DARC Internal Report No. 7: Collection of results on background error covariance models

Ross N. Bannister
National Centre for Earth Observation (NCEO)
(formerly Data Assimilation Research Centre (DARC))
Dept. of Meteorology, University of Reading, RG6 6BB, UK r.n.bannister@reading.ac.uk
July 2006
(last modified 10th August 2009)

This document is a collection of discussion points and results that I have encountered or derived while studying the theory of background error covariances. The sections are in no particular order.

Contents

1. Inhomogeneity of the ECMWF wavelet transform

1.1 Introduction

The ECMWF wavelet transform is the spatial/univariate part of the control variable transform, $\mathbf{B}_s^{1/2}$. It is of the form

$$\mathbf{B}_s^{1/2}\chi' = \Sigma_B \sum_{j=1}^{K} \psi_j \otimes \mathbf{C}_j^{1/2}\chi'_j, \tag{1.1}$$

where Σ_B is the diagonal matrix of grid-point standard deviations. Other symbols are associated with the fact that the control vector χ' has contributions from K different bands in the following way

$$\chi' = \begin{pmatrix} \chi'_1 \\ \chi'_2 \\ \cdots \\ \chi'_K \end{pmatrix}. \tag{1.2}$$

Each band is associated with the data assimilation state for a range of scales. Band j is represented by control subvector χ'_j; it is a function of longitude, latitude and height. Matrix $\mathbf{C}_j^{1/2}$ is a vertical covariance matrix which operates on χ'_j, and represents a different vertical covariance at different horizontal positions. ψ_j is wavelet j and is a function of horizontal position. Fisher (2003) gives some example plots of wavelet functions, illustrating the different characteristic scales for a number of bands. Fisher (2003) also explains the form of (1.1). The general idea of this control variable transform is to allow the vertical correlations to be a

function of position (as $\mathbf{C}_j^{1/2}$ above does explicitly), and concurrently, to allow them to be scale dependent also (as $\mathbf{C}_j^{1/2}$ is a function of band, j).

This section sets out to show that this transform - which implies a covariance matrix - does not yield correlation functions that are homogeneous. This is important in meteorological data assimilation, which requires inhomogeneous correlation functions.

1.2 Implied covariance matrix formula

The control variable transform, $\mathbf{B}_s^{1/2}$ relates the control vector to an incremental vector in the same space as the model variables, ie

$$\boldsymbol{x}' = \mathbf{B}_s^{1/2}\chi'. \tag{1.3}$$

The control vector is meant to comprise elements that have no forecast error correlation, and whose variances are unity, thus $\langle \chi'\chi'^T \rangle$, where angular brackets indicate average over possible values. The covariances that are implied by (1.1) are then

$$\mathbf{B}_s = \langle \boldsymbol{x}'\boldsymbol{x}'^T \rangle = \mathbf{B}_s^{1/2}\mathbf{B}_s^{T/2}. \tag{1.4}$$

1.3 Convolution theorem

In order to evaluate (1.4), the transpose of (1.1) is required. What is the transpose of the convolution operator? The need to do a convolution can be avoided by judicious use of Fourier transforms. Consider the convolution of a function, $f(x)$, by another, $g(x)$. The convolution theorem says that a convolution is simplified in Fourier space. In Fourier space a convolution is equivalent to the product of the two functions' Fourier transforms, $\hat{f}(k)$ and $\hat{g}(k)$. Algebraically

$$\mathbf{S}(f(x) \otimes g(x)) = \hat{f}(k)\hat{g}(k), \tag{1.5}$$

$$\hat{f}(k) = \mathbf{S}f(x), \tag{1.6}$$

$$\hat{g}(k) = \mathbf{S}g(x), \tag{1.7}$$

where $\mathbf{S}$ is the Fourier transform operator. Thus using the vector representation of functions, (1.5) is

$$\mathbf{S}(\boldsymbol{f} \otimes \boldsymbol{g}) = \hat{\mathbf{F}}\mathbf{S}\boldsymbol{g}, \tag{1.8}$$

where $\hat{\mathbf{F}}$ is the diagonal matrix whose diagonal elements are those of the Fourier transformed vector $\mathbf{S}\boldsymbol{f}$. The right hand side of (1.8) is then equivalent to the product of the Fourier transforms $\mathbf{S}\boldsymbol{f}$ and $\mathbf{S}\boldsymbol{g}$ as in (1.5). This can be applied to (1.1)

$$\mathbf{B}_s^{1/2}\chi' = \Sigma_B\mathbf{S}^{-1}\sum_{j=1}^{K}\hat{\Psi}_j\mathbf{S}\mathbf{C}_j^{1/2}\chi_j'. \tag{1.9}$$

The matrix $\hat{\Psi}_j$ is akin to $\hat{\mathbf{F}}$ in (1.8); it is the diagonal matrix whose diagonal elements comprise those of the vector $\mathbf{S}\psi_j$ (the hat on $\hat{\Psi}_j$ reminds us that this matrix is in the Fourier representation).

1.4 Wavelet implied covariances

Putting together (1.4) and (1.9) for the implied covariances of the wavelet scheme gives

$$\begin{aligned}\mathbf{B}_s &= \Sigma_B\mathbf{S}^{-1}\sum_{j=1}^{K}\hat{\Psi}_j\mathbf{S}\mathbf{C}_j^{1/2}\left(\Sigma_B\mathbf{S}^{-1}\sum_{j=1}^{K}\hat{\Psi}_j\mathbf{S}\mathbf{C}_j^{1/2}\right)^T, \\ &= \Sigma_B\mathbf{S}^{-1}\left(\sum_{j=1}^{K}\hat{\Psi}_j\mathbf{S}\mathbf{C}_j\mathbf{S}^T\Psi_j\right)\mathbf{S}^{-T}\Sigma_B. \end{aligned} \tag{1.10}$$

The correlation part of this implied covariance is (1.10) without the background standard deviation terms

$$\text{cor} \;=\; \mathbf{S}^{-1}\left(\sum_{j=1}^{K}\hat{\Psi}_j\mathbf{S}\mathbf{C}_j\mathbf{S}^T\hat{\Psi}_j\right)\mathbf{S}^{-T}. \tag{1.11}$$

To look for inhomogeneity it is sufficient to examine the properties of (1.11) in a simplified set-up. We choose a 1-d horizontal space, r, plus height, z, framework. The spectral transform is taken as the Fourier transform in r, whose inverse is the same as the transpose. Let wavenumbers be denoted by k. This correlation operator acting on the input state $x(r, z)$ is then

$$x'(r'', z') \;=\; \int \mathrm{d}k'\, \exp(ik'r'') \sum_{j=1}^{K}\hat{\Psi}_j(k') \int \mathrm{d}r'\, \exp(-ik'r') \;\times$$

$$\int \mathrm{d}z C_j(z', z; r') \int \mathrm{d}k\, \exp(ikr')\hat{\Psi}_j(k) \int \mathrm{d}r\, \exp(-ikr)\, x(r, z), \tag{1.12}$$

where $x'(r'', z')$ is the output state. Let the input state, $x(r, z)$ be prepared as the delta-function with its peak at position r_0, z_0

$$x(r, z) \;=\; \delta(r - r_0, z - z_0). \tag{1.13}$$

On substitution into (1.12), this becomes

$$x'(r'', z') \;=\; \int \mathrm{d}k'\, \exp(ik'r'') \sum_{j=1}^{K}\hat{\Psi}_j(k') \int \mathrm{d}r'\, \exp(-ik'r') \;\times$$

$$C_j(z', z_0; r') \int \mathrm{d}k\, \exp(ikr')\hat{\Psi}_j(k)\, \exp(-ikr_0),$$

$$= \sum_{j=1}^{K} \int \mathrm{d}r' \int \mathrm{d}k'\hat{\Psi}_j(k')\, \exp(ik'(r'' - r')) \;\times$$

$$C_j(z', z_0; r') \int \mathrm{d}k\hat{\Psi}_j(k)\, \exp(ik(r' - r_0)). \tag{1.14}$$

There are now just two integrals over wavenumber, which can be rewritten in the following way

$$\int \mathrm{d}k'\hat{\Psi}_j(k')\, \exp(ik'(r'' - r')) \;=\; \Psi_j(r'' - r'), \tag{1.15}$$

$$\int \mathrm{d}k\hat{\Psi}_j(k)\, \exp(ik(r' - r_0)) \;=\; \Psi_j(r' - r_0). \tag{1.16}$$

Substituting these into (1.14) gives

$$x'(r'', z') \;=\; \sum_{j=1}^{K} \int \mathrm{d}r'\Psi_j(r'' - r')\, C_j(z', z_0; r')\, \Psi_j(r' - r_0). \tag{1.17}$$

This may be described in the following way. The delta-function at r_0, z_0 is coupled to a new arbitrary position at r' via the wavelet function $\Psi_j(r' - r_0)$. This is multiplied by the vertical correlation function from level z_0 to z' at the arbitrary position r'. This is multiplied by the wavelet function that couples arbitrary position r' to the new position at r'', where r'' is the argument of the output function. This is summed over all arbitrary positions and all wavelet bands. Thus the value of this correlation between r_0, z_0 and r'', z'' may be thought as having contributions due to all possible routes between these positions that go via one other horizontal position at r', and through all possible bands.

1.5 Is this correlation function homogeneous?

The correlation function of (1.17) is not a function of $r'' - r_0$ (ie the horizontal distance between the delta-function peak, r_0, and the argument point, r''). This means that it is not homogeneous. This is true even if the vertical correlations lose their position dependence, ie from (1.17)

$$x'(r'', z') = \sum_{j=1}^{K} C_j(z', z_0) \int \mathrm{d}r' \Psi_j(r'' - r') \Psi_j(r' - r_0). \qquad (1.18)$$

The known exception is in the case of only one band, $K = 1$. In this case, the Fourier transforms of the wavelet functions are constant (say unity). Then, from (1.15) and (1.16), and using orthogonality of exponentials

$$\int \mathrm{d}k' \exp(ik'(r'' - r')) = \Psi_j(r'' - r') = \delta(r'' - r') = \delta(r' - r''), \qquad (1.19)$$

$$\int \mathrm{d}k \exp(ik(r' - r_0)) = \Psi_j(r' - r_0) = \delta(r' - r_0). \qquad (1.20)$$

In (1.19), the fact that a delta-function is even is used in the last equality of that equation. Putting (1.19) and (1.20) into (1.18) gives

$$x'(r'', z') = C(z', z_0) \int \mathrm{d}r' \delta(r' - r'') \delta(r' - r_0),$$

$$= C(z', z_0) \delta_{r'', r_0}, \qquad (1.21)$$

where the kronecker delta-function in (1.21) evaluates to zero, except where $r'' = r_0$ where it is unity. This is homogeneous!

2. The Met Office 'errors-of-the-day' scheme

2.1 Introduction

Barker et al. (2005) suggest a modified form of the variational assimilation cost function that has an extra contribution associated with flow dependent structures. These structures or 'errors-of−the-day' (EOTD) are found by an error breeding technique developed by Toth and Kalnay (1993, 1997), and for this reason these structures are called bred modes. They are regarded as the fastest growing modes in the non-linear forecast system. The modification to be described below is an attempt to incorporate flow dependency into the specification of forecast error covariances in Var.

The way that bred modes are computed is described in the above references, but for this document we assume that a set of bred modes, $\{\boldsymbol{v}_k\}$, is available for use in the variational scheme. Here we look at the modified scheme and at the way that it changes the implied covariances.

The modified cost function has three contributions (Barker et al 2005)

$$J = J_B + J_O + J_E. \qquad (2.1)$$

J_B and J_O are the usual background and observation terms respectively. J_E is the EOTD term. The background and observation terms have the same form as the standard scheme, and the background term uses the same control variables, χ. New control variables, χ_k, arise in association with the EOTD term (let there be n control variables for n bred modes, ie $1 \leqslant k \leqslant n$). In control space, J_B and J_E have the following forms

$$J_B = \chi^T \chi, \qquad J_E = \sum_{k=1}^{n} \chi_k^T \chi_k, \qquad (2.2)$$

ie all control variables are mutually uncorrelated and have unit variance. χ and $\{\chi_k\}$ each have blocks that are associated with each Met Office control parameter, ie streamfunction, unbalanced pressure, velocity potential and relative humidity. For simplicity, we will concentrate here on only one parameter (say streamfunction, ψ). Let ψ be denoted by the state $\boldsymbol{v}_p$, and let χ and χ_k now represent those portions

of the original control variables that are concerned only with ψ. A similar analysis can be done for the other control parameters. $\boldsymbol{v}_p$ is found from χ and $\{\chi_k\}$ by the control variable transform

$$\boldsymbol{v}_p = \beta_0 \mathbf{U}_v^0 \mathbf{U}_h^0 \chi + \beta_1 \sum_{k=1}^{n} (\mathbf{U}_v^1 \mathbf{U}_h^1 \chi_k) \circ \boldsymbol{v}_k. \tag{2.3}$$

The new parameters, β_0 and β_1 are weights, $\mathbf{U}_v^0 \mathbf{U}_h^0$ and $\mathbf{U}_v^1 \mathbf{U}_h^1$ are the spatial (horizontal and vertical) transforms for the background and EOTD contributions respectively ($\mathbf{U}_v^0$ and $\mathbf{U}_h^0$ are the same as in the standard scheme), $\boldsymbol{v}_k$ is the kth EOTD mode, suitably normalized (these are given) and $\circ$ denotes the Shur (element-by-element) product. The need to use a Shur product can be removed by assembling the components of $\boldsymbol{v}_k$ into the diagonal elements of the diagonal matrix $\mathbf{V}_k$. Then (2.3) can be written

$$\boldsymbol{v}_p = \beta_0 \mathbf{U}_v^0 \mathbf{U}_h^0 \chi + \beta_1 \sum_{k=1}^{n} \mathbf{V}_k \mathbf{U}_v^1 \mathbf{U}_h^1 \chi_k, \tag{2.4}$$

which is easier to work with.

2.2 The control variable transform in matrix form

Barker et al. (2005) use EOTD control vectors, $\{\chi_k\}$, that are two-dimensional, having no vertical dependence. The vertical transform, $\mathbf{U}_v^1$ in (2.4) is then a 'null' transform; apart from a position dependent scaling, it just copies the two-dimensional input field to three-dimensions (see section 2.4). Equation (2.4) has the following matrix structure

$$\boldsymbol{v}_p = \left(\beta_0 \mathbf{U}_v^0 \mathbf{U}_h^0 \;\middle|\; \beta_1 \mathbf{V}_1 \mathbf{U}_v^1 \mathbf{U}_h^1 \;\middle|\; \beta_1 \mathbf{V}_2 \mathbf{U}_v^1 \mathbf{U}_h^1 \;\middle|\; \cdots \;\middle|\; \beta_1 \mathbf{V}_n \mathbf{U}_v^1 \mathbf{U}_h^1 \right) \left(\begin{array}{c} \chi \\ \hline \chi_1 \\ \hline \chi_2 \\ \hline \cdots \\ \hline \chi_n \end{array} \right). \tag{2.5}$$

χ is three-dimensional but $\{\chi_k\}$ are two-dimensional, as stated above.

2.3 The implied covariances

The control vector's elements are uncorrelated and so the implied covariance matrix of (2.5) is

$$\mathbf{B} = \left(\beta_0 \mathbf{U}_v^0 \mathbf{U}_h^0 \;\middle|\; \beta_1 \mathbf{V}_1 \mathbf{U}_v^1 \mathbf{U}_h^1 \;\middle|\; \beta_1 \mathbf{V}_2 \mathbf{U}_v^1 \mathbf{U}_h^1 \;\middle|\; \cdots \;\middle|\; \beta_1 \mathbf{V}_n \mathbf{U}_v^1 \mathbf{U}_h^1 \right) \left(\begin{array}{c} \beta_0 \mathbf{U}_h^{0T} \mathbf{U}_v^{0T} \\ \hline \beta_1 \mathbf{U}_h^{1T} \mathbf{U}_v^{1T} \mathbf{V}_1 \\ \hline \beta_1 \mathbf{U}_h^{1T} \mathbf{U}_v^{1T} \mathbf{V}_2 \\ \hline \cdots \\ \hline \beta_1 \mathbf{U}_h^{1T} \mathbf{U}_v^{1T} \mathbf{V}_n \end{array} \right),$$

$$= \beta_0^2 \mathbf{U}_v^0 \mathbf{U}_h^0 \mathbf{U}_h^{0T} \mathbf{U}_v^{0T} + \beta_1^2 \sum_{k=1}^{n} \mathbf{V}_k \mathbf{U}_v^1 \mathbf{U}_h^1 \mathbf{U}_h^{1T} \mathbf{U}_v^{1T} \mathbf{V}_k. \tag{2.6}$$

The first term is the implied background error covariance matrix of the standard Var. scheme, and the remaining terms are modifications due to the extra EOTD variables. The modifications are flow dependent, which comes from the bred vectors in $\mathbf{V}_k$.

2.4 The EOTD horizontal and vertical transforms

In order to look at the structure functions associated with the EOTD terms in (2.6), we need to consider the horizontal and vertical transforms, $\mathbf{U}_h^1$ and $\mathbf{U}_v^1$ which appear in (2.6).

In the usual formulation, the combination of horizontal transforms, $\mathbf{U}_h^1\mathbf{U}_h^{1T}$ represents an isotropic covariance matrix in the horizontal, which shall be denoted by $\mathbf{C}$. The vertical transforms, $\mathbf{U}_h^1$ and $\mathbf{U}_h^{1T}$ do not appear together in (2.6) and so here we need to consider the form of the vertical transform.

The vertical transform acts on a two-dimensional field - call this field $x(\lambda, \phi)$, (a function of longitude and latitude) - and outputs a three dimensional field - call this field $x'(\lambda, \phi, z)$ (a function of longitude, latitude and model level). Each level is just a copy of the two-dimensional input vector, but multiplied by a longitude and latitude dependent scalar. Mathematically this is

$$x'(\lambda, \phi, z) = U_v^1(\lambda, \phi)\, x(\lambda, \phi),$$

$$= \sum_{\lambda,\phi} U_v^1(\lambda', \phi')\, \delta_{\lambda\lambda'}\delta_{\phi\phi'}\, x(\lambda', \phi'). \qquad (2.7)$$

This operator will be used in section 2.5 when investigating (2.6). In (2.6) its transpose is also required. A mechanical means of forming the transpose is to calculate the operator that links derivatives with respect to the input state, $x(\lambda', \phi')$ with those with respect to the output state, $x'(\lambda, \phi, z)$ (Bannister 2008b). By the chain rule this is

$$\frac{\partial}{\partial x(\lambda', \phi')} = \sum_{\lambda\phi z} \frac{\partial x'(\lambda, \phi, z)}{\partial x(\lambda', \phi')} \frac{\partial}{\partial x'(\lambda, \phi, z)}. \qquad (2.8)$$

The operator that is represented by this is the transpose of $\mathbf{U}_v^1$. The derivatives in (2.8) can be found from (2.7)

$$\frac{\partial x'(\lambda, \phi, z)}{\partial x(\lambda', \phi')} = U_v^1(\lambda, \phi)\, \delta_{\lambda\lambda'}\delta_{\phi\phi'} \qquad (2.9)$$

Inserting (2.9) into (2.8) gives a form for the transpose operator

$$\frac{\partial}{\partial x(\lambda', \phi')} = \sum_{\lambda\phi z} U_v^1(\lambda, \phi)\, \delta_{\lambda\lambda'}\delta_{\phi\phi'} \frac{\partial}{\partial x'(\lambda, \phi, z)},$$

$$= U_v^1(\lambda', \phi') \sum_z \frac{\partial}{\partial x'(\lambda', \phi', z)}, \qquad (2.10)$$

which will be used when investigating (2.6).

2.5 Examination of a bred vector contribution to B

In this section we look at the structure functions associated with one of the EOTD terms in (2.6), by using the forms of the operators defined in section 2.4. This may help us understand how the EOTD modifications work. The contribution to the implied covariances (2.6) of the kth bred vector is, $\beta_1^2\mathbf{V}_k\mathbf{U}_v^1\mathbf{U}_h^1\mathbf{U}_h^{1T}\mathbf{U}_v^{1T}\mathbf{V}_k$. Acting on an input state, $x(\lambda, \phi, z)$ to give an output state $x'(\lambda', \phi', z')$, this operator can be expanded as follows

$$x(\lambda', \phi', z') = \beta_1^2 v_k(\lambda', \phi', z')\, U_v^1(\lambda', \phi') \times$$

$$\sum_{\lambda\phi} C(\lambda', \phi'; \lambda, \phi)\, U_v^1(\lambda, \phi) \sum_z v_k(\lambda, \phi, z)\, x(\lambda, \phi, z). \qquad (2.11)$$

The structure functions can be considered by letting this covariance operator act on a delta function at position λ_0, ϕ_0, z_0, ie

$$\text{let } x(\lambda, \phi, z) = \delta_{\lambda\lambda_0}\delta_{\phi\phi_0}\delta_{zz_0},$$

$$x(\lambda', \phi', z') = \beta_1^2 v_k(\lambda', \phi', z') U_v^1(\lambda', \phi') C(\lambda', \phi'; \lambda_0, \phi_0) U_v^1(\lambda_0, \phi_0) v_k(\lambda_0, \phi_0, z_0). \quad (2.12)$$

This gives the implied covariance between a point at λ_0, ϕ_0, z_0 with another at λ', ϕ', z' due to the kth EOTD term. The following points can be made.

- The implied covariance is subject to three contributions: (i) a standard correlation model, $C(\lambda', \phi'; \lambda_0, \phi_0)$, which is a decaying function of the horizontal distance between the two points, (ii) flow dependent modulation functions that are each a function of the start and end points, $v_k(\lambda', \phi', z') U_v^1(\lambda', \phi')$ and $U_v^1(\lambda_0, \phi_0) v_k(\lambda_0, \phi_0, z_0)$, and (iii) a parameter β_1^2.
- In this formulation where the $\{\chi_k\}$ control variables are not a function of z, the standard correlation model, $C(\lambda', \phi'; \lambda_0, \phi_0)$, is a function of the horizontal distance between the two points, but not on the vertical distance. Thus all levels in (2.12) have unit correlation at a single horizontal location.
- Apart from the lack of vertical correlation, which can be solved by adding z dependence to the $\{\chi_k\}$, this result is similar to the Riishøjgaard flow-dependent background error covariance model (Riishøjgaard 1998), but with one other difference. In Riishøjgaard's model, the prescribed correlations are modulated by a function that is the difference between the background state at positions λ_0, ϕ_0, z_0 and λ', ϕ', z', but here it is modulated by the Bred vector values at these positions.

3. Why does the NMC method give twice the covariance values of background errors?

3.1 What is the NMC method?

The 'National Meteorological Center' (NMC) method is a means of estimating the statistics of forecast errors by analysing the statistics of forecast *differences*. The covariance matrix of forecast errors is defined by the outer product

$$\mathbf{B} = \langle (\boldsymbol{x} - \boldsymbol{x}_t)(\boldsymbol{x} - \boldsymbol{x}_t)^T \rangle, \quad (3.1)$$

where $\boldsymbol{x}$ is the state of interest and $\boldsymbol{x}_t$ is the unknown true state. This is the background error covariance matrix (the so-called **B**-matrix) used in data assimilation. The true state is never known, and so (3.1) is impossible to calculate. Parrish and Derber (1992) introduced a way of approximating (3.1) by instead analysing the differences between two forecasts initialised 24 hours apart, but valid at the same time, e.g.

$$\mathbf{B} \approx \frac{1}{2} \langle (\boldsymbol{x}_{48} - \boldsymbol{x}_{24})(\boldsymbol{x}_{48} - \boldsymbol{x}_{24})^T \rangle, \quad (3.2)$$

where $\boldsymbol{x}_{48}$ is a 48-hour forecast and $\boldsymbol{x}_{24}$ is a 24-hour forecast. It is assumed that the differences $\boldsymbol{x}_{48} - \boldsymbol{x}_{24}$ are a surrogate for $\boldsymbol{x} - \boldsymbol{x}_t$, and have similar error structures. Here we look at the details that lead to this assumption. In particular we show that forecast differences have twice the covariance of forecast errors - and hence the factor of ½ in (3.2). Here is a complicated argument towards this factor of ½ (a simpler argument is given in Bannister (2008a).

3.2 The problem

Let two forecasts be $\boldsymbol{x}_{48}$ and $\boldsymbol{x}_{24}$, both valid at the same time. Each has a forecast random error, ε_{48} and ε_{24}, and a bias, $\boldsymbol{b}_{48}$ and $\boldsymbol{b}_{24}$. Hólm et al. (2002) discuss the relationship between the probability density function (PDF) of the difference between these forecasts (as characterised by (3.2) when the PDF is Gaussian) and the PDFs of each forecast (as characterised by (3.1) where $\boldsymbol{x}$ is a short-range

forecast). Each forecast can be written as

$$\boldsymbol{x}_{48} = \boldsymbol{x}_t + \varepsilon_{48} + \boldsymbol{b}_{48}, \tag{3.3}$$

$$\boldsymbol{x}_{24} = \boldsymbol{x}_t + \varepsilon_{24} + \boldsymbol{b}_{24}. \tag{3.4}$$

The difference between the forecasts is $\delta\boldsymbol{x} = \boldsymbol{x}_{48} - \boldsymbol{x}_{24} = \varepsilon_{48} - \varepsilon_{24} + \boldsymbol{b}_{48} - \boldsymbol{b}_{24}$. By assuming that the biases are the same, the difference simplifies to

$$\delta\boldsymbol{x} = \boldsymbol{x}_{48} - \boldsymbol{x}_{24} = \varepsilon_{48} - \varepsilon_{24}. \tag{3.5}$$

It is a reasonable assumption that the biases are the same since (i) both forecasts are valid at the same time, and (ii) each forecast has initial conditions - each taken from analyses - 24 hours apart, thus eliminating any diurnal cycle in analysis biases. This is the reason why the 24 hour difference is used in the NMC method (in data assimilation, the background state is often a six-hour forecast, and so some centres use the difference between a 30-hour and a six-hour forecast). We review the question: how is the PDF of $\delta\boldsymbol{x}$ related to the PDFs of ε_{48} and ε_{24}?

3.3 The PDFs

A Gaussian has the following form (for the vector error quantity μ)

$$P(\mu) \propto \exp -\frac{\mu^T \mathbf{E}^{-1} \mu}{2}, \tag{3.6}$$

where **E** is the error covariance. Let the PDF of errors in the two forecasts be $P_{48}(\varepsilon_{48})$ and $P_{24}(\varepsilon_{24})$. Assuming that they are equal and Gaussian, let the bias-corrected PDFs have error covariance **E**, just like in (3.6)

$$P_{48}(\varepsilon_{48}) \propto \exp -\frac{\varepsilon_{48}^T \mathbf{E}^{-1} \varepsilon_{48}}{2}, \tag{3.7}$$

$$P_{24}(\varepsilon_{24}) \propto \exp -\frac{\varepsilon_{24}^T \mathbf{E}^{-1} \varepsilon_{24}}{2}, \tag{3.8}$$

$$\mathbf{E} = \langle \varepsilon_1 \varepsilon_1^T \rangle = \langle \varepsilon_2 \varepsilon_2^T \rangle. \tag{3.9}$$

This gives $P_{48}(\varepsilon) = P_{24}(\varepsilon) \equiv P(\varepsilon)$. Let the PDF of the forecast difference be $P_D(\delta\boldsymbol{x})$. Let this also be Gaussian and have error covariance $\mathbf{E}_D$

$$P_D(\delta\boldsymbol{x}) \propto \exp -\frac{\delta\boldsymbol{x}^T \mathbf{E}_D^{-1} \delta\boldsymbol{x}}{2}, \tag{3.10}$$

$$\mathbf{E}_D = \langle \delta\boldsymbol{x} \delta\boldsymbol{x}^T \rangle. \tag{3.11}$$

3.4 The relationship between the PDFs

The relationship between $P(\varepsilon)$ and $P_D(\delta\boldsymbol{x})$ is the following (Hólm et al. 2002)

$$P_D(\delta\boldsymbol{x}) = \int_{-\infty}^{\infty} \mathrm{d}\,\varepsilon_{48} P(\varepsilon_{48}) P(\varepsilon_{24} \mid \delta\boldsymbol{x} = \varepsilon_{48} - \varepsilon_{24}). \tag{3.12}$$

This is the combined probability density of the 48-hour forecast error is ε_{48} and the 24-hour forecast error is ε_{24} given that the difference between the forecasts is $\delta\boldsymbol{x}$ (see (3.5)). This is integrated over all possibilities of ε_{48}. By using (3.5), and by simplifying the notation $\varepsilon_{48} \rightarrow \varepsilon$, (3.12) is

$$\begin{aligned} P_D(\delta\boldsymbol{x}) &= \int_{-\infty}^{\infty} \mathrm{d}\,\varepsilon P(\varepsilon) P(\varepsilon - \delta\boldsymbol{x}), \\ &= \int_{-\infty}^{\infty} \mathrm{d}\,\varepsilon P(\varepsilon) P(\delta\boldsymbol{x} - \varepsilon), \end{aligned} \tag{3.13}$$

where in the last line, the even property of P is used. Thus the PDF of the forecast difference is a convolution of the PDF of each forecast with itself.

3.5 Use of the convolution theorem of Fourier transforms

Since (3.13) is a convolution, the relationship between P_D and P can be investigated further using the convolution theorem of Fourier transforms. By indicating the Fourier transform of each PDF with a 'hat', and by using k as the argument in Fourier space, the convolution theorem applied to (3.13) is expressed as

$$\hat{P}_D(k) = 2\pi\hat{P}^2(k). \tag{3.14}$$

The squared function on the right hand side is a consequence of the fact that the convolution of P in (3.13) is carried out with itself. Equation (3.14) is a one-dimensional result. For simplicity, we shall prove results in a one-dimensional phase space and extend conclusions to arbitrary dimensions (the extension to arbitrary dimensions may be made trivially by working in a phase space representation where the error covariances are diagonal; the multidimensional problem then reduces to the product of many one-dimensional problems). As a consequence of working in one-dimension, the error covariance **E** in (3.7) and (3.8) may be written as the scalar σ, and $\mathbf{E}_D$ in (3.10) may be written as σ_D.

If $P(\varepsilon)$ of Gaussian form is then $P(\varepsilon) = \exp(-\varepsilon^2/2\sigma)$, then the Fourier transform may be found from the general result

$$\text{''}\exp(-\alpha x^2) \quad \begin{matrix}\xrightarrow{\text{FT}}\\ \xleftarrow{\text{IFT}}\end{matrix} \quad \frac{1}{\sqrt{4\pi\alpha}}\exp(-k^2/4\alpha), \quad \alpha > 0\text{''}. \tag{3.15}$$

By applying (3.15) to Fourier transform the one-dimensional Gaussian PDFs ($\alpha = 1/2\sigma$) gives , $\hat{P}(k) = \exp(-k^2\sigma/2)/\sqrt{2\pi/\sigma}$. Putting this into (3.14) gives

$$\hat{P}_D(k) = \sigma\exp(-k^2\sigma). \tag{3.16}$$

This is a way of performing the convolution. Equation (3.16) needs to be transformed back to the real phase space, by applying (3.15) in reverse to do the inverse Fourier transform ($1/4\alpha = \sigma$, $\alpha = 1/4\sigma$) giving

$$P_D(\delta x) = \sqrt{\pi\sigma}\exp-\frac{\delta x^2}{4\sigma} \propto \exp-\frac{\delta x^2}{2\cdot 2\sigma}, \tag{3.17}$$

ie $\sigma_D = 2\sigma$. Extending this result to multidimensional phase space gives

$$\mathbf{E}_D = 2\mathbf{E}. \tag{3.18}$$

The forecast difference statistics have twice the covariance as the forecast error statistics. This is why the factor of ½ is present in (3.2).

4. Imposing homogeneity and isotropy of structure functions with a Fourier transform

4.1 The spectral representation

Consider the function x, which can be written in terms of its Fourier transform, $\tilde{x}$ as follows

$$x(\boldsymbol{r}) = \frac{1}{\sqrt{A}}\sum_{k}\tilde{x}(\boldsymbol{k})\ \exp i\boldsymbol{k}\cdot\boldsymbol{r}. \tag{4.1}$$

Here $\boldsymbol{k} = (k_x, k_y)$ is the wavevector, $\boldsymbol{r} = (x, y)$ is the position vector and A is the area of the two-dimensional domain. Let $B(\boldsymbol{r}, \boldsymbol{r}')$ be the covariance between the field x at positions $\boldsymbol{r}$ and $\boldsymbol{r}'$ as follows (assume that x has zero mean at each position)

$$B(\boldsymbol{r}, \boldsymbol{r}') = \langle x(\boldsymbol{r})x^*(\boldsymbol{r}')\rangle, \tag{4.2}$$

where the angled brackets indicate average and the * indicates complex conjugate

(this is normally ommited as fields have real values, but here complex numbers are introduced as a result of the Fourier transform in (4.1)). Substitute (4.1) into (4.2)

$$B(\boldsymbol{r}, \boldsymbol{r}') = \langle \frac{1}{\sqrt{A}} \sum_{k} \tilde{x}(\boldsymbol{k}) \exp i\boldsymbol{k} \cdot \boldsymbol{r} \frac{1}{\sqrt{A}} \sum_{k'} \tilde{x}(\boldsymbol{k}') \exp -i\boldsymbol{k}' \cdot \boldsymbol{r}' \rangle,$$

$$= \frac{1}{A} \sum_{k,k'} \langle \tilde{x}(\boldsymbol{k}) \tilde{x}(\boldsymbol{k}') \rangle \exp i\boldsymbol{k} \cdot \boldsymbol{r} \exp -i\boldsymbol{k}' \cdot \boldsymbol{r}'. \quad (4.3)$$

In (4.3), $\langle \tilde{x}(\boldsymbol{k}) \tilde{x}(\boldsymbol{k}') \rangle$ is the covariance in spectral representation. It is convenient to write $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{s}$, where $\boldsymbol{s}$ is a separation. Equation (4.3) becomes

$$B(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{s}) = \frac{1}{A} \sum_{k,k'} \langle \tilde{x}(\boldsymbol{k}) \tilde{x}(\boldsymbol{k}') \rangle \exp i\boldsymbol{k} \cdot \boldsymbol{r} \exp -i\boldsymbol{k}' \cdot (\boldsymbol{r} + \boldsymbol{s}). \quad (4.4)$$

Equation (4.4) is the starting point for the discussion on homogeneity below.

4.2 Homogeneity

Homogeneity is the property of structure (or correlation) functions whose shape does not depend upon position. This means that if B is homogeneous there would be no $\boldsymbol{r}$ dependence. Consider the case when the covariance in the spectral representation is diagonal, ie let

$$\langle \tilde{x}(\boldsymbol{k}) \tilde{x}(\boldsymbol{k}') \rangle = V^s(\boldsymbol{k}) \delta_{kk'}. \quad (4.5)$$

Substituting (4.5) into (4.4) gives

$$B(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{s}) = \frac{1}{A} \sum_{k,k'} V^s(\boldsymbol{k}) \delta_{kk'} \exp i\boldsymbol{k} \cdot \boldsymbol{r} \exp -i\boldsymbol{k}' \cdot (\boldsymbol{r} + \boldsymbol{s}),$$

$$= \frac{1}{A} \sum_{k} V^s(\boldsymbol{k}) \exp i\boldsymbol{k} \cdot \boldsymbol{r} \exp -i\boldsymbol{k} \cdot (\boldsymbol{r} + \boldsymbol{s}),$$

$$= \frac{1}{A} \sum_{k} V^s(\boldsymbol{k}) \exp -i\boldsymbol{k} \cdot \boldsymbol{s}. \quad (4.6)$$

In (4.6) the dependence on $\boldsymbol{r}$ has disappeared. Hence imposing a covariance that is diagonal in spectral space implies homogeneity in real space. This is a method often used in covariance modelling to impose homogenity. The term $V^s(\boldsymbol{k})$ is called the variance spectrum. Equation (4.6) is now the starting point for the discussion on isotropy below.

4.3 Isotropy

Isotropy is the property of structure (or correlation) functions whose shape is a function only of the distance between two points and not on the relative orientation. In terms of the separation vector, an isotropic structure function is one that depends only on s (= $|\boldsymbol{s}|$).

How isotropy can be imposed can be seen by taking the continuous limit. Then, the summation in (4.6) becomes an integral in spectral space as follows

$$B(\boldsymbol{r}, \boldsymbol{r} + \boldsymbol{s}) = \frac{1}{A} \int_{k_x,k_y} dk_x\, dk_y\, V^s(\boldsymbol{k}) \exp -i\boldsymbol{k} \cdot \boldsymbol{s},$$

$$= \frac{1}{A} \int_{k,\phi^k} dk\, d\phi^k\, k\, V^s(\boldsymbol{k}) \exp -i\boldsymbol{k} \cdot \boldsymbol{s},$$

$$= \frac{1}{A} \int_{k} dk\, k \int_{\phi^k=0}^{2\pi} d\phi^k\, V^s(\boldsymbol{k}) \exp -i\boldsymbol{k} \cdot \boldsymbol{s}. \quad (4.7)$$

In the last two lines, the integral over spectral space has been written in polar co-ordinates where $k = |\boldsymbol{k}|$, ϕ^k is the angle between $\boldsymbol{k}$ and the k_x-axis. The integration limits for the ϕ^k variable has been given explicitly (0 to 2π). Let the wave and

position vectors be written as the following

$$\boldsymbol{k} = (k\cos\phi^k,\ k\sin\phi^k), \qquad \boldsymbol{s} = (s\cos\phi^s,\ s\sin\phi^s), \tag{4.8}$$

where ϕ^s is the angle between $\boldsymbol{s}$ and the s_x-axis (the s_x and k_x axes are coincident). The scalar product in (4.6) is then (using the cosine formula)

$$\boldsymbol{k}\cdot\boldsymbol{s} = ks(\cos\phi^k\cos\phi^s + \sin\phi^k\sin\phi^s) = ks\cos(\phi^k - \phi^s). \tag{4.9}$$

Substituting (4.9) into (4.7) gives

$$B(\boldsymbol{r},\ \boldsymbol{r}+\boldsymbol{s}) = \frac{1}{A}\int_k dk\,k\int_{\phi^k=0}^{2\pi} d\phi^k\,V^s(\boldsymbol{k})\ \exp-[iks\cos(\phi^k - \phi^s)]. \tag{4.10}$$

Consider the case when the covariance in the spectral representation is itself isotropic, ie, $V^s(\boldsymbol{k}) = V^s(k)$. Then

$$B(\boldsymbol{r},\ \boldsymbol{r}+\boldsymbol{s}) = \frac{1}{A}\int_k dk\,k\int_{\phi^k=0}^{2\pi} d\phi^k\,V^s(k)\ \exp-[iks\cos(\phi^k - \phi^s)],$$

$$\frac{1}{A}\int_k dk\ kV^s(k)\int_{\phi^k=0}^{2\pi} d\phi^k\ \exp-[iks\cos(\phi^k - \phi^s)]. \tag{4.11}$$

It should be easy to see that any integral of the form

$$\int_{\phi=0}^{2\pi} f[\cos(\phi - \theta)], \tag{4.12}$$

is independent of θ for any function f (all θ does is to shift the start phase of the integral). This means that (4.11) is independent of ϕ^s. Hence imposing a covariance that is isotropic in spectral space is also isotropic in real space. Further details can be found in Berre (2000).

5. A simplified Kalman filter (reduced rank Kalman filter)

5.1 What is a simplified Kalman filter

Fisher (1998) has developed a framework for a reduced rank Kalman filter (RRKF) for use in a variational environment. A RRKF in this context may be regarded as a modification to the existing **B**-matrix in Var. that allows the dynamical evolution of a subspace of the state vector. The particular way that this is done by Fisher, and the approximations that are made, are outlined below.

5.2 Definition of the subspace by Hessian singular vectors

The first stage is to identify the subspace that will be treated with explicit flow dependence. Let the dimension of the subspace be K, which can be chosen arbitrarily, but restricted in practice by cost. Fisher defines the subspace by the K most rapidly growing Hessian singular vectors. The reason why they are chosen to be singular vectors of the Hessian will become evident later. In order to write down the problem that must be solved, we introduce two norms, which we assume are available.

- Let the covariance matrix $\mathbf{P}^a$ be the error covariance of the analysis of the previous cycle. In order for Fisher's method to work, it must be possible to act with the matrix $\mathbf{P}^{a-1}$ (or an approximation of it).
- Let the matrix **W** be the norm used to measure the size of a perturbation. It must be possible to act with the matrix $\mathbf{W}^{-1}$.

Let the time of the previous data assimilation cycle be $t = -T$ and the time of the present analysis be $t = 0$. States that have no time label are valid at $t = 0$ by default.

Let the tangent linear model, $\mathbf{M}_{0\leftarrow -T}$ act on perturbations at time $t = -T$ and give

a perturbation at time $t = 0$

$$\delta\boldsymbol{x} = \mathbf{M}_{0\leftarrow -T}\delta\boldsymbol{x}(-T). \tag{5.1}$$

If $\delta\boldsymbol{x}(-T)$ were known, then the size of $\delta\boldsymbol{x}$ according to the $\mathbf{W}$-norm would be J_1

$$J_1 = \delta\boldsymbol{x}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x} = \delta\boldsymbol{x}^{T}(-T)\,\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{M}_{0\leftarrow -T}\delta\boldsymbol{x}(-T). \tag{5.2}$$

The Hessian singular vectors are defined as those $\delta\boldsymbol{x}(-T)$ that maximise J_1 subject to the constraint that $\delta\boldsymbol{x}(-T)$ is distributed according to $\mathbf{P}^{a}$, ie

$$\delta\boldsymbol{x}^{\mathrm{T}}(-T)\,\mathbf{P}^{a^{-1}}\delta\boldsymbol{x}(-T) - \text{const} = 0, \tag{5.3}$$

for an arbitrary constant, 'const'. The constrained optimisation problem may therefore be posed as

$$\nabla_{x(-T)}\left[J_1 - \lambda\left(\delta\boldsymbol{x}^{\mathrm{T}}(-T)\,\mathbf{P}^{a^{-1}}\delta\boldsymbol{x}(-T) - \text{const}\right)\right] = 0, \tag{5.4}$$

where λ is the Lagrange multiplier. Applying the derivative operator and setting the solutions to $\delta\boldsymbol{x}_k$ (with associated Lagrange multiplier λ_k) gives

$$\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{M}_{0\leftarrow -T}\delta\boldsymbol{x}_k(-T) = \lambda_k\mathbf{P}^{a^{-1}}\delta\boldsymbol{x}_k(-T), \tag{5.5}$$

which is a generalised eigenvalue problem. The $\delta\boldsymbol{x}_k(-T)$ are the Hessian singular vectors. The set of vectors $\mathbf{P}^{a^{-1/2}}\delta\boldsymbol{x}_k(-T)$ are eigenvectors of the matrix $(\mathbf{W}^{-1/2}\mathbf{M}_{0\leftarrow -T}\mathbf{P}^{a^{1/2}})^{\mathrm{T}}(\mathbf{W}^{-1/2}\mathbf{M}_{0\leftarrow -T}\mathbf{P}^{a^{1/2}})$ and are equivalently the right singular vectors of the matrix $\mathbf{W}^{-1/2}\mathbf{M}_{0\leftarrow -T}\mathbf{P}^{a^{1/2}}$. Let $\boldsymbol{s}_k = \mathbf{M}_{0\leftarrow -T}\delta\boldsymbol{x}_k(-T)$. Those $\boldsymbol{s}_k$ with the largest λ_k define the subspace whose background errors be treated explicitly by the RRKF.

A general perturbation at $t = 0$, $\delta\boldsymbol{x}$ has a part $\delta\boldsymbol{x}_s$ that lies in that subspace, which can be found as a linear combination of the $\boldsymbol{s}_k$. Identification of this subspace can be simplified by first constructing an orthogonalized and normalized set of vectors, $\tilde{\boldsymbol{s}}_k$ (e.g. by the Gramm-Schmidt procedure). Then

$$\delta\boldsymbol{x}_s = \tilde{\mathbf{S}}\delta\boldsymbol{a}, \tag{5.6}$$

where $\tilde{\mathbf{S}}$ is the $n \times K$ matrix of $\tilde{\boldsymbol{s}}_k$ vectors (the vector space has n dimensions) and $\delta\boldsymbol{a}$ is the K-element vector of (as yet unknown) coefficients. Orthogonalization should be done with respect to an inner product that non-dimensionalises the components (e.g. the $\mathbf{W}^{-1}$ inner product) such that $\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\tilde{\mathbf{S}} = \mathbf{I}$. The benefit of first orthogonalising the vectors is to allow $\delta\boldsymbol{a}$ to be found easily from $\delta\boldsymbol{x}$

$$\delta\boldsymbol{a} = \tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x}. \tag{5.7}$$

Note that $\delta\boldsymbol{a}$ contains only information about the part of $\delta\boldsymbol{x}$ that is spanned by the singular vector subspace. Hence in general, $\delta\boldsymbol{x}$ is not recovered with (5.6). This means that $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1} \neq \mathbf{I}$. However, $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}$ can be used as a filter to remove parts of the space which is not spanned by the singular vectors, i.e. $\delta\boldsymbol{x}_s = \tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x}$. The part of $\delta\boldsymbol{x}$ that is not within the singular vector subspace is the residual $\delta\bar{\boldsymbol{x}}_s$

$$\delta\bar{\boldsymbol{x}}_s = \delta\boldsymbol{x} - \delta\boldsymbol{x}_s. \tag{5.8}$$

This is orthogonal to $\delta\boldsymbol{x}_s$ under the $\mathbf{W}^{-1}$ norm, which may be proved as follows

$$\begin{aligned}
\delta\bar{\boldsymbol{x}}_s^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x}_s &= (\delta\boldsymbol{x} - \delta\boldsymbol{x}_s)^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x}_s,\\
&= (\delta\boldsymbol{x} - \tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x})^{\mathrm{T}}\mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x},\\
&= \delta\boldsymbol{x}^{\mathrm{T}}(\mathbf{I} - \tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1})^{\mathrm{T}}\mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\delta\boldsymbol{x},\\
&= \delta\boldsymbol{x}^{\mathrm{T}}(\mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1} - \mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1})^{\mathrm{T}}\delta\boldsymbol{x},\\
&= \delta\boldsymbol{x}^{\mathrm{T}}(\mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1} - \mathbf{W}^{-1}\tilde{\mathbf{S}}\tilde{\mathbf{S}}^{\mathrm{T}}\mathbf{W}^{-1})^{\mathrm{T}}\delta\boldsymbol{x},\\
&= 0, \qquad \text{as required.}
\end{aligned}$$

Note that there are other ways of defining a subspace. In Sec. 5.8, we include an

appendix on how ensembles could be used in this manner.

5.3 The background cost function in the new subspace

The usual background cost function in VAR is

$$J_b = \frac{1}{2}(\delta\boldsymbol{x} - \delta\boldsymbol{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\boldsymbol{x} - \delta\boldsymbol{x}^b), \tag{5.9}$$

where $\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^g$, $\delta\boldsymbol{x}^b = \boldsymbol{x}^b - \boldsymbol{x}^g$ and $\boldsymbol{x}^g$ is a reference (or guess) state. The **B**-matrix in (5.9) is static. Equation (5.9) may be written in terms of the components $\delta\boldsymbol{x}_s$ and $\delta\bar{\boldsymbol{x}}_s$ by substituting (5.8) into (5.9). This gives three parts: (i) the part that involves only the special subspace that has been identified from the K singular vectors, (ii) the part that couples this subspace with the rest of the state, and (iii) the part that involves only the rest of the state

$$J_b = \frac{1}{2}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b) + (\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b) +$$

$$\frac{1}{2}(\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b). \tag{5.10}$$

This cost function is identical to (5.9). The RRKF is constructed by imposing a flow dependent error covariance matrix for the first two terms ($\mathbf{B} \rightarrow \mathbf{P}^f$) but keeping the static **B**-matrix in the last term

$$J_b \rightarrow \frac{1}{2}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b)^{\mathrm{T}}\mathbf{P}^{f^{-1}}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b) + \alpha(\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b)^{\mathrm{T}}\mathbf{P}^{f^{-1}}(\delta\boldsymbol{x}_s - \delta\boldsymbol{x}_s^b) +$$

$$\frac{1}{2}(\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b)^{\mathrm{T}}\mathbf{B}^{-1}(\delta\bar{\boldsymbol{x}}_s - \delta\bar{\boldsymbol{x}}_s^b). \tag{5.11}$$

The factor α, added by Fisher (1998), is to help ensure that J_b is convex.

5.4 Control variable transforms stage

It is usual in VAR to make a change of variable from model variables in $\delta\boldsymbol{x}$ to control variables, which are conventionally named χ. In the RRKF there are two control variable transforms (**X** and **L**) as follows

$$\delta\boldsymbol{x} = \mathbf{L}\mathbf{X}\chi, \tag{5.12}$$

where **X** is an orthogonal matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{I}$ (see below) and **L** is the usual control variable transform used in VAR where $\mathbf{L}\mathbf{X}(\mathbf{L}\mathbf{X})^{\mathrm{T}} = \mathbf{L}\mathbf{L}^{\mathrm{T}} = \mathbf{B}$. Substituting (5.12) into (5.11) and using the property mentioned above gives

$$J_b = \frac{1}{2}(\chi_s - \chi_s^b)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{L}\mathbf{X}(\chi_s - \chi_s^b) + \alpha(\bar{\chi}_s - \bar{\chi}_s^b)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{L}\mathbf{X}(\chi_s - \chi_s^b) +$$

$$\frac{1}{2}(\bar{\chi}_s - \bar{\chi}_s^b)^{\mathrm{T}}(\bar{\chi}_s - \bar{\chi}_s^b), \tag{5.13}$$

where $\chi_s = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\boldsymbol{x}_s$, $\chi_s^b = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\boldsymbol{x}_s^b$, $\bar{\chi}_s = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\bar{\boldsymbol{x}}_s$ and $\bar{\chi}_s^b = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\bar{\boldsymbol{x}}_s^b$. The matrix **X** is not present in standard VAR, but is introduced in (5.12) to isolate the special subspace identified in section 5.2 from the remainder. As it stands, (5.13) looks complicated to treat. Let

$$\mathbf{P}_\chi^{f^{-1}} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{L}\mathbf{X}, \tag{5.14}$$

then (5.13) is

$$J_b = \frac{1}{2}(\chi_s - \chi_s^b)^{\mathrm{T}}\mathbf{P}_\chi^{f^{-1}}(\chi_s - \chi_s^b) + \alpha(\bar{\chi}_s - \bar{\chi}_s^b)^{\mathrm{T}}\mathbf{P}_\chi^{f^{-1}}(\chi_s - \chi_s^b) +$$

$$\frac{1}{2}(\bar{\chi}_s - \bar{\chi}_s^b)^{\mathrm{T}}(\bar{\chi}_s - \bar{\chi}_s^b), \tag{5.15}$$

and we shall seek to determine explicitly the part of $\mathbf{P}_\chi^{f^{-1}}$ that is important.

It is possible to define a suitable **X** using a sequence of Householder

transformations (see Sec. 5.6).

5.5 Determination of P_χ^{f-1}

The key to simplifying (5.15) is in the design of $\mathbf{X}$. Let $\mathbf{X}$ be designed such that:

- $\mathbf{X}^{\mathrm{T}}$ acting on any vector in the subspace $\mathbf{L}^{-1}\boldsymbol{s}_k$ gives a vector that is non-zero only in the first K elements

$$\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\boldsymbol{s}_k = \begin{pmatrix} \alpha_1 \\ \cdots \\ \alpha_K \\ 0 \\ \cdots \\ 0 \end{pmatrix}, \tag{5.16}$$

- and $\mathbf{X}^{\mathrm{T}}$ acting on a vector in the orthogonal complement gives a vector that is non-zero only in the remaining $n - K$ elements.

How this design specification of $\mathbf{X}$ can be achieved is covered in section 5.6. The vector $\mathbf{L}^{-1}\delta\boldsymbol{x}_s$ spans the space defined by the K vectors $\mathbf{L}^{-1}\boldsymbol{s}_k$, and consequently, χ_s and χ_s^b comprise only the first K elements and this must be maintained in the minimization of J_B. This means that in (5.15), only the first K columns of $\mathbf{P}_\chi^{f-1}$ need to be known. Following Fisher (1998) let

$$\mathbf{Z} = \mathbf{P}^{f^{-1}}\mathbf{S}, \tag{5.17}$$

where $\mathbf{S}$ is the $n \times K$ matrix whose columns are the $\boldsymbol{s}_k$ and $\mathbf{Z}$ is the $n \times K$ result after acting with the inverse of the flow-dependent error covariance matrix. Let us develop this expression using the definition (5.14) along the way

$$\begin{aligned} \mathbf{Z} &= \mathbf{P}^{f^{-1}}\mathbf{L}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S}, \\ \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z} &= \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{L}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S}, \\ &= \mathbf{P}_\chi^{f^{-1}}\hat{\mathbf{I}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S}, \\ \mathbf{X}^{\mathrm{T}}_{(n\times n)}\mathbf{L}^{\mathrm{T}}_{(n\times n)}\mathbf{Z}_{(n\times K)} &= \mathbf{P}^{f^{-1}}_{\chi(n\times K)}\hat{\mathbf{I}}_{(K\times n)}\mathbf{X}^{\mathrm{T}}_{(n\times n)}\mathbf{L}^{-1}_{(n\times n)}\mathbf{S}_{(n\times K)}, \end{aligned} \tag{5.18}$$

where $\hat{\mathbf{I}}$ is the compound operator

$$\hat{\mathbf{I}}_{(K\times n)} = (\mathbf{I}_{(K\times K)}\mathbf{0}_{(K\times n-K)}). \tag{5.19}$$

This operator is included to remove the superfluous zero-elements for rows $i > K$ of $\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S}$ (by the design of $\mathbf{X}$). In (5.19) and in the line after (5.18), labels have been added to the matrices to indicate their dimensions. Equation (5.18) allows us to write

$$\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z}\,(\hat{\mathbf{I}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S})^{-1} = \mathbf{P}_\chi^{f^{-1}}, \tag{5.20}$$

where the operator inverted is a calculable $K \times K$ matrix, which we assume is non-singular. Note that (5.20) is for only part of the inverse covariance matrix and so is not symmetric. The matrix yet unknown is $\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z}$ which, as shown below, is a by-product the Hessian singular vector calculation shown in section 5.2. Firstly, by the definition of $\mathbf{Z}$, (5.17)

$$\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{S}, \tag{5.21}$$

the right hand side of which can be found from (5.5), as follows. Let columns of a new matrix, $\mathbf{S}_{-T}$, be those K states at $t = -T$ that evolve into the columns of $\mathbf{S}$ (the columns of $\mathbf{S}_{-T}$ are the states $\delta\boldsymbol{x}_k(-T)$ in (5.5)

$$\mathbf{S} = \mathbf{M}_{0\leftarrow -T}\mathbf{S}_{-T}. \tag{5.22}$$

This is useful in the derivation to follow. First write (5.5) in complete matrix form

$$\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{M}_{0\leftarrow -T}\mathbf{S}_{-T} = \mathbf{P}^{a^{-1}}\mathbf{S}_{-T}\Lambda, \tag{5.23}$$

where Λ is the diagonal matrix of λ_k. Also important is the propagation of the error covariances (ignoring the model error contribution)

$$\mathbf{P}^{f} = \mathbf{M}_{0\leftarrow -T}\mathbf{P}^{a}\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}. \tag{5.24}$$

These equations can be manipulated to give the matrix in (5.21) required to solve (5.20). Starting from (5.23)

$$\begin{aligned}
\mathbf{P}^{a}\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1} &= \mathbf{S}_{-T},\\
\mathbf{M}_{0\leftarrow -T}\mathbf{P}^{a}\mathbf{M}_{0\leftarrow -T}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1} &= \mathbf{M}_{0\leftarrow -T}\mathbf{S}_{-T},\\
\mathbf{P}^{f}\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1} &= \mathbf{S},\\
\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1} &= \mathbf{P}^{f^{-1}}\mathbf{S},\\
\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1} &= \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{f^{-1}}\mathbf{S},\\
&= \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z} \qquad \text{by (5.21)},\\
\therefore \qquad \mathbf{P}_{\chi}^{f^{-1}} &= \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{S}\Lambda^{-1}(\hat{\mathbf{I}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{S})^{-1} \qquad \text{by (5.20)}.
\end{aligned} \tag{5.25}$$

The right hand side of (5.25) is known and thus all relevant elements of the background cost function (5.15) are calculable.

5.6 Design of X

It remains to be shown how **X** can be formulated to achieve property (5.16). Fisher (1998) states that this is achieved with a sequence of Householder transformations. A single Householder transformation, **P**, (e.g. Press et al. 1986) may be written as follows

$$\mathbf{P} = \mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{T}\boldsymbol{u}}, \tag{5.26}$$

where

$$\boldsymbol{u} = \boldsymbol{x} \mp |\boldsymbol{x}|\ \boldsymbol{e}_1. \tag{5.27}$$

The vector $\boldsymbol{x}$ is arbitrary and $\boldsymbol{e}_1$ is a vector which is zero valued except for the first element, which has unit value (the properties of the Householder transformation hold for a general single element being chosen instead, although here we always choose the first element). **P** is useful because it has the following useful properties.

- The Householder transformation is orthogonal

$$\begin{aligned}
\mathbf{P}\mathbf{P}^{T} &= \left(\mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)\left(\mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)^{\mathrm{T}},\\
&= \mathbf{I} - 4\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} + 4\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}},\\
&= \mathbf{I} - 4\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} + 4\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} = \mathbf{I}.
\end{aligned} \tag{5.28}$$

- When acting on the state $\boldsymbol{x}$, which is used to define **P** in (5.26) and (5.27), the result is a vector with all but the first element zero

$$\begin{aligned}
\mathbf{P}\boldsymbol{x} &= \left(\mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)\boldsymbol{x},\\
&= \left(\mathbf{I} - 2\frac{\boldsymbol{u}(\boldsymbol{x} \mp |\boldsymbol{x}|\ \boldsymbol{e}_1)^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)\boldsymbol{x},\\
&= \left(\mathbf{I} - 2\frac{\boldsymbol{u}(\boldsymbol{x} \mp |\boldsymbol{x}|\ \boldsymbol{e}_1)^{\mathrm{T}}}{2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp 2\ |\boldsymbol{x}|\ x_1}\right)\boldsymbol{x},
\end{aligned}$$

$$= \boldsymbol{x} - 2\frac{\boldsymbol{u}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp |\boldsymbol{x}|\ x_1)}{2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp 2\ |\boldsymbol{x}|\ x_1},$$

$$= \boldsymbol{x} - \boldsymbol{u},$$

$$= \pm\ |\boldsymbol{x}|\ \boldsymbol{e}_1. \tag{5.29}$$

- When acting on a state $\bar{\boldsymbol{x}}$, which is orthogonal to the state $\boldsymbol{x}$, which is used to define $\mathbf{P}$ in (5.26) and (5.27), the result is a vector with zero in the first element

$$\text{Note} \qquad \boldsymbol{u}^{\mathrm{T}}\bar{\boldsymbol{x}} = \boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} \mp |\boldsymbol{x}|\ \boldsymbol{e}_1^{\mathrm{T}}\bar{\boldsymbol{x}},$$

$$= \mp\ |\boldsymbol{x}|\ \bar{x}_1, \tag{5.30}$$

$$\text{then} \qquad \mathbf{P}\bar{\boldsymbol{x}} = \left(\mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)\bar{\boldsymbol{x}},$$

$$= \bar{\boldsymbol{x}} \pm 2\frac{|\boldsymbol{x}|\ \bar{x}_1(\boldsymbol{x} \mp |\boldsymbol{x}|\ \boldsymbol{e}_1)}{2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp 2\ |\boldsymbol{x}|\ x_1},$$

$$= \frac{|\boldsymbol{x}|^2\bar{\boldsymbol{x}} \mp |\boldsymbol{x}|\ x_1\bar{\boldsymbol{x}} \pm |\boldsymbol{x}|\ \bar{x}_1\boldsymbol{x} - \bar{x}_1\ |\boldsymbol{x}|^2\boldsymbol{e}_1}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp |\boldsymbol{x}|\ x_1}. \tag{5.31}$$

This does not have weight in element 1. To show this, do a scalar product with $\boldsymbol{e}_1$

$$\boldsymbol{e}_1^{\mathrm{T}}\mathbf{P}\bar{\boldsymbol{x}} = \frac{|\boldsymbol{x}|^2\bar{x}_1 \mp |\boldsymbol{x}|\ x_1\bar{x}_1 \pm |\boldsymbol{x}|\ \bar{x}_1x_1 - \bar{x}_1\ |\boldsymbol{x}|^2}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \mp |\boldsymbol{x}|\ x_1} = 0, \tag{5.32}$$

In these equations, $\bar{x}_1$ and x_1 are the first components of $\bar{\boldsymbol{x}}$ and $\boldsymbol{x}$ respectively. The first property gives $\mathbf{P} = \mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}$. These properties can be combined to give $\mathbf{X}$ in the following way. Defining $\mathbf{R}^{(0)} = \mathbf{L}^{-1}\mathbf{S}$, let $\mathbf{X}^{\mathrm{T}}\mathbf{R}^{(0)}$ be a vector of two components

$$\mathbf{X}^{\mathrm{T}}\mathbf{R}^{(0)} = \begin{pmatrix}\mathbf{A}\\ \mathbf{0}\end{pmatrix}, \tag{5.33}$$

where $\mathbf{A}$ is a $K \times K$ matrix consistent with the required property of (5.16). In fact, by the way that $\mathbf{X}^{\mathrm{T}}$ is to be formed, $\mathbf{A}$ will turn out to be upper triangular. Let

$$\mathbf{X}^{\mathrm{T}}\mathbf{R}^{(0)} = \mathbf{P}_K \ldots\ \mathbf{P}_k \ldots\ \mathbf{P}_2\mathbf{P}_1\mathbf{R}^{(0)}. \tag{5.34}$$

Each $\mathbf{P}_l$ transformation is Householder-like according to the following, e.g. for $\mathbf{P}_1$

$$\mathbf{P}_1\mathbf{R}^{(0)} = \left(\mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\right)\mathbf{R}^{(0)},\ \boldsymbol{u} = \boldsymbol{r}_1^{(0)} - |\boldsymbol{r}_1^{(0)}|\ \boldsymbol{e}_1^{(n)}, \tag{5.35}$$

where $\boldsymbol{r}_1^{(0)}$ is the first column of $\mathbf{R}^{(0)}$ and $\boldsymbol{e}_1^{(n)}$ is the n-element vector with all but the first element zero (which is unity). This generates a new matrix $\mathbf{R}^{(1)} = \mathbf{P}_1\mathbf{R}^{(0)}$ which has the form

$$\mathbf{R}^{(1)} = \begin{pmatrix} r_{11}^{(1)} & r_{12}^{(1)} & r_{13}^{(1)} & \ldots & r_{1K}^{(1)} \\ 0 & r_{22}^{(1)} & r_{23}^{(1)} & \ldots & r_{2K}^{(1)} \\ 0 & r_{32}^{(1)} & r_{33}^{(1)} & \ldots & r_{3K}^{(1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & r_{k2}^{(1)} & r_{k3}^{(1)} & \ldots & r_{kK}^{(1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & r_{n2}^{(1)} & r_{n3}^{(1)} & \ldots & r_{nK}^{(1)} \end{pmatrix}, \tag{5.36}$$

having only the first element non-zero of the first column (since $\mathbf{P}_1$ is designed in terms of the first column of $\mathbf{R}^{(0)}$. The aim now is to act with a $n-1 \times n-1$ element Householder operator on $\mathbf{R}^{(1)}$ excluding the first row and first column

$$\mathbf{P}_2\mathbf{R}^{(1)} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} \end{pmatrix}\mathbf{R}^{(1)}, \quad \boldsymbol{u} = \boldsymbol{r}_2^{(1)} - |\boldsymbol{r}_2^{(1)}|\, \boldsymbol{e}_1^{(n-1)}, \tag{5.37}$$

where the partitioned-off part of $\mathbf{P}_2$ is a $n-1 \times n-1$ matrix, $\boldsymbol{r}_2^{(1)}$ is the $n-1$-element second column of $\mathbf{R}^{(1)}$ (excluding the first component) and $\boldsymbol{e}_1^{(n-1)}$ is the $n-1$-element vector with all but the first element zero (which is unity). This generates a new matrix $\mathbf{R}^{(2)} = \mathbf{P}_2\mathbf{R}^{(1)}$ which has the form

$$\mathbf{R}^{(2)} = \begin{pmatrix} r_{11}^{(1)} & r_{12}^{(1)} & r_{13}^{(1)} & \dots & r_{1K}^{(1)} \\ 0 & r_{22}^{(2)} & r_{23}^{(2)} & \dots & r_{2K}^{(2)} \\ 0 & 0 & r_{33}^{(2)} & \dots & r_{3K}^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & r_{k3}^{(2)} & \dots & r_{kK}^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & r_{n3}^{(2)} & \dots & r_{nK}^{(2)} \end{pmatrix}. \tag{5.38}$$

The kth operator, $\mathbf{P}_k$, has the form

$$\mathbf{P}_k\mathbf{R}^{(k-1)} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}}{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}} \end{pmatrix}\mathbf{R}^{(k-1)}, \quad \boldsymbol{u} = \boldsymbol{r}_k^{(k-1)} - |\boldsymbol{r}_k^{(k-1)}|\, \boldsymbol{e}_1^{(n-k+1)}, \tag{5.39}$$

where the partitioned-off part of $\mathbf{P}_k$ is a $n-k+1 \times n-k+1$ matrix, $\boldsymbol{r}_k^{(k-1)}$ is the $n-k+1$-element kth column of $\mathbf{R}^{(k-1)}$ (excluding the first $k-1$ components) and $\boldsymbol{e}_1^{(n-k+1)}$ is the $n-k+1$-element vector with all but the first element zero (which is unity). After all K operators have acted, the result is $\mathbf{R}^{(k)} = \mathbf{X}^{\mathrm{T}}\mathbf{R}^{(0)}$

$$\mathbf{R}^{(k)} = \left(\begin{array}{ccccc} r_{11}^{(1)} & r_{12}^{(1)} & r_{13}^{(1)} & \dots & r_{1K}^{(1)} \\ 0 & r_{22}^{(2)} & r_{23}^{(2)} & \dots & r_{2K}^{(2)} \\ 0 & 0 & r_{33}^{(3)} & \dots & r_{3K}^{(3)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & r_{KK}^{(k)} \\ \hline \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{array}\right). \tag{5.40}$$

where the top section is the matrix $\mathbf{A}$ in (5.33), and the bottom section comprises zeros.

It should also be shown that $\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{I}$. From (5.34) this is easy to show given that each pair has the property that $\mathbf{P}_k^{\mathrm{T}}\mathbf{P}_k = \mathbf{I}$

$$\mathbf{P}_1^{\mathrm{T}}\mathbf{P}_2^{\mathrm{T}}\dots\, \mathbf{P}_k^{\mathrm{T}}\dots\, \mathbf{P}_K^{\mathrm{T}}\mathbf{P}_K\dots\, \mathbf{P}_k\dots\, \mathbf{P}_2\mathbf{P}_1 = \mathbf{I}. \tag{5.41}$$

It remains to be shown that the string of Householder operators $\mathbf{P}_K\dots\, \mathbf{P}_k\dots\, \mathbf{P}_2\mathbf{P}_1$ acting on a state, $\bar{\boldsymbol{r}}^{(0)}$ (which is orthogonal to all columns of $\mathbf{R}^{(0)}$) gives a state that is zero in the first K elements. All $\mathbf{P}_k$ are formed in the same way as shown above (ie with respect to the $\mathbf{R}^{(k)}$ matrices).

First let $\bar{\boldsymbol{r}}^{(1)} = \mathbf{P}_1\bar{\boldsymbol{r}}^{(0)}$. Since $\bar{\boldsymbol{r}}^{(0)}$ is orthogonal to $\boldsymbol{r}_1^{(0)}$ (the latter is the vector used to define $\mathbf{P}_1$ in (5.35)), and by property (5.32), the vector $\bar{\boldsymbol{r}}^{(1)}$ has zero in the first element. Next, let $\bar{\boldsymbol{r}}^{(2)} = \mathbf{P}_2\bar{\boldsymbol{r}}^{(1)}$. By a similar argument, if vector formed from the last $n-1$ components of $\bar{\boldsymbol{r}}^{(1)}$ is orthogonal to the vector formed from the last $n-1$ components of $\boldsymbol{r}_2^{(1)}$ (the latter is the vector used to define $\mathbf{P}_2$ in (5.37)), and by property (5.32), the vector $\bar{\boldsymbol{r}}^{(2)}$ will have zero in the first two elements. Because the

first element of $\mathbf{P}_1\bar{\boldsymbol{r}}^{(0)}$ is always zero, the remaining $n-1$-component inner product in question is equal to the full n-component inner product as the first element contributes zero. The orthogonality test is therefore satisfied if the following n-component inner product is zero

$$(\mathbf{P}_1\bar{\boldsymbol{r}}^{(0)})^T\,\mathbf{P}_1\boldsymbol{r}_2^{(0)} \;=\; \bar{\boldsymbol{r}}^{(0)^{\mathrm{T}}}\mathbf{P}_1^{\mathrm{T}}\mathbf{P}_1\boldsymbol{r}_2^{(0)} \;=\; \bar{\boldsymbol{r}}^{(0)^{\mathrm{T}}}\boldsymbol{r}_2^{(0)} \;=\; 0. \tag{5.42}$$

This is satisfied because the vector $\bar{\boldsymbol{r}}^{(0)}$ is orthogonal to $\boldsymbol{r}_2^{(0)}$ by definition. These arguments continue for all K operators. The final result is a vector of zeros in the first K components.

5.7 Summary of the RRKF

There are a number of issues relating to the RRKF.

- The analysis error estimation in the Hessian singular vector calculation may be inadequate. The modes that actually dominate the background error at $t = 0$ may not be those that dominate the singular vector calculation at $t = -T$. For example, the actual fastest growing modes may be well constrained by the previous cycle's analysis, and so not be given prominence in the Hessian singular vector calculation, or the modes that do emerge from the Hessian singular vector calculation may saturate in the non-linear system and be overtaken by other modes.
- The number of modes used (10-25) may be inadequate. Ehrendorfer and Bouttier (1998) find that in some studies as many as 100 singular vectors are needed to account for just over half of the forecast error over two days. Furthermore, the blending of flow dependent and static error covariances may lead to physically unrealistic structure functions (Beck and Ehrendorfer 2005).
- No account is taken of model error in the propagation of covariance information from the previous analysis to the current background, although Beck and Ehrendorfer (2005) show how model error can be added.
- The RRKF makes fresh estimates of the flow-dependent subspace for every cycle, and so the propagation of flow dependent information through the system is largely new for every cycle.
- The method is only as good as the knowledge of the Hessian in the Hessian singular vector calculation (Barkmeijer et al. 1998).

5.8 Appendix on the use of ensembles to define the subspace

6. Homogeneous structure functions in real and spectral-spaces

6.1 Real space

An error covariance that gives separable structure functions in real space can be written as a convolution. A convolution in 1-D has the form

$$g(x) \;=\; \int \mathrm{d}x' C(x;\,x') f(x'), \tag{6.1}$$

If $f(x) \;=\; \delta(x - x_0)$, then the result of (1) is

$$g(x) = \int \mathrm{d}x' C(x; x') \delta(x' - x_0),$$
$$= C(x; x_0), \tag{6.2}$$

which is the structure function associated with the source point x_0.

Let our 2-D structure functions be the separable in horizontal and vertical positions

$$R(r; r_0, z_0) Z(z; r_0, z_0), \tag{6.3}$$

where r (z) is the horizontal (vertical) position. The real-space correlation model is thus

$$g(r, z) = \int \int \mathrm{d}r' \, \mathrm{d}z' R(r; r', z') Z(z; r', z') f(r', z'). \tag{6.4}$$

This can be written in the more familiar way using the matrix notation

$$\vec{g} = \mathbf{C}\vec{f}. \tag{6.5}$$

6.2 Spectral space

In spectral space, first write f and g as linear combinations of horizontal basis functions

$$f(r, z) = \int \mathrm{d}k' \bar{f}(k', z) F_h(r, k'), \tag{6.6}$$

$$g(r, z) = \int \mathrm{d}k' \bar{g}(k', z) F_h(r, k'). \tag{6.7}$$

In these equations, k is the horizontal wavenumber. Note the orthogonality relations of the horizontal basis functions,

$$\int \mathrm{d}r F_h^*(r, k) F_h(r, k') = \delta(k - k'), \tag{6.8}$$

Inserting (6.6) and (6.7) into (6.4) gives,

$$\int \mathrm{d}k' \bar{g}(k', z) F_h(r, k') = \int \int \mathrm{d}r' \, \mathrm{d}z' R(r; r', z') Z(z; r', z')$$
$$\int \mathrm{d}k' \bar{f}(k', z') F_h(r', k'). \tag{6.9}$$

Multiplying (6.9) by $F_h^*(r, k)$ and integrating over r gives

$$\int \mathrm{d}r \int \mathrm{d}k' \bar{g}(k', z) F_h^*(r, k) F_h(r, k') = \int \mathrm{d}r \int \int \mathrm{d}r' \, \mathrm{d}z' R(r; r', z') Z(z; r', z')$$
$$\int \mathrm{d}k' \bar{f}(k', z') F_h^*(r, k) F_h(r', k'). \tag{6.10}$$

Orthogonality (6.8) results in the left hand side simplifying,

$$\bar{g}(k, z) = \int \mathrm{d}r \int \int \mathrm{d}r' \, \mathrm{d}z' R(r; r', z') Z(z; r', z')$$
$$\int \mathrm{d}k' \bar{f}(k', z') F_h^*(r, k) F_h(r', k'). \tag{6.11}$$

This is the correlation (6.4), but in spectral space. Rearrange the order of the integrals in (6.11)

$$\bar{g}(k, z) = \int \mathrm{d}k' [\int \mathrm{d}r \int \int \mathrm{d}r' \, \mathrm{d}z' R(r; r', z') Z(z; r', z') \times \times$$
$$F_h^*(r, k) F_h(r', k')] \; \bar{f}(k', z'). \tag{6.12}$$

The term in large brackets constitutes the covariance model operator in spectral space. The question now is: is this model separable in spectral space? This is easy to show by acting with (6.12) on the state

$$\bar{f}(k', z') = \delta(k' - k_0) \delta(z' - z_0). \tag{6.13}$$

If the result, $\bar{g}(k, z)$ is separable, then we have proved the result

$$\bar{g}(k, z) = \int \mathrm{d}r \int \mathrm{d}r' R(r; r', z_0) F_h^*(r, k) F_h(r', k_0) Z(z; r', z_0). \tag{6.14}$$

This is not separable in general - (6.14) is not of the form of a product of a function that depends upon k, but not z, with another that depends upon z, but not k. If Z becomes horizontally homogeneous (ie is no longer a function of r') then (6.14) becomes

$$\bar{g}(k, z) = \int \mathrm{d}r \int \mathrm{d}r' R(r; r', z_0) F_h^*(r, k) F_h(r', k_0) \times Z(z; z_0), \quad (6.15)$$

which is separable in spectral space.

6.3 Conclusion

A correlation that is separable in real space (and has vertical structure functions that are horizontally homogeneous) is also separable in spectral space.

7. How do observations affect background error covariance lengthscales?

7.1 Introduction

It is often reported that the correlation lengthscale of background errors is reduced in the presence of observations (Ingleby 2001). This note discusses the mechanism of how this occurs. We look at the simpler problem of analysing how the presence of a single observation affects the correlation lengthscale of the analysis state, rather than the correlation lengthscale of the subsequent forecast (it is reasonable to expect that the qualitative characteristics of the analysis will be carried forward in the subsequent forecast).

7.2 Simple analysis

Consider a background error covariance matrix, **B**. Let it be homogeneous and isotropic, so that its representation in spectral space is diagonal (and depend upon total wavenumber only). Its spectral representation shall be used below. Let the observation system be denoted by the Jacobian, **H**, and let the error covariance of the observations be **R**. The error covariance matrix of the analysed state, **A**, is the inverse Hessian

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}. \quad (7.1)$$

The space in which the analysis state, the **B**-matrix and the right-space of **H** is usually spatial. It is more convenient for this analysis to look at **A** in spectral space. Let the following operators be the forward and inverse Fourier transform operators

$$\text{Forward FT:} \quad \mathbf{F}^{-1}, \quad (7.2)$$

$$\text{Inverse FT:} \quad \mathbf{F}. \quad (7.3)$$

The Fourier transform is orthogonal, $\mathbf{F}^T = \mathbf{F}^{-1}$. In one-dimension, the matrix elements of **F** are proportional to simple plane waves

$$\mathbf{F}_{mq} = \frac{1}{\sqrt{N}} \exp ix_m k_q, \quad (7.4)$$

for N grid-points. The spectral-space version of (8.1) is $\mathbf{F}^T\mathbf{A}\mathbf{F}$

$$\begin{aligned} \mathbf{F}^T\mathbf{A}\mathbf{F} &= \mathbf{F}^T(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{F}, \\ &= (\mathbf{F}^T\mathbf{B}^{-1}\mathbf{F} + \mathbf{F}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{F})^{-1}, \\ &= ([\mathbf{F}^T\mathbf{B}\mathbf{F}]^{-1} + [\mathbf{H}\mathbf{F}]^T\mathbf{R}^{-1}[\mathbf{H}\mathbf{F}])^{-1}. \end{aligned} \quad (7.5)$$

The reason for converting to spectral space is for simplicity - we shall assume homogeneity throughout (and so the covariances in spectral space are diagonal) and we can infer lengthscales from the variance spectra - see below.

The background term in spectral space, $\mathbf{F}^T\mathbf{B}\mathbf{F}$, is diagonal, with diagonal elements

$$[\mathbf{F}^T\mathbf{B}\mathbf{F}]_{qq} = \sigma_B^2(q). \tag{7.6}$$

Consider (7.5) with just one observation of grid-point l. The Jacobian is then

$$\mathbf{H} = (0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0), \tag{7.7}$$

with the '1' at position l. The operator $\mathbf{HF}$, which is the single row Jacobian acting on a spectral-space state, is made up of elements

$$[\mathbf{HF}]_{1q} = \frac{1}{\sqrt{N}} \exp ix_l k_q. \tag{7.8}$$

By taking the transpose operator to include an additional complex conjugate operation, the operator $[\mathbf{HF}]^T\mathbf{R}^{-1}[\mathbf{HF}]$ is then

$$\begin{aligned}[[\mathbf{HF}]^T\mathbf{R}^{-1}[\mathbf{HF}]]_{qk} &= \frac{1}{\sqrt{N}} \exp -ix_l k_q \times \frac{1}{\sigma_{Ob}^2} \times \frac{1}{\sqrt{N}} \exp ix_l k_k, \\ &= \frac{\exp ix_l (k_k - k_q)}{N\sigma_{Ob}^2},\end{aligned} \tag{7.9}$$

where σ_{Ob} is the observation standard deviation. Enforcing a homogeneous model, the contribution, $[\mathbf{HF}]^T\mathbf{R}^{-1}[\mathbf{HF}]$, is diagonal. Ignoring off-diagonal elements (setting them to zero), leaves the diagonal elements

$$[[\mathbf{HF}]^T\mathbf{R}^{-1}[\mathbf{HF}]]_{qq} = \frac{1}{N\sigma_{Ob}^2}, \tag{7.10}$$

which is a constant. Wavenumber component q of (7.5) is thus

$$\begin{aligned}[\mathbf{F}^T\mathbf{A}\mathbf{F}]_{qq} &= \left(\frac{1}{\sigma_B^2(q)} + \frac{1}{N\sigma_{Ob}^2}\right)^{-1}, \\ &= \left(\frac{N\sigma_{Ob}^2 + \sigma_B^2(q)}{\sigma_B^2(q)N\sigma_{Ob}^2}\right)^{-1} = \frac{\sigma_B^2(q)N\sigma_{Ob}^2}{N\sigma_{Ob}^2 + \sigma_B^2(q)}.\end{aligned} \tag{7.11}$$

Equation (7.11) says that those modes of the background state that have a small variance, ie $\sigma_B^2(q) << N\sigma_{Ob}^2$, will be unaffected by the observation. Those modes that have a large variance, ie $\sigma_B^2(q) >> N\sigma_{Ob}^2$, will have its variance reduced to a value, at most $N\sigma_{Ob}^2$. This is illustrated in Fig. 7.1. The variance of the longer modes (small wavenumbers) have reduced in value. As this has not affected the small modes (large wavenumbers), this has the effect to broaden the correlation spectra associated with the variance spectra in Fig. 7.1. This will shorten the lengthscale in positional space. In other words, the longer scales are analysed more than smaller scales, because they started with larger variance in **B**.

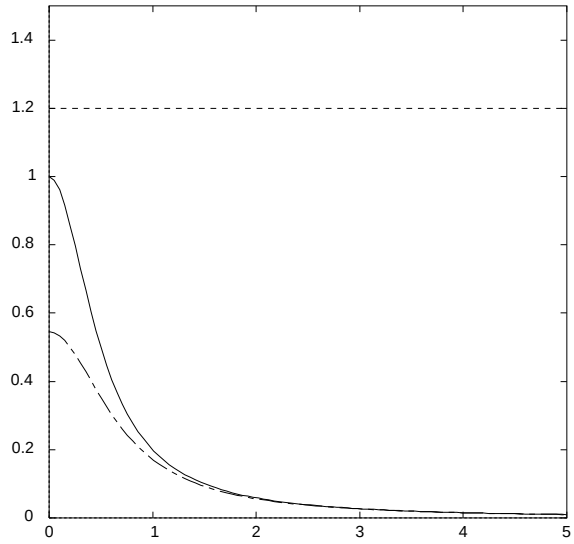

Figure 7.1: Variance spectra: **B**-matrix in spectral space (continuous curve), $N\sigma_{Ob}^2$ (straight dotted line) and **A**-matrix in spectral space (dash-dotted line).

7.3 Numerical study - dropping homogeneity

By dropping the homogeneity assumption, off-diagonal elements in (7.9) will be present. This is result in an **A**-matrix that has a reduced lengthscale, but only in the vicinity of the observation. This can be shown numerically.

Consider the one-dimensional system ($0 \leqslant x \leqslant 1$) with a number of observations near the middle of the domain. Let observation i make a direct measurement at position x. The row in the Jacobian will be zero apart from the elements corresponding to grid-points immediately before (x_1) and after (x_2) the observation. These will have elements

$$1 - \frac{x - x_1}{x_2 - x_1} \qquad \text{and} \qquad \frac{x - x_1}{x_2 - x_1}, \tag{7.12}$$

by assuming linear interpolation.

In this numerical study, we need not invoke spectral space, and deal with small matrices directly. We use 30 points and 5 observations near the centre of the domain. The background error covariances have the simple form

$$\mathbf{B} \ : \ \mathrm{COV}_B(\Delta x) \ = \ \frac{\sigma_B^2}{1 \ + \ (\Delta x / L)^2}, \tag{7.13}$$

where Δx is distance, $\sigma_B = 0.1$ and the correlation length parameter $L = 0.2$, and we invoke periodic boundary conditions for a well-behaved background error covariance matrix. The analysis error covariance matrix is then,

$$\mathbf{A} \ = \ (\mathbf{B}^{-1} \ + \ \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1} . \tag{7.14}$$

Plotted in Fig. 7.2 are the background error and analysis error correlation matrices,

$$\mathrm{COR}_B \ = \ \Sigma_B^{-1}\mathbf{B}\Sigma_B^{-1}, \tag{7.15}$$

$$\mathrm{COR}_A \ = \ \Sigma_A^{-1}\mathbf{A}\Sigma_A^{-1}, \tag{7.16}$$

where Σ_B is the background error standard deviation matrix $\Sigma_B = \sigma_B\mathbf{I}$, and Σ_A is the analysis error standard deviation matrix.

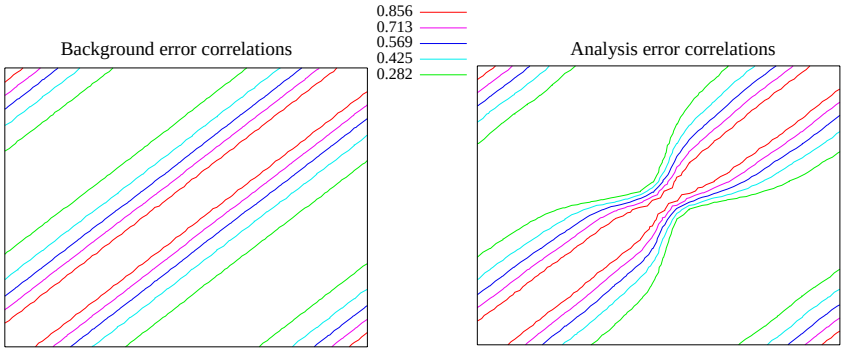

Figure 7.2: The background error correlations from (7.13) and (7.15) (left) and the analysis error correlations from (7.14) and (7.16) after the assimilation of five observations near the centre of the domain.

The analysis state shows clearly the reduction of correlation lengthscale in the analysis in the vicinity of the observation locations.

References

Bannister R.N., 2008a, A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, Q. J. Royal. Met. Soc. 134, 1951-1970.

Bannister R.N., 2008b, A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance stastics, Q. J. Royal. Met. Soc. 134, 1971-1996.

Barker D.M., Lorenc A.C., Semple A., 2005, The use of synoptically-dependent background error structures in 3dVar, Var. Scientific Documentation Paper No. 26, Met Office, Exeter, 11 pp.

Barkmeijer J.M., van Gijzen, Bouttier F., 1998, Singular vectors and estimates of the analysis error covariance metric, Q. J. Royal. Met. Soc. 124, 1695-1713.

Beck A., Ehrendorfer M., 2005, Singular-vector-based covariance propagation in a quasigeostrophic assimilation system, Mon. Wea. Rev. 133, 1295-1310.

Berre L., 2000, Estimation of synoptic and mesoscale forecast error covariances in a limited-area model, Mon. Wea Rev. 128, 644-667.

Ehrendorfer M., Bouttier F., 1998, An explicit low resolution extended Kalman filter: Implementation and preliminary eperimentation, ECMWF Tech. Memo. 259.

Fisher M., 1998, Development of a simplified Kalman filter, ECMWF Tech. Memo. No. 260, 1-16.

Fisher M., 2003, Background error covariance modelling, in Seminar Proceedings, Recent developments in data assimilation for atmosphere and ocean, 8-12 Sept. 2003, 45-63.

Hello G, Bouttier F., 2001, Using adjoint sensitivity as a local structure function in variational data assimilation, Nonlinear Processes in Geophys. 8, 347-355.

Hólm E., Andersson E., Beljaars A., Lopez P., Mahfouf J.F., Simmons A., Thépaut J.N., 2002,

Assimilation and modelling of the hydrological cycle: ECMWF's status and plans, ECMWF Tech. Memo. 383, 1-55.

Ingleby N.B., 2001, The statistical structure of forecast errors and its representation in the Met Office global 3-dimensional variational data assimilation system. Royal. Met. Soc. 127, 209-231.

Parrish D.F., Derber J.C., 1992, The National Meteorological Center's spectral statistical interpolation analysis system, Mon. Wea. Rev., 120, 1747-1763.

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 1986, Numerical Recipes.

Riishøjgaard L.P., 1998, A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, Tellus, 50A, 42-57.

Toth Z., Kalnay E., 1993, Ensemble forecasting at NMC: the generation of perturbations, Bull. Amer. Meteor. Soc., 74, 2317-2330.

Toth Z., Kalnay E., 1997, Ensemble forecasting at NCEP: the breeding method, Mon. Wea. Rev., 125, 3297-3318.