimation of time-averaged spatial variations of background error variances and also for the estimation of variance fields for a specific date. This is likely to be quite important for the comparison and validation of ensemble-based variance fields with innovation-based variance fields. In particular, such a comparison is believed to be essential for the study and representation of model error covariances in ensemble assimilation systems.

A similar issue can be raised for the estimation of ensemble-based correlation functions. Experimental studies in both idealized and realistic frameworks indicate again that sampling noise tends to be smaller scale than the signal of interest in correlation length scale fields. This supports the idea of applying spatial averaging techniques for the estimation of correlation functions. This can be done by using a wavelet diagonal approach, which amounts to calculating weighted spatial averages of correlation functions. Experimental results indicate that this technique allows sampling noise effects to be attenuated to a large extent, while representing the flow-dependent signal of interest.
While these spatial averaging techniques have been experimented mostly in the context of variational data assimilation for NWP, it has also been observed that there are links with related approaches in different contexts such as EnKF and probabilistic forecasting. This corresponds to connections with methodologies for sampling noise studies (with respect to correlations in EnKF), temporal averaging approaches, Schur filtering of correlations, and spatial BMA techniques in probabilistic forecasting. Although not mentioned in the present review, there may also be relations with ergodic space and/or time averages in turbulence (e.g., Boer 1983; Monin and Yaglom 1971) and also with spatial interpolation of analysis weights in the context of the local ensemble transform Kalman filter (Yang et al. 2009).

In connection with the temporal averaging approach, while we have presented 2D horizontal filtering techniques, the 4D nature of the atmosphere and of the covariance estimation may suggest that the use of 4D filtering techniques should be considered in order to further optimize the covariance estimation.

Finally, it is worth noting that there is a strong analogy between the covariance calculation posed as a linear estimation problem, and the determination of the atmospheric state with similar linear estimation techniques (which are at the heart of data assimilation systems). This suggests that the use of data assimilation techniques may be considered to combine ensemble- and innovation-based covariances in an optimized way.

Acknowledgments. We thank the three anonymous reviewers for their constructive suggestions on the manuscript. Special thanks to Susan Becker for improving the English usage.

APPENDIX

Interpretation of the Spectral Linear Regression in Terms of Amplitude and Phase Information

Because of the spectral shape of $\rho$ (see Figs. 5 and 6), the regression presented in section 4d corresponds to a multiplication of the raw spectral field by a low-pass filter. In accordance with the discussion of spatial structures in section 3, this reflects the idea of extracting the large-scale signal of interest, while filtering out small-scale sampling noise. Knowing that spectral coefficients such as $\mathbf{s}$ contain both amplitude and phase information, one can note that there are two reasons for filtering the raw variance field.

A first reason is that the amplitudes of spectral coefficients $\mathbf{s}$ are too large compared to $\mathbf{s}^*$, as shown by $p(\mathbf{s}) = p(\mathbf{s}^*) + p(s) > p(\mathbf{s}^*)$ [Eq. (10)]. The level of amplitude accuracy (of $\mathbf{s}$ with respect to $\mathbf{s}^*$) can thus be measured by the ratio $\sqrt{\frac{p(\mathbf{s}^*)}{p(\mathbf{s})}}$. If sampling noise is negligible, $\sqrt{\frac{p(\mathbf{s}^*)}{p(\mathbf{s})}}$ will be close to 1. Conversely, if sampling noise is predominant, $\sqrt{\frac{p(\mathbf{s}^*)}{p(\mathbf{s})}}$ will be close to 0.

A second reason is that spatial structures in the raw variance field are partly out of phase compared to spatial structures in the signal, again because of sampling noise. The amount of phase accuracy can be measured by the spectral correlations between $\mathbf{s}$ and $\mathbf{s}^*$, namely $\text{cor}(\mathbf{s}, \mathbf{s}^*) = \text{cor}(\mathbf{s}, \mathbf{s}^*) / \sqrt{\text{cor}(\mathbf{s}, \mathbf{s}) \text{cor}(\mathbf{s}, \mathbf{s}^*)}$. Typically, if sampling noise is small, $\mathbf{s} \approx \mathbf{s}^*$. This means that spectral coefficients in $\mathbf{s}$ will not only have the exact amplitude $\mathbf{p}(\mathbf{s}) \approx \mathbf{p}(\mathbf{s}^*)$ but will also have the exact phase, and thus $\text{cor}(\mathbf{s}, \mathbf{s}^*) \approx 1$. Conversely, if sampling noise makes a very large contribution, spectral coefficients $\mathbf{s}$ will have a random phase uncorrelated with the signal, and thus $\text{cor}(\mathbf{s}, \mathbf{s}^*) \approx 0$.

Moreover, it can be shown easily that the linear regression coefficient $\rho$ is, in fact, the product between these two accuracy quantities, namely the amplitude accuracy $\sqrt{\text{cor}(\mathbf{s}, \mathbf{s}^*)}$ and the phase accuracy $\text{cor}(\mathbf{s}, \mathbf{s}^*)$. This can be shown by expanding Eq. (18) as follows:

$$\rho = \frac{\text{cor}(\mathbf{s}, \mathbf{s}^*)}{\mathbf{p}(\mathbf{s})} = \frac{\text{cor}(\mathbf{s}, \mathbf{s}^*) / \sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)}}{\sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)}} = \text{cor}(\mathbf{s}, \mathbf{s}^*) / \sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)}. \quad (A1)$$

This expression allows the linear regression to be interpreted as a two-step procedure, which accounts for both amplitude and phase errors in the raw variance field. The first step is a scaling that accounts for the amplitude accuracy, measured by $\sqrt{\text{cor}(\mathbf{s}, \mathbf{s}^*) / \mathbf{p}(\mathbf{s})}$. The next step is a second scaling that accounts for the phase accuracy, measured by $\text{cor}(\mathbf{s}, \mathbf{s}^*) / \sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)}$. Finally, it can easily be shown that these two scaling factors are, in fact, equal: $\text{cor}(\mathbf{s}, \mathbf{s}^*) = \text{cor}(\mathbf{s}, \mathbf{s}^*) / \sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)} = \mathbf{p}(\mathbf{s}) / \sqrt{\mathbf{p}(\mathbf{s}) \mathbf{p}(\mathbf{s}^*)}$. This can be seen as an indication that amplitude and phase errors are of equal importance in the raw variance field.

REFERENCES


