RMetS
Royal Meteorological Society

Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office

A. M. Clayton*, A. C. Lorenc and D. M. Barker
Met Office, Exeter, UK

*Correspondence to: A. M. Clayton, Met Office, FitzRoy Road, Exeter EX1 3PB, UK.
E-mail: adam.clayton@metoffice.gov.uk

We describe the development and testing of the hybrid ensemble/4D-Var global data assimilation system that was implemented operationally at the Met Office in July 2011, giving an average reduction of RMS errors of just under 1%. The scheme uses the extended control variable technique to implement a hybrid background error covariance that combines the standard climatological covariance with a covariance derived from the 23-member operational ensemble MOGREPS-G. Unique features of the Met Office scheme include application of a horizontal 'anti-aliasing' filter to the ensemble error modes, a vertical localization scheme based uniquely on a modification of the climatological stream function covariance, and inflation of the climatological covariance to maintain the analysis fit to observations. Findings during development include a significantly greater impact of the scheme in 3D-Var than 4D-Var, a clear positive impact from the combination of the anti-aliasing filter and vertical localization, and a relatively small sensitivity to full coupling of the ensemble and 4D-Var systems. Supplementary experiments suggest that the ability of the ensemble to capture coherent 'Errors of the Day' is key to the improvements in forecast skill.

A particular problem encountered during development was significantly poorer tropical verification scores when measured against own analyses. In contrast, verification against independent (ECMWF) analyses gave scores that were much more consistent with those against observations.

Key Words: background error covariance; covariance localization; MOGREPS

Received 22 February 2012; Revised 27 July 2012; Accepted 13 September 2012; Published online in Wiley Online Library 12 November 2012

Citation: Clayton A. M., Lorenc A. C. Barker D. M. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**: 1445–1461. DOI:10.1002/qj.2054

1. Introduction

The success of a data assimilation system relies heavily on the characterization of the background errors statistics, i.e. the statistics of the short-range forecasts that data assimilation seeks to correct. In the variational (VAR) data assimilation systems now used at all of today's major operational forecasting centres, background errors are typically based on highly-parametrized models of the error covariance, with the parameters obtained from climatological error statistics. One weakness of this approach is the difficulty of representing 'Errors of the Day' – the variations in error due to the locations of recent instabilities and observations. Thus the impact of observational information around, for example, frontal structures, is often highly suboptimal. By incorporating more sophisticated balance relationships into the covariance model – e.g. the nonlinear or omega balance equations (Fisher, 2003) – it is possible to improve the modelling of error structures that can be diagnosed directly from the model state. In four-dimensional variational

(4D-Var) systems the inclusion of a linear forecast model also allows a degree of additional implicit flow dependence to develop within the assimilation window, but this information is not carried forward to following cycles. Arguably, however, the most promising source of Errors of the Day information is a suitably designed ensemble prediction system (EPS) that accounts properly for the spatial and temporal characteristics of the observation network, and propagates error structures using a full nonlinear forecast model.

At the Met Office, ways to incorporate ensemble error structures were sought early in the development of its 3D-Var system. Using the extended control variable method of Lorenc (2003), experiments were run to test the impact of blending in a single error mode generated by a two-member error breeding system (Barker, 1999). With just a single mode, however, the impact of this 'hybrid' covariance on forecast performance was found to be negligible. The work was therefore put aside until a more sophisticated ensemble system became available.

In the meantime, the development and growing maturity of ensemble data assimilation techniques, such as the ensemble Kalman filter (EnKF; Evensen, 1994), increasingly proved their value in providing realistic estimates of short-range forecast error, with a natural inclusion of Errors of the Day. In a hybrid 3D-Var system coupled to an EnKF, based on a quasi-geostrophic model in a perfect model framework, Hamill and Snyder (2000) found best results when the standard quasi-static background error covariance was replaced almost entirely with the ensemble covariance. For small ensemble sizes, however, optimal performance was obtained with a reweighting towards the climatological covariance. Etherton and Bishop (2004) found similar results with a barotropic vorticity model, but found that when model error was introduced it was better to give more weight to the static covariance, presumably because of its better representation of model error. Buehner (2005) brought the hybrid scheme forward into a quasi-operational 3D-Var setting, but found the impacts to be rather small, suggesting that in the real world the effects of model and sampling error largely outweigh the benefits of capturing flow-dependent covariance structures.

Wang *et al.* (2008a, 2008b) studied the impact of hybrid covariances in a limited-area configuration of the Weather Research and Forecasting (WRF) 3D-Var system (Barker *et al.*, 2004) coupled to an ensemble transform Kalman filter (ETKF; Bishop *et al.*, 2001). In a perfect model setting, a blend of static and ensemble covariances was again found to give optimal results, particularly in data-sparse regions. Using real observations, hybrid covariances were again found to give the best results, but with a smaller impact, and an optimal weighting more towards the static covariance. Recently, Zhang and Zhang (2012) used a similar configuration of WRF to test a 4D-Var/EnKF hybrid, finding a significant improvement over the standard 4D-Var scheme.

All of the hybrid schemes mentioned above implement a covariance that is a simple linear combination of the climatological and modified ensemble covariances. Despite the modifications, designed to compensate for the small ensemble size, the ensemble covariances are still recognizably based on those used in an ensemble Kalman filter. An alternative approach is to use contemporary ensemble information to generate parameters for the standard error covariance model, rather than using only climatological training data. This approach is able to apply more rigorous filtering techniques to the estimation of the selected parameters. For instance, based on earlier work by Raynaud *et al.* (2009), Bonavita *et al.* (2012) used an independent ensemble of 4D-Vars to provide flow-dependent background error *variances* (not covariances) for use within the covariance model of the European Centre for Medium Range Weather Forecasts (ECMWF) deterministic 4D-Var system, giving substantial overall improvements in forecast quality. The estimated parameters – in this case variances – are more accurately determined from the ensemble than in our method. The approach is limited, however, by the flexibility of the particular covariance model on which it is based. For instance the Fisher (2003) model does not have parameters which might be determined from the ensemble to specify any three-dimensional anisotropy, or variations in the inter-variable correlations, both of which occur naturally in the ensemble-based covariances used in our approach. Anisotropic correlations appear to be important in Figure 9 below, and flow-dependent inter-variable correlations were shown to be important by Montmerle and Berre (2010). Our approach starts from the ensemble covariance and modifies it by 'localization' to remove aspects likely to be spurious. Possible improvements to the localization method are discussed in section 5, but in the first instance it is designed to remove spurious features rather than to optimally determine particular aspects of the covariance. The alternative approach starts from a model of the climatological error covariance and modifies it to determine a few parameters from the ensemble; a range of alternative covariance models are possible (e.g. that of Purser *et al.*, 2003, does allow anisotropy). While the chosen characteristics can be optimally estimated from the ensemble plus prior climatology, other aspects are determined by the chosen covariance model rather than the ensemble.*

At the Met Office, work on a global hybrid ensemble/variational data assimilation was resumed in late 2008, after the operational implementation of 4D-Var in 2004, and in particular the development and implementation of 'MOGREPS' (Bowler *et al.*, 2008), a global and regional ensemble prediction system based on the ETKF. The development and implementation of this hybrid ensemble/4D-Var system are the subject of this paper, which is organized as follows. Section 2 describes the formulation of the hybrid 4D-Var system, including details of the ensemble system providing the Errors of the Day, and its coupling with 4D-Var. In section 3, we describe the initial development path of the system, from a basic configuration including horizontal localization alone, to the configuration that formed the basis for pre-operational trials. Results of these trials are described in section 4, and we conclude with a discussion in section 5.

2. Hybrid 4D-Var formulation

2.1. 4D-Var

The Met Office's global 4D-Var scheme (Rawlins *et al.*, 2007) is based on the incremental formulation of Courtier *et al.* (1994), which linearizes the variational problem around a latest 'guess' trajectory produced by the full nonlinear

*Nonlinear balance transforms may, for instance, give some anisotropy, but that is determined by the linearization-state flow rather than the ensemble covariance.

forecast model. To further reduce computational costs, the analysis equations are solved in the state space of a simplified, lower-resolution 'Perturbation Forecast' (PF) model, with a simplification operator S and the approximate inverse $\mathbf{S}^{-I}$ of its linearization $\mathbf{S}$ enabling transformations between the full and low-resolution model grids. Using the notation of Rawlins *et al.* (2007), in which underlining is used to denote four-dimensional quantities, the variational problem is to find the increment $\delta\mathbf{w}$ which minimizes the cost function J given by

$$\begin{aligned} J(\delta\mathbf{w}) &= J_b + J_o + J_c \\ &= \frac{1}{2}\left(\delta\mathbf{w} - \delta\mathbf{w}^b\right)^T \mathbf{B}^{-1}\left(\delta\mathbf{w} - \delta\mathbf{w}^b\right) \\ &\quad + \frac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \left(\underline{\mathbf{E}} + \underline{\mathbf{F}}\right)^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right) \\ &\quad + J_c. \end{aligned} \tag{1}$$

In the background term J_b, all quantities are valid at the start of the 4D-Var window, with $\delta\mathbf{w}^b$ the difference between the simplified background and guess states (so that $\delta\mathbf{w} - \delta\mathbf{w}^b$ is the total difference w.r.t. the background), and $\mathbf{B}$ the background error covariance. However, since we currently run without updates to the background trajectory (i.e. a single 'outer loop'), $\delta\mathbf{w}^b = 0$, and J_b simplifies to $\frac{1}{2}\delta\mathbf{w}^T\mathbf{B}^{-1}\delta\mathbf{w}$.

In the observation term J_b, $\underline{\mathbf{y}}$ is the vector of observations, $\underline{\mathbf{E}}$ and $\underline{\mathbf{F}}$ the covariances of instrument and representativity error, and $\underline{\mathbf{y}}$ the model equivalent of $\underline{\mathbf{y}}^o$, given by

$$\underline{\mathbf{y}} = V\left[L\left\{M(\mathbf{x})\right\} + \mathbf{S}_c^{-I}\widetilde{\mathbf{L}}\widetilde{\mathbf{M}}\delta\mathbf{w}\right], \tag{2}$$

with the quantities on the right-hand side as follows:

- $\mathbf{x}$: full-resolution guess state at the beginning of the window;
- M: full nonlinear forecast model, producing a four-dimensional trajectory $\underline{\mathbf{x}}$ spanning the window;
- L: horizontal and time interpolation of $\underline{\mathbf{x}}$ to columns of model variables at the observation positions;
- $\widetilde{\mathbf{M}}$: linear PF model, producing a four-dimensional increment trajectory $\delta\underline{\mathbf{w}}$;
- $\widetilde{\mathbf{L}}$: linear horizontal and time interpolation of $\delta\underline{\mathbf{w}}$ to columns of increments at the observation positions;
- $\mathbf{S}_c^{-I}$: column version of the inverse simplification operator $\mathbf{S}^{-I}$;
- V: observation operator, acting on columns of model variables.

We use the strong-constraint form of 4D-Var in which the model is assumed perfect, so the expression for $\underline{\mathbf{y}}$ includes no model error terms.

Compared to the initial (2004) operational implementation described in Rawlins *et al.* (2007), the cost function J also includes an imbalance penalty J_c: a penalty on the increments from a low-pass digital time filter, similar to that of Gauthier and Thépaut (2001) but using a different norm to measure the size of the filter increments.[†]

[†]While Gauthier and Thépaut (2001) use the full energy norm, the Met Office scheme uses only the 'elastic' term that depends on the pressure increment, and applies this to both pressure and pressure-tendency.

The focus of the hybrid system is the modelling of the background error covariance $\mathbf{B}$. All other aspects of 4D-Var are unaffected.

2.2. Climatological covariance

In the Met Office's standard (non-hybrid) 4D-Var scheme, $\mathbf{B}$ is approximated by a static covariance $\mathbf{B}_c$ which is kept constant from cycle to cycle. Thus we aim for a $\mathbf{B}_c$ which captures the main climatological features of the background error covariance. To improve the conditioning of the minimization problem, we require not $\mathbf{B}_c$ itself, but its square root $\mathbf{U}$. Equation (1) can then be rewritten in terms of a new control vector $\mathbf{v}$ defined by

$$\delta\mathbf{w} = \mathbf{U}\mathbf{v}, \tag{3}$$

with J becoming a function of $\mathbf{v}$ alone, and the background term J_b simplifying to $\frac{1}{2}\mathbf{v}^T\mathbf{v}$.

$\mathbf{U}$ is designed indirectly, as the approximate inverse of a matrix $\mathbf{T}$ designed to transform samples of the background error covariance on the model grid to a set of scalar variables – the elements of $\mathbf{v}$ – which can be considered approximately uncorrelated with unit variance. The process of removing inter-variable correlations is divided into stages by expressing $\mathbf{T}$ as the product of three transforms:

$$\mathbf{T} = \mathbf{T}_h\mathbf{T}_v\mathbf{T}_p. \tag{4}$$

Here, $\mathbf{T}_p$ transforms to four control variable fields with approximately uncorrelated errors: increments to stream function ψ; velocity potential χ; a geostrophically unbalanced pressure p_u defined with respect to a generalized form of the geostrophic relationship, and a humidity field μ. Intra-field correlations are then reduced via a transform $\mathbf{T}_v$ which projects onto a global set of approximately uncorrelated vertical modes, and via a subsequent horizontal transform $\mathbf{T}_h$ which projects onto global spherical harmonic functions. The parameters required by the transform are derived from a suitable collection of training data: a set of several hundred model increments chosen to be representative of the climatologically-averaged background error.

The transform $\mathbf{U}$ is constructed by finding approximate or exact inverses $\mathbf{U}_h$, $\mathbf{U}_v$ and $\mathbf{U}_p$ of $\mathbf{T}_h$, $\mathbf{T}_v$ and $\mathbf{T}_p$:

$$\mathbf{U} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h. \tag{5}$$

The assumption that the control vector elements are uncorrelated with unit variance then defines $\mathbf{B}_c$ via Eq. 3. We note that the assumption of uncorrelated horizontal spectral coefficients leads to homogeneous and isotropic correlations of the vertical mode coefficients. The scalings of the vertical modes are allowed to vary spatially, but only as a function of latitude. Thus $\mathbf{B}_c$ is zonally uniform. These and other assumptions built into the design of $\mathbf{B}_c$ mean that it is at best a highly simplified parametrization of the true climatological error covariance. For a fuller description of $\mathbf{B}_c$, see Lorenc *et al.* (2000) and Ingleby (2001). Further discussion in the context of the designs used at other centres is contained in Bannister (2008).

2.3. Ensemble covariances

Apart from its simplified representation of the 'true' climatologically averaged background error covariance, the

main weakness of $\mathbf{B}_c$ is its failure to represent Errors of the Day; i.e. the variations in background error statistics due to the locations of recent atmospheric instabilities, or variations in the observation network. In 4D-Var, the implicit propagation of $\mathbf{B}_c$ by the PF model leads to the generation of flow-dependent covariance structures later in the data assimilation window, but these structures are not carried over to the beginning of the following cycle, which starts again from $\mathbf{B}_c$.

The idea of hybrid data assimilation systems is to remedy this situation by blending in error structures from an ensemble prediction system (EPS). The operational EPS at the Met Office is called MOGREPS; the Met Office Global and Regional Ensemble Prediction System (Bowler *et al.*, 2008). The global component of this system (MOGREPS-G) is based on the ensemble transform Kalman filter (ETKF) of Wang *et al.* (2004), and for the experiments described in this paper included the horizontal localization scheme described in Bowler *et al.* (2009), the stochastic kinetic energy backscatter (SKEB) scheme described in Tennant *et al.* (2011), and the level-dependent inflation scheme described in Flowerdew and Bowler (2012). One of the important design features of MOGREPS is its focus on short-range forecast errors, giving hope of reasonable covariance estimates at the times required by the hybrid system. In particular, unlike ensemble systems based on singular vectors for example, MOGREPS takes into account the distribution of recent observations.

Rather than creating its own ensemble-mean analyses, analysis perturbations generated by MOGREPS-G are recentred around analyses produced by the deterministic global 4D-Var system. Thus MOGREPS-G is dependent on 4D-Var. Since the ensemble forecasts use a lower-resolution model configuration than the deterministic forecasts, this necessitates interpolation of the analyses onto the ensemble grid. For the experiments described in this paper, MOGREPS-G was run with 23 perturbed members, with updates at 0000 and 1200 UTC each day.

In the hybrid system, 4D-Var becomes dependent on forecast data from MOGREPS-G, so the two systems become fully coupled. Since the 4D-Var system uses a 6 h assimilation cycle, each MOGREPS-G cycle is required to provide data at two forecast ranges, coinciding with the start of the 4D-Var analysis windows, as illustrated schematically in Figure 1. Thus the 0600 and 1800 UTC hybrid analyses use $T+3$ forecast data, and the 0000 and 1200 UTC analyses use $T+9$ data. The background states at the beginning of each 4D-Var window are nominally $T+3$ forecasts, so in theory we might expect the 0000 and 1200 UTC analyses to be disadvantaged by using ensemble data at the wrong forecast range. In practice, however, the variance and length scale differences between $T+3$ and $T+9$ MOGREPS forecast errors (not shown) are small enough for us to expect only a minor impact of this factor on forecast performance.

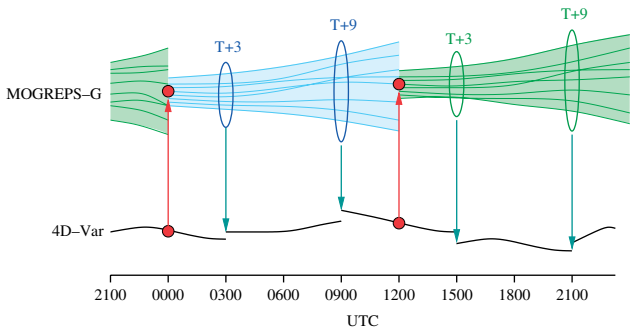

Figure 1. Schematic showing the coupling between MOGREPS-G and 4D-Var. The line segments at the bottom represent the 4D-Var assimilation windows, the large dots the analyses around which the ensemble perturbations are recentred, and the vertical arrows the exchange of information between the two systems. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

At the start of each 4D-Var window, the necessary ensemble forecast fields are taken from MOGREPS-G, and interpolated onto the analysis grid used by 4D-Var using the simplification operator S, producing states $\mathbf{w}_k$, where k is the member index. It is convenient to introduce a rectangular matrix $\mathbf{W}$ whose columns are scaled differences between the ensemble forecasts and the ensemble mean:

$$\begin{aligned}\mathbf{W} &= \frac{1}{\sqrt{K-1}}\left(\mathbf{w}_1-\bar{\mathbf{w}}, \mathbf{w}_2-\bar{\mathbf{w}}, \ldots, \mathbf{w}_K-\bar{\mathbf{w}}\right)\\ &= \left(\mathbf{w}'_1, \mathbf{w}'_2, \ldots, \mathbf{w}'_K\right),\end{aligned} \tag{6}$$

where K is the number of perturbed members. The ensemble covariance $\mathbf{P}_e^f$ is then given by

$$\mathbf{P}_e^f = \mathbf{W}\mathbf{W}^T. \tag{7}$$

The main problem with this raw ensemble covariance is its low rank (K) and the presence of sampling error. To remedy both problems, Houtekamer and Mitchell (2001) suggested replacing Eq. (7) by

$$\mathbf{B}_e = \mathbf{P}_e^f \circ \mathbf{C}, \tag{8}$$

where $\mathbf{C}$ is a 'localization' covariance (normally correlation) matrix, and the operator $\circ$ denotes the element-by-element product of two same-sized matrices known as the Schur product. The basic aim of $\mathbf{C}$ is to downweight small ensemble correlations that are likely to be dominated by sampling noise, while leaving larger more robust correlations relatively unaffected. The simplest designs of $\mathbf{C}$ are localizations in a literal sense; applying a downweighting factor that decreases monotonically from unity as physical separation increases, on the assumption that ensemble correlations are likely to reduce in magnitude with separation. The localization scheme used within our hybrid system relaxes this assumption in the vertical, but like other simple spatial schemes does not address the issue of sampling noise in the ensemble variances – only in the correlations.

2.4. *Hybrid covariances*

The hybrid 4D-Var system seeks to implement a background error covariance $\mathbf{B}$ that is a linear combination of the climatological and ensemble covariances described above; i.e. a covariance of the form

$$\mathbf{B} = \beta_c^2\mathbf{B}_c + \beta_e^2\mathbf{B}_e, \tag{9}$$

where β_c^2 and β_e^2 are scalar weights. To accomplish this, we use the extended control variable method described in section 5 of Lorenc (2003). In this formulation, the climatological contribution $\mathbf{Uv}$ to $\delta\mathbf{w}$ is multiplied by β_c, and the introduction of each ensemble error mode $\mathbf{w}'_k$ is

controlled via its Schur product with a 3-dimensional scalar field $\boldsymbol{\alpha}_k$:

$$\delta\mathbf{w} = \beta_c\mathbf{U}_p\mathbf{U}_v\mathbf{U}_h\mathbf{v} + \beta_e\sum_{k=1}^{K}\mathbf{w}'_k \circ \boldsymbol{\alpha}_k. \quad (10)$$

The VAR cost function is then modified to

$$J(\mathbf{v},\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_K) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \sum_{k=1}^{K}\boldsymbol{\alpha}_k^T\mathbf{C}^{-1}\boldsymbol{\alpha}_k + J_o + J_c. \quad (11)$$

To improve the conditioning of the minimization problem, Eq. (11) is rewritten in terms of an 'alpha' control vector $\mathbf{v}^\alpha$ which is the concatenation of the K vectors $\mathbf{v}_k^\alpha$ defined by

$$\boldsymbol{\alpha}_k = \mathbf{U}^\alpha\mathbf{v}_k^\alpha, \quad (12)$$

where $\mathbf{U}^\alpha = \mathbf{C}^{1/2}$. Substituting for $\boldsymbol{\alpha}_k$ in Eq. (11), we then have

$$J\left(\mathbf{v},\mathbf{v}^\alpha\right) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\left(\mathbf{v}^\alpha\right)^T\left(\mathbf{v}^\alpha\right) + J_o + J_c. \quad (13)$$

The proof that Eqs (10), (12) and (13) implement the hybrid covariance given by Eq. (9) is given by Wang *et al.* (2007).

We note that localization is being performed in model space. For non-local observations such as satellite radiances, Campbell *et al.* (2010) show that model space localization is superior to the observation-space localization schemes typically used with the ensemble Kalman filter, such as that of Houtekamer and Mitchell (2001).

2.5. *Modelling of the localization covariance* **C**

The localization covariance **C** is determined by the definition of its square root $\mathbf{U}^\alpha$. As for **U**, we model this via a series of transforms:

$$\mathbf{U}^\alpha = \mathbf{U}_v^\alpha\mathbf{U}_h^\alpha, \quad (14)$$

where $\mathbf{U}_h^\alpha$ and $\mathbf{U}_v^\alpha$ define localizations in the horizontal and vertical respectively.

The horizontal transform $\mathbf{U}_h^\alpha$ makes use of the spectral transform built into the standard horizontal transform $\mathbf{U}_h$. This is used to model a homogeneous and isotropic Gaussian correlation of the form

$$\mu(z) = \exp\left(-\frac{z^2}{2L^2}\right), \quad (15)$$

where z is the horizontal separation between the two grid points in question and L is the parameter we use to specify the length scale. To improve the efficiency of the transform and reduce the length of $\mathbf{v}^\alpha$, we truncate the spectrum used to represent the function, discarding high-wavenumber modes that make insignificant contributions.

As an aside, it is worth noting that there are a number of ways to specify the length scales of Gaussian-like correlation functions, and it is important to take these differences into account when comparing localization schemes. Many authors use the fifth-order piecewise rational function given by equation (4.10) of Gaspari and Cohn (1999), which is approximately Gaussian, but with exactly zero correlations beyond a certain distance $2c$. It is this distance that is normally quoted as the length scale, and it can be shown that the L of the corresponding Gaussian function is given by $L = \frac{1}{2}\sqrt{0.3}(2c) \approx 0.27 \times (2c)$. Alternatively, Wang *et al.* (2008a) specify the length scale via the e-folding distance S_e of the Gaussian function. In this case, $L = \sqrt{0.5}S_e \approx 0.71 \times S_e$.

The design of the vertical localization scheme is described in section 3.2, with further modifications described in section 4. The basic approach is to choose a suitable 'target' vertical localization matrix $\mathbf{C}_{v,\text{target}}$, and then model it via an approximate square root $\mathbf{U}_v^\alpha$, keeping the number of columns (vertical modes) to a minimum to reduce memory usage and the cost of transform. This is done by using a standard algorithm (DSYEV in the LAPACK library) to calculate the empirical orthogonal functions (EOFs) of $\mathbf{C}_{v,\text{target}}$ with respect to a mass-weighted inner product, producing an ordered series of vertical modes explaining the maximum remaining variance. After removing any EOFs with non-positive eigenvalues, we retain sufficient EOFs to explain most (∼95%) of the variance, and restore the variance that is lost by applying level-dependent rescalings to the set of retained modes.[‡] Currently we assign the same horizontal localization scale to each vertical mode, leading to a separable representation of **C**.

Compared to the rather complex design of $\mathbf{B}_c$, we note that **C** can be much simpler. Rather than providing a full representation of the background error covariance, it is only required to improve the properties of the covariance $\mathbf{P}_e^f$ provided directly by the ensemble. Most of the interesting properties of $\mathbf{B}_e$, such as possible flow-dependent tilt, or inter-variable correlations, are inherited from $\mathbf{P}_e^f$, not from **C**.

2.6. *Balance-preserving localization*

In the Met Office system, the ensemble error modes $\mathbf{w}'_k$ contain increments to the wind components u and v and pressure field p on fixed-height levels, plus increments to potential temperature θ and a total humidity variable q^T. As shown in section 3(c) of Lorenc (2003), application of the Gaussian horizontal localization function (15) in $u/v/p$ space leads to the generation of sub-geostrophic wind increments, reducing the degree of geostrophic balance. This follows from the form of the incremental geostrophic balance equation, which is approximately given by

$$\mathbf{v}'_g = \mathbf{k}\wedge\frac{1}{\rho f}\nabla_h p', \quad (16)$$

where f is the Coriolis parameter and ρ is the density from the linearization state. We see that there is a horizontal derivative on p', but not on the wind increments. Thus, on localization, the change to the left-hand side depends on the local value of the localization field, while the change to the

[‡]Note that the positive-definiteness attained by using only EOFs with positive eigenvalues is unaffected by the rescaling, so the localization transform $\mathbf{U}_v^\alpha$ gives rise to a genuine covariance.

right-hand side depends on its horizontal gradient, leading to a reduced degree of geostrophic balance. Similarly, in terms of p' and θ', the incremental form of the hydrostatic equation is

$$\frac{\partial p'}{\partial z} = Ap' - B\theta', \tag{17}$$

where A and B are linearization constants and z is the height above the model surface – the vertical grid coordinate. Again, there is a spatial gradient on one side of the equation, but not on the other, so vertical localization in p/θ space will tend to upset hydrostatic balance.

To overcome these problems, we choose instead to localize in the space of the control variable fields ψ, χ, p_u and μ. This is accomplished by replacing Eq. (10) with

$$\delta\mathbf{w} = \mathbf{U}_p \left\{ \beta_c \mathbf{U}_v \mathbf{U}_h \mathbf{v} + \beta_e \sum_{k=1}^{K} \left(\mathbf{T}_p \mathbf{w}'_k \right) \circ \boldsymbol{\alpha}_k \right\}, \tag{18}$$

where $\mathbf{T}_p$ is the parameter transform introduced in Eq. 4. Under this scheme, the geostrophically balanced portion of the pressure increment in each error mode $\mathbf{w}'_k$ is discarded during the transform by $\mathbf{T}_p$, and recalculated by $\mathbf{U}_p$ *after* Schur multiplication by the localizing field $\boldsymbol{\alpha}_k$. Thus the geostrophic portion of each error mode remains geostrophic after localization, leading to better-balanced analysis increments. Note that the geostrophically unbalanced pressure increment p'_u is not discarded, but directly localized. (We are not imposing geostrophic balance on the error modes.)

By design, $\mathbf{T}_p$ is insensitive to hydrostatically unbalanced pressure increments, so any such increments present within the ensemble error modes $\mathbf{w}'_k$ are ignored, whether or not localization is applied. As with geostrophically balanced pressure, hydrostatic pressure increments are discarded during the transform by $\mathbf{T}_p$, and recalculated by $\mathbf{U}_p$ after localization. Thus Eq. 18 imposes hydrostatic balance.

One balance that is not preserved under (vertical) localization is the approximate cancellation of mass convergence and divergence into a model column – the so-called 'Dines compensation' effect. We plan to address this issue in a future version of the hybrid scheme.

2.7. *Inter-variable localization*

We note that a common scalar localization field $\boldsymbol{\alpha}_k$ is used for each of the control variable fields in $\mathbf{T}_p\mathbf{w}'_k$. This implies no 'localization' of the correlations between the control variable fields, so we are accepting the presence of sampling noise in the inter-variable correlations. The simplest alternative is to use independent alpha fields $\boldsymbol{\alpha}_k^{\psi}$, $\boldsymbol{\alpha}_k^{\chi}$, $\boldsymbol{\alpha}_k^{p_u}$ and $\boldsymbol{\alpha}_k^{\mu}$ for each control variable field, which has the effect of removing inter-variable correlations entirely, and also gives the opportunity to use different localization scales for different fields. This is the approach taken in Buehner (2005), albeit with common localization scales. A test of this strategy in our own hybrid system gave slightly worse results (not shown), so it seems that retaining inter-variable relationships slightly outweighs the impact of sampling noise.

2.8. *Quality control*

The Met Office performs its observational quality control in a pre-processing step (Lorenc and Hammon, 1988), including tests performed during 1D-Var retrievals (Pavelin *et al.*, 2008). These use estimates of variances and vertical correlations based on $\mathbf{B}_c$ with some added flow dependence (Parrett, 1992), so they are not affected by this development. If we were to incorporate ensemble-based variances, we would use the variance estimation approach of Raynaud *et al.* (2009). On the other hand, if we adopted variational QC (Ingleby and Lorenc, 1993; Andersson and Järvinen, 1999), then our hybrid covariances would automatically be used in quality control.

3. Initial development path

In this section, we outline the key steps made during initial scientific development and trialling of the hybrid system, starting from a basic system with no vertical localization, and ending with the configuration that formed the basis for the pre-operational trials described in section 4. During this process, the main workhorse was a low-resolution trial configuration designed to be computationally cheap enough to allow multiple trials and fast turnaround, but close enough to the operational system for the findings to be relevant. The main details of the trial configuration were as follows:

Period:
  5–31 May 2008
Deterministic forecast grid:
  $288 \times 217 \times 38$
MOGREPS grid:
  $288 \times 217 \times 38$
VAR grid:
  $216 \times 163 \times 38$
Humidity control variable μ:
  total relative humidity
MOGREPS/VAR coupling?
  No
Training data for B_c:
  $T + 30 - T + 6$ forecasts

The deterministic and MOGREPS forecasts were carried out using 38-level configurations of the Met Office Unified Model (Cullen, 1993), with the top level 40 km above sea level. Global configurations of the model use regular latitude–longitude grids in the horizontal, so the $288 \times 217 \times 38$ grid has spacings of $1\frac{1}{4}^{\circ}$ in the east–west direction, and $\frac{5}{6}^{\circ}$ in the north–south direction.

Note that for these early tests we decided to run without coupling between MOGREPS and VAR. Instead, the MOGREPS ensemble was run in advance of the hybrid VAR trials, with the perturbations recentred around analyses from a 4D-Var trial that had been run for other purposes. The MOGREPS run was started on 1 May to allow 4 days of spin-up, and the forecast data required for the hybrid system saved to an archive. The hybrid trials then retrieved these data as required. The main advantage of running uncoupled is that it is only necessary to run the ensemble once for each trial period. The disadvantage is that the influence of the hybrid analyses on MOGREPS, and the impact of keeping the MOGREPS and hybrid analyses in step are excluded. For early development of the hybrid system, we judged that these impacts would likely be small compared to the impact of the hybrid scheme itself. This judgement was later confirmed by one of the tests described in section 4.

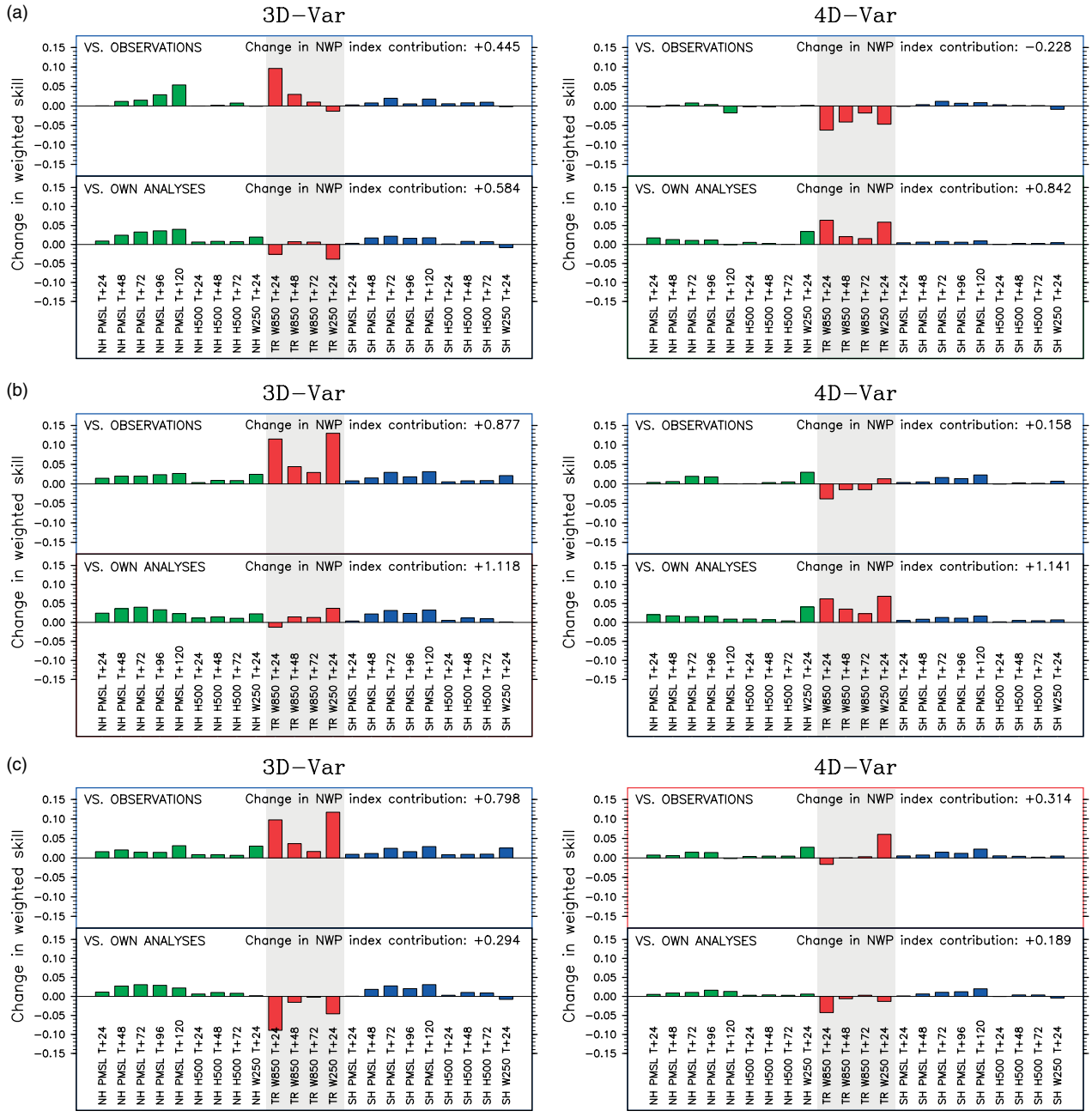

Figure 2. Weighted skill and NWP index changes relative to non-hybrid controls for the sequence of 3D-Var (left) and 4D-Var (right) trials described in sections 3.1–3.3. (a) 50% $\mathbf{B}_c$/50% $\mathbf{B}_e$ hybrid with horizontal localization alone ($L = 1500$ km). (b) As (a), but with the horizontal anti-aliasing filter and vertical localization. (c) As (b), but with 80% $\mathbf{B}_c$. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

As well as 4D-Var, we also ran 3D-Var hybrid trials, using a set-up as close as possible to the 4D-Var configuration. One of the necessary changes was to use $T + 6$ and $T + 12$ forecast data from MOGREPS, rather than the $T + 3$ and $T + 9$ forecasts used for 4D-Var. Because of the lack of a 3D-Var J_c term to control imbalances within the VAR minimization, we also introduced an external digital filtering initialization scheme.

3.1. *Initial trials with horizontal localization alone*

For the first pair of trials, the hybrid scheme was run with horizontal localization alone ($\mathbf{U}_v^{\alpha}$ a column vector with all elements equal to 1), using a localization scale L of 1500 km. The weighting factors β_c and β_e in Eq. (9) were both set to $\sqrt{0.5}$, giving a 50%–50% hybrid of the climatological and ensemble background error covariances.

As a summary of the verification results, Figure 2(a) presents changes to weighted skill scores for the fields and forecast ranges used within the so-called 'global NWP index', which is the basic overall measure of global forecast performance used at the Met Office. The fields chosen for this index are those judged to be of most relevance to the Met Office's global NWP customers, and are weighted according to importance (Table 1). For example, Northern Hemisphere forecasts are given twice the weight of Southern Hemisphere forecasts. The weighted skill for a particular field is given by $w_i\left(1 - r_f^2/r_p^2\right)$, where r_f is the RMS error of the forecast, r_p that of the corresponding persistence forecast, and w_i the weight given in the table. The skill scores

Table 1. Fields and weights used in the NWP index.

Region	Field	$T+24$	$T+48$	$T+72$	$T+96$	$T+120$
NH	PMSL	10	8	6	4	4
	H500	6	4	2	—	—
	W250	12	—	—	—	—
TR	W850	5	3	2	—	—
	W250	6	—	—	—	—
SH	PMSL	5	4	3	2	2
	H500	3	2	1	—	—
	W250	6	—	—	—	—

PMSL is pressure at mean sea level, H500 geopotential height at 500 hPa, and W250 and W850 the winds at 250 hPa and 850 hPa. The Tropics are defined as the region between 20°S and 20°N.

are calculated against surface and radiosonde observations, and against the analyses produced by the trial itself. For verification against analyses, all fields are first interpolated onto a regular lower-resolution verification grid. The plots also show changes to the NWP index itself, which is a function combining the individual skill scores, as explained in Appendix A of Rawlins *et al.* (2007).

Except for slight degradations of the $T+24$ wind scores against analyses in the Tropics, the 3D-Var hybrid trial shows consistent improvements in forecast skill over the non-hybrid control. Typically, combined NWP developments at the Met Office increase the global NWP index by approximately 2.5 points per year, so the total NWP index improvement of around 1 point – obtained by summing the two individual contributions of 0.445 and 0.584 – represents a substantial improvement. In particular, it is well above the 0.3-point improvement that we judge from experience to be empirically significant for a one-month trial. Thus, had we been running a 3D-Var scheme operationally, the results might have justified an immediate move to full pre-operational trials. The scores for 4D-Var, however, were less encouraging, with small but mostly positive changes in the Extratropics, but significantly negative scores against radiosonde wind observations in the Tropics. This slightly disappointing result led us to seek further developments aimed at a more decisive improvement in 4D-Var performance.

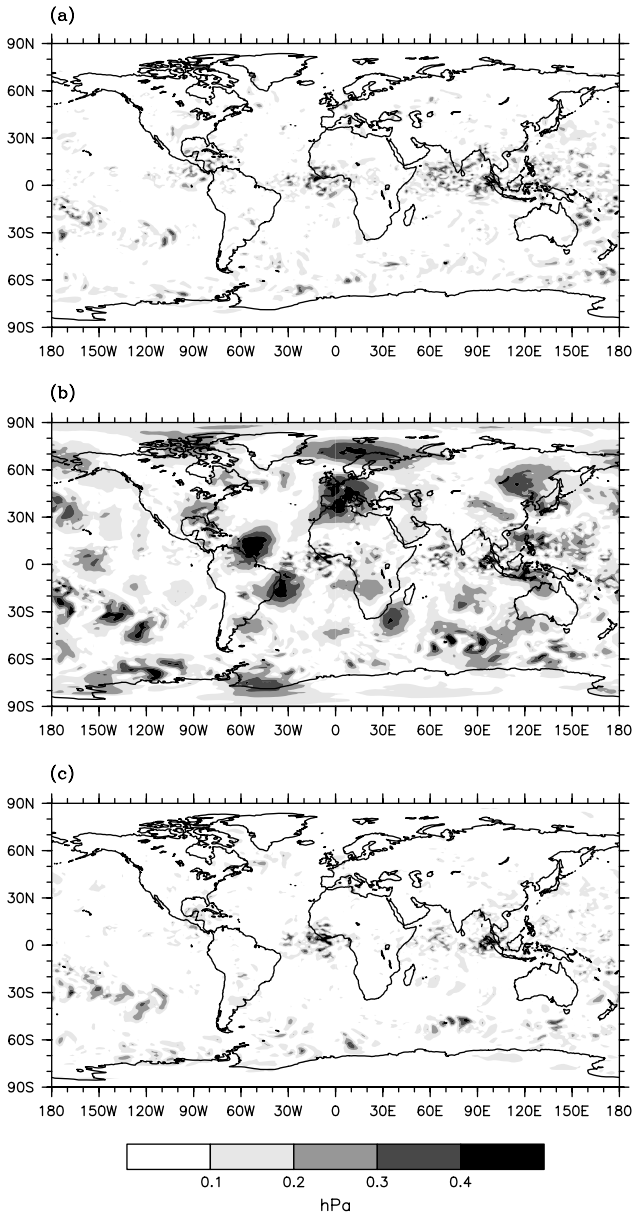

Figure 3. Absolute high-pass time filter output for level-one pressure for samples from the ensemble covariance $\mathbf{B}_e$. (a) No localization. (b) Horizontal localization with $L = 1500$ km. (c) As (b), but after application of a horizontal 'anti-aliasing' filter to the transformed error modes $\mathbf{T}_p\mathbf{w}'_k$.

3.2. *Anti-aliasing filter and vertical localization*

The next pair of trials included two changes: application of a horizontal 'anti-aliasing' filter to the MOGREPS error modes and addition of a vertical localization scheme.

The first of these developments was motivated by studies of the degree of balance of random samples from the ensemble covariance $\mathbf{B}_e$; i.e. the balance of increments of the form $\delta\mathbf{w} = \mathbf{B}_e^{1/2}\mathbf{v}^\alpha = \mathbf{U}_p\left\{\sum_{k=1}^{K}\left(\mathbf{T}_p\mathbf{w}'_k\right)\circ\mathbf{U}^\alpha\mathbf{v}_k^\alpha\right\}$, with the elements of the $\mathbf{v}_k^\alpha$ random samples from a Gaussian distribution with zero mean and unit variance. As a basic balance diagnostic, we integrated these samples forwards using the PF model, and applied a high-pass digital time filter with a cut-off period of approximately 4 h to reveal the high frequency part of the solution, including the presence of high-frequency gravity-inertia waves. Figure 3(a) shows absolute filter output for pressure at model level one (just above the surface) for a particular random sample with no covariance localization applied. Most of the high-frequency activity is in the Tropics, and can be attributed to gravity-inertia waves generated by the convection scheme used within the MOGREPS forecast model. In the Extratropics, we see relatively little sign of imbalance – only the kind of structures we would expect from the movement of increments to weather systems. Thus, except for a small amount of intrinsic model noise in the Tropics, the unlocalized ensemble covariance is well balanced.

Figure 3(b) shows the same diagnostic, but after application of the horizontal localization scheme used within the trials of the previous section. Here we see a significant amount of high-frequency activity throughout the domain, clearly due to the presence of gravity-inertia wave activity. As explained in section 2.6, the localization scheme applied to the ensemble covariance is designed to approximately maintain geostrophic balances, so the result was initially a surprise. On investigation, it was found that the

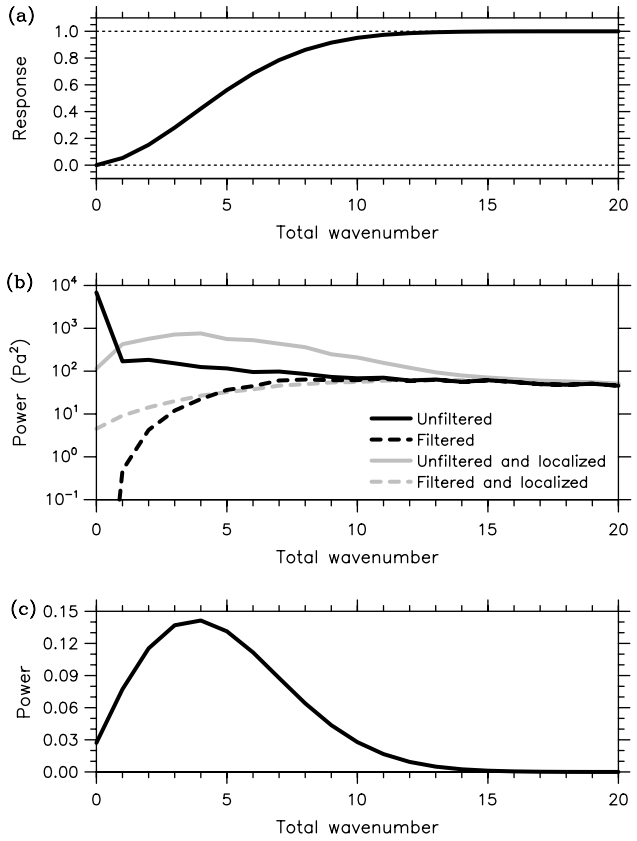

Figure 4. The anti-aliasing filter (a) and its effect (b) on the horizontal power spectra of localized and unlocalized level-one unbalanced pressure (p_u) ensemble perturbations. For reference, (c) shows the spectrum of the localization field $\boldsymbol{\alpha}$, normalized to unit variance. See section 3.2 for full details.

problems arose mainly from localization of increments to the geostrophically unbalanced pressure p_u. These increments include the globally averaged pressure increment – which does not project appreciably onto gravity-inertia wave structures – and also contributions to high-order balances not captured by the geostrophic relationship used to define p_u. Modification of these structures under localization can upset existing balances and create high-frequency gravity-inertia waves.

The power spectrum for level-one p_u – averaged over the 23 ensemble members used for the experiment – is shown by the solid black line in Figure 4(b). We see significant power in the lower wavenumbers, particularly wavenumber 0 – the global average.[§] The black dashed line shows the average spectrum for 1000 samples from the localized ensemble covariance; i.e. a good approximation to the spectrum after localization. We see that the power in wavenumber 0 has been reduced, but that power has been increased from wavenumber 1 to around wavenumber 14. This is the result of aliasing from low wavenumbers to the scales present in the localization covariance, whose spectrum is shown in Figure 4(c). This aliasing effect is most easily understood for the globally averaged pressure increments. These are introduced into the analysis via their Schur product with the corresponding alpha field $\boldsymbol{\alpha}_k$ (Eq. 18), producing a pressure field proportional to $\boldsymbol{\alpha}_k$. Since there are no corresponding changes to the wind field, these structures are largely radiated away as gravity-inertia waves with scales similar to those of alpha itself.

In order to reduce this problem, we implemented a high-pass 'anti-aliasing' horizontal filter for the transformed ensemble modes $\mathbf{T}_p\mathbf{w}'_k$, to remove large-scale power from the ensemble covariance in $\psi/\chi/p_u/\mu$ space. The filter is based on the same Gaussian function (Eq. (15)) as used for horizontal localization, but with the length scale L multiplied by a tunable factor $1/F$. The square root of this function's power spectrum is subtracted from unity to create a high-pass filter function for application to the spectral coefficients of each horizontal field. Figure 4(a) shows the response of this filter for the chosen value of L (1500 km), with $F = \sqrt{2}$.[¶] The effect on the level-one p_u spectrum is shown by the grey lines in Figure 4(b). We see that the aliasing of power onto the scales of the localization spectrum is much reduced. The impact of this filter on balance is shown in Figure 3(c), which is the same as (b) but includes the effects of the filter. We see that the spurious gravity-wave activity has been almost completely removed.[‖] All subsequent experiments discussed in this paper include the filtering, with $F = \sqrt{2}$ throughout.

Although initially motivated by balance considerations, we would argue that the aliasing of large scales onto those of the localization spectrum is a more general issue, potentially damaging the quality of the localized covariances. Thus, when the ensemble has significant power in scales larger than the localization scale, use of an anti-aliasing filter might be a useful addition regardless of any connection with imbalance. In the case of MOGREPS-G, independent (currently unpublished) work suggests that length scales are unrealistically large. If and when these problems are solved, we may reconsider the use of the anti-aliasing filter, but until then it is probably a sensible addition to compensate for flaws in the ensemble.

The second development was to introduce a vertical localization scheme. Generalization of the Gaussian horizontal localization scheme into the vertical is problematic for two reasons. Firstly, the magnitude of vertical correlations can no longer be assumed to reduce smoothly with separation. For example, temperature and divergence errors often have large negative autocorrelations with other levels (Ingleby, 2001). Secondly, it is challenging to find a vertical coordinate with respect to which correlation length scales are approximately constant as a function of level. Buehner (2005) used a Gaussian-like correlation function defined with respect to pressure scale height, which is probably a reasonable compromise. We have chosen instead to sidestep these issues and work directly in correlation space. We choose a representative variable, obtain its globally averaged vertical correlations, and then modify them to produce an appropriate localization covariance.

As the representative variable, we have chosen stream function ψ, firstly because its vertical correlation scales

[§]The peak in power at wavenumber 0 is unrealistic – MOGREPS-G is exaggerating the uncertainty in global-average pressure. For the pre-operational experiments described in section 4, code was added to the ETKF algorithm to remove globally averaged pressure increments at level one, and propagate the correction hydrostatically to higher levels. This has the desired effect of removing the peak at wavenumber 0.

[¶]Our intention was to use $F = 1$, but due to an error in the original code we effectively ended up using the shorter scale, giving a greater degree of filtering than intended.

[‖]Note that filtering out wavenumber 0 alone ($F = 1/\infty$) is insufficient – a significant amount of imbalance is still created by aliasing from the remaining low wavenumbers.

are at least as large as for the other control variables, and will therefore not lead to over-localization, and secondly because it is dominant in determining the structure of the extratropical analysis increments via the geostrophic relation. (We note that Bishop and Hodyss (2011) also base their localization scheme on ψ correlations). The globally averaged vertical error correlations are taken from $\mathbf{B}_c$, and modified by taking their absolute value, and raising to the power $1/R^2$, where R is a parameter. That is, if c_{avg} is the average ψ correlation between two levels, the corresponding element c_{loc} of the target localization covariance $\mathbf{C}_{v,\text{target}}$ is given by

$$c_{\text{loc}} = \left|c_{\text{avg}}\right|^{\frac{1}{R^2}} . \tag{19}$$

The idea is to replicate the relationship between the underlying and localization correlations if a Gaussian correlation function with length scale L is localized with a Gaussian function with scale $L \times R$. Thus we refer to R as the 'pseudo scale ratio'. For our initial tests, we chose to set R to 2 and to truncate at six vertical modes. The resulting localization covariance $\mathbf{C}$ is shown in Figure 5(a), with the average ψ correlations on which it is based shown in Figure 5(b).

The impact of these two changes on forecast skill is summarized in Figure 2(b), which shows consistent performance gains over the previous trials for both 3D and 4D-Var. In particular, for 4D-Var there are small but consistent improvements in the Extratropics, and the previous poor scores against observations in the Tropics are significantly reduced. We note that, despite the large performance improvement for 3D-Var, there is still a significant performance deficit relative to non-hybrid 4D-Var (not shown).

3.3. *Correcting for insufficient degrees of freedom*

In choosing the weights β_c and β_e in Eq. (9), most other authors have chosen to impose the condition that the fractions used of the climatological and localized ensemble covariance sum to unity; i.e. $\beta_c^2 + \beta_e^2 = 1$. Assuming the climatological and ensemble variances are similar, this preserves the total background error variance relative to the non-hybrid system. However, preserving background error variance does not necessarily preserve the analysis fit to observations. Because localized ensemble covariances tend to be of lower rank than climatological covariances, they provide a reduced ability to fit the full range of observations, particularly in data-dense areas. This effect is illustrated in Figure 6(a), which shows the analysis observation penalty J_o when using $\mathbf{B}_e$ alone, for a selection of horizontal localization scales and ensemble sizes. We included the vertical localization scheme described above, but not the anti-aliasing filter, and so that the results were not complicated by differences in the underlying covariance the ensemble data was obtained by taking random samples from the climatological covariance $\mathbf{B}_c$; i.e. by obtaining increments of the form $\delta\mathbf{w} = \mathbf{B}_c^{1/2}\mathbf{v} = \mathbf{U}\mathbf{v}$, with the elements of $\mathbf{v}$ random samples from a Gaussian distribution with zero mean and unit variance. The horizontal dashed lines show the initial and final penalties for the non-hybrid analysis, so we see that for the ensemble size (23) and localization scale ($L = 1500$ km) chosen for our experiments the ensemble fit to observations is significantly reduced. To get a comparable

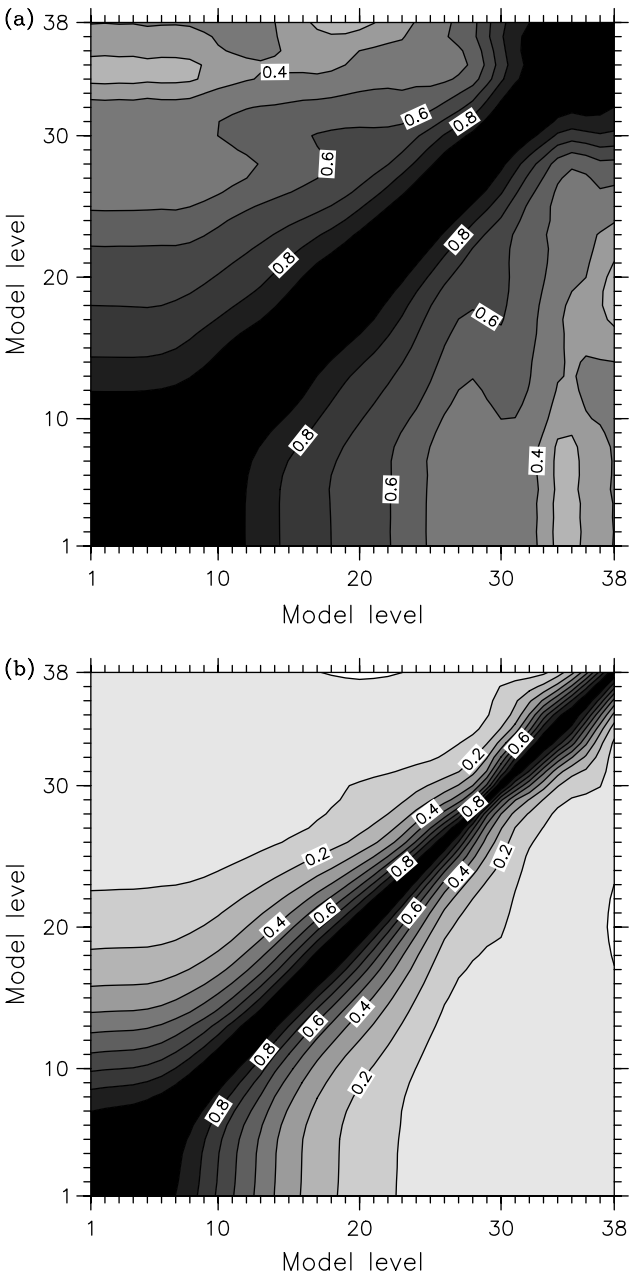

Figure 5. (a) Vertical localization covariance for the 38-level model configuration. (b) Globally averaged vertical correlations from $\mathbf{B}_c$ for stream function, on which the localization covariance is based.

fit, we would have to significantly increase the ensemble size and reduce the localization scale. Switching to a 50%–50% hybrid (Figure 6(b)), the relative underfit of the observations is reduced, but for the chosen settings the analysis J_o is still approximately 8% higher than for the non-hybrid configuration.

As noted by Wang *et al.* (2008b), the analysis fit to the assimilated observations is not a good guide to analysis quality. Nevertheless, for our first operational implementation of the hybrid system we decided that it would be sensible to keep the analysis fit to observations similar, particularly with such a small ensemble size. To achieve this, we decided to fix the ensemble covariance percentage at 50%, and inflate the climatological percentage until the non-hybrid analysis fit to observations was achieved. Based on representative analysis cases, the required percentage was found to be around 80%, implying that

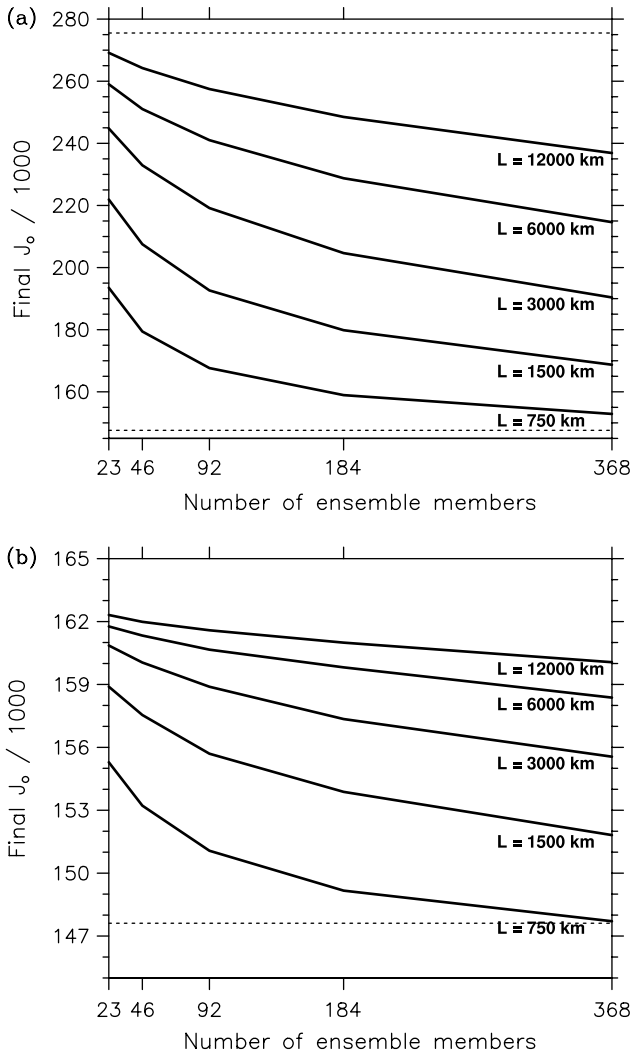

Figure 6. Final (i.e. analysis) observation penalty J_o as a function of the ensemble size and horizontal localization scale L: (a) 100% ensemble covariance alone; (b) 50%–50% hybrid covariance (with a different vertical scale). The horizontal dashed lines are the initial and final penalties for the non-hybrid analysis. (3D-Var experiments, including vertical localization but no anti-alias filtering.)

although $\mathbf{B}_c$ and $\mathbf{B}_e$ have similar variances, 50% of $\mathbf{B}_e$ can compensate for the removal of only around 20% of $\mathbf{B}_c$. This 'inflation' of $\mathbf{B}_e$ is analogous to the covariance inflation used in ensemble systems to counteract filter divergence.

Results from trials of the 80%–50% hybrid scheme are summarized in Figure 2(c). In the 4D-Var trial, the skill scores against observations are relatively unaffected in the Extratropics, but improved in the Tropics. The scores against analyses, however, are generally reduced, particularly in the Tropics. As discussed in section 4.2, however, we were sceptical that these changes to scores against analyses were a true reflection of forecast skill, so on the basis of the improved tropical scores against observations we decided to proceed with the 80%–50% hybrid.

3.4. *Experiments with alternative ensemble modes*

The results of the previous section indicate clear benefits of the hybrid system in the low-resolution test configuration, particularly for 3D-Var. However, it is not immediately clear what aspects of the ensemble covariances lead to better performance – whether it is their average character (e.g. length scales or regionally varying variances) that lead to the improvements, or whether it is the ability to capture 'Errors of the Day' that is most important. To shed light on this, we used the 3D-Var hybrid configuration of the previous section to run two further experiments. In the first, the MOGREPS modes $\mathbf{w}'_k$ were substituted with random samples from the climatological covariance $\mathbf{B}_c$, using a different random number seed for each sample used within the trial. As shown in Figure 7(a), this led to a large overall drop in performance relative to the control non-hybrid system, presumably due to the introduction of sampling error and the artefacts of localization. For the second experiment, we reintroduced the MOGREPS modes, but for each analysis replaced each mode with a randomly selected mode for the same cycle but a different day of the trial, at least 7 days away from the correct time. Thus the average character of the ensemble covariances was retained, but the flow dependence was lost. The results of this experiment (Figure 7(b)) show a more neutral impact relative to the non-hybrid system, but a far worse performance compared to use of modes at the correct time (left-hand plot in Figure 2(c)). We conclude that the ability of the ensemble to capture 'Errors of the Day' is crucial to the performance gains noted above.

3.5. *Tuning of the horizontal localization scale*

The choice of horizontal localization scale ($L = 1500$ km) for the preceding trials was little more than an educated guess, based on a knowledge of the climatological error correlation scales for stream function. Given the characteristics of the unprocessed ensemble error covariances, theoretical guidance exists for determining appropriate localization functions (e.g. Hamill *et al.*, 2001). However, in the current hybrid system, we use a single 'one-size-fits-all' horizontal localization function for all four control variable fields, regardless of horizontal location or model level. Thus a scale that suits a particular variable in one region may not be optimal for other regions or other variables. Inevitably, then, some degree of experimentation is necessary to determine the optimal choice.

To explore the sensitivity to horizontal localization scale, the 50%–50% 3D-Var hybrid configuration of section 3.2 – which included vertical localization – was rerun four times with L values of 900, 1200, 1800 and 2100 km. Verification results (not shown) showed clear benefits from each reduction of the localization scale, except for the reduction from 1200 to 900 km, which had only a marginal further impact. For the pre-operational trials described in the next section, we thus decided to set L to 1200 km.

4. Pre-operational trials

For pre-operational testing, the trial configuration was updated to be as close as possible to that which would likely be implemented operationally. The main details were as follows:

Periods:
17 Dec. 2009–15 Jan. 2010
2 Jun.–1 Jul. 2010
Deterministic forecast grid:
640 × 481 × 70
MOGREPS grid:
432 × 325 × 70

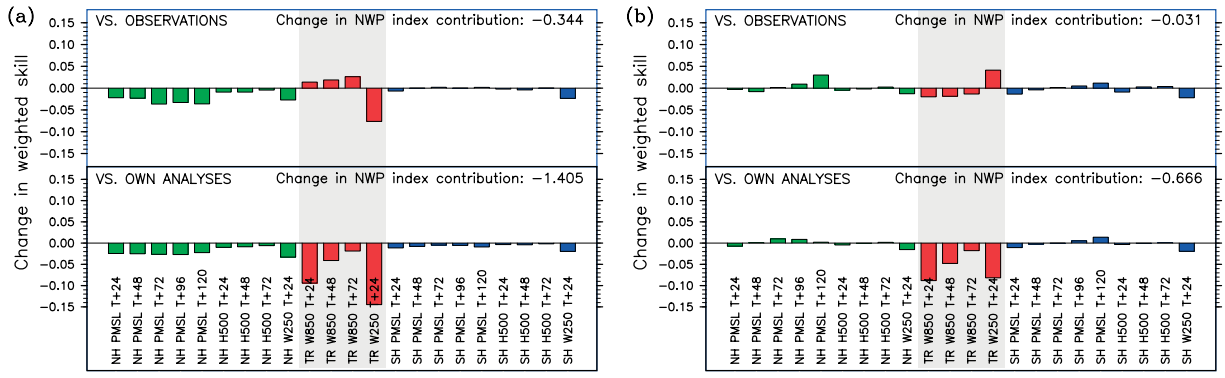

Figure 7. As left-hand plot in Figure 2(c), but with the MOGREPS error modes substituted with (a) random samples for the climatological covariance $\mathbf{B}_c$, and (b) random modes for the same time of day, but the day displaced by at least 7 days. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

VAR grid:
$216 \times 163 \times 70/432 \times 325 \times 70$
Humidity control variable μ:
nonlinear variable of Ingleby *et al.* (2012).
MOGREPS/VAR coupling?
No (Dec./Jan.); yes (Jun.)
Training data for B_c:
ECMWF ensemble of 4D-Vars

The only significant compromise was a lower-resolution configuration of the deterministic forecast model: 640 × 481 × 70 rather than the operational resolution of 1024 × 769 × 70. Apart from the inclusion of a coupled trial for the June period, other noteworthy changes relative to the configuration of the previous section were: (a) the move to a higher-resolution two-stage 4D-Var analysis, in which 30 iterations of a lower-resolution analysis were used to calculate preconditioning vectors for a further 30 iterations at high resolution; (b) replacement of the total relative humidity control variable with a nonlinear variable that has more symmetric statistics for low and high background humidities (Ingleby *et al.* (2012)); and (c) a change to the training data for the climatological covariance, moving from $T + 30$–$T + 6$ forecast differences to forecast data based on ECMWF's 10-member ensemble of 4D-Vars (Fisher, 2003).

The only change to the hybrid configuration developed at low resolution was a relaxation to full climatological covariances in the upper model layers. The reason for this was twofold. Firstly, the 70-level model has a much higher model top: 80 km above mean sea-level, compared to 40 km for the 38-level configuration. With such a major extension – bringing in most of the mesosphere – there was a risk that the characteristics of the upper-level ensemble covariances would prove problematic, particularly considering the lack of observations to constrain the ensemble at such heights. Secondly, above the tropopause horizontal correlation scales increase significantly with height (Ingleby, 2001), so the scheme's 'one-size-fits-all' localization scale would likely be significantly shorter than appropriate in the upper model levels. To reduce the risk of these issues degrading performance, we allowed the weighting factors β_c and β_e to vary with model level, and implemented a smooth relaxation to $\beta_c^2 = 1.0$ / $\beta_e^2 = 0.0$ between 16 and 21 km above mean sea level. The effective vertical localization covariance for the 70-level configuration is shown in Figure 8.

In summary, the hybrid settings for the pre-operational trials were as follows:

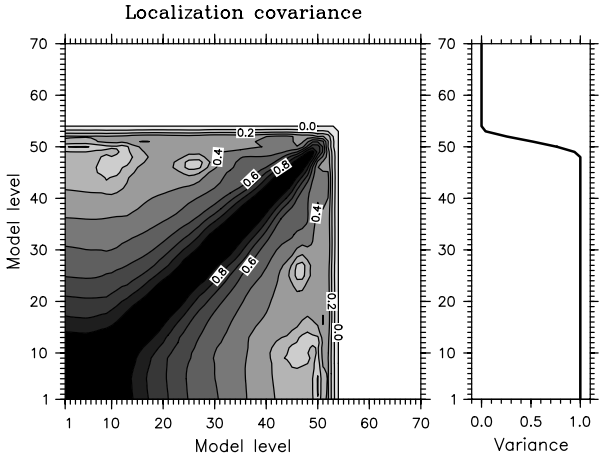

Figure 8. Vertical localization covariance (left) and variance (right) used with the 70-level model configuration.

Horizontal loc. scale:
$L = 1200$ km
Horizontal filtering:
$F = \sqrt{2}$
Vertical loc. scheme
as described in section 3.2
(truncation at 8 modes)
β_c^2:β_e^2
0.8:0.5 below 16 km
1.0:0.0 above 21 km

4.1. *Single-observation tests*

As an illustration of the characteristics of the hybrid covariances, Figure 9 shows results from the assimilation of a single zonal wind observation (black dot) placed in a frontal region off the west coast of North America, on a model level close to 500 hPa. When the observation is placed at the beginning of the 6 h 4D-Var window, the non-hybrid analysis increment to zonal wind at the same time and model level is insensitive to the underlying front. In the hybrid system, however, the inclusion of the ensemble covariance leads to some stretching of the increment along the front, which is what would be expected if there is uncertainty in its position. Thus in the hybrid system we see flow dependence right from the beginning of the analysis window, though some of the detail will certainly be due to sampling error.

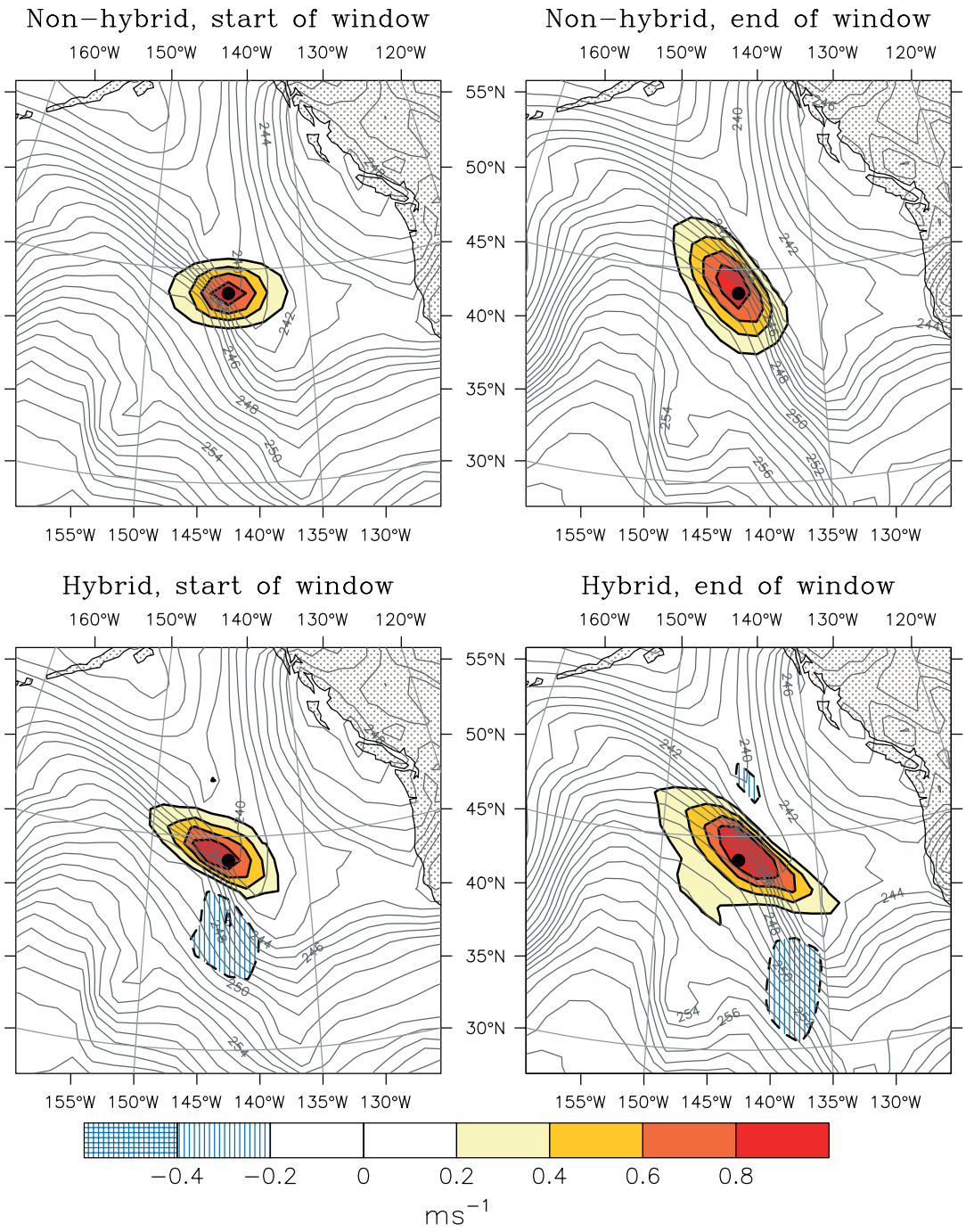

Figure 9. Zonal wind responses (filled thick contours, with negative contours dashed) to a single zonal wind observation at the start (left-hand plots) and end (right-hand plots) of the 6 h 4D-Var window. The plots are for the same time and model level (≈ 500 hPa) as the observation. Upper plots are for the non-hybrid configuration; lower plots for the hybrid configuration used within the pre-operational trials. The observation location is marked with a black dot at the centre of each plot. The unfilled contours show the background temperature field. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

When the observation is placed at the end of the window, the implicit time propagation of the initial covariances by the PF model leads to more realistic-looking structures in the non-hybrid analysis, but the basic characteristics of the hybrid analysis are relatively unchanged.

4.2. *Trial verification*

The changes to weighted-skill scores for the two trial periods are shown in Figure 10. For verification against observations, we see consistent improvements over the non-hybrid system, particularly for the June period. The contributions to the NWP index are significant: +0.606 for Dec./Jan. and +0.793 for June – not far short of what is typically achieved in a whole year of global NWP development at the Met Office. In the Extratropics, the scores against own analyses are generally improved, the only significant exception being 250 hPa winds at $T + 24$. However, the tropical wind skill scores are much worse compared to the non-hybrid system. In particular, for the winter period the tropical scores against own analyses are enough to produce an overall negative impact on the NWP index.

For evaluating changes to the data assimilation system, verification against the analyses produced by the trial itself

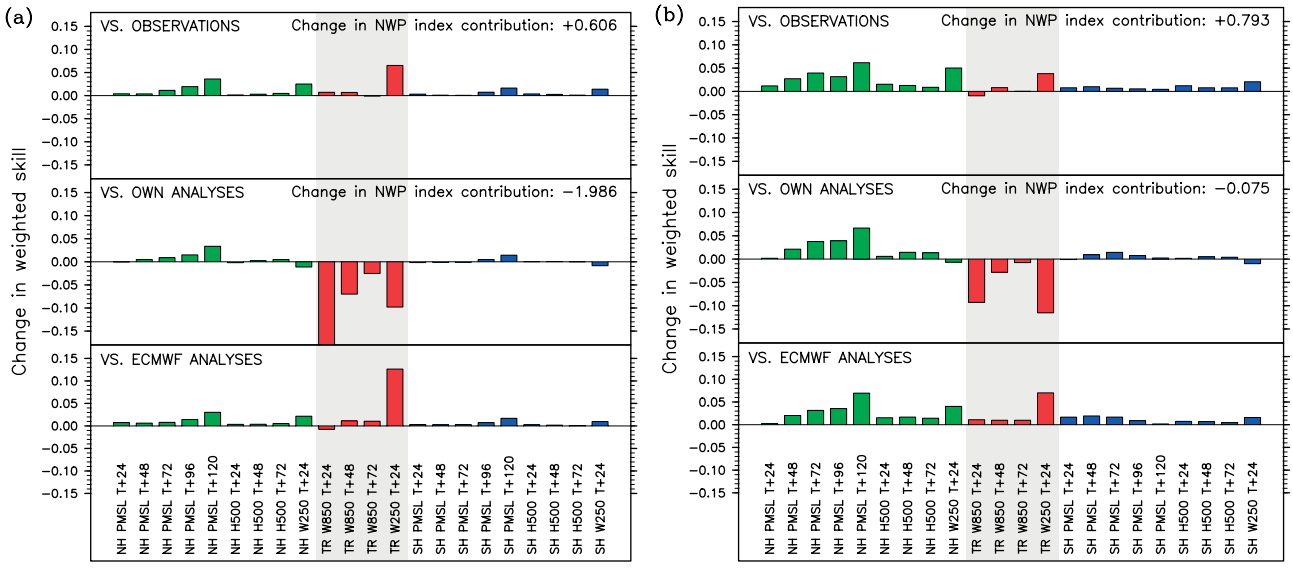

Figure 10. Changes to weighted skill scores *versus* non-hybrid controls for (a) the December 2009–January 2010 period, and (b) the June 2010 period. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

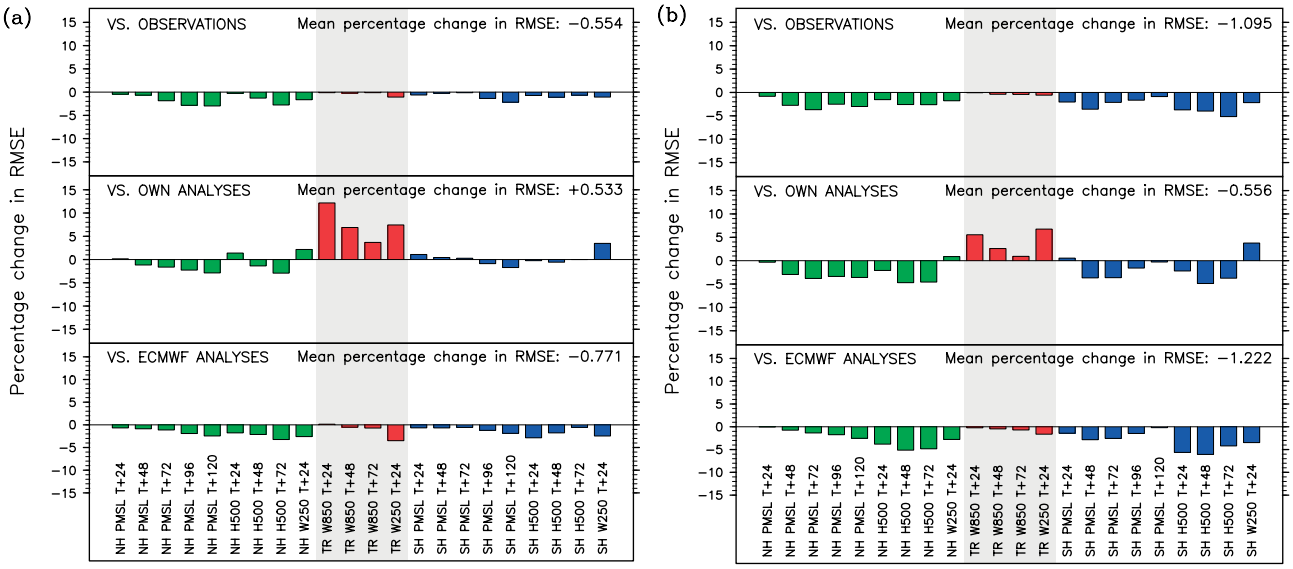

Figure 11. As Figure 10, but showing percentage changes to RMS error rather than changes to weighted skill scores. This figure is available in colour online at wileyonlinelibrary.com/journal/qj

has long been recognized as problematic. To take an extreme case, if data assimilation added no increments at all, skill scores against own analyses would be perfect, despite the forecasts eventually becoming useless. More generally, it is often found that a data assimilation change that decreases the average size of the analysis increments leads to better skill scores against analyses, even when the corresponding scores against observations are degraded. The reason is most easily seen for the verification of the background forecasts, whose errors *are* the analysis increments, but the effect is seen at longer forecast ranges too, albeit with diminishing impact. The effect is generally strongest in the Tropics, where analysis increments tend to persist more strongly through the forecast.

Compared to the $\mathbf{B}_c$ used for the low-resolution trials, the $\mathbf{B}_c$ used here – based on data from the ECMWF ensemble of analyses – has significantly lower variances and correlation scales in the Tropics (not shown). Introduction of the ensemble covariances through the hybrid system partially reverses this change, leading to larger analysis increments in the Tropics compared with the non-hybrid system. We suspect this is a key factor in the poorer tropical scores against own analyses. One way to avoid this issue is to verify against an independent set of analyses. In Figure 10, we have included skill scores measured against operational analyses from the ECMWF 4D-Var system. We see that the poor tropical scores (and the skill reductions for $T + 24$ 250 hPa winds in the Extratropics) are replaced by neutral or positive scores, giving a much closer correspondence with the score changes against observations. This was sufficient to persuade us that the poor tropical scores are indeed an artefact of the verification measure.

Figure 11 shows changes to RMS error scores for the fields used in the NWP index. The main difference from the skill scores presented in Figure 10 is the lack of weighting and the removal of the influence of persistence scores. Apart

Table 2. Number of better/neutral/worse RMS error scores for the two pre-operational trial periods. 'Neutral' is defined as a change of less than 2%.

	vs. observations			vs. own analyses			vs. ECMWF analyses		
	NH	Tropics	SH	NH	Tropics	SH	NH	Tropics	SH
Dec./Jan.	29/94/0	6/117/0	12/109/2	16/ 91/16	7/69/47	3/106/14	35/79/0	39/75/0	14/100/0
June	34/89/0	9/114/0	46/74/3	49/63/11	9/86/28	18/82/23	63/51/0	29/85/0	47/65/2

from one tropical field, RMS errors against observations and ECMWF analyses are all reduced, with an average reduction of 0.9% across the two trial periods.

A broader summary of the impact of the hybrid scheme is presented in Table 2, which shows the number of 'better' or 'worse' RMS error changes for a wider range of fields – including temperature and relative humidity – and a wider range of pressure levels and forecast ranges. Here, 'better'/'worse' corresponds to a decrease/increase in RMS error of at least 2%. For verification against observations or ECMWF analyses, there are far more better scores than worse. Similarly, for the fit of the background $(T+6)$ forecasts to the categories of observations used within the analysis – e.g. radiances from selected satellite channels; radiosonde temperature observations at selected levels – there were far more better scores than worse (not shown), with an overall reduction of approximately 0.5% in the initial total observation penalty J_o.

One notable feature of the verification scores is the larger relative impact of the hybrid system for the June 2010 period. One of the differences between this and the December/January period was that the hybrid system was run fully coupled to the ensemble; i.e. with the analysis ensemble centred around the hybrid 4D-Var analysis, and the resulting ensemble forecasts being fed back into the two subsequent 4D-Var analyses. However, a rerun with the coupling removed gave very similar results (not shown), so this appears not to have been a significant factor in the difference. Currently, the cause of the seasonal difference is unknown.

As an aside, we note that this apparent lack of sensitivity to coupling has important practical implications. It means that in testing future 4D-Var developments it will normally be sufficient to use stored ensemble data, with no necessity to rerun the ensemble in conjunction with each 4D-Var trial. (Conversely, coupling has been found to have negligible impact on the characteristics of the ensemble.)

4.3. *Vertical mode smoothing*

Putting aside the issues of verifying against own analyses, the hybrid system brings consistent improvements across a broad range of fields. One exception was a slight degradation in the verification scores against screen-level temperature observations. On investigation, we found a significant increase in lower-level temperature variances when moving to the hybrid system, which we traced to the impact of vertical localization. In the Met Office system, ensemble temperature covariances are not localized directly – their modification under localization is a consequence of their relationship to the variables that *are* directly localized. Increments to temperature T' are derived from the linearized form of the hydrostatic relation, which when combined with the

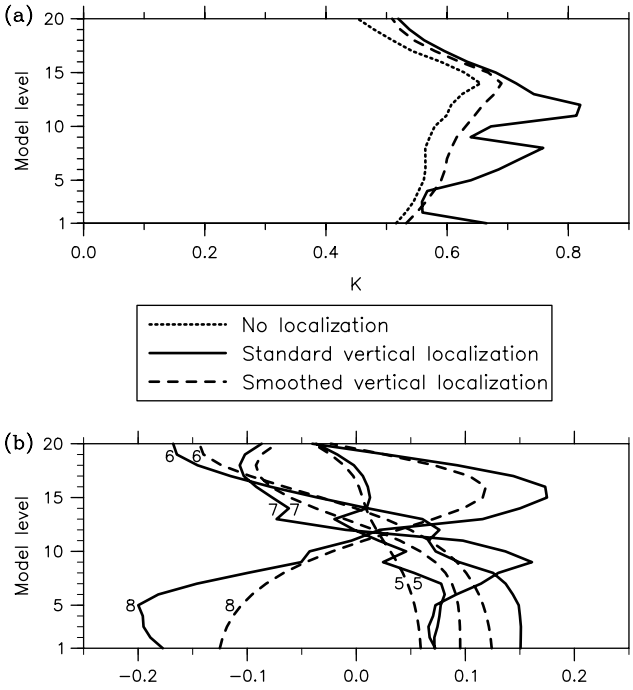

Figure 12. (a) The effect of vertical localization on implied average temperature standard deviations for model levels 20 (≈700 hPa) and below. The dotted and solid lines show the variances before and after application of vertical localization. The dashed line shows the standard deviations after smoothing of the localization covariance. (b) The effect of smoothing on the last four of the eight vertical modes defining the vertical localization. The solid lines show the standard modes, and the dashed lines the modes for the smoothed localization covariance.

equation of state $p = \rho RT$ can be written as

$$T' = \frac{RT^2}{g\rho}\frac{\partial p'}{\partial z} + \frac{p'}{p}, \tag{20}$$

where the unprimed variables correspond to the linearization state, and z is the height above the model surface. The second term on the right-hand side is much smaller than the first, so to a good approximation increments to temperature are determined by the vertical gradient of the pressure increments p'. These are composed of a geostrophically unbalanced part p'_u and a geostrophic part p'_g, which through the 'local balance equation' $p' = \rho f \psi'$ (Gauthier *et al.*, 1999) can be seen to be approximately proportional to the stream function increment ψ'. p'_u and ψ' are localized together, so to a good approximation the effect of localization on p' is equivalent to the direct application of the localization covariance. We then see that any sharp vertical gradients in the localization covariance will induce spurious increments in T'.

The dotted line in Figure 12(a) shows the globally averaged temperature standard deviation as a function of model level up to level 20 (≈700 hPa) for 50 random samples from the climatological covariance $\mathbf{B}_c$ – enough samples

for the effects of sampling error to be insignificant. After application of the chosen vertical localization scheme (solid line), the standard deviation increases due to the addition of spurious increments generated from vertical gradients in the localization covariance. We see that the bottom model level is one of those most badly affected, but there are similar problems higher up. To reduce the impact of this problem, we decided to apply some vertical smoothing to the vertical mode matrix $\mathbf{U}_v^{\alpha}$ by applying a Gaussian filter with a length scale L of 500 m. The impact of this smoothing on the standard deviations is shown by the dashed line. We see that the spurious temperature variance has been reduced considerably. The smoothing mostly affects the last four of the eight vertical modes, shown in Figure 12(b).

To test the impact of this smoothing, the first 13 days of the December/January trial were rerun with the smoothing included. The result was an approximate reversal of the poor skill scores against screen-level temperature observations, so the smoothing was accepted as a beneficial change to the system.

5. Discussion and further work

The encouraging performance of the hybrid system in the pre-operational trials led to its inclusion in a package of global model changes that was subjected to further testing, and made operational at the Met Office on 20 July 2011 – the first implementation at a major NWP centre of a hybrid ensemble/variational scheme making use of the complete ensemble covariance structures.

The use of ensemble data within the deterministic forecasting system has added an extra focus to the development of MOGREPS. As well as estimating the local properties of forecast errors out to the medium range, the ensemble is now required to provide realistic estimates of error *covariances* at the short forecast ranges required by hybrid 4D-Var. Thus changes that have a positive impact on traditional measures of ensemble performance now need additional testing for their impact on deterministic forecasts through the hybrid 4D-Var system, with all the additional resources that entails. However, to the extent that improvements in traditional measures pass through to very short-range covariance estimates, this dual use of the ensemble is a good thing, extracting additional benefit from the resources assigned to it.

Since the original implementation of the hybrid system described here, MOGREPS-G has been switched operationally from a 12 h to a 6 h cycle, motivated mainly by the boundary data requirements of a new UK ensemble, but also allowing hybrid 4D-Var to use $T+3$ rather than $T+9$ forecast data for the 0000 and 1200 UTC analyses; i.e. data at the 'correct' forecast range. This change entailed a number of adjustments to the ETKF algorithm – particularly the inflation scheme – and a reduction in the number of perturbed members from 23 to 22. However, despite these rather substantial changes to MOGREPS-G, the overall impact on deterministic forecast quality via hybrid 4D-Var was found to be approximately neutral.

Compared to ensembles at other operational NWP centres, which typically have O(100) members, the Met Office's current 22-member ensemble is rather small. We plan to remedy this by increasing the ensemble size within the cycling part of MOGREPS. Traditional ensemble products require a relatively small number of members, so in the 6 h cycling system the additional members will only be required to run to $T+9$ – the end of the following observation window. As larger ensembles become available to the hybrid system, it should become beneficial to put more weight on the ensemble covariance, and reduce the degree of localization.

Within the hybrid system itself, the main issue is how to improve the current rather simple localization scheme, which consists of separate static spatial localizations in the horizontal and vertical. Our scheme already has an improvement to basic localization, applying it to transformed variables. Improvements to the variational parameter transform such as Ingleby *et al.* (2012) improve the definition of the balance not damaged by our localization. We mentioned in section 2.7 that inter-variable localization was not beneficial in the current system – this decision may be revisited if we extend the scheme to the assimilation of constituents such as ozone. Buehner (2012) demonstrated benefits from combined spatial/spectral localization – this has been included in our system but not tested in time for the experiments described here. Eventually, it is likely that some form of adaptive localization (e.g. Bishop and Hodyss, 2011) will be desirable.

Another issue is the effect of localization on the dynamical properties of the ensemble error modes. We saw in section 3.2 that it was necessary to introduce an anti-aliasing filter to reduce imbalances caused by horizontal localization. As noted in section 2.6, however, there remains an issue with the vertical localization scheme, which tends to upset the approximate balance between convergence and divergence into a model column. We intend to address this shortly.

Finally, we note that the implementation of the hybrid system is only the first step in a path towards greater synergy between ensemble and deterministic forecasting at the Met Office. In particular, we are currently developing a '4D-Ensemble-Var'** system (Liu *et al.*, 2008; Buehner *et al.*, 2010), which generalizes the hybrid VAR code to make use of ensemble data spread throughout the data assimilation window. By eliminating the need for a linear 'perturbation forecast' model, this system should be much more scalable on massively parallel computer architectures, and should be easily extendable to generate its own analysis ensemble, creating a unified deterministic/ensemble forecasting system. Lessons learned from hybrid 4D-Var – particularly its localization scheme – should be of direct value to the development of this system.

Acknowledgements

We would like to thank the large number of scientific and technical staff involved in the development and implementation of the hybrid system, in particular Neill Bowler, Peter Jermey, Rick Rawlins and Mike Thurlow. Neill Bowler also provided significant input into the development of the manuscript. We are also grateful to the two anonymous reviewers, whose comments helped us improve several important aspects of the presentation.

**The system is called 'En4DVar' by Liu *et al.* (2008) and 'En-4D-Var' by Buehner *et al.* (2010). We prefer the name '4D-Ensemble-Var' because the key feature is the 4-dimensional use of the ensemble. It is also more consistent with the '4DEnKF' terminology of Hunt *et al.* (2004).

References

Andersson E, Järvinen H. 1999. Variational quality control. *Q. J. R. Meteorol. Soc.* **125**: 697–722.

Bannister RN. 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Q. J. R. Meteorol. Soc.* **134**: 1971–1996.

Barker DM. 1999. The use of synoptically-dependent error structures in 3DVAR. VAR Scientific Documentation Paper No. 25, Met Office, Exeter, UK.

Barker DM, Huang W, Guo YR, Bourgeois AJ, Xiao QN. 2004. A three-dimensional variational data assimilation system for MM5: implementation and initial results. *Mon. Weather Rev.* **132**: 897–914.

Bishop CH, Hodyss D. 2011. Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Weather Rev.* **139**: 1241–1255.

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Weather Rev.* **129**: 420–436.

Bonavita M, Isaksen L, Holm E. 2012. On the use of EDA background error variances in the ECMWF 4D-Var. ECMWF Tech. Memo. 664, ECMWF, Reading, UK.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 703–722.

Bowler NE, Arribas A, Beare SE, Mylne KR, Shutts GJ. 2009. The local ETKF and SKEB: upgrades to the MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **135**: 767–776.

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**: 1013–1043.

Buehner M. 2012. Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon. Weather Rev.* **140**: 617–636.

Buehner M, Houtekamer PL, Charette C, Mitchell HL, He B. 2010. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Weather Rev.* **138**: 1550–1566.

Campbell WF, Bishop CH, Hodyss D. 2010. Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Weather Rev.* **138**: 282–290.

Courtier P, Thépaut JN, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1387.

Cullen MJP. 1993. The unified forecast/climate model. *Meteorol. Mag.* **122**: 81–94.

Etherton BJ, Bishop CH. 2004. Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. *Mon. Weather Rev.* **132**: 1065–1080.

Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**: 10143–10162.

Fisher M. 2003. 'Background error covariance modelling'. In *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean*, 8–12 September 2003. ECMWF: Reading, UK; 45–64.

Flowerdew J, Bowler NE. 2012. Online calibration of the vertical distribution of ensemble spread. *Q. J. R. Meteorol. Soc.*, DOI: 10.1002/qj.2072.

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

Gauthier P, Thépaut JN. 2001. Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Weather Rev.* **129**: 2089–2102.

Gauthier P, Charette C, Fillion L, Koclas P, Laroche S. 1999. Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos. Ocean* **37**: 103–156.

Hamill TM, Snyder C. 2000. A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Weather Rev.* **128**: 2905–2919.

Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **129**: 2776–2790.

Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.

Hunt BR, Kalnay E, Kostelich EJ, Ott E, Patil DJ, Sauer T, Szunyogh I, Yorke JA, Zimin AV. 2004. Four-dimensional ensemble Kalman filtering. *Tellus A* **56**: 273–277.

Ingleby NB. 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.

Ingleby N, Lorenc A. 1993. Bayesian quality control using multivariate normal distributions. *Q. J. R. Meteorol. Soc.* **119**: 1195–1225.

Ingleby B, Lorenc AC, Ngan K, Rawlins F, Jackson DR. 2012. Improved variational analyses using a nonlinear humidity control variable. *Q. J. R. Meteorol. Soc.*, DOI: 10.1002/qj.2073 (in press).

Liu C, Xiao Q, Wang B. 2008. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Weather Rev.* **136**: 3363–3373.

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

Lorenc AC, Hammon O. 1988. Objective quality control of observations using Bayesian methods: theory, and a practical implementation. *Q. J. R. Meteorol. Soc.* **114**: 515–543.

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **126**: 2991–3012.

Montmerle T, Berre L. 2010. Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Q. J. R. Meteorol. Soc.* **136**: 1408–1420.

Parrett C. 1992. 'Background errors for the quality control and assimilation of atmospheric observations in the Unified Model: the situation in July 1992'. Met Office Short Range Forecasting Tech. Rept. 22, Met Office, Exeter, UK.

Pavelin EG, English SJ, Eyre JR. 2008. The assimilation of cloud-affected infrared satellite radiances for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **134**: 737–749.

Purser RJ, Wu WS, Parrish DF, Roberts NM. 2003. Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. *Mon. Weather Rev.* **131**: 1536–1548.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.

Raynaud L, Berre L, Desroziers G. 2009. Objective filtering of ensemble-based background-error variances. *Q. J. R. Meteorol. Soc.* **135**: 1177–1199.

Tennant W, Shutts G, Arribas A, Thompson S. 2011. Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Weather Rev.* **139**: 1190–1206.

Wang X, Bishop CH, Julier SJ. 2004. Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Weather Rev.* **132**: 1590–1605.

Wang X, Snyder C, Hamill TM. 2007. On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. *Mon. Weather Rev.* **135**: 222–227.

Wang X, Barker DM, Snyder C, Hamill TM. 2008a. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **136**: 5116–5131.

Wang X, Barker DM, Snyder C, Hamill TM. 2008b. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: Real observation experiments. *Mon. Weather Rev.* **136**: 5132–5147.

Zhang M, Zhang F. 2012. E4DVar: coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model. *Mon. Weather Rev.* **140**: 587–600.