Ensemble-based background-error covariances in variational data assimilation

Mark Buehner

Environment Canada, 2121 Trans-Canada Hwy, Dorval, QC, H9P 1J3, CANADA

1 Introduction

Since early 2005, Environment Canada has been operationally running both a global four-dimensional variational analysis system (4D-Var) and an ensemble Kalman filter (EnKF). The 4D-Var is used to provide global "deterministic" analyses for producing the medium-range global forecasts and also for initializing the high-resolution regional system for forecasts up the 48h (Gauthier et al. 2007). The EnKF is used to obtain the ensemble of initial states for the ensemble prediction system (EPS) from which probabilistic forecast products are produced (Houtekamer *et al.* 2005). Both systems employ the same Global Environmental Multi-scale (GEM) model (Coté *et al.* 1998), though with different spatial resolution. In addition, both systems assimilate nearly the same set of observations. In fact, the quality control decisions and observation bias corrections computed by the 4D-Var system are used for the EnKF.

The co-existence of these two operational systems at a numerical weather prediction (NWP) centre provides a unique opportunity. This paper first discusses approaches for incorporating background-error covariances estimated from the EnKF ensembles in the variational assimilation system (Section 2). The techniques of spatial and spectral localization are briefly described and demonstrated with a simple one-dimensional problem. Then, the impact of localizing ensemble-based covariances in a 3D-Var system are shown in Section 3. Finally, the conclusions are given.

2 Localized ensemble-based correlations

By necessity, the size of the ensemble of background states used to estimate background-error covariances is always much smaller than the dimension of the system. Consequently, no covariance information can be directly estimated along the vast majority of possible directions in phase space. The background state will not be corrected by the analysis system along all such directions unless additional hypotheses concerning the structure of the background-error covariances are adopted. For example, covariances (or correlations) are often assumed to be horizontally homogeneous and isotropic. The result is a matrix that is diagonal in spectral space (Courtier *et al.* 1998). Alternatively, one may assume the amplitude of spatial correlations gradually approaches zero at a specified separation distance (e.g. Gaspari and Cohn 1999; Houtekamer and Mitchell 2001), or that the correlations are diagonal when transformed into a space defined by an expansion of wavelet functions (Fisher and Andersson 2001; Pannekoucke *et al.* 2007). The use of any of these hypotheses results in a significant increase in the rank of the correlation matrix that can be estimated from an ensemble of a given size. However, increased rank is alone not sufficient to guarantee an improved covariance estimate. The ability to obtain a better estimate of the true covariances depends on how consistent the hypothesis is with the true covariances and also on the pre-existing level of sampling error.

For the sake of clarity, in this section only the estimation of univariate spatial background-error correlations along a single spatial dimension is considered. The extension to three-dimensional multivariate error covariances is relatively straightforward (Buehner and Charron, 2007).

2.1 Spatial localization

Spatial localization relies on the hypothesis that the amplitude of the correlations generally decreases with increasing separation distance and may even reach zero at a finite distance. To reduce estimation error, the Schur product (element-wise multiplication) between the estimated correlation matrix and a matrix with a prescribed localization weights that monotonically decrease as a function of separation distance is computed. The result is a valid (positive semi-definite) localized version of the original correlations as long as the matrix of localization weights is positive semi-definite (Gaspari and Cohn 1999). The Schur product is expressed as

$$C_{gl}(x_1, x_2) = C(x_1, x_2) L_g(x_1, x_2) \tag{1}$$

where C_{gl} denotes the spatially localized version of the estimated correlations and L_g denotes the localization weights in the spatial domain.

For application in variational assimilation the same localization can be incorporated in the square-root of the background-error covariance matrix used to transform the control vector α into the analysis increment $\Delta\mathbf{x}$ (Lorenc 2003; Buehner 2005):

$$\Delta\mathbf{x} = \mathbf{B}^{1/2}\alpha, \tag{2}$$

where bolded upper-case (lower-case) symbols represent matrices (vectors). Briefly, a separate portion of the control vector exists for each ensemble member and is allowed to vary as a function of location (denoted by α_n for the nth ensemble member). The fields are constrained to have the correlation structure prescribed by the matrix of localization weights through multiplication by its square-root $\mathbf{L}_g^{1/2}$. Then the analysis increment is constructed by computing the Schur product (denoted by $\circ$) between the spatially correlated portions of the control vector with the corresponding ensemble members (where the ensemble mean is assumed to be zero) and the contributions from all ensemble members are summed:

$$\Delta\mathbf{x} = \frac{1}{\sqrt{N_{ens}}} \sum_{n=1}^{N_{ens}} \mathbf{e}_n \circ \left(\mathbf{L}_g^{1/2} \alpha_n \right). \tag{3}$$

When no localization is applied (corresponding to a matrix $\mathbf{L}_g$ containing all ones) the control vector has no spatial dependence and therefore the ensemble contributes by the same amount at all grid-points. In that case only a single scalar element of the control vector is required for each ensemble member to determine the contribution of that member to the analysis increment.

2.2 Spectral localization

Spectral localization relies on the hypothesis that the amplitude of the correlations in spectral space generally decreases as the absolute difference in wavenumber between pairs of spectral components increases. The validity of this hypothesis is likely to be highly case dependent. However, it is certainly more easily satisfied than the hypothesis currently employed at many operational NWP centres of assuming the correlations are diagonal in spectral space. Spectrally localized correlations are denoted as

$$\hat{C}_{sl}(k_1, k_2) = \hat{C}(k_1, k_2) \hat{L}_s^*(k_1, k_2), \tag{4}$$

where the $\hat{}$ denotes that the quantity is in spectral space, $\hat{C}_{sl}(k_1, k_2)$ represents the spectrally localized correlation between wavenumbers k_1 and k_2, $\hat{L}_s$ represents the function used for localizing the estimated correlations and the superscript asterisk denotes the complex conjugate.

Buehner and Charron (2007) demonstrated a straightforward interpretation of spectral localization. They showed that, under certain conditions, spectral localization is equivalent to spatial averaging of the correlations in grid-point space. The more extreme the spectral localization, the larger the region over which the spatial averaging is performed on the correlations. The limiting case of forcing the correlations to be diagonal

in spectral space consequently corresponds with averaging the correlations over the entire domain, resulting in homogeneous correlations.

Implementing spectrally localized ensemble correlations in a variational assimilation scheme is similar to the approach described earlier for spatial localization. As with spatial localization the control vector is divided into separate portions, one for each ensemble member, and is made to vary as a function of wavenumber. Then these vectors are each multiplied by the square-root of the localization matrix $\hat{\mathbf{L}}_s$. Next, element-wise multiplication is performed between these portions of the control vector and the spectrally transformed ensemble members $\hat{\mathbf{e}}_n$ (again with zero ensemble mean). After summing the contributions from each ensemble member, the result is transformed back into grid-point space. Thus, the transformation from control vector to analysis increment is given in matrix-vector form as

$$\Delta\mathbf{x} = \frac{1}{\sqrt{N_{ens}}}\mathbf{S}^{-1}\sum_{n=1}^{N_{ens}}\hat{\mathbf{e}}_n \circ \left(\hat{\mathbf{L}}_s^{1/2}\hat{\alpha}_n\right), \tag{5}$$

where $\mathbf{S}^{-1}$ is the inverse spectral transform and $\hat{\alpha}_n$ is the portion of the spectral-space control vector corresponding to the nth ensemble member.

2.3 Impact of localization on estimated 1-D correlation functions

The impact of spectral and spatial localization is demonstrated in this section by estimating one-dimensional spatial correlations in an idealized experimental setting. The ensemble of random error realizations is generated using a prescribed 'true' correlation matrix. The true correlations are specified using the Gaussian function with a smoothly varying length scale, l_{corr}. To localize correlations in both spectral and grid-point space, the fifth-order piecewise rational function given as equation (4.10) in Gaspari and Cohn (1999) was used. The localization 'radius' is defined as the 'distance' at which the function becomes zero.

Figures 1(a) shows the correlations with respect to a grid-point near the centre of the domain. The true correlations are shown (dotted) along with the correlations estimated from 30 random realizations (solid). The estimated correlations are clearly affected by sampling error at all locations. The remaining panels show the impact of spatial and spectral localization when applied separately and together with the optimal localization radii. Spatial localization clearly reduces sampling error by suppressing the spurious remote correlations where the true correlations are close to zero. However, the original correlations are little changed near the peak. As mentioned in the previous section, spectral localization is equivalent to a smoothing in grid-point space of the correlation functions corresponding to nearby locations. This is demonstrated in Figures 1(c) where spectral localization leads to a general reduction in sampling error at all locations, even though the overall magnitude of the reduction appears to be less than from the application of spatial localization. Finally, the combined effect of applying both types of localization (Figure 1(d)) appears to produce the estimate with the lowest sampling error. Comparing these results with those from Pannekoucke *et al.* (2007) shows that the application of both spectral and spatial localization produces qualitatively similar results as when correlations are constrained to be diagonal in a space defined by an expansion of wavelet functions.

Figure 2 compares the computed correlation length scale as a function of location for the true and estimated correlation matrices. In each case the length scale was computed as

$$l_{corr} = \left(-\frac{d^2C}{dx^2}\right)^{-1/2}. \tag{6}$$

The length scale from the true correlation matrix varies smoothly in space. The correlation length scale computed from the correlation matrix estimated without localization (Figure 2(a)) is relatively noisy, but generally follows the main variations of the length scale computed from the true correlation matrix. The application of spatial localization (Figure 2(b)) only slightly reduces the larger values. In contrast, applying spectral localization removes most of the rapid spatial variations in the length scale (Figure 2(c)). For comparison the length scales from using more extreme spectral localization (dashed) is also shown.

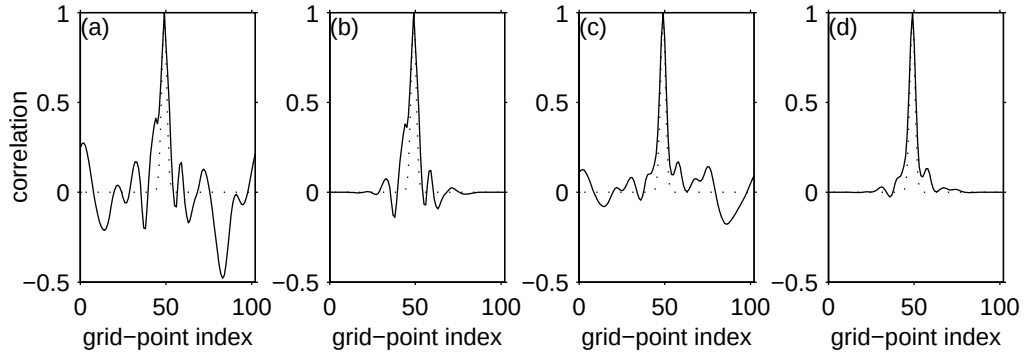

Figure 1: True (dotted) and estimated (solid) correlation function for a location near the centre of the domain. (a) Original sample estimate, (b) sample estimate with optimal amount of spatial localization, (c) sample estimate with optimal amount of spectral localization, (d) sample estimate with optimal amount of spatial and spectral localization.

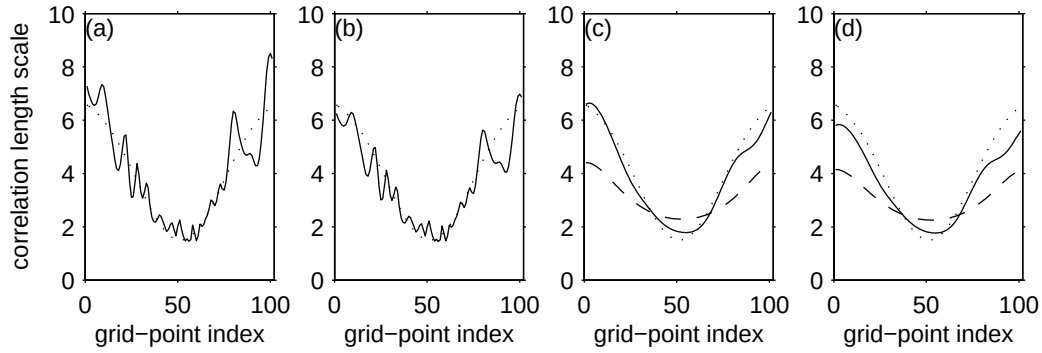

Figure 2: Correlations length scales computed from the true (dotted) and estimated (solid) correlation matrices. The four panels correspond to the same correlation estimates as in the previous figure. For comparison, the correlation length scales are also computed when more extreme spectral localization is applied (dashed).

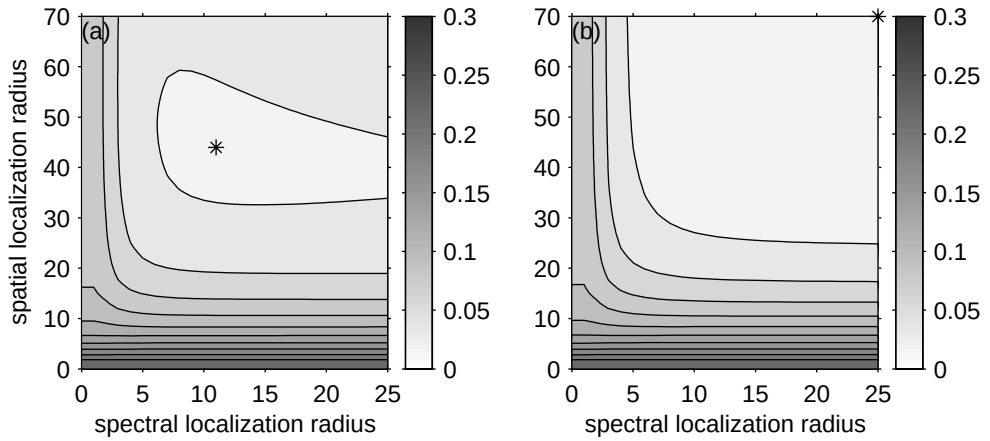

Figure 3: (a) Root-mean-square difference between elements in localized version of the estimated correlation matrix and the true correlation matrix as a function of both spatial and spectral localization radii. (b) is similar to (a), but shows the difference between the true correlation matrix with and without localization. The asterisk denotes the location of the minimum and the contour interval is 0.02.

To evaluate the ability of both types of localization to reduce sampling error, the rms difference between the elements of the estimated and true correlation matrices is computed as a function of the two localization radii and is shown in Figure 3(a). These results confirm what was already seen from the previous figures, that the combined use of spectral and spatial localization results in the best estimate of the correlations (the asterisk indicates the location of the minimum). Figure 3(b) shows the impact of applying localization to the true correlation matrix. In this case no local minimum exists and the rms difference between the true correlation matrix with and without localization monotonically decreases as the localization radii increase.

3 Localization in a realistic NWP context

3.1 Spectral localization in 3D-Var

In this section, results are presented from applying spectral localization in a realistic NWP context to specify the background-error correlations in 3D-var. The impact on the spatial structure of the background-error covariances is demonstrated by performing experiments in which only a single observation is assimilated. The use of spectrally localized background-error correlations are compared with the homogeneous and isotropic correlations used operationally, as described by Gauthier *et al.* (1998). The ensemble of background-error samples was obtained using a Monte Carlo approach applied to the 3D-Var system as described by Buehner *et al.* (2005).

The same fifth-order piecewise rational function used in the previous section is also used to specify the homogeneous spectral localization function. The constraint of homogeneity for the localization function is not necessary for implementation in the variational assimilation system, but is used in these initial experiments for the sake of simplicity. In addition, spectral localization is applied with respect to only the total horizontal wavenumber. No spatial localization is applied.

Figure 4 shows the zonal cross-section of the meridional wind analysis increment from assimilating a single zonal wind observation located over the southern Pacific ocean at 60° S, 180° E and 300 hPa. The result has been normalized so that the increment at the location of the observation is equal to one. The result is therefore similar to the zonal wind background-error correlations. Note how the use of horizontally homogeneous and isotropic correlations (Figure 4(a)) produces a spatially smooth covariance structure. Conversely, when no

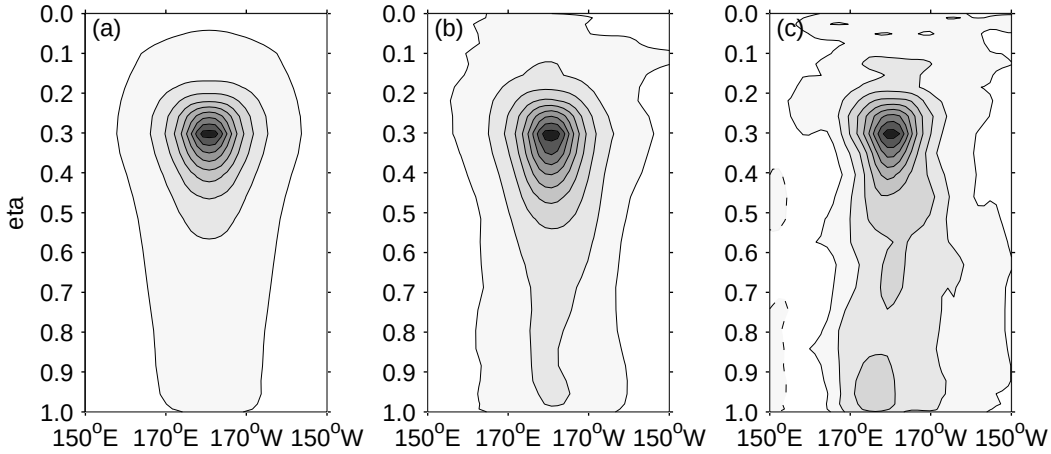

Figure 4: Zonal cross-sections of the zonal wind covariances with respect to 60° S, 180° E and 300 hPa with (a) homogeneous and isotropic correlations, (b) correlations with spectral localization and (c) correlations with no localization. The covariances are normalized to have a maximum value of one, the contour interval is 0.1 and dashed contours denote negative values.

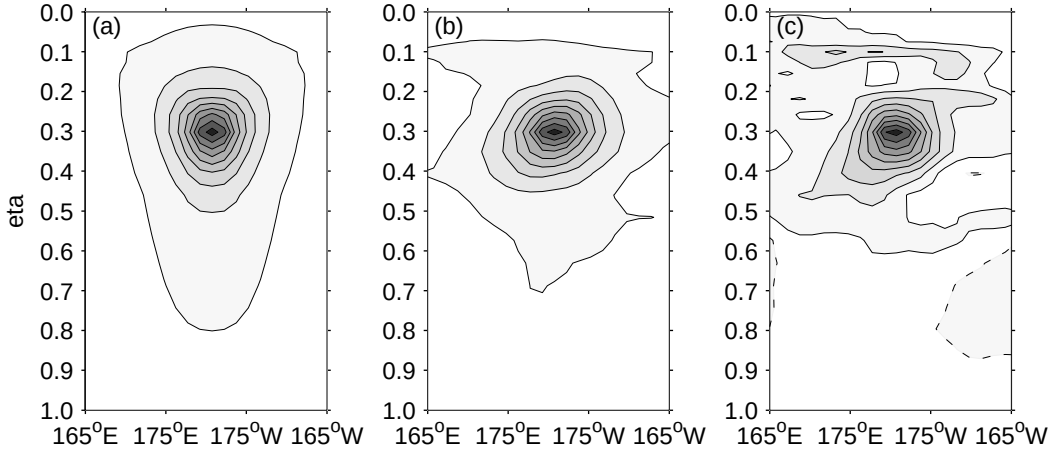

Figure 5: Same as the previous figure, but for a location at the equator.

localization is applied the covariances are quite noisy (Figure 4(c)). When spectral localization is applied with a localization radius of 10 (Figure 4(b)), the covariance structure is somewhat more noisy than with the diagonal spectral correlations, but significantly smoother than when no localization is applied. With no localization, the correlation structure is sharper in the zonal direction and broader in the vertical direction relative to the homogeneous correlations. The spectrally localized correlations appear to also exhibit this difference with the homogeneous correlation, though to a lesser degree.

Figure 5 shows the same type of result as Figure 4 except the location has been shifted to the equator. Again the spectrally localized correlations result in a spatially smoother covariance structure (Figure 5(b)) than the simple sample estimate (Figure 5(c)), but somewhat noisier than when employing horizontally homogeneous and isotropic correlations (Figure 5(a)). Now the correlations with spectral localization or no localization both exhibit a correlation structure that is broader in the zonal direction and sharper in the vertical direction. This is the opposite difference compared to the covariances at 60º S and presumably is a robust result related to differences in the extra-tropical versus tropical atmospheric dynamics. Similar latitudinal variations were also demonstrated by Ingleby (2001).

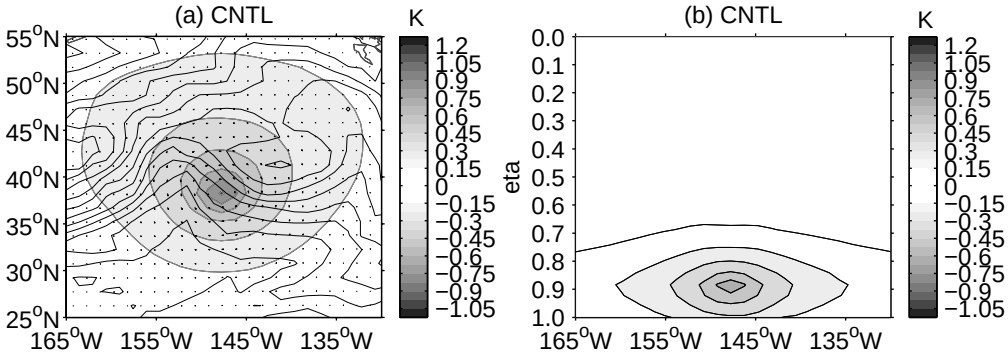

Figure 6: Analysis increment of temperature (shaded contours) and wind (vectors) from a single temperature observation near 900 hPa located in a strong near-surface temperature front at 12 UTC, 27 May 2002. Panel (a) shows temperature and wind increments near 900 hPa and (b) shows the vertical-zonal cross-section of the temperature increment along 40° N. The ensemble mean background temperature field is shown in the left panel as black unshaded contours with a contour interval 10 times larger than for the temperature increment. The CNTL background-error covariances are used for generating this analysis increment.

3.2 Case study with ensemble-based background-error covariances

To demonstrate the ability of the EnKF to provide suitable flow-dependent background-error covariances, a single temperature observation 1 K greater than the background temperature near 900 hPa was assimilated within a strong near-surface temperature front that appeared over the North Pacific on 27 May 2002 at 12 UTC. The background-error std dev for temperature and zonal wind estimated from the EnKF for this case (not shown) is elevated both in the vicinity of the front and close to the associated low pressure centre that is located to the north-west of the front at 45° N, 170° W. During the preceding two days this low pressure centre deepens rapidly as it progresses eastward across the Pacific Ocean.

The analysis increment produced using the background-error covariances from the operational 3D-Var (Figure 6) is clearly unaffected by the local meteorological conditions (the background temperature is shown in dark contours). The temperature increment decays in a nearly isotropic fashion away from the observation location and the wind increment is nearly zero at the location of the temperature observation. In contrast, when using the covariances estimated from the EnKF (Figure 7) the temperature increment is slightly elongated along the front and the wind increment is larger with vectors oriented parallel with the background temperature gradient at the observation location. For comparison, the same experiment was performed using the EnKF covariances from the previous day (not shown). For that case the gradient in the background temperature field is oriented almost perpendicular to the original case and again the analysis increment for wind is parallel with the background temperature gradient.

The 3D-Var was used to produce global analyses for 27 May 2002 at 12 UTC using the full set of operational observations. Analyses were obtained using the operational static error covariances (CNTL) and EnKF-based flow-dependent error covariances with spatial localization (ENKF-ENS). The statistics were computed for the difference between the two analyses and between the two resulting forecasts (not shown) for the North Pacific region (20° N to 65° N and 140° E to 120° W). The differences in the analyses due to the change in background-error covariances have a std dev near 1 m s^{-1} for the wind components, 0.5 K for the temperature field and less than 3 K for the dew-point depression (ES). After 2 days the maximum difference in the forecasts is seen for the winds near the level of the extra-tropical jet where the std dev has grown by about a factor of two. After 5 days the differences in all the variables have grown by about a factor of four, except for the winds at the jet level for which they have grown by about a factor of seven.

Figure 8 shows the statistics of the differences between forecasts and the operational analyses. Even though the difference between the two background-error covariance matrices produces a large difference in the resulting

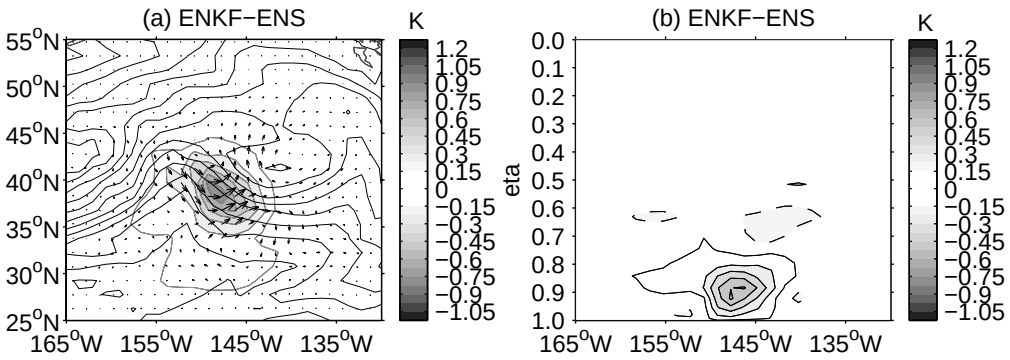

Figure 7: Same as Figure 6, but using ENKF-ENS background-error covariances estimated from the ensemble of EnKF background states valid at 12 UTC, 27 May 2002 with spatially localization.

forecasts, the agreement of the forecasts with the verifying analyses is much more similar. As expected the largest difference for the winds occurs at the jet level where the std dev of the 5 day forecast from the analysis with EnKF covariances is about 1 m s^{-1} smaller than for the control. In general, the analysis using the EnKF background-error covariances produces an improved forecast for this case and this region.

4 Conclusions

The localization of background-error correlations in both the spectral and spatial domains when estimated from small ensembles was presented. The two types of localization have a complementary effect on reducing sampling error. In the context of a 3D-Var system used for global NWP, spectral localization is shown to produce smoother background-error covariances than those estimated without localization. At the same time, the spectrally localized correlations are able to resolve large-scale differences between the covariance functions in the tropics and extra-tropics that are consistent with expectations. This is in contrast to the homogeneous and isotropic (i.e. spectrally diagonal) correlations currently used operationally. While the approach for implementing spectral localization in the variational analysis is relatively computationally inexpensive, it would likely be substantially more expensive to combine spectral and spatial localization. Additional research is required to obtain a practical implementation of simultaneous spectral and spatial correlation localization in a realistic NWP context.

Results from a previous study using EnKF-based background-error covariances in 3D-Var show promise. Since the time of those experiments both the EnKF and variational data assimilation system have been significantly improved. The EnKF system now relies less on additive random perturbations for representing model error. Also, these perturbations are applied just after the analysis step, instead of before. Consequently, these perturbations are now evolved through the model dynamics for 6 hours to the time of the next background ensemble. Finally, the EnKF now employs a four-dimensional analysis procedure, in which background-error correlations are computed between the time of the observation and the central analysis time. The variational system also has been extended from 3D-Var to 4D-Var (Gauthier *et al.* 2007). All these factors suggest that it may be worthwhile to reassess the impact of using flow-dependent EnKF-based background-error covariances for the variational assimilation system.

Acknowledgements

The author wishes to acknowledge the contributions of Peter Houtekamer, Pierre Gauthier and Martin Charron.

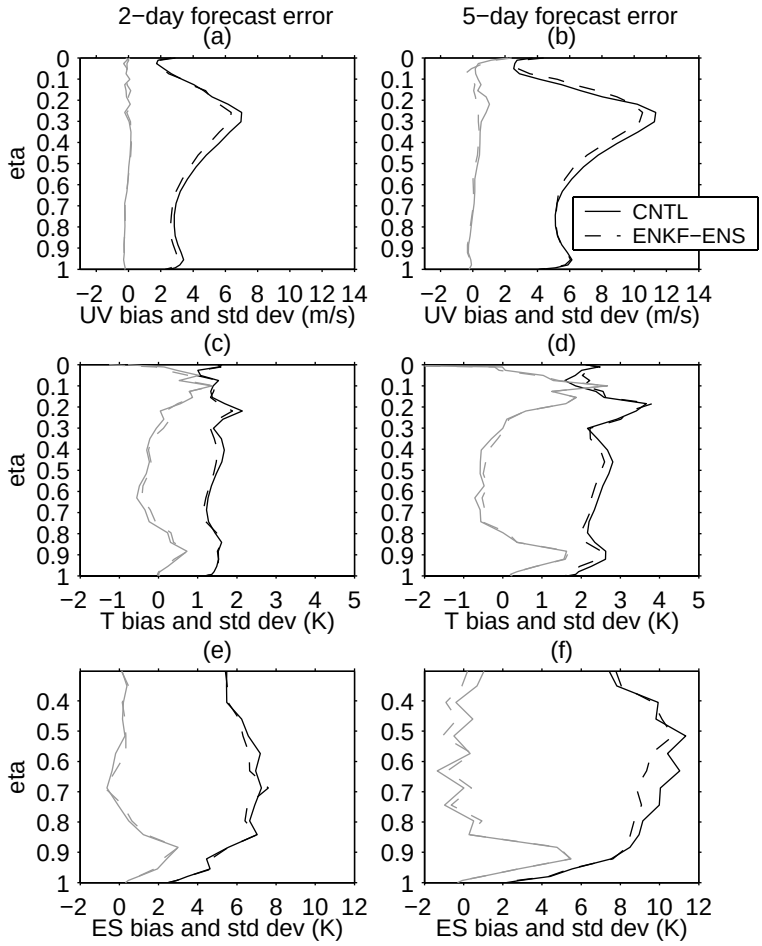

Figure 8: Bias (gray curves) and std dev (black curves) for the North Pacific region of the difference between the analyses from the CNTL forecast-analysis experiment and the 2 and 5 day forecasts from the analysis produced using either the operational background-error covariances (CNTL) or the flow-dependent covariances estimated from the EnKF (ENKF-ENS) at 12 UTC, 27 May 2002.

References

Buehner, M., 2005. Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**: 1013–1044

Buehner, M., P. Gauthier, and Z. Liu, 2005. Evaluation of new estimates of background and observation error covariances for variational assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3373–3383

Buehner, M. and M. Charron, 2007. Spectral and spatial localization of background-error correlations for data assimilation. *Q. J. R. Meteorol. Soc.* **133**: 615–630

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

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998. The operational CMC-MRB global environmental multiscale (GEM) model. part I: Design considerations and formulation. *Mon. Weather Rev.* **126**: 1373–1395

Fisher, M. and E. Andersson, 2001. Developments in 4D-Var and Kalman filtering. *ECMWF research department technical memorandum*, **347**

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

Gauthier, P., M. Buehner, and L. Fillion, 1998. 'Background-error statistics modelling in a 3D variational data assimilation scheme.' Pp. 131–145 in Proceedings of the ECMWF workshop on diagnosis of data assimilation systems, 2–4 November, 1998, Reading, UK

Gauthier, P., M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007. Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.* **135**: 2339–2354

Houtekamer, P.L. and H.L. Mitchell, 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137

Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005. Atmospheric data assimilation with an ensemble Kalman filter: results with real observations. *Mon. Weather Rev.* **133**: 604–620

Ingleby, B.N., 2001. The statistical structure of forecast errors and its representation in the Met. Office global 3-D variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **127**: 209–231

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

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