QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
Q. J. R. Meteorol. Soc. **134**: 1951–1970 (2008)
Published online 24 November 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.339

RMetS
Royal Meteorological Society

Review
A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances

R. N. Bannister*
Data Assimilation Research Centre, University of Reading, UK

ABSTRACT: This article reviews the characteristics of forecast error statistics in meteorological data assimilation from the substantial literature on this subject. It is shown how forecast error statistics appear in the data assimilation problem through the background error covariance matrix, **B**. The mathematical and physical properties of the covariances are surveyed in relation to a number of leading systems that are in use for operational weather forecasting. Different studies emphasize different aspects of **B**, and the known ways that **B** can impact the assimilation are brought together.
Treating **B** practically in data assimilation is problematic. One such problem is in the numerical measurement of **B**, and five calibration methods are reviewed, including analysis of innovations, analysis of forecast differences and ensemble methods. Another problem is the prohibitive size of **B**. This needs special treatment in data assimilation, and is covered in a companion article (Part II).
Examples are drawn from the literature that show the univariate and multivariate structure of the **B**-matrix, in terms of variances and correlations, which are interpreted in terms of the properties of the atmosphere. The need for an accurate quantification of forecast error statistics is emphasized.

KEY WORDS background error statistics; balance; calibration; multivariate; separability

Received 12 February 2008; Revised 22 August 2008; Accepted 2 October 2008

1. Introduction

Variational data assimilation (VAR) is the method of choice in many numerical weather prediction centres to estimate the state of the atmosphere for weather prediction (Rabier, 2005). VAR can assimilate efficiently direct observations of meteorological fields from *in situ* instruments on sondes, aircraft and in weather stations, and indirect observations from satellites and from ground-based radar (Schlatter, 2000). Despite this wealth of observational information available, VAR needs to use an a priori (or background) state. This is of great importance to data assimilation problem as it provides information otherwise missing from observations, and provides a realistic reference state needed to form the nonlinear observation operators used to assimilate many of the indirect observations.

As with all information, the background state is prone to error, and VAR must account for this. This is most conveniently achieved through the so-called background error covariance statistics, which are represented by the matrix **B**. These days, the background state is derived from a short numerical forecast, and so background errors share properties with forecast errors. Here, background errors are referred to specifically as forecast errors, to distinguish them from errors in other types of background state used historically in data assimilation (e.g. climatology). Accurate knowledge of forecast error statistics is considered to be extremely important to the success of the assimilation process, and so they have received much attention over the last decade or so. Consequently, there is now a substantial literature on forecast error statistics, covering their nature and the subtle ways that they affect the assimilation, the way that they are measured, and how they can be modelled for use in an operational setting.

There are now many leading centres around the world that use VAR for weather forecasting, and there are often differences in the way that forecast error statistics are measured, described and used by each. Despite this, forecast error statistics can possess interesting properties that are common to all centres, but the methods that are used to describe them can also have some common difficulties. It is therefore timely to review the state of research in this area. This article and its companion (Bannister, 2008, hereafter referred to as Part II) concern the nature of forecast error covariance statistics and the special ways that VAR handles them. Part I has the following structure. In section 2, we introduce the forecast error covariance matrix in the cost function, define its structure and describe some of its properties. In section 3,

**Correspondence to: R. N. Bannister Department of Meteorology, University of Reading, Earley Gate, Whiteknights, Reading, Berkshire, RG6 6BB, UK. E-mail: r.n.bannister@reading.ac.uk

we show how this matrix is important in data assimilation. In section 4, we highlight the leading practical problems when dealing with **B**. In section 5, we review the leading methods of measuring the matrix. In section 6, we show what forecast error statistics typically look like, and in section 7 we give some concluding remarks. In Part II, we show the techniques that are used to model them in VAR.

2. Forecast error statistics in variational data assimilation

2.1. Variational assimilation

The objective of VAR is to minimize a cost function, shown here in the incremental formulation (Courtier, Thépaut and Hollingsworth, 1994), and using the Ide *et al.* (1997) notation where possible:

$$J[\delta\mathbf{x}, \mathbf{x}^{\mathrm{g}}] = \frac{1}{2}\left(\delta\mathbf{x} - \delta\mathbf{x}^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\delta\mathbf{x} - \delta\mathbf{x}^{\mathrm{b}}\right) + \frac{1}{2}\left\{\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{g}} + \delta\mathbf{x})\right\}^{\mathrm{T}}\mathbf{R}^{-1} \times \left\{\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{g}} + \delta\mathbf{x})\right\}. \quad (1)$$

States denoted with $\mathbf{x}$ (details below) each comprise three-dimensional fields describing the meteorological state (e.g. winds, pressure, temperature, humidity, etc.) and each state consists of n elements in total. The cost, J, is minimized with respect to the increment $\delta\mathbf{x}$. At the minimum, $\delta\mathbf{x}$ describes the analysis, $\mathbf{x}^{\mathrm{a}}$, specified with respect to a reference (or 'guess') state, $\mathbf{x}^{\mathrm{g}}$, i.e.

$$\mathbf{x}^{\mathrm{a}} = \mathbf{x}^{\mathrm{g}} + \delta\mathbf{x}. \quad (2)$$

Similarly $\delta\mathbf{x}^{\mathrm{b}}$ is the incremental description of the background, $\mathbf{x}^{\mathrm{b}} = \mathbf{x}^{\mathrm{g}} + \delta\mathbf{x}^{\mathrm{b}}$ (often the reference state is the background state and so $\delta\mathbf{x}^{\mathrm{b}} = 0$). In (1), $\mathbf{y}^{\mathrm{o}}$ are the observations, H is the observation operator (computing the model's version of the observations from $\mathbf{x}$), $\mathbf{B}$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix. In this formulation, $\mathbf{x}^{\mathrm{b}}$, $\mathbf{x}^{\mathrm{g}}$, $\delta\mathbf{x}$, $\delta\mathbf{x}^{\mathrm{b}}$ and $\mathbf{B}$ are valid at $t = 0$. By incorporating in H a forecast model step (e.g. $H \rightarrow H_t M_{t\leftarrow 0}$ for observations at time t where $M_{t\leftarrow 0}$ is the forecast step), as done in, for example, Johnson, Hoskins and Nichols (2005) and using observations distributed over a time window $t > 0$, (1) is a four-dimensional cost function (4d-VAR). Otherwise, (1) is a three-dimensional cost function (3d-VAR). Often extra terms are appended to the cost function, for instance to allow for the correction of systematic errors in the observation processing (e.g. Dee, 2004) or in the case of 4d-VAR, to correct for forecast model error (e.g. Zupanski, 1997). The topic of this article is the background error covariance matrix, which is described in section 2.2.

2.2. Definition and structure of the **B**-matrix

The background error covariance matrix describes the probability distribution function (PDF) of forecast errors. The PDF is taken to be a simple Gaussian-shaped distribution (see section 2.3), as a more general description of the PDF is unfeasible for the large n used in current operational systems ($n \sim 10^7$–10^8 elements).

In the Kalman filter equations (Kalman, 1960), forecast error covariances are denoted by $\mathbf{P}^{\mathrm{f}}$. These change from cycle to cycle, and are denoted by

$$\mathbf{P}^{\mathrm{f}} = \mathbf{M}\mathbf{P}^{\mathrm{a}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}, \quad (3)$$

where $\mathbf{P}^{\mathrm{a}}$ is the error covariance matrix that emerges from the analysis at the previous cycle, $\mathbf{M}$ is the tangent linearization of forecast model that produced the forecast (typically over a 6 or 12 h period) and $\mathbf{Q}$ is the model error covariance matrix. If $\mathbf{P}^{\mathrm{f}}$ were known, then that would be the optimal choice of matrix to describe forecast error covariances in VAR. Unfortunately, as $\mathbf{P}^{\mathrm{f}}$ is an $n \times n$ matrix, the operation count needed to solve the Kalman filter equations, including (3), is prohibitive. For this reason, the background error covariance matrix used in VAR is not $\mathbf{P}^{\mathrm{f}}$, but is the approximation, denoted by $\mathbf{B}$ in (1). This is not fully recomputed on a synoptic basis, but is hoped to contain some properties that $\mathbf{P}^{\mathrm{f}}$ has on average.

In (1), $\mathbf{B}$ is based on the following definition

$$\mathbf{B} = \langle\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\rangle. \quad (4)$$

The angled brackets denote an average over a suitable population of forecast errors, $\boldsymbol{\eta} = \mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}$, where $\mathbf{x}^{\mathrm{t}}$ is the true state of the atmosphere. Equation (4) requires that errors in the forecast are unbiased (i.e. $\langle\boldsymbol{\eta}\rangle = 0$), which we assume is the case. There are further difficulties in evaluating (4). First, the truth is unknown, and secondly $\mathbf{B}$ remains a matrix that is too large to treat explicitly. The first difficulty is important when trying to estimate $\mathbf{B}$, and is discussed in section 4. The second problem is dealt with using the covariance modelling techniques shown in Part II. Even though in the approximation $\mathbf{B}$ is static, the covariance modelling sometimes introduces quasi-flow dependence to the assimilation.

The basic structure of $\mathbf{B}$ is shown in Figure 1 for analysis variables u, v, p and θ (zonal wind, meridional wind, pressure and potential temperature, respectively). The element at row i and column j, B_{ij}, represents the covariance between errors in elements i and j of the background. Elements i and j of the model state vector each represent a particular field at a particular position (e.g. latitude, longitude and model level in a grid-point representation or total wave number, meridional wave number and model level in a global spectral representation). From (4), the individual element B_{ij} is, generally, $B_{ij} = \langle\eta_i\eta_j\rangle$ (e.g. element 'a' in Figure 1). For $i = j$, the element is on the diagonal of $\mathbf{B}$ and, in this special case, the element is called a variance (e.g. element 'b'). The square root of a variance is called the standard deviation. If i and j represent the same field, but are at different positions, the element is an autocovariance (e.g. element 'c'), which comprises a submatrix called an autocovariance matrix. These are the block-diagonal submatrices that are shaded in Figure 1. Off-diagonal submatrices represent cross or multivariate covariances.

DOI: 10.1002/qj

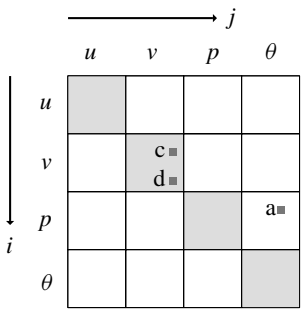

Figure 1. Basic structure of the **B**-matrix for a system with four variables: zonal wind (u), meridional wind (v), pressure (p) and potential temperature (θ). Each variable is a discrete three-dimensional field whose values at each position are represented as a vector and whose covariances are represented as a submatrix in the above. Submatrices that are the autocovariances of a single field between pairs of positions in space are the shaded block diagonal matrices, and submatrices that are the cross covariances between different variables and between pairs of positions in space (multivariate covariances) are unshaded. Example individual elements (dark shaded squares labelled a–c) are discussed in the text.

2.3. How the **B**-matrix enters the cost function

The matrix $\mathbf{B}$ describes an approximate PDF, P^{f}, of average forecast error. This PDF is the probability that the state $\mathbf{x}$ is the truth, $\mathbf{x}^{\mathrm{t}}$, in the absence of new observational information (Lorenc, 1986). By assuming that $\mathbf{x}^{\mathrm{b}}$ is the mean of a Gaussian distribution, then P^{f} is

$$P^{\mathrm{f}}(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})\right\}. \quad (5)$$

In this multidimensional Gaussian, $\mathbf{B}$ takes an analogous role to the simple variance that appears in the single variable form of the Gaussian (e.g. Barlow, 1989). In the Bayesian framework, P^{f} is known as the prior distribution as it describes probability before observations are considered. In this framework, observations are considered by introducing the following conditional PDF

$$P^{\mathrm{o}}(\mathbf{y}^{\mathrm{o}}|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left\{\mathbf{y}^{\mathrm{o}}-H(\mathbf{x})\right\}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}^{\mathrm{o}}-H(\mathbf{x})\right\}\right], \quad (6)$$

which is read as the probability that the observations are those measured, $\mathbf{y}^{\mathrm{o}}$, conditioned on the state of the atmosphere being $\mathbf{x}$ (and assuming that the forward model, H, is exact). Terms (5) and (6) appear in the Bayes rule, which gives the posterior distribution $P^{\mathrm{a}}(\mathbf{x}|\mathbf{y}^{\mathrm{o}}) \propto P^{\mathrm{f}}(\mathbf{x})P^{\mathrm{o}}(\mathbf{y}^{\mathrm{o}}|\mathbf{x})$ (Kalnay, 2003). The probability $P^{\mathrm{a}}(\mathbf{x}|\mathbf{y}^{\mathrm{o}})$ is the PDF of the analysis, which is the probability that $\mathbf{x}$ is the true state given the a priori and the new observational information. Substituting (5) and (6) into the Bayes rule gives

$$P^{\mathrm{a}}(\mathbf{x}|\mathbf{y}^{\mathrm{o}}) \propto \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^{\mathrm{b}}) - \frac{1}{2} \times \left\{\mathbf{y}^{\mathrm{o}}-H(\mathbf{x})\right\}^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}^{\mathrm{o}}-H(\mathbf{x})\right\}\right]. \quad (7)$$

The variational cost function is defined as $J = -\ln P^{\mathrm{a}}(\mathbf{x}|\mathbf{y}^{\mathrm{o}})$, which when applied to (7) and written in incremental form with respect to a reference state $\mathbf{x}^{\mathrm{g}}$ results in (1). The most likely state (the maximum probability) is the minimum 'cost' defined in this way where $\mathbf{x} = \mathbf{x}^{\mathrm{a}}$. This simple analysis gives the useful interpretation that $\mathbf{B}$ defines the shape of the prior PDF.

Note that the Bayesian optimal estimate may be more formally found by defining the 'posterior expected loss', $\rho(\tilde{\mathbf{x}}|\mathbf{y}^{\mathrm{o}})$

$$\rho(\tilde{\mathbf{x}}|\mathbf{y}^{\mathrm{o}}) = \int L(\mathbf{x}-\tilde{\mathbf{x}})P^{\mathrm{a}}(\mathbf{x}|\mathbf{y}^{\mathrm{o}})\,\mathrm{d}\mathbf{x}, \quad (8)$$

where $L(\mathbf{x}-\tilde{\mathbf{x}})$ is called a loss function (e.g. Lee, 1997), or sometimes a cost function (e.g. Lewis, Lakshmivarahan, and Dhall, 2006), although the latter presents a clash of terminology with J. The optimal estimate is that $\tilde{\mathbf{x}}$ for which $\rho(\tilde{\mathbf{x}}|\mathbf{y}^{\mathrm{o}})$ is minimized. For the case of Gaussian statistics, and for any quadratic form of L, the optimal estimate found from (8), $\tilde{\mathbf{x}}$, is the same as the $\mathbf{x}^{\mathrm{a}}$ found by minimizing (7).

2.4. Mathematical properties of the **B**-matrix

The covariance matrix $\mathbf{B}$ has fundamental properties, some of which are outlined below.

- In a given representation, $\mathbf{B}$ has a correlation part, $\mathbf{C}$ (a non-diagonal matrix of correlations between elements of $\mathbf{x}$) and a variance part, $\mathbf{\Sigma}^2$ ($\mathbf{\Sigma}^2$ is a diagonal matrix of variances of the elements of $\mathbf{x}$). The covariance matrix is formed by multiplying respective columns and rows of $\mathbf{C}$ by the square roots of the variances (i.e. standard deviations, $\mathbf{\Sigma}$)

$$\mathbf{B} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}. \quad (9)$$

- A covariance matrix is a rank 2 tensor and can be written in a new representation using the transform shown in (10). An orthogonal transform, $\mathbf{V}$, is introduced to transform a vector between its $\mathbf{x}$-representation and an alternative $\hat{\mathbf{x}}$-representation, $\boldsymbol{\eta} = \mathbf{V}\hat{\boldsymbol{\eta}}$ (we assume that $\mathbf{V}$ is an orthogonal matrix with a Euclidean inner product for simplicity, $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$). Given that the background error covariance matrix is $\mathbf{B}$ in the $\mathbf{x}$ representation (4), in the $\hat{\mathbf{x}}$ representation it is, $\hat{\mathbf{B}}$:

$$\hat{\mathbf{B}} = \langle\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}^{\mathrm{T}}\rangle = \mathbf{V}^{\mathrm{T}}\mathbf{B}\mathbf{V}. \quad (10)$$

- A special representation of $\mathbf{B}$ is the eigenrepresentation, where $\hat{\mathbf{B}}$ is diagonalized. In this representation, the eigenvectors are the columns of $\mathbf{V}$, and the eigenvalues are the diagonal elements of $\hat{\mathbf{B}}$. As there are no covariances between eigenvectors, the $\mathbf{B}$-matrix in its eigenrepresentation is diagonal. Thus, the component of a perturbation that projects onto those eigenvectors with large variances are known with less accuracy than those that project onto eigenvectors with small variances, and so are constrained to the background less tightly in the minimization. The eigenvectors involving the

DOI: 10.1002/qj

dynamical variables are sometimes associated with physical modes (Phillips, 1986), where those with large variances are Rossby-like ('slow' structures that dominate the weather) and those with small variances are inertio-gravity-like. Although the size of **B** means that it is practically impossible to be diagonalized, the diagonal representation of sub-matrices of **B** is useful in the process of modelling background error covariance models (see Part II).

- The **B**-matrix is square and symmetric, $B_{ij} = B_{ji}$. The eigenvalues of **B** are therefore real-valued and the eigenvectors are mutually orthogonal (e.g. Lanczos, 1988). This property is used in Part II as a tool in the construction of control variable transforms that model **B**.
- Covariance matrices are positive semi-definite. Consequently, all eigenvalues of **B** are greater than or equal to zero (even though individual off-diagonal matrix elements need not be positive). The positive semi-definite requirement forms part of the definition of a covariance matrix (Gaspari and Cohn, 1999). This is important as it means that the background term in the cost function, the first term of (1), is convex or flat in all directions in state space. A negative eigenvalue would imply that the background part of the cost function would be partially concave and the minimum of the whole cost function may then not exist.

3. How the B-matrix manifests itself in the analysis

The background error covariance matrix has a profound impact on the analysis in VAR, which is why it is considered important to use as realistic a form of **B** as possible. The impact of **B** on the analysis can be understood by using the expression for the analysis found from the best linear unbiased estimator (BLUE; Kalnay, 2003; Ehrendorfer, 2007):

$$\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}\left\{\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}})\right\}. \quad (11)$$

The BLUE yields an analysis that is equivalent to the process of minimizing (1) in VAR when H is linear. It is approximate, but still useful, when H is nonlinear, provided that its linearization, $\mathbf{H} = \partial\mathbf{y}/\partial\mathbf{x}$ (found at $\mathbf{x}^{\mathrm{b}}$) is a good approximation, that is, $H(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}) \approx H(\mathbf{x}^{\mathrm{b}}) + \mathbf{H}\delta\mathbf{x}$. Unlike for VAR, however, (11) is an explicit expression for $\mathbf{x}^{\mathrm{a}}$, allowing us to understand more clearly the role of **B**, even though it is not possible to evaluate (11) explicitly at operational resolutions. The role of **B** in this equation is discussed below and for simplicity (unless specified otherwise) we consider the time-independent case when H and **H** involve no time evolution, and so the solution of (11) would be comparable to 3d-VAR.

3.1. The B-matrix spreads out information in the vertical and horizontal directions in space and weights the importance of the a priori state

Background errors at different locations can be correlated and appear in **B** as off-diagonal matrix elements. These cause information to be transferred between these locations during the assimilation. This effect can be seen in (11) by introducing a single observation of one element of **x**. In this case, $\mathbf{y}^{\mathrm{o}} = y^{\mathrm{o}}$, $\mathbf{R} = \sigma^2$ (both become scalars), and H and **H** each become a row vector of zeroes apart from the element of the state vector that is being observed (kth element), which has value 1 in H and **H**. From (11), the analysis increment at element l, $x_l^{\mathrm{a}} - x_l^{\mathrm{b}}$, is

$$x_l^{\mathrm{a}} - x_l^{\mathrm{b}} = B_{lk}\frac{y - x_k^{\mathrm{b}}}{B_{kk} + \sigma^2}. \quad (12)$$

Even though the observation has been made of element k, the information has been spread to other positions l because of **B**, with a magnitude proportional to the matrix element. In general, the **B**-matrix is not the only means by which information is spread in data assimilation. An observation operator that depends upon more than one element of the state vector will result in these elements being coupled in the analysis, but the assimilation of an observation of a single model quantity isolates the spreading effects of **B**. This effect helps smooth the fields in a way that is consistent with the structure of the **B**-matrix. However, an ill-formed **B**-matrix can permit the possibility of anomalies appearing in $\mathbf{x}^{\mathrm{a}}$. Such problems are sometimes encountered, for example, in the anomalous appearance of water vapour concentration in otherwise dry areas. Derber and Bouttier (1999) report that in the European Centre for Medium Range Weather Forecasts (ECMWF) system, the vertical covariances of water vapour have to be modified from their diagnosed values in order to prevent observations made in the troposphere unduly moistening the stratosphere.

The non-local effect of **B** emphasizes the need in global data assimilation systems to solve the problem globally, and not in separate regions, as was done in the past for computational reasons with optimal interpolation schemes. This produced separate analyses that needed to be stitched together. In recent years, the application of VAR has allowed the global problem to be solved in one go. Single-observation VAR experiments remain a common and straightforward way to diagnose the structure of background error covariances, which in VAR are usually not known explicitly. Figure 2 shows the analysis increment, $\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}}$, found from running such an experiment (with the Met Office 3d-VAR system), assimilating a synthetic observation of pressure. Even though the observation measures the values of pressure at one point, the data assimilation has spread out this information. The analysis increment response is proportional to the matrix elements B_{lk} (where l is the field position index in Figure 2, and k is the observation position index at the centre of the shaded regions shown), as found in (12). This traces a column of **B** and is called a structure function. Structure functions have characteristic length-scales in the horizontal and vertical directions, and also include the response in variables other than those measured (see section 3.2). Structure functions, calculated by single-observation (or pseudo-observation) experiments are common basic diagnostics used in studies of VAR systems; for example, see

DOI: 10.1002/qj

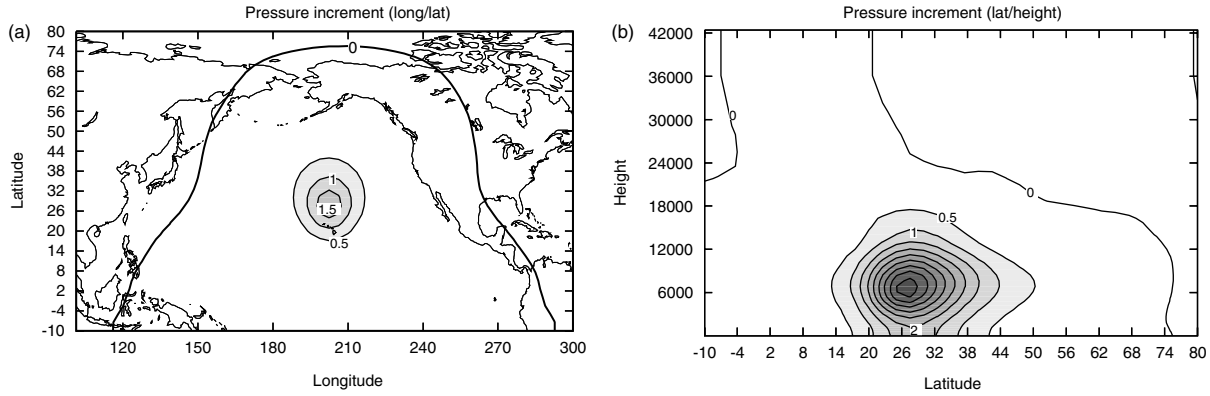

Figure 2. Analysis increments found from a VAR run of the Met Office 3d-VAR system by assimilating a single synthetic pressure observation of a value of 1 hPa above the background value (and with an error of 0.1 hPa), at a height of about 5.5 km over the Central Pacific Ocean. (a) A longitude/latitude cross-section at 5.5 km; (b) a latitude/height cross-section at 200° longitude.

Parrish and Derber (1992) for the global system of the National Meteorological Centre (NMC), Courtier *et al.* (1998) and Derber and Bouttier (1999) for the ECMWF global system, Gustafsson *et al.* (2001) for the High Resolution Limited Area Model (HIRLAM) system, Barker *et al.* (2004) for the National Centers for Atmospheric Research (NCAR) mesoscale system, and Weaver *et al.* (2005) for an ocean VAR system. A challenge of VAR is to represent realistic structure functions, given that it is not possible to hold an explicit matrix (see Part II).

As well as the non-local effect, (12) contains local variance weighting, which can be analysed by setting $l = k$. If $\sigma^2 \ll B_{kk}$, (12) says that $x_l^{\mathrm{a}} \rightarrow y$, and conversely if $\sigma^2 \gg B_{kk}$, then $x_l^{\mathrm{a}} \rightarrow x_l^{\mathrm{b}}$. This is a very important role of **B**, which allows background information to be weighted against observational information according to their respective uncertainties.

3.2. The **B**-matrix spreads information to other variables and imposes balance

The **B**-matrix permits multivariate error correlations (see Figure 1), which is a generalization of the spatial spreading described above. The background states used in operational schemes are produced from model integrations, and consequently the equations of motion influence strongly the way in which errors between (and within) variables are correlated. It is believed that the atmosphere is predominantly in a state of hydrostatic and geostrophic balance on most scales resolved by global models. The PDF that is represented by the **B**-matrix must allow only those likely background states that contain this near-balanced property. Consequently, a perturbation of a given background state at a given location should be accompanied by adjustments elsewhere that are in near balance with the perturbation. For geostrophic balance, for example, the covariances yield an effect like that of the geostrophic adjustment problem (geostrophic adjustment is discussed by Haltiner and Williams, 1980).

The **B**-matrix contains balance information in statistical form. This occurs automatically when computing the **B**-matrix from atmospheric states via (4). In practice, the **B**-matrix cannot be stored explicitly, but the near-balanced property is used in Part II to construct models of **B** by using forms of balance relations.

By way of example, horizontal structure functions consistent with geostrophic adjustment in an idealized setting are shown in Bartello and Mitchell (1992), Pailleux (1997) and Kalnay (2003). References that show multivariate structure functions in more realistic settings are provided in section 3.1. Some of these articles show structure functions sourced near the equator where geostrophy is not prevalent, emphasizing the complicated nature of covariances. Figure 3 shows some multivariate structure functions from the Met Office's **B**-matrix for fields other than pressure for the single-observation experiment of pressure introduced in Figure 2 (at a midlatitude location). The winds (Figures 3a and 3b) have a dipole structure consistent with the expected near-geostrophic anticyclonic response to the adjustment of the pressure field.

The potential temperature has a simple response in the horizontal (bottom left), but a more complicated structure in the vertical (bottom right). According to hydrostatic balance, the potential temperature, $\delta\theta$, and pressure, δp, increments are related via $\delta\theta \sim \theta^{\mathrm{b}} \partial \delta p / \partial z$, where θ^{b} is the background potential temperature. Using the pressure increment information in Figure 2, and the knowledge that θ^{b} grows in the stratosphere, this helps to explain the change in sign of $\delta\theta$ where δp peaks and the growth of $\delta\theta$ in the stratosphere. The stratospheric peak in $\delta\theta$ is an example of a strong long-range covariance pattern.

3.3. The analysis increment lies in the subspace spanned by **B**

The **B**-matrix is the last operator that appears in the computation of the analysis increment in (11), and restricts the 'directions' that the analysis increment can take in state space to those that are spanned by **B**. Consider a population of N representative forecast errors, $\boldsymbol{\eta}_i$, used in the computation of **B** via (4):

$$\mathbf{B} = \frac{1}{N-1} \sum_{i=1}^{N} \boldsymbol{\eta}_i \boldsymbol{\eta}_i^{\mathrm{T}}. \tag{13}$$

DOI: 10.1002/qj

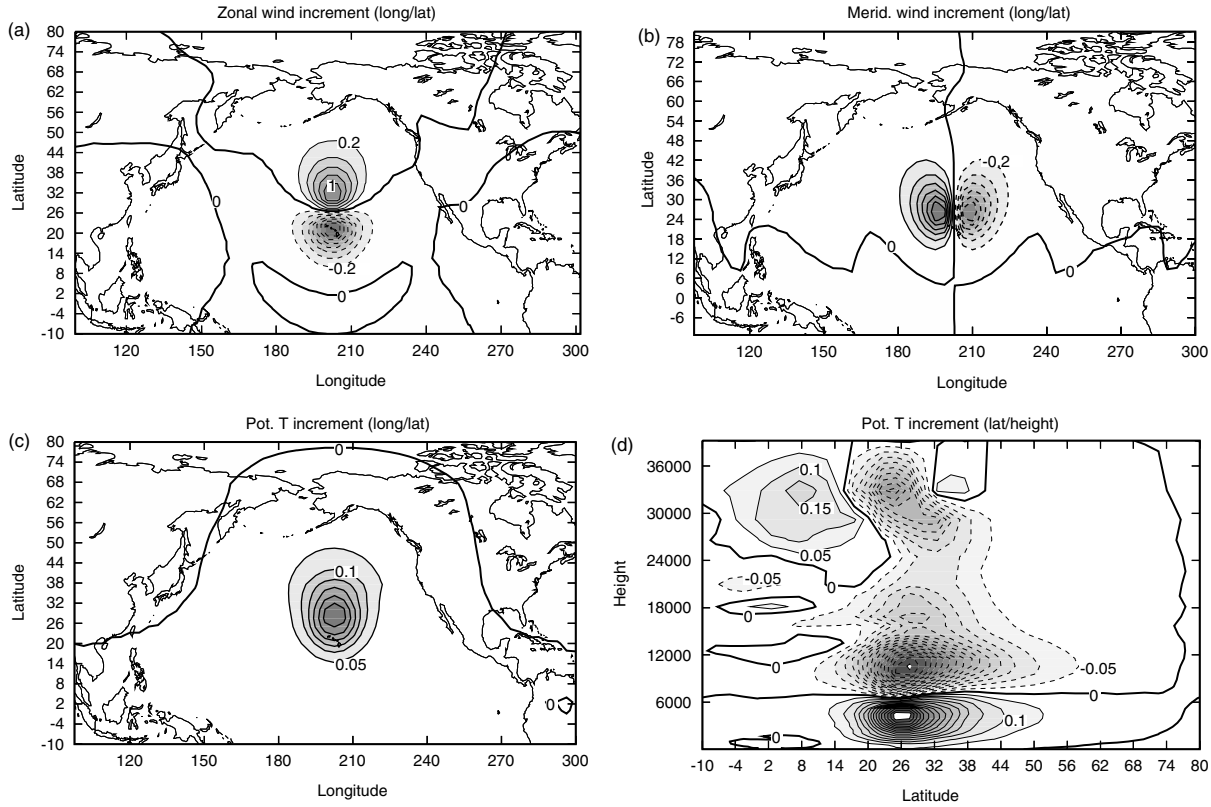

Figure 3. Analysis increments found from the same VAR run of Figure 2 (synthetic pressure observation) to show the structure of the multivariate structure functions. Longitude/latitude cross-sections of (a) zonal wind and (b) meridional wind increments at 5.5 km. (c) Latitude/longitude plot of potential temperature increment at 5.5 km, and (d) a latitude/height cross-section at longitude 200° of the same quantity. In all panels, the shading indicates the weight of each structure function. Continuous (dashed) contours represent positive (negative) values.

This reveals the mechanics of computing $\mathbf{B}\delta\mathbf{x}$, with $\mathbf{B}$ computed in this way. The result is a linear combination of $\boldsymbol{\eta}_i$ with weights $\boldsymbol{\eta}_i^{\mathrm{T}}\delta\mathbf{x}$. Thus, $\mathbf{B}\delta\mathbf{x}$, and hence the analysis increment in (11), will include only contributions from directions in state space that are spanned by $\boldsymbol{\eta}_i$. Directions not spanned by $\boldsymbol{\eta}_i$ are referred to as the null-space of $\mathbf{B}$ (Lorenc, 2003a). If there is an insufficient population of forecast errors, this effect will artificially reduce the rank of $\mathbf{B}$.

The eigenvalue spectrum of $\mathbf{B}$ can also provides insight into the role that $\mathbf{B}$ plays. Replacing $\mathbf{B}$ with its diagonal form, $\hat{\mathbf{B}}$ from (10), $\mathbf{B} = \mathbf{V}\hat{\mathbf{B}}\mathbf{V}^{\mathrm{T}}$, shows that (1) $\delta\mathbf{x}$ is projected onto the eigenvectors, (2) the projections are weighted with the eigenvalues, and (3) the answer is projected back into state space. The eigenvectors that are accompanied by non-zero eigenvalues are in the subspace spanned by $\mathbf{B}$, but unlike $\boldsymbol{\eta}_i$, form an orthonormal set, allowing them to be considered independently. Those eigenvectors that have high variance (labelled Rossby-like modes in section 2.4) are given more weight as a result of the assimilation (11) than those that have low variance (labelled inertio-gravity-like modes). This qualitative correspondence is an interpretation of how $\mathbf{B}$ imposes partial (but not total) balance to the increment by allowing mainly balanced, but some unbalanced modes, and excluding altogether modes that lie in the null space of $\mathbf{B}$. The presence of a strict null-space in the $\mathbf{B}$-matrix found from (13) is probably an artefact of an inadequate population of forecast error states, which will occur in any practical situation regarding an operational $\mathbf{B}$-matrix.

3.4. $\mathbf{B}$ provides a means by which observations can act in synergy

Another manifestation of the $\mathbf{B}$-matrix relates to how observations 'see' each other. $\mathbf{B}$ allows observations to reinforce each other in a way that improves the analysis to a degree that is greater than their individual contributions. This effect has been demonstrated by Lorenc (1981), who showed how three observations (wind, layer thickness and height) assimilated together in the ECMWF's early data assimilation system reduced the analysis error to a degree that is greater than the aggregate effects of each observation assimilated separately.

There are two parts in (11) where the $\mathbf{B}$-matrix allows observations to see each other. The first effect is an extension of points 3.1 and 3.2 above, and is explained by Lorenc (1986), who wrote the transpose of (11)

$$(\mathbf{x}^{\mathrm{a}} - \mathbf{x}^{\mathrm{b}})^{\mathrm{T}} = \left\{\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}})\right\}^{\mathrm{T}} (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{H}\mathbf{B}, \quad (14)$$

which gives the analysis increment as a row vector. It is useful to think of $\mathbf{B}$ on the right-hand side of (14) as a matrix of structure functions comprising the columns of $\mathbf{B}$. Each column gives the background error covariances associated with each element of the state space (one such structure function is shown in Figures

DOI: 10.1002/qj

2 and 3). These structure functions are acted upon in (14) with the observation operator **H**, which yields a new matrix whose columns are the observations associated with each structure function, which is then acted upon by $(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1}$. The resulting matrix provides a basis of vectors (rows) in which to specify the analysis increment. The analysis increment is a linear combination of these rows with weights that are elements of the innovation vector, $\mathbf{y}^{\mathrm{o}}-H(\mathbf{x}^{\mathrm{b}})$. The structure functions are non-local, and so the analysis increment at a particular grid point will be influenced not only by those observations that observe that point, but by other observations that observe those parts of the state space 'touched' by its structure function.

The second effect is due to the term $(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1}$ of (11) and (14), which is mentioned above. The rows of **H** are the weighting functions of the linearized observation operators. The matrix element $(\mathbf{HBH}^{\mathrm{T}})_{ij}$ is the overlap between weighting functions corresponding to observations i and j (rows i and j of **H**), after weighting function j is acted upon by **B**. This means that a non-zero element can result even if the weighting functions themselves do not overlap. Matrix $\mathbf{HBH}^{\mathrm{T}}$ is added to **R**, so it behaves as an effective additional error covariance between observation errors, which are themselves known to have an important impact on the analysis (e.g. Liu and Rabier, 2002). Thus, **B** can allow observations to 'constructively interfere'.

3.5. Note about background error propagation in 4d-VAR

The above arguments are applied to observations assimilated at the same time that $\mathbf{x}^{\mathrm{b}}$ and **B** are valid, which is the underlying assumption behind 3d-VAR, even though in practice, observations can be made, several hours from the validity time of $\mathbf{x}^{\mathrm{b}}$. By design, the state vector in 4d-VAR is propagated explicitly by the forecast model to the appropriate time of each observation. An important consequence is that the **B**-matrix is also propagated by the forecast model from the background state's validity time (t_0) to the time of each observation (e.g. t), but this occurs implicitly in 4d-VAR (Fisher, 2001). This can be demonstrated for the special case when all observations in 4d-VAR are made at the same time, $t > t_0$. In this case, (11) is equivalent to the incremental 4d-VAR analysis (Courtier *et al.*, 1994) if the observation operator, H, includes propagation by the model. Let $H \rightarrow H_t M_{t\leftarrow t_0}$, where M represents a forecast model run from t_0 to t. The linearization $\mathbf{H} \rightarrow \mathbf{H}_t\mathbf{M}_{t\leftarrow t_0}$ follows and with the matrix identity $\mathbf{BH}^{\mathrm{T}}(\mathbf{R}+\mathbf{HBH}^{\mathrm{T}})^{-1} = (\mathbf{B}^{-1}+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$ the following is found:

$$(\mathbf{I}+\mathbf{MBM}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_t)\mathbf{M}(\mathbf{x}^{\mathrm{a}}-\mathbf{x}^{\mathrm{b}}) = \mathbf{MBM}^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}^{-1}\left\{\mathbf{y}^{\mathrm{o}}-H_tM(\mathbf{x}^{\mathrm{b}})\right\} \qquad (15)$$

(dropping the forecast model's $t \leftarrow t_0$ label for convenience). All occurrences of **B** in (15) appear as $\mathbf{MBM}^{\mathrm{T}}$, which is identified as the propagation of **B** (which is itself valid at t_0) to time t by the linearized forecast model. This property, which is implicit in 4d-VAR, is one reason why 4d-VAR is superior to 3d-VAR (Lorenc and Rawlins, 2005), as it allows maximum gain from the manifestations of the **B**-matrix (sections 3.1–3.4) as the **B**-matrix will be more appropriate at each subsequent time than if the **B**-matrix was not propagated. This is especially beneficial in cases where the evolving **B**-matrix becomes significantly different to the static **B**-matrix (e.g. in the presence of fronts or baroclinic growth). Even in 4d-VAR however, the **B** matrix does not evolve indefinitely, as it reverts back to the static matrix at the start of the next cycle (6 or 12 h later).

Thépaut *et al.* (1996) give examples of such evolved structure functions in 4d-VAR in the case of baroclinic growth. Other examples are shown in Rabier *et al.* (1997). The implied **B**-matrix propagation is believed to hold generally in 4d-VAR, which is connected to the equivalence of 4d-VAR to the Kalman smoother (Ménard and Daley, 1996; Li and Navon, 2001; Fisher, Leutbecher and Kelly, 2005), which holds under special conditions (i.e. linear observation operators, linear forecast model, and the same background error covariance matrix at the start of the assimilation).

4. Practical difficulties with the B-matrix

The previous section shows that the **B**-matrix is a key element of data assimilation. As such, care should be taken to ensure that **B** is accurate. There are, however, a number of practical difficulties that arise when dealing with **B** (e.g. insisting upon an explicit definition and treatment of **B** would limit unacceptably the resolution of the data assimilation problems that computers could deal with). Three of the main difficulties are as follows.

1. The first problem is the unknown 'true' state, needed to measure errors against. This is the most fundamental difficulty, but surrogate states can be used to mimic forecast errors. These include the use of forecast differences (instead of forecast errors), or the substitution of the 'true' state by an ensemble mean, to measure errors against, as is done in the context of the ensemble Kalman filter (e.g. Evensen, 2003). Methods that use forecast differences and ensembles are discussed in section 5.
2. The second problem is that of dealing with the unfeasibly large size of the **B**-matrix. This can be dealt with by using covariance modelling techniques, which are discussed in Part II.
3. The third problem is a consequence of the lack of availability of a sufficiently large population of error (or surrogate) states to calculate **B** to full rank (see section 3.3). If there are n elements in the state vector, then at least n independent error estimates are needed to calculate **B** to full rank. Usually, however, the number of estimates available is only a very small fraction of n. There are well-known consequences of using a reduced rank **B**-matrix.

DOI: 10.1002/qj

First, the analysis increments are forced to lie in a subspace, which is much reduced from that found using a full rank **B**-matrix. Secondly, spurious long-range correlations can appear (Hamill *et al.*, 2000; Hamill, Whitaker and Snyder, 2001; Houtekamer and Mitchell 2001; Lorenc 2003b). The use of the covariance modelling techniques discussed in Part II partially alleviates this problem, as the missing information is made up for by the covariance models (e.g. by imposing homogeneity, closeness to balance, etc.).

5. Measuring the B-matrix

In this section we show how to measure some important properties of forecast error covariances in ways that are meant to overcome problem 1 of section 4. In a data assimilation system, the process of determining the **B**-matrix is sometimes called the 'calibration step'. There have been a number of strategies for measuring **B** without the need for the 'truth' to be known, and most methods make an assumption of ergodicity (e.g. Courtier *et al.*, 1998). The leading methods are now described.

5.1. Analysis of innovations

Reliable estimates of forecast errors of well-observed quantities can be gained by analysing innovations (Rutherford, 1972). Innovations are the so-called observation-minus-background quantities familiar in data assimilation. The analysis starts with the relation between the observations, $\mathbf{y}^{\mathrm{o}}$, the 'true' state of the atmosphere, $\mathbf{x}^{\mathrm{t}}$, and observational error, $\boldsymbol{\varepsilon}^{\mathrm{o}}$:

$$\mathbf{y}^{\mathrm{o}} = H(\mathbf{x}^{\mathrm{t}}) + \boldsymbol{\varepsilon}^{\mathrm{o}}. \tag{16}$$

Because it is unknown, the true state of the atmosphere must be eliminated. Using the definition of forecast error, $\boldsymbol{\eta}$ ($\boldsymbol{\eta} = \mathbf{x}^{\mathrm{b}} - \mathbf{x}^{\mathrm{t}}$), and then linearizing H about $\mathbf{x}^{\mathrm{b}}$, $\mathbf{y}^{\mathrm{o}} = H(\mathbf{x}^{\mathrm{b}} - \boldsymbol{\eta}) + \boldsymbol{\varepsilon}^{\mathrm{o}} \approx H(\mathbf{x}^{\mathrm{b}}) - \mathbf{H}\boldsymbol{\eta} + \boldsymbol{\varepsilon}^{\mathrm{o}}$, gives

$$\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}}) \approx \boldsymbol{\varepsilon}^{\mathrm{o}} - \mathbf{H}\boldsymbol{\eta}, \tag{17}$$

which relates innovations (on the left-hand side) with observation and background errors. This method allows us to study forecast errors of (and between) those model variables that are directly observed. Let $y^{\mathrm{o}}(v, \mathbf{r})$ be an observation of model variable v at position $\mathbf{r}$. Directly observing model variables means that we can substitute $H(\mathbf{x}^{\mathrm{b}})$ with the grid-box value of $\mathbf{x}^{\mathrm{b}}$ of v at $\mathbf{r}$ ($= x^{\mathrm{b}}(v, \mathbf{r})$) and $\mathbf{H}\boldsymbol{\eta}$ with the forecast error of v at $\mathbf{r}$ ($= \eta(v, \mathbf{r})$). The appropriate component of (17) is then

$$y^{\mathrm{o}}(v, \mathbf{r}) - x^{\mathrm{b}}(v, \mathbf{r}) = \varepsilon^{\mathrm{o}}(v, \mathbf{r}) - \eta(v, \mathbf{r}). \tag{18}$$

The background error covariance between variable v_1 at position $\mathbf{r}$ and variable v_2 at position $\mathbf{r} + \Delta\mathbf{r}$ is then found from the product of (18):

$$\begin{aligned} &\langle \{y^{\mathrm{o}}(v_1, \mathbf{r}) - x^{\mathrm{b}}(v_1, \mathbf{r})\} \\ &\qquad \times \{y^{\mathrm{o}}(v_2, \mathbf{r} + \Delta\mathbf{r}) - x^{\mathrm{b}}(v_2, \mathbf{r} + \Delta\mathbf{r})\} \rangle \\ &= \langle \{\varepsilon^{\mathrm{o}}(v_1, \mathbf{r}) - \eta(v_1, \mathbf{r})\} \\ &\qquad \times \{\varepsilon^{\mathrm{o}}(v_2, \mathbf{r} + \Delta\mathbf{r}) - \eta(v_2, \mathbf{r} + \Delta r)\} \rangle \\ &\approx \langle \varepsilon^{\mathrm{o}}(v_1, \mathbf{r}) \varepsilon^{\mathrm{o}}(v_2, \mathbf{r} + \Delta\mathbf{r}) \rangle \\ &\qquad + \langle \eta(v_1, \mathbf{r}) \eta(v_2, \mathbf{r} + \Delta\mathbf{r}) \rangle. \end{aligned} \tag{19}$$

Here, we have assumed that observation and background errors are uncorrelated. In the case of different variables ($v_1 \neq v_2$) or where observations are made at different positions ($\Delta r > 0$), the first term in (19) may be assumed to be zero (as, for direct observations, measurements may be made independently of each other). This leaves information only on background error covariances.

The averaging, indicated by angled brackets, can be done over different times (by assuming ergodicity) and over different $\mathbf{r}$ (by assuming homogeneity over the averaging region). Spatial averaging may be limited to within a latitude band. This method was used to diagnose the forecast error statistics for the ECMWF optimal interpolation system, as reported by Hollingsworth and Lönnberg (1986), Lönnberg and Hollingsworth (1986), and more recently for VAR by Järvinen (2001). A sample of the results of Hollingsworth and Lönnberg is shown in Figure 4, which shows the horizontal autocorrelations of wind-component errors found from

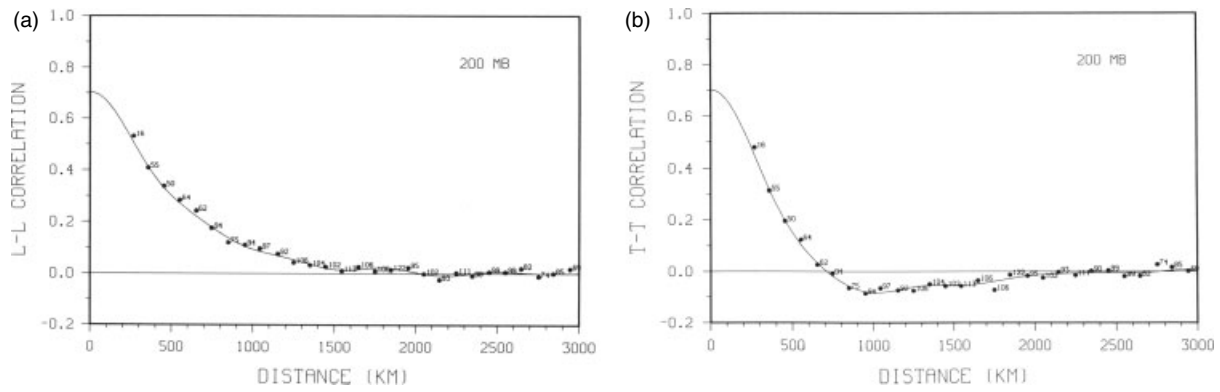

Figure 4. Horizontal autocorrelations of errors in two components of wind as a function of distance. (a) Longitudinal–longitudinal wind correlation. (b) Transverse–transverse wind correlation. ©Blackwell Publishing 1986, from Hollingsworth and Lönnberg (1986). Reproduced by permission of Blackwell Publishing.

DOI: 10.1002/qj

radiosonde observations. There is a lack of measured correlations (dots) for small $\Delta\mathbf{r}$ owing to the absence of radiosonde stations close together. Correlations have been (perhaps subjectively) extrapolated to small $\Delta\mathbf{r}$ where the first term in (19) becomes important.

This method of measuring forecast error statistics is reasonably successful, but it has its limitations. It relies on the presence of sufficient numbers of independent *in situ* observations of all quantities present in the background state. This method may be adequate for those regions of the atmosphere that are well observed by radiosondes (i.e. the continental Northern Hemisphere troposphere) but inadequate elsewhere. It also relies upon the availability of unbiased forecasts and observations. This said, the analysis of innovations has been used to good effect recently to tune the background error variances of an existing data assimilation scheme (Buehner, Gauthier and Liu, 2005).

5.2. Differences of varying length forecasts: 'NMC method'

The need to observe directly all quantities in all regions of the atmosphere can be avoided by analysing forecast differences. A common method is to take the difference between pairs of forecasts of different lengths, but each valid at the same time. Parrish and Derber (1992) considered the following:

$$\mathbf{B} \approx \frac{1}{2}\langle(\mathbf{x}^{48} - \mathbf{x}^{24})(\mathbf{x}^{48} - \mathbf{x}^{24})^{\mathrm{T}}\rangle, \tag{20}$$

(note the factor of 1/2). Here, $\mathbf{x}^{48}$ and $\mathbf{x}^{24}$ are, respectively, 48 h and 24 h forecasts valid at the same time, which are launched from analyses that are 24 h apart, as in the following:

$$\mathbf{x}^{48} = M_{48\leftarrow 0}\mathbf{x}^{\mathrm{a}}(t=0), \tag{21a}$$

$$\mathbf{x}^{24} = M_{48\leftarrow 24}\mathbf{x}^{\mathrm{a}}(t=24), \tag{21b}$$

Here, M is the full, nonlinear forecast model and $\mathbf{x}^{\mathrm{a}}$ is the analysis at the specified times. The difference $\mathbf{x}^{48} - \mathbf{x}^{24}$ is a proxy for forecast error (i.e. it is assumed that it contains univariate and multivariate structures that resemble those of forecast errors, although the theoretical justification for this is not rigorous; Bouttier, 1996). The angled brackets represent an average over different cases in time (typically over one month). This approach has become known as the 'NMC' or the 'NCEP' method, named after the National Meteorological Center (now the National Center for Environmental Prediction).

We now examine the relationship between the covariances of forecast errors and those of forecast differences. The two forecasts in (21a) and (21b) can be written in terms of their errors

$$\mathbf{x}^{48} = \mathbf{x}^{\mathrm{t}} + \boldsymbol{\eta}^{48} + \mathbf{b}^{48}, \tag{22a}$$

$$\mathbf{x}^{24} = \mathbf{x}^{\mathrm{t}} + \boldsymbol{\eta}^{24} + \mathbf{b}^{24}, \tag{22b}$$

where $\mathbf{x}^{\mathrm{t}}$ is the true state at the validity time, $\boldsymbol{\eta}^{48/24}$ is the random error and $\mathbf{b}^{48/24}$ is the bias present in each forecast. Assuming that $\mathbf{b}^{48} = \mathbf{b}^{24}$, the forecast difference is

$$\delta\mathbf{x} = \mathbf{x}^{48} - \mathbf{x}^{24} = \boldsymbol{\eta}^{48} - \boldsymbol{\eta}^{24}. \tag{23}$$

Assuming that the forecast errors $\boldsymbol{\eta}^{48}$ and $\boldsymbol{\eta}^{24}$ are uncorrelated and that they each have the same error covariances as $\boldsymbol{\eta}$, the error covariance matrix of $\delta\mathbf{x}$ is twice $\mathbf{B}$:

$$\begin{aligned}\langle\delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}}\rangle &= \langle(\boldsymbol{\eta}^{48} - \boldsymbol{\eta}^{24})(\boldsymbol{\eta}^{48} - \boldsymbol{\eta}^{24})^{\mathrm{T}}\rangle \\ &= \langle\boldsymbol{\eta}^{48}\boldsymbol{\eta}^{48\,\mathrm{T}}\rangle + \langle\boldsymbol{\eta}^{24}\boldsymbol{\eta}^{24\,\mathrm{T}}\rangle \\ &\approx 2\langle\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}}\rangle = 2\mathbf{B}.\end{aligned} \tag{24}$$

This is the reason for the factor of 1/2 in (20). An alternative derivation is given in Bannister (2006).

Even though the NMC method has been used by most forecast centres, there are known difficulties that exist (in addition to the rank deficiency).

- The background is usually a 6 h forecast, yet the NMC method uses longer (48 and 24 h) forecast differences. The reason for using forecasts staggered by 24 h is to eliminate errors in modelling the diurnal cycle, but the need for these particular forecast lengths is not necessary. The Met Office, for instance, has recently used 30 and 6 h forecast differences (M. Cullen and D. Pearson, personal communication) and Berre (2000) has used 36 and 12 h forecast differences.
- The NMC method has problems in poorly observed regions and at large scales (Berre, 2000; Fisher, 2003; Berre, Ştefănescu and Pereira, 2006). In poorly observed regions, there are likely to be only very small differences between the forecasts, and so here we expect the method to underestimate variances. At large scales, it is found that the NMC method overestimates covariances. This is not only because of the use of forecasts that are temporally longer than those used for the *a-priori* state, but also because of the inherent properties of the method itself, which would show even if the system were static (Berre *et al.*, 2006). Generally, the NMC statistics often require adjustment to compensate for these and other problems (e.g. Derber and Bouttier, 1999; Ingleby, 2001). The NMC method also often overestimates vertical correlation length-scales (Fisher, 2003).
- Forecast differences are usually calculated over a reasonably long period of time (e.g. a month or a season of forecast differences). The variation from one case to the next originates from the natural variability of the model. This makes the NMC method suitable only for estimation of climatological covariances (Bouttier, 1996).

5.3. The lagged NMC method

Forecasts with limited area models (LAMs) have an additional source of error that comes from the larger

DOI: 10.1002/qj

scale model providing its lateral boundary conditions (LBCs). The lagged NMC method for LAMs (Široká *et al.*, 2003; Sadiki and Fischer, 2005; Berre *et al.*, 2006) is a modification to the standard NMC method, which is meant to eliminate this source of error.

In the standard NMC method for LAMs, each forecast in (21a) and (21b) uses LBCs from runs of the larger scale model, which is run in parallel with the LAM forecasts from analyses at the same times as in (21a) and (21b). (The lead times used for LAM statistics are typically shorter than those for global models, but here we refer to the same lead times for convenience.) In the lagged NMC method, the shorter forecast, $\mathbf{x}^{24}$, now uses the same lateral boundary conditions as those used by the longer forecast, $\mathbf{x}^{48}$, where they overlap. In order to preserve consistency between the LBCs and the initial conditions for $\mathbf{x}^{24}$, the initial conditions are now also taken from the larger scale model at $t = 24$. This has to be reconfigured to the LAM grid, and initialized. Mathematically, (21a) and (21b) turn into

$$\mathbf{x}^{48} = M^{\mathrm{LBC}}_{48\leftarrow 0}\mathbf{x}^{\mathrm{a}}(t=0), \tag{25a}$$

$$\mathbf{x}^{24} = M^{\mathrm{LBC}}_{48\leftarrow 24}\mathbf{x}^{1}(t=24), \tag{25b}$$

where

$$\mathbf{x}^{1}(t=24) = DR\hat{M}_{24\leftarrow 0}\hat{\mathbf{x}}^{\mathrm{a}}(t=0). \tag{25c}$$

In (25c), which gives the initial conditions, $\mathbf{x}^1(t = 24)$, for the shorter forecast, $\hat{\mathbf{x}}^{\mathrm{a}}(t = 0)$ is the large-scale analysis at $t = 0$, $\hat{M}_{24\leftarrow 0}$ is the 24 h large-scale model, R is reconfiguration from the large-scale grid to the small-scale grid and D is initialization by digital filtering (Lynch and Huang, 1992). In (25a) and (25b), the 'LBC' superscript indicates that the same LBCs are used for both LAM runs.

Variances of the new differences, $\mathbf{x}^{48} - \mathbf{x}^{24}$, are meant to have smaller values at large scales than in the standard NMC method. Berre *et al.* (2006), who compared the standard and lagged NMC methods using M as the Aladin LAM and $\hat{M}$ as the Arpège global model, with 36 and 12 h forecasts instead of 48 and 24 h, found that the length-scales in the lagged NMC method were indeed shorter than those found from the standard NMC. Široká *et al.* (2003) also found that the lagged NMC-derived vertical correlation length-scales are broader than those from the standard NMC.

5.4. The ensemble (Monte Carlo) method

A favoured method of producing forecast error statistics is by the generation of an ensemble of forecasts, which is intended to have the same spread as the true forecast error PDF. There are two general approaches for generating the ensemble. The first is to use the ensemble produced as part of an existing, but separate, ensemble-based system, such as an ensemble Kalman filter (e.g. Evensen, 2003) or an ensemble transform Kalman filter (Bishop, Etherton and Majumdar, 2001). The second is to feed an ensemble of states through an existing VAR/forecast system in order to simulate all important sources of error (e.g. observation and background error in the assimilation and model error in the forecast). This is a popular method based on the 'system simulation' approach of Houtekamer *et al.* (1996). Most attention here is given to the second approach, but many of the statements apply also to the first (see Ehrendorfer, 2007, for a review of ensemble-based data assimilation methods). Buehner (2005) compares both approaches with the NMC method.

An ensemble of N members can be used to estimate $\mathbf{B}$ in a flow-dependent way by taking covariance statistics of the differences between each member and the mean:

$$\mathbf{B} \approx \langle (\mathbf{x}^{\mathrm{b}} - \langle \mathbf{x}^{\mathrm{b}} \rangle)(\mathbf{x}^{\mathrm{b}} - \langle \mathbf{x}^{\mathrm{b}} \rangle)^{\mathrm{T}} \rangle. \tag{26}$$

The averaging (angled brackets) may be made over N members at a particular time for an estimate of the instantaneous $\mathbf{B}$-matrix or additionally over a period of many assimilation cycles for an estimate of the climatological $\mathbf{B}$-matrix. In (26), the ensemble mean, $\langle \mathbf{x}^{\mathrm{b}} \rangle$, represents the 'truth' (cf. (4)), but is itself not a real state. The ensemble can be used to estimate $\mathbf{B}$ in an alternative way that avoids reference to $\langle \mathbf{x}^{\mathrm{b}} \rangle$ by taking differences between pairs of ensemble members:

$$\mathbf{B} \approx \frac{1}{2}\langle (\mathbf{x}^{\mathrm{b}}_i - \mathbf{x}^{\mathrm{b}}_j)(\mathbf{x}^{\mathrm{b}}_i - \mathbf{x}^{\mathrm{b}}_j)^{\mathrm{T}} \rangle. \tag{27}$$

In (27), the indexing i and j represents a pair of members and there are $N - 1$ such independent pairs. The factor of $1/2$ is present for the same reason as in the NMC method (20). Again, the averaging may be carried out instantaneously over the $N - 1$ pairs or additionally over many assimilation cycles. Recently, the ECMWF have started to use (27) (averaging additionally over one month) to determine their $\mathbf{B}$-matrix (replacing the NMC method) in the way described in Fisher and Andersson (2001), Fisher (2003) and Žagar, Andersson and Fisher (2005), and Météo-France (Berre, 2006) use a similar ensemble-based method. Isaksen, Fisher and Berner (2007) have used the method to investigate flow-dependent variances in the ECMWF system. Typically, ensembles of sizes $N \sim \mathcal{O}(10)$ to $\mathcal{O}(100)$ are affordable at present.

Some sources of error have been studied enough to simulate reasonably (e.g. observation and, to some extent, background errors). (The need for background error statistics in this procedure is a chicken-and-egg situation, as the aim of the calculation is to produce such error statistics. In practice, this method is a means of refining such statistics and one may perturb the background state of each member using existing background error estimates, for example, from the NMC method.) At the ECMWF, these are perturbed in the ensemble according to their appropriate distributions, and fed into an assimilation system to produce ensembles of perturbed analyses, which can then each be propagated by the forecast model to give the $\mathbf{x}^{\mathrm{b}}$ members. At the start of this process, an

DOI: 10.1002/qj

existing estimate of background error is needed, but thereafter background errors emerge autonomously (Lorenc, 2003b).

Another important source of error, which needs to be included in the forecast stage, is model error. This is more difficult to simulate than observation and background error as it is largely unknown. Houtekamer *et al.* (1996) simulate model error by varying elements of the forecast model in controlled ways. They perturb important auxiliary fields used by the model (sea-surface temperature, model orography, surface roughness length and albedo), and vary the model's parametrization schemes (diffusion, gravity wave drag, radiation and convection schemes). Varying the convection scheme is especially important in representing model error in the Tropics. In recent experiments at the ECMWF, a stochastic parametrization used in their model (called the spectral backscatter scheme; Shutts, 2005) has been shown to be a realistic source of model error in the Tropics (Isaksen *et al.*, 2007).

Figure 5 shows part of the propagation of two ensemble members through the ECMWF forecast/analysis system (Fisher, 2003) for a large number of cycles (typically over a month). Observations are perturbed in each analysis step according to their error statistics and preliminary knowledge of forecast error is needed to set the initial separation of the members. Because covariance statistics are averaged in time, only climatological covariances are diagnosed.

The ensemble method is able to yield results that are more realistic than the NMC method (e.g. yield shorter length-scales), and less susceptible to problems related to poorly observed regions. The most obvious advantage of this method over NMC is that statistics are computed from forecasts of the same lead time as for a background (6 or 12 h). In addition, Berre *et al.* (2006) have shown that, unlike the standard NMC method, the differences between the two members obey the same equations of propagation as forecast errors. To demonstrate the difference between the NMC and ensemble methods, Figure 6 plots the horizontal correlation function of 500 hPa geopotential error. The length-scale found from the ensemble is shorter than that found from NMC, as expected. Considering the distance at which correlations drop by one-half, NMC gives 680 km and the ensemble gives 550 km. The new method compares more favourably with the value found from the innovation method (section 5.1) giving 540 km (Lönnberg and Hollingsworth, 1986).

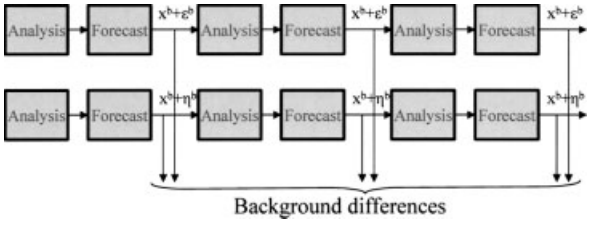

Figure 5. The propagation of two members of an ensemble where $\boldsymbol{\varepsilon}^b$ and $\boldsymbol{\eta}^b$ are two perturbations to the forecasts, $\mathbf{x}^b$. ©ECMWF 2003, from Fisher (2003). Reproduced by permission of the ECMWF.

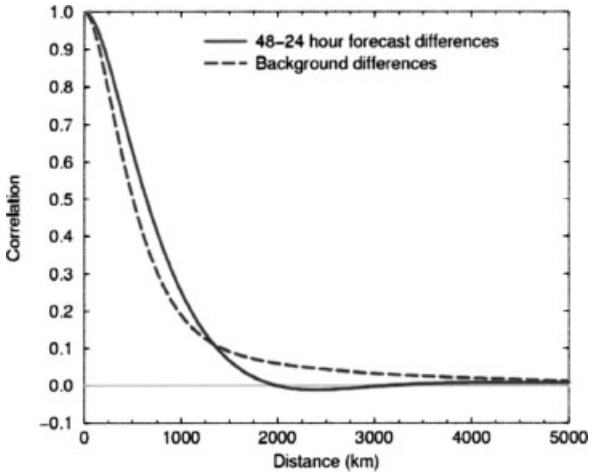

Figure 6. Comparison of horizontal autocorrelations for 500 hPa geopotential error found from the NMC (continuous lines) and ensemble (dashed line) methods of measuring background errors. ©ECMWF 2003, from Fisher (2003). Reproduced by permission of the ECMWF.

Using (26) or (27), it is possible to study flow-dependent forecast errors, as unlike the NMC method it is not always necessary to perform time averaging. This could allow valuable flow dependency to be added to VAR. This has been used by Buehner (2005) and Buehner *et al.* (2005) in this way with the Canadian 3d-VAR system, who found that it gives smaller forecast errors and improved forecasts. Other problems may arise when generating instantaneous covariances. For $N \ll n$ (where n is the size of the state space), spurious long-range correlations may appear because of undersampling, and these need to be eliminated (see references in point 3 of section 4).

5.5. Error estimates using forecast time lags

The above methods rely on the existence of a calibrated analysis system to continually produce the initial conditions for the forecast models, and such an analysis system needs the **B**-matrix in the first place. This dilemma can be dealt with by first estimating **B** crudely and then using that matrix to compute a refined version. In Polavarapu *et al.* (2005), covariances for the Canadian Middle Atmosphere Model (CMAM) data assimilation system based its error statistics on time lag forecasts. They used the following surrogate for **B**

$$\mathbf{B} \approx \frac{1}{2}\langle\{\mathbf{x}^b(t+6) - \mathbf{x}^b(t)\}\{\mathbf{x}^b(t+6) - \mathbf{x}^b(t)\}^T\rangle, \quad (28)$$

where the 6 h time lag forecast difference, $\mathbf{x}^b(t+6) - \mathbf{x}^b(t)$, is the proxy used for the 6 h forecast error (assuming that mean values have been subtracted). This method is sometimes called the 'Canadian Quick' (CQ) method.

The CQ method is the least well justified of the methods discussed, but (28) is easy to calculate from a long free run of a model, which is often all that is available when developing statistics from scratch. In common with the NMC method, this method is

DOI: 10.1002/qj

suitable only for climatological covariances, and has some drawbacks of its own. The basic assumption of the method is that forecast errors resemble forecast tendencies so that variances will be assigned according to how quickly the fields are changing. Thus, spuriously small variances will be assigned to regions that evolve slowly over the lag period. The aliasing effect of the 6 h lag will also assign small variances to oscillations that have a similar period. This is important, for example, in models with periodic boundary conditions where a feature propagates through the domain in the same time as the forecast lag (Katz, 2007). Lahoz *et al.* (2006) actually found that the analyses of temperature above 10 hPa gave better results (in terms of comparison with independent temperature measurements) when using temperature covariances derived with the CQ method than those derived with the NMC method. This probably suggests severe problems with the NMC method in the stratosphere rather than evidence that the CQ method is a suitable alternative.

6. Example measured properties of forecast errors

There have been a number of studies of the stationary three-dimensional and multivariate properties of forecast errors for both global (e.g. Derber and Bouttier, 1999; Ingleby, 2001) and mesoscale (e.g. Berre, 2000) systems. The **B**-matrix is dependent upon the forecast model, the season and the method of calculation (see section 5). Consequently, here we can only give an impression of the properties of **B** and draw from a few important examples that hold across all models. Results reviewed in this section are derived using the NMC method (section 5.2) unless otherwise stated.

6.1. Example variance functions

The simplest aspects of **B** are the variances (section 2.2). These diagonal elements of **B**, which are sometimes specified as standard deviations (the square root of variance), measure how well known forecast quantities are. When **B** is in a spatial representation, variances are a function of position. Alternatively, when **B** is in a spectral representation (i.e. $\hat{\mathbf{B}}$ in (10) when $\mathbf{V}^{\mathrm{T}}$ is a Fourier transform), variances are a function of scale. In position and spectral spaces, most of the variation occurs with latitude and total wave number, respectively, and height. Figure 7 plots the standard deviation of temperature in Northern Hemisphere winter as a function of latitude and model level for a version of the Met Office forecast model (Figure 7a; Ingleby, 2001), and as a function of total wave number and model level for a version of the ECMWF model (Figure 7b; Derber and Bouttier, 1999). Note that the level labelling is not the same for each panel (see caption).

Figure 7(a) shows that temperature variances are largely symmetric between each hemisphere. Peak values are found in the middle troposphere at midlatitudes (peaks around 500 and 250 hPa), at the top of the model and near the surface in the Northern Hemisphere. These regions are where the forecast of temperature is least accurate. Figure 7(b) is of the spectral space variances, showing the peaks in the middle troposphere. Most of the uncertainty at these levels is associated with forecast errors at large scales, between wave numbers 5 and 20.

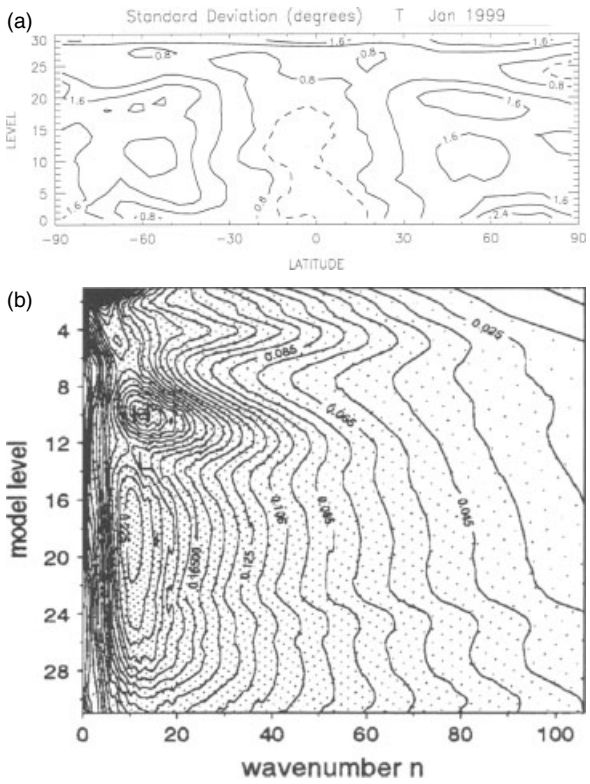

Figure 7. Standard deviations of temperature forecast error as a function of latitude and model level (a) and of total wave number and model level (b). (a) An analysis of a version of the Unified Model of the Met Office, where model levels (L) correspond roughly to L4 (850 hPa), L11 (500 hPa), L19 (250 hPa) and L25 (100 hPa). (b) An analysis of a version of the ECMWF model, where the model levels correspond roughly to L21 (650 hPa), L13 (300 hPa), L12 (250 hPa) and L10 (200 hPa). (a) ©Crown copyright 2001 from Ingleby (2001). Data supplied by the Met Office. (b) ©Blackwell Publishing 1999, from Derber and Bouttier (1999). Reproduced by permission of Blackwell Publishing.

6.2. Example horizontal correlations

The shape and extent of horizontal correlations are important as they prescribe the degree of horizontal spreading of the observational information (section 3.1). The study by Hollingsworth and Lönnberg (1986) looked at midlatitude wind errors forecast by the ECMWF model using the method of analysis of innovations (section 5.1). Figure 4 gives two results from this article that show the horizontal correlations of horizontal wind components at 200 hPa. These correlations represent horizontal contributions to the off-diagonal elements of **B**. Hollingsworth and Lönnberg did not make specific reference to the zonal and meridional winds, but instead to the longitudinal (i.e. radial) and transverse winds relative to a given point and direction.

Figure 4(a) is the error correlation of wind in the longitudinal direction between two points as a function of their

DOI: 10.1002/qj

separation. This could represent zonal wind error correlation as a function of longitudinal distance or meridional wind error correlation as a function of latitudinal distance. Correlations show a length-scale of around 600 km and are virtually always positive. Figure 4(b) is the error correlation for the transverse component of wind. This could represent zonal wind error correlation as a function of latitudinal distance or meridional wind error correlation as a function of longitudinal distance. The character of the correlation is different, showing negative correlations and an apparent shorter length-scale. Daley (1991, section 5.2) explains that these two results are consistent with atmospheric motion that is homogeneous and close to non-divergence, such as that motion which is in geostrophic balance. Under this theory, the theoretical longitudinal–longitudinal, $\rho_{\mathrm{ll}}(\Delta r)$, and transverse–transverse, $\rho_{\mathrm{tt}}(\Delta r)$, wind correlation functions are related to the stream function correlation function, $\rho_{\psi\psi}(\Delta r)$, for non-divergent conditions by the following expressions (Daley, 1991, section 5.2)

$$\rho_{\mathrm{ll}}(\Delta r) = -\frac{L_\psi^2}{\Delta r}\frac{\mathrm{d}\rho_{\psi\psi}(\Delta r)}{\mathrm{d}\Delta r}, \tag{29a}$$

$$\rho_{\mathrm{tt}}(\Delta r) = -L_\psi^2\frac{\mathrm{d}^2\rho_{\psi\psi}(\Delta r)}{\mathrm{d}\Delta r^2}. \tag{29b}$$

Here, L_ψ is the natural length-scale for stream function errors and Δr is horizontal separation. By assuming that error correlations of the stream function obey a simple isotropic correlation model, $\rho_{\psi\psi}(\Delta r) = \exp(-\Delta r^2/2L_\psi^2)$ (Daley, 1991, section 4.3), these formulae qualitatively give similar shaped correlation functions to those plotted in Figure 4, including the negative correlation part of $\rho_{\mathrm{tt}}(\Delta r)$.

The length-scales revealed in Figure 4 are expected to vary with latitude and level in a manner commensurate with the Rossby radius of deformation, L_{R}, which is a horizontal length-scale that is important to geostrophic adjustment processes. The Rossby radius of deformation valid at midlatitudes is

$$L_{\mathrm{R}} = \frac{ND}{f}, \tag{30}$$

where N is the Brunt–Väisälä frequency, D is the vertical length-scale and f is the Coriolis parameter. Length-scales would therefore be longer at low latitudes (where f is smaller) than at high latitudes, and longer in the stratosphere (where N is larger) than in the troposphere. Ingleby (2001) finds clear length-scale variations with latitude for the zonal and meridional wind variables u and v, and the mass variables p and T for a level around 250 hPa. Ingleby's results are reproduced in Figure 8. Figures 8(a) and 8(b) are, respectively, the u and v horizontal correlation functions in the east–west direction. At latitude $\sim 40^\circ$, these results are analogous to Figures 4(a) and 4(b), respectively, and show the negative correlations present in v. Length-scales are clearly lengthened towards the equator. This is true for mass variables also (Figures 8c and 8d). These have length-scales that are most similar to u near the poles, but show length-scales that are many times longer in the Tropics (Ingleby points out that T and p exhibit biases at large scales).

Evidence that length-scales increase with height is also well documented (e.g. Lönnberg and Hollingsworth, 1986; Phillips, 1986; Rabier *et al.*, 1998; Ingleby, 2001). Figure 9 is an example taken from Rabier *et al.* (1998), showing the variation of temperature, pressure and specific humidity correlation length-scales as a function of pressure (the 'modified pressure' shown in the Figure 9 is not discussed here and the specific humidity length-scale shown, as Rabier *et al.* note, is not reliable in the stratosphere). On the whole, horizontal length-scales increase with the height of those variables that may be characterized by quasi-geostrophic motion. This is expected because the Brunt–Väisälä frequency increases with height in (30).

6.3. Separability and non-separability of correlations

A property of pressure and temperature errors, evident from Figure 9, is that the horizontal length-scales of pressure (or equivalently the geopotential height field, Φ) are much larger than those of temperature. This has consequences regarding the structure of the correlation function that is allowed, given that the atmosphere is predominantly in a state of hydrostatic balance:

$$\frac{\partial\Phi}{\partial z} = \frac{RT}{H}. \tag{31}$$

In (31), which applies also to error perturbations, R is the specific gas constant and H is the scale height. Hydrostatic balance provides a qualitative reason why geopotential height errors are larger scale than temperature errors. The vertical derivative in (31) means that vertical length-scales of temperature should be shorter than those of geopotential height. The effect of three-dimensional isotropy, where errors with a large vertical length-scale also have a large horizontal length-scale, means that also the horizontal length-scales of temperature should be shorter than of geopotential height.

Equation (31) also puts restrictions on the separable nature of correlation functions of temperature and height. Let us propose for now that the correlation functions for temperature and height are separable. A separable function is one that is the product of two functions, one depending upon horizontal position, r, but not vertical position, z, with the other depending upon vertical but not horizontal position. Error correlation functions are columns of the matrix $\mathbf{C}$ in (9), which can be derived from $\mathbf{B}$ via (4) and (9). The correlations for temperature error, δT between positions (r_1, z_1) and (r_2, z_2) (and for simplicity assuming temperature standard deviations, σ_T,

DOI: 10.1002/qj

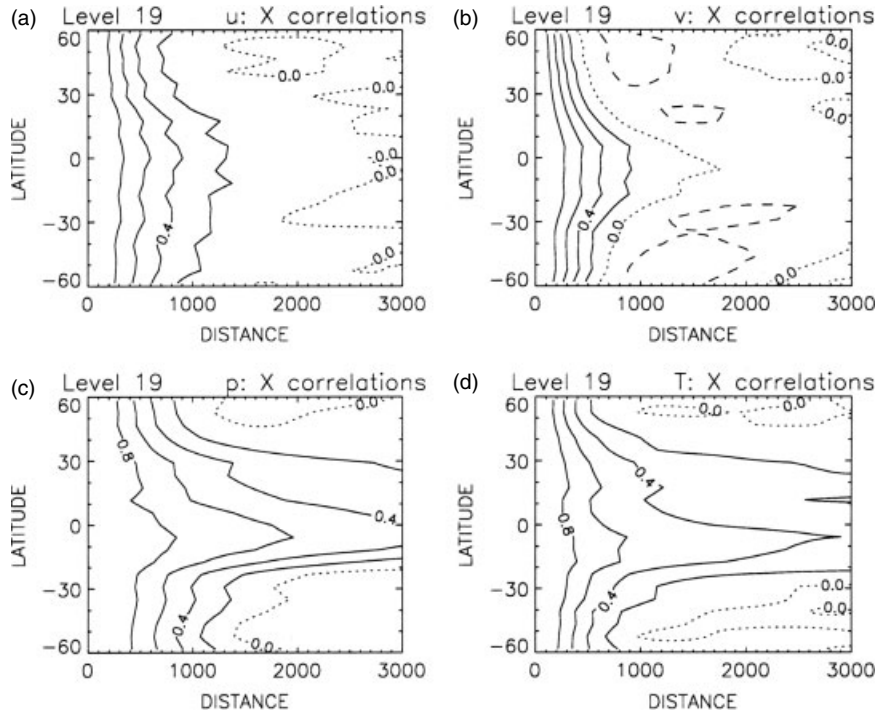

Figure 8. Horizontal east–west error correlations for (a) zonal wind, (b) meridional wind, (c) pressure and (d) temperature errors. Positive (negative) correlations are continuous (dashed) lines, and the zero correlation line is dotted. ©Crown copyright 2001 from Ingleby (2001). Data supplied by the Met Office.

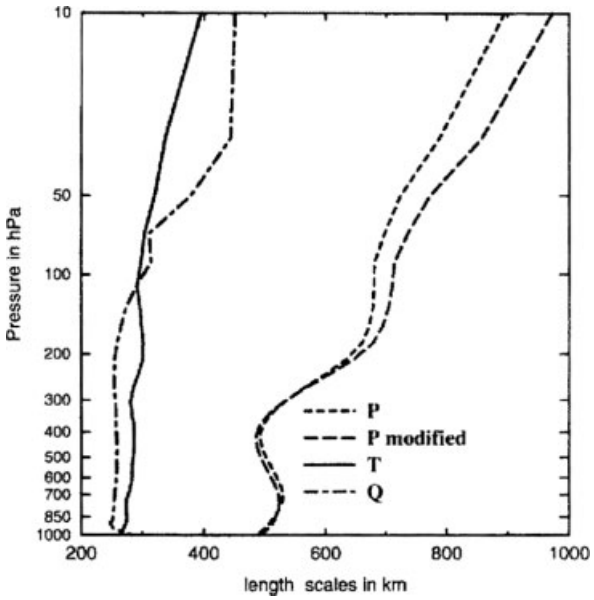

Figure 9. Variation of the horizontal length-scale of temperature, pressure and specific humidity errors as a function of pressure. ©Royal Meteorological Society 1998, from Rabier *et al.* (1998). Reproduced with permission of the Royal Meteorological Society.

are homogeneous) gives

$$\begin{aligned} \mathrm{C}_T(r_2-r_1, z_2-z_1) &= \frac{1}{\sigma_T^2}\langle \delta T(r_1,z_1)\delta T(r_2,z_2)\rangle \\ &= R_T(r_2-r_1) \\ &\quad \times Z_T(z_2-z_1). \end{aligned} \tag{32}$$

In (32), the correlation function is assumed to be separable and made up of the horizontal and vertical functions R_T and Z_T, respectively. Substituting (31) into (32) gives

$$\begin{aligned} \mathrm{C}_T(r_2-r_1, z_2-z_1) &= \frac{H^2}{R^2}\frac{1}{\sigma_T^2} \\ &\quad \times \left\langle \frac{\partial \delta\Phi(r_1,z_1)}{\partial z_1}\frac{\partial \delta\Phi(r_2,z_2)}{\partial z_2}\right\rangle \\ &= \frac{H^2}{R^2}\frac{1}{\sigma_T^2}\frac{\partial^2}{\partial z_1 \partial z_2} \\ &\quad \times \langle \delta\Phi(r_1,z_1)\delta\Phi(r_2,z_2)\rangle\,. \end{aligned} \tag{33}$$

The equivalent of (32), but for correlations of $\delta\Phi$ is

$$\begin{aligned} \mathrm{C}_\Phi(r_2-r_1, z_2-z_1) &= \frac{1}{\sigma_\Phi^2}\langle \delta\Phi(r_1,z_1)\delta\Phi(r_2,z_2)\rangle \\ &= R_\Phi(r_2-r_1) \\ &\quad \times Z_\Phi(z_2-z_1). \end{aligned} \tag{34}$$

Putting (32)–(34) together results in

$$\begin{aligned} R_T(r_2-r_1)Z_T(z_2-z_1) &= \frac{H^2}{R^2}\frac{\sigma_\Phi^2}{\sigma_T^2}R_\Phi(r_2-r_1) \\ &\quad \times \frac{\partial^2 Z_\Phi(z_2-z_1)}{\partial z_1 \partial z_2}. \end{aligned} \tag{35}$$

The horizontal parts of each side of (35) are related only via a factor, implying that (for separable correlation functions and conditions of hydrostatic balance) the horizontal length-scales of height and temperature errors

DOI: 10.1002/qj

should be the same. As it is observed in Figure 9 that mass length-scales are significantly longer than temperature length-scales, the conclusion is reached that the separability assumption cannot hold. Separability of variables was common in early operational data assimilation schemes, but non-separability is an important consideration when building models of background error covariances for use in modern VAR systems (see Part II).

6.4. Example vertical correlations

There are other manifestations of non-separability, which are exhibited also by other variables. Ingleby (2001), for example, shows error correlations between temperature at ~ 500 hPa and other model levels as a function of wave number (Figure 10a). Apart from very small wave numbers, the vertical length-scale shortens with increasing wave number or, equivalently, with decreasing horizontal length-scale. This effect is another example of three-dimensional isotropy (see section 6.3).

Vertical correlations depend on position as well as scale. Figure 10(b), from Ingleby (2001), shows how vertical correlations of temperature error vary with latitude. Vertical length-scales appear narrow in the Tropics, increase poleward with a peak at 30–45° in each hemisphere, and reduce slightly towards the poles. There are significant negative correlations with levels in the upper troposphere and lower stratosphere, which are presumably due to hydrostatic processes (see Figure 3d). Some of these characteristics are robust with regard to the method used to calculate the correlations. Belo Pereira and Berre (2006) compare such correlations of temperature error between the NMC method (section 5.2), as used by Ingleby, and the ensemble method (section 5.4). Length-scales from the ensemble method are narrower overall compared to NMC, but still show that the region of narrowest correlation lengths is in the Tropics. The reason for the shorter vertical length-scales in the Tropics compared to the extratropics has not been discussed in the literature, but a possible explanation is the higher baroclinicity of weather systems in the extratropics than those in the Tropics. Baroclinic systems may be expected to have broader vertical length-scales than barotropic systems.

The property of three-dimensional isotropy is violated in the Tropics where vertical length-scales are short (Figure 10b) and horizontal length-scales are long (Figure 8d). Close to the equator, (30) is not valid, but the following can be applied instead

$$L_R = \sqrt{\frac{NDa}{\Omega}}, \tag{36}$$

(Ingleby, 2001), where a is the Earth's radius and Ω is its angular rotation rate. Because a is very large, (36) allows L_R to be large while D is short. The notion of three-dimensional isotropy is restored poleward of the Tropics.

6.5. Other influences of horizontal length-scales

Geostrophic and hydrostatic adjustment processes are important factors that determine the lengths and structures of error correlations (e.g. (30), (31) and (36)), and are used in Part II to formulate models of forecast errors. However, there are also other factors. For instance,

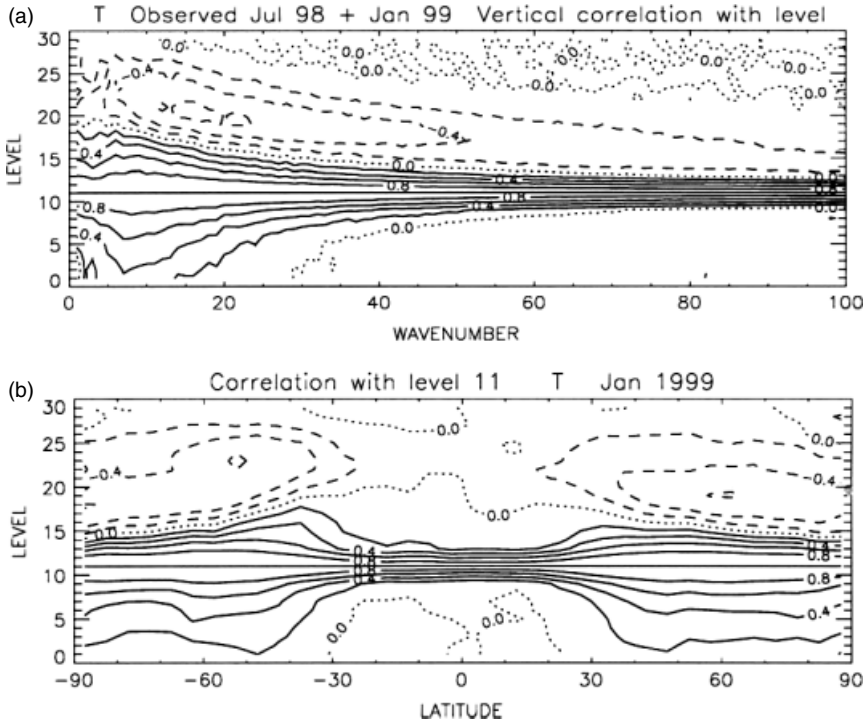

Figure 10. Vertical correlations of temperature error at ~ 500 hPa with other levels as a function of (a) horizontal wave number and (b) latitude. ©Crown copyright 2001, from Ingleby (2001). Data supplied by the Met Office.

DOI: 10.1002/qj

the large value of N in (30) in the stratosphere is only partly responsible for the long length-scales found there. Wave breaking in the stratosphere by planetary waves emanating from the troposphere is also thought to contribute to the long length-scales. In this mechanism, only planetary scale waves are usually permitted to reach stratospheric altitudes owing to the very low transmissivity of the stratosphere to short waves (Charney and Drazin, 1961).

Observation density is also known to affect forecast error correlation length-scales (Belo Pereira and Berre, 2006). Even though length-scales of balanced variables at midlatitudes are determined largely by (30), forecast errors over land and in the Northern Hemisphere are found to have shorter length-scales than those over the sea and in the Southern Hemisphere (at equivalent heights or pressures). This is because of the higher observation density over Northern Hemisphere continents than over other parts of the world (the advent of satellite data has helped partly to homogenize the observation density in recent years).

The way that the observation density affects the error correlation length can be understood by considering how the assimilation of a single observation of an arbitrary scalar quantity modifies the variance spectrum of forecast errors. In the spectral representation, the variance spectrum of $\mathbf{B}$ for the quantity in question comprises the diagonal elements, $\sigma_{\mathrm{b}}^2(k)$, where k is wave number. With off-diagonal elements zero, the variance spectrum is equivalent to a homogeneous background error covariance in positional space (homogeneity is often enforced in background error covariances; see Part II). The variance spectrum of the analysis error, $\sigma_{\mathrm{a}}^2(k)$, after assimilation of the observation, is found by looking at the equation for analysis error (the inverse Hessian), $\mathbf{P}^{\mathrm{a}} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}$ in spectral space. For a single observation and maintaining homogeneity, $\sigma_{\mathrm{a}}^2(k)$ (the diagonal elements of $\mathbf{P}^{\mathrm{a}}$) become

$$\sigma_{\mathrm{a}}^2(k) = \left\{ \frac{1}{\sigma_{\mathrm{b}}^2(k)} + \frac{1}{N\sigma_{\mathrm{Ob}}^2} \right\}^{-1} = \frac{N\sigma_{\mathrm{Ob}}^2 \sigma_{\mathrm{b}}^2(k)}{N\sigma_{\mathrm{Ob}}^2 + \sigma_{\mathrm{b}}^2(k)}, \tag{37}$$

where the observation has error variance σ_{Ob}^2, and the spectrum comprises N wave numbers (see Bannister, 2006, for a derivation of this result). Equation (37) is applied to a simple one-dimensional system that has the red forecast error variance spectrum in Figure 11 (continuous curve). (It is often found that background error spectra tail-off at large wave numbers (e.g. Berre, 2000). Assimilating the observation gives the analysis error variance spectrum (short-dashed curve), which has been found from (37) using those parameters described in the caption of Figure 11. The high wave numbers, which are forecast well (where $\sigma_{\mathrm{b}}^2(k) \ll N\sigma_{\mathrm{Ob}}^2$) are unaffected by the observation, but variances of other wave numbers are reduced, allowing the observation to reduce the variance of the large scales more than for the small scales. The widths of the analysis and background spectra can be compared by normalizing $\sigma_{\mathrm{a}}^2(k)$ (long-dashed curve). This is broader than the background spectrum, which indicates a shorter length-scale in positional space.

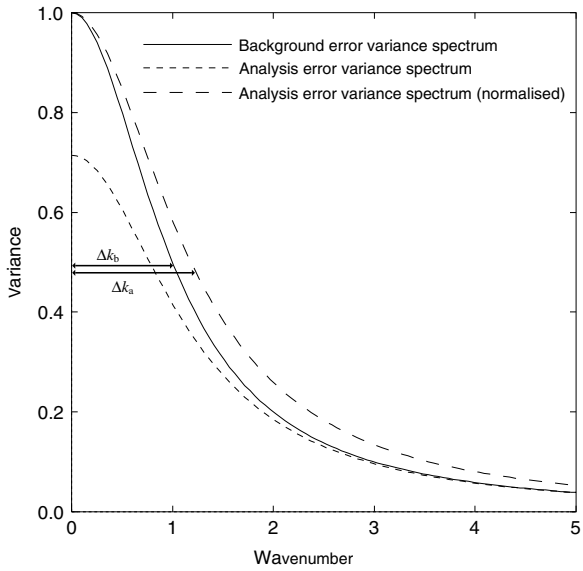

Figure 11. Example error variance spectra for a background state and the analysis state after assimilation of an observation from (37) (with and without normalization). The parameters used in (37) are: $\sigma_{\mathrm{b}}^2(k) = \sigma_{\mathrm{b}}^2(0)/(1+k^2)$, $\sigma_{\mathrm{b}}^2(0) = 1$, $N = 100$ and $\sigma_{\mathrm{Ob}}^2 = 0.025$ (σ_{Ob}^2 is extremely small to exaggerate the effect of a single observation).

The analysis error covariance can be propagated forward by the forecast model using (3) to give a revised forecast error covariance. If the model run is sufficiently long, then the length-scales in the new forecast error covariance matrix, $\mathbf{P}^{\mathrm{f}}$, will revert to their natural values described by $\mathbf{B}$. However, in a short forecast, the shorter length-scales in $\mathbf{P}^{\mathrm{a}}$ are likely to give rise to length-scales that are shorter in $\mathbf{P}^{\mathrm{f}}$ also. Based on this analysis, the more observations that are present in the calibration process of $\mathbf{B}$, the shorter the characteristic length-scales. Any substantial change in the observation network would therefore require a recalibration of $\mathbf{B}$.

This analysis is an oversimplification of a real case; for instance, unlike in Figure 11, forecast spectra rarely decrease monotonically with wave number. We have also paid no attention to regional dependence on length-scale, although it is possible to study this with a toy system (Bannister, 2006).

6.6. Example multivariate covariances

It has already been shown in Figure 3 how geostrophic and hydrostatic relations between mass and wind variables lead to multivariate couplings in $\mathbf{B}$. The analysis increments shown in Figures 2 and 3 are proportional to the structure function (column of $\mathbf{B}$), in this example associated with pressure at a particular location 5.5 km above the Pacific Ocean. The multivariate covariances allow information to be gained on variables other than those measured. In the example shown, the pressure observation made in Figure 2 will yield wind and temperature contributions to the analysis increment, as illustrated in

DOI: 10.1002/qj

Figure 3. It will also allow nearby wind and temperature observations to influence pressure, and to act synergistically with the pressure observation (section 3.4).

There are many other examples of multivariate covariances in the literature (see previous references in this section). An example is drawn from Berre (2000) and involves a moisture variable (as moisture has not yet been discussed). The coupling between errors in dynamic, thermodynamic and moisture variables is complicated because of the range of diabatic and adiabatic, and local and non-local, processes that can govern them. Figure 12 shows the vertical covariance between temperature (T) errors and specific humidity (q) errors, averaged over the domain of the regional model used (2300×2300 km^2 over France). It shows that T and q errors are coupled significantly in the vertical in a complicated, non-local fashion. Figure 12 shows that q covaries positively with T for T perturbations up to the $\sim$ 800 hPa level. Above this, q covaries negatively within a few tens of hPa about the T perturbation. It also shows that the level of peak q response does not coincide exactly with the level of the T perturbation (and vice versa).

Berre suggests two separate processes that may account for these covariances. The first is associated with so-called 'balanced' motion in the atmosphere. The term 'balance' here refers to processes associated with vorticity perturbations, and other perturbations that are correlated with vorticity, such as part of the divergence perturbation (such concepts arise in the discussion of background error covariance models in Part II). The poleward advection of warm and moist air (e.g. in warm cyclonic conveyor belts) and the associated convective moistening of the troposphere brought about by warming help explain the positive covariances in Figure 12. Berre goes on to suggest that a second process is responsible for the negative covariances, close to the diagonal of Figure 12 above the 800 hPa level. This is the latent heating of the atmosphere as a consequence of dehydration through condensation (e.g. Holton, 1992, section 11.3), which creates a negative relationship between temperature and specific humidity.

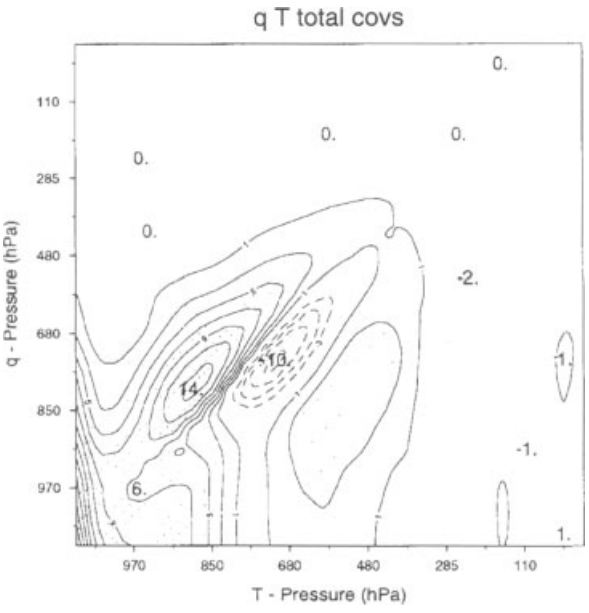

Figure 12. Midlatitude regional mean of the vertical error covariance matrix between temperature and specific humidity. Contours for negative covariances are dashed. ©American Meteorological Society (AMS) 2000, from Berre (2000). Reproduced by permission of the AMS.

The error covariances in Figure 12 are for a midlatitude piece of the atmosphere, and averaged over three months (December to February) for a particular year. The geographical and temporal variations of such covariances are also likely to be significant. The tropical covariances between temperature and specific humidity are likely to be very different, owing to the different relative importance of the moist processes there, and there is likely to be important flow dependency (e.g. whether the air is rising or falling and whether the sky is cloudy or cloud-free). These complexities serve to highlight the huge amount of information needed to represent background error covariances in VAR. Even though it is possible to capture some flow dependences in a rudimentary way in VAR, for example, through 'cycling' of variances (Derber and Bouttier, 1999) or enforcing relationships between errors through the use of flow-dependent balance relations (Fisher, 2003), a significant focus is still on gaining a good representation of the 'climatological' component of background errors, which should contain as many of the characteristics referred to in this section as possible.

7. Concluding remarks

Even though the Earth's atmosphere is relatively well observed, an *a-priori* state is still necessary for data assimilation, and comes in the form of a short numerical forecast. The uncertainty of the *a-priori* state is described by a PDF, which is approximated by a static (or quasi-static) **B**-matrix. This use of a covariance matrix assumes that the forecast error statistics are normally distributed. In this article, we have reviewed many aspects of the **B**-matrix in the context of meteorological data assimilation.

The **B**-matrix can manifest itself in the assimilation in subtle ways: it spreads information both univariately and multivariately in ways that, in large-scale systems, are consistent with near-balanced motion (geostrophic and hydrostatic), it defines a subspace for the analysis increment, and it acts as a means in which many observations can interfere constructively. These manifestations serve to highlight the importance of **B** in VAR, but **B** remains a matrix that is very difficult to deal with, owing to difficulties with measuring it and because of its prohibitive size.

Many techniques have been developed to allow measurement of **B**, including the analysis of innovations, the analysis of forecast differences with NMC or time-lag methods, and most recently with ensemble-based approaches. All of these methods have their difficulties and limitations. The popular NMC method is believed to give results that are physically reasonable, but performs poorly for large scales, and in those parts of the

DOI: 10.1002/qj

atmosphere that are void of observations. The ensemble methods provide potentially the most realistic simulation of forecast errors and are beginning to be used by some operational centres, but these are expensive to use.

To date, most knowledge about forecast error covariances in weather forecasting models has been derived from the NMC method. A small collection of plots has been drawn from the literature to illustrate some important physical properties of the stationary component of **B**, including variances, univariate horizontal and vertical correlations, and multivariate covariances. These results show that errors statistics have a complicated three-dimensional structure, which is determined by a combination of factors, including dynamics, observation density and other forcing mechanisms. The results have focused on dynamical fields (wind, pressure, temperature and specific humidity), but lately models have been extended to include prediction (and hence assimilation) of chemical constituents, particularly ozone, which needs to be included in the **B**-matrix (Dethof and Hólm, 2004). Lahoz, Fonteyn and Swinbank (2007) review general issues of chemical data assimilation, including those related to forecast errors. There is also a body of work with respect to errors in oceanic forecasts for VAR (e.g. Weaver *et al.*, 2005).

How the **B**-matrix is used in a pragmatic and compact way in VAR is covered in Part II, and techniques and examples are drawn from knowledge of current synoptic-scale atmospheric models. Models are becoming more complex, and often are coupled to other models, for example, atmosphere-ocean (Knopf *et al.*, 2005; Saha *et al.*, 2006) and atmosphere-chemistry (Eyring *et al.*, 2006). They are also gaining finer resolution, for example, to simulate convective scale phenomena explicitly for meteorological (Zhang, 2005) and air-quality prediction (Seaman, 2003). For these developments to be useful, each requires a data assimilation system that is appropriate. This includes a suitable representation of forecast errors, which will be strongly dependent on the nature of the system. There is major interest in understanding the structure of forecast errors in these next generation of Earth-system models, which applies whether the assimilation system is of a VAR or an ensemble Kalman filter type.

There are many related topics that are outside the scope of this article, but should be mentioned for completeness. Non-Gaussian errors are not captured with a covariance matrix and so are not accounted for in ordinary assimilation systems. PDFs that are non-Gaussian in shape, but which are modelled as Gaussian (via their first and second moments, as in (1)) may give rise to an incorrect analysis. Despite this, issues of non-Gaussianity have been left largely unaddressed in operational systems, apart from some notable studies, for example, in connection with atmospheric humidity (Hólm *et al.*, 2002). Fletcher and Zupanski (2006) have looked specifically at log-normal distributions. There are contributions of Gaussian errors that are normally unaccounted for in VAR that are not discussed, namely proper flow dependences (e.g. Zhang, 2005) and four-dimensional aspects (Lorenc, 2003a) of forecast error statistics, and covariances between the background and the observations (Lorenc, 1986). By improving the representation of errors in data assimilation (e.g. by using a better **B**-matrix) may lead to improvements in the quality of knowledge (and its measure of uncertainty) in geophysical systems for applications such as weather forecasting, pollution and flood forecasting.

Acknowledgements

I would like to thank the many people who have influenced my understanding of data assimilation over the past few years: Phil Andrews, Sue Ballard, Paul Berrisford, Roger Brugge, Mike Cullen, Sarah Dance, Mark Dixon, Martin Ehrendorfer, John Eyre, Mike Fisher, Alan Geer, Bruce Ingleby, David Jackson, Mike Keil, William Lahoz, Amos Lawless, Andrew Lorenc, Stefano Migliorini, Alan O'Neill, Nancy Nichols, Dave Pearson, Ian Roulstone, Olaf Stiller, Hamish Struthers, Sean Swarbrick, Richard Swinbank and Matt Szyndel. Particular thanks go to Martin Ehrendorfer, Mike Fisher, Bruce Ingleby, Andrew Lorenc and Alan O'Neill for reading the draft manuscript. I would like to thank three anonymous reviewers for their valuable comments and also the authors of the work cited in this article. I gratefully acknowledge the AMS, Blackwell Publishing, the ECMWF, the Met Office and the Royal Meteorological Society for permission to reproduce figures. The author is funded by the Natural Environment Research Council.

References

Bannister RN. 2006. Collection of results on background error covariance models. DARC Internal Report, **7** (available from Data Assimilation Research Centre, Department of Meteorology, University of Reading, Reading, RG6 6BB, UK; http://darc.nerc.ac.uk/v2/internalreports.html).

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**. DOI: 10.1002/qj.340.

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.

Barlow R. 1989. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. John Wiley: New York.

Bartello P, Mitchell HL. 1992. A continuous three-dimensional model of short-range forecast error covariances. *Tellus* **44A**: 217–235.

Belo Pereira MB, Berre L. 2006. The use of an ensemble approach to study the background error covariances in a global NWP model. *Mon. Weather Rev.* **134**: 2466–2489.

Berre L. 2000. Estimation of synoptic and mesoscale forecast error covariances in a limited area model. *Mon. Weather Rev.* **128**: 644–667.

Berre L, Ştefănescu SE, Pereira MB. 2006. The representation of the analysis effect in three error simulation techniques. *Tellus* **58A**: 196–209.

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.

Bouttier F. 1996. 'Application of Kalman filtering to numerical weather prediction'. Pp. 61–90 in Proceedings of the 1996 ECMWF seminar on data assimilation, 2–6 September 1996. ECMWF: Reading, UK.

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, Gauthier P, Liu Z. 2005. Evaluation of new estimates of background and observation error covariances for variational assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3373–3384.
Charney JG, Drazin PG. 1961. Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.* **66**: 83–109.
Courtier P, Thépaut J-N, Hollingsworth A. 1994. A strategy for operational implementation of 4d-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1387.
Courtier P, Andersson E, Heckley W, Pailleux J, Vasiljevic D, Hollingsworth A, Fisher M, Rabier F. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.* **124**: 1783–1807.
Daley R. 1991. *Atmospheric Data Analysis*. Cambridge University Press: UK.
Dee DP. 2004. 'Variational bias correction of radiance data in the ECMWF system'. Pp. 97–112 in ECMWF Workshop on Assimilation of high spectral resolution sounders in NWP, 28 June–1 July 2004. ECMWF: Reading, UK.
Dethof A, Hólm EV. 2004. Ozone assimilation in the ERA-40 reanalysis project. *Q. J. R. Meteorol. Soc.* **130**: 2851–2872.
Derber J, Bouttier F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus* **51A**: 195–221.
Ehrendorfer M. 2007. A review of issues in ensemble-based Kalman filtering *Meteorol. Z.* **16**: 795–818.
Evensen G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics* **53**: 343–367.
Eyring V, Butchart N, Waugh DW, Akiyoshi H, Austin J, Bekki S, Bodeker GE, Boville BA, Bruhl C, Chipperfield MP, Cordero E, Dameris M, Deushi M, Fioletov VE, Frith SM, Garcia RR, Gettelman A, Giorgetta MA, Grewe V, Jourdain L, Kinnison DE, Mancini E, Manzini E, Marchand M, Marsh DR, Nagashima T, Newman PA, Nielsen JE, Pawson S, Pitari G, Plummer DA, Rozanov E, Schraner M, Shepherd TG, Shibata K, Stolarski RS, Struthers H, Tian W, Yoshiki M. 2006. Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.* **111**(:): D22308 (doi:10.1029/2006JD007327).
Fisher M. 2001. 'Assimilation techniques (4): 4dVar'. ECMWF Meteorological Training Course Lecture Series (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).
Fisher M. 2003. 'Background error covariance modelling'. Pp. 45–64 in ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8–12 September 2003. ECMWF: Reading UK.
Fisher M, Andersson E. 2001. Developments in 4d-Var and Kalman filtering. ECMWF Tech. Memo., **347** (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).
Fisher M, Leutbecher M, Kelly GA. 2005. On the equivalence between Kalman smoothing and weak constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3235–3246.
Fletcher SJ, Zupanski M. 2006. A data assimilation method for log-normally distributed observational errors. *Q. J. R. Meteorol. Soc.* **132**: 2505–2519.
Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **125**: 723–757.
Gustafsson N, Berre L, Hornquist S, Huang XY, Lindskog M, Navascues B, Mogensen KS, Thornsteinsson S. 2001. Three-dimensional variational data assimilation for a limited area model. *Tellus* **53A**: 425–446.
Haltiner GJ, Williams RT. 1980. *Numerical Prediction and Dynamic Meteorology*, 2nd edition. John Wiley: London.
Hamill TM, Mullen SL, Snyder C, Toth Z, Baumhefner DP. 2000. Ensemble forecasting in the short to medium range: Report from a workshop. *Bull. Am. Meteorol. Soc.* **81**: 2653–2664.
Hamill TM, Whitaker JS, Snyder C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filer. *Mon. Weather Rev.* **129**: 2776–2790.
Hollingsworth A, Lönnberg P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus* **38A**: 111–136.
Hólm E, Andersson E, Beljaars A, Lopez P, Mahfouf JF, Simmons A, Thépaut J-N. 2002. Assimilation and modelling of the hydrological cycle: ECMWF's status and plans. ECMWF Tech. Memo., **383** (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).
Holton JR. 1992. *An Introduction to Dynamic Meteorology*. Academic Press: New York.
Houtekamer PL, Mitchell HL. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**: 123–137.
Houtekamer PL, Lefaivre L, Derome J, Ritchie H, Mitchell HL. 1996. A system simulation approach to ensemble prediction. *Mon. Weather Rev.* **124**: 1225–1242.
Ide K, Coutier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn* **75**: 181–189.
Ingleby NB. 2001. The statistical structure of forecast errors and its representation in the Met Office global three-dimensional variational data assimilation system. *Q. J. R. Meteorol. Soc.* **127**: 209–231.
Isaksen L, Fisher M, Berner J. 2007. 'Use of analysis ensembles in estimating flow-dependent background error variance'. Pp. 65–86 in ECWMF Workshop Proceedings on Flow-dependent Aspects of Data Assimilation, 11–13 June 2007. ECMWF: Reading, UK.
Järvinen H. 2001. Temporal evolution of innovation and residual statistics in the ECMWF variational data assimilation systems. *Tellus* **53A**: 333–347.
Johnson C, Hoskins BJ, Nichols NK. 2005. A singular vector perspective of 4d-Var: Filtering and interpolation. *Q. J. R. Meteorol. Soc.* **131**: 1–19.
Kalman RE. 1960. A new approach to linear filter and prediction problems. *J. Basic. Eng.* **82**: 35–45.
Kalnay E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press: Cambridge, UK.
Katz D. 2007. *The application of PV-based control variable transformations in variational data assimilation*. PhD Thesis, Reading University, Department of Mathematics.
Knopf B, Held H, Schellnhuber HJ. 2005. Forced versus coupled dynamics in Earth system modelling and prediction. *Nonlinear Processes in Geophys.* **12**: 311–320.
Lahoz WA, Geer AJ, Bekki S, Bormann N, Ceccherini S, Elbern H, Errera Q, Eskes HJ, Fonteyn D, Jackson DR, Khattatov B, Massart S, Peuch V-H, Rharmill S, Ridolfi M, Segers A, Talagrand O, Thornton HE, Vik AF, von Clarmann T. 2006. The assimilation of Envisat Data (ASSET) project. *Atmos. Chem. Phys. Discuss.* **6**: 12769–12824.
Lahoz WA, Fonteyn D, Swinbank R. 2007. Data assimilation of atmospheric constituents: A review. *Atmos. Chem. Phys. Discuss.* **7**: 9561–9633.
Lanczos C. 1988. *Applied Analysis*. Dover Publications: New York.
Lee PM. 1997. *Bayesian Statistics: An Introduction*. Arnold Publishers: London.
Lewis JM, Lakshmivarahan S, Dhall SK. 2006. *Dynamic Data Assimilation: A Least-squares Approach*. Cambridge University Press: Cambridge, UK.
Li ZJ, Navon IM. 2001. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.* **127**: 661–683.
Lönnberg P, Hollingsworth A. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors. *Tellus* **38A**: 137–161.
Lorenc AC. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Weather Rev.* **109**: 701–721.
Lorenc AC. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**: 1177–1194.
Lorenc AC. 2003a. Modelling of error covariances by 4d-Var assimilation. *Q. J. R. Meteorol. Soc.* **129**: 3167–3182.
Lorenc AC. 2003b. The potential of the ensemble Kalman filter for NWP – a comparison with 4d-Var. *Q. J. R. Meteorol. Soc.* **129**: 3183–3203.
Lorenc AC, Rawlins F. 2005. Why does 4D-Var beat 3D-Var? *Q. J. R. Meteorol. Soc.* **131**: 3247–3257.
Lui Z-Q, Rabier F. 2002. The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc.* **128**: 1367–1386.
Lynch P, Huang X-Y. 1992. Initialization of the HIRLAM model using a digital filter. *Mon. Weather Rev.* **120**: 1019–1034.
Ménard R, Daley R. 1996. The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus* **48A**: 221–237.
Pailleux J. 1997. Large-scale data assimilation systems: atmospheric applications to numerical weather prediction. *J. Meterol. Soc. Jpn* **75**: 347–358.

Parrish DF, Derber JC. 1992. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.* **120**: 1747–1763.

Phillips NA. 1986. The spatial statistics of random geostrophic modes and first-guess errors. *Tellus* **38A**: 314–332.

Polavarapu S, Ren S, Rochon Y, Sankey D, Ek N, Koshyk J, Tarasick D. 2005. Data assimilation with the Canadian middle atmosphere model. *Atmos.–Ocean* **43**: 77–100.

Rabier F. 2005. Overview of global data assimilation developments in numerical weather prediction centres. *Q. J. R. Meteorol. Soc.* **131**: 3215–3233.

Rabier F, Mahfouf J-F, Fisher M, Järvinen H, Simmons AJ, Andersson E, Bouttier F, Courtier P, Hamrud M, Haseler J, Hollingsworth A, Isaksen L, Klinker E, Saarinen S, Temperton C, Thépaut J-N, Undén P, Vasiljevic D. 1997. Recent experimentation on 4D-Var and first results from a Simplified Kalman Filter. ECMWF Tech. Memo., **240** (available from ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK).

Rabier F, McNally A, Andersson E, Courtier P, Undén P, Eyre J, Hollingsworth A, Bouttier F. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.* **124**: 1809–1829.

Rutherford ID. 1972. Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.* **29**: 809–815.

Sadiki W, Fischer C. 2005. A posteriori validation applied to the 3D-VAR Arpège and Aladin data assimilation systems *Tellus* **57A**: 21–34.

Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool HM, Pan HL, Moorthi S, Behringer D, Stokes D, Pena M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP Climate Forecast System. *J. Clim.* **19**: 3483–3517.

Schlatter TW. 2000. Variational assimilation of meteorological observations in the lower atmosphere: a tutorial on how it works. *J. Atmos. Solar–Terr. Phys.* **62**: 1057–1070.

Seaman NL. 2003. Future directions of meteorology related to air quality research. *Environ. Int.* **29**: 245–252.

Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **131**: 3079–3102.

Široká M, Fischer C, Cassé V, Brožkova R, Geleyn J-F. 2003. The definition of mesoscale selective forecast error covariances for a limited area variational analysis. *Meteorol. Atmos. Phys.* **82**: 227–244.

Thépaut J-N, Courtier P, Belaud G, Lemaitre G. 1996. Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Q. J. R. Meteorol. Soc.* **122**: 535–561.

Weaver AT, Deltel C, Machu E, Ricci S, Daget N. 2005. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.* **131**: 3605–3626.

Žagar N, Andersson E, Fisher M. 2005. Balanced tropical data assimilation based on a study of equatorial waves in ECMWF short-range forecast errors. *Q. J. R. Meteorol. Soc.* **131**: 987–1011.

Zhang F. 2005. Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts. *Mon. Weather Rev.* **133**: 2876–2893.

Zupanski D. 1997. A general weak constraint applicable to operational 4d-Var data assimilation systems. *Mon. Weather Rev.* **125**: 2274–2291.

DOI: 10.1002/qj