A method for merging flow-dependent forecast error statistics from an ensemble with static statistics for use in high resolution variational data assimilation

R. E. Petrie and R. N. Bannister

Department of Meteorology, Earley Gate, University of Reading, Reading, RG6 6BB, United Kingdom

Abstract

The background error covariance matrix, $\mathbf{B}$, is often used in variational data assimilation for numerical weather prediction as a static and hence poor approximation to the fully dynamic forecast error covariance matrix, $\mathbf{P}^{\mathrm{f}}$. In this paper the concept of an Ensemble Reduced Rank Kalman Filter (EnRRKF) is outlined. In the EnRRKF the forecast error statistics in a subspace defined by an ensemble of states forecast by the dynamic model are found. These statistics are merged in a formal way with the static statistics, which apply in the remainder of the space. The combined statistics may then be used in a variational data assimilation setting. It is hoped that the nonlinear error growth of small-scale weather systems will be accurately captured by the EnRRKF, to produce accurate analyses and ultimately improved forecasts of extreme events.

Keywords: Data assimilation, Background error covariance, Flow-dependence, Hybrid methods

1. INTRODUCTION

Data Assimilation for numerical weather prediction (NWP) is the process of combining observational data with a prior estimate of the atmospheric state to produce an analysis state that is the best fit to both. The analysis state is then used as the initial conditions for an NWP forecast. Data assimilation is carried out before every forecast to allow new observations to generate realistic initial conditions. The prior estimate (which is also known as the background state) is found from a short forecast starting from a previous analysis.

The specification of the error statistics of the observations and of the background state are crucial in data assimilation and are represented as observation and forecast error covariance matrices respectively. For instance the variance information, comprising the diagonal elements of the error covariance matrices, represent the degree of confidence in the data which allows the data assimilation scheme to determine whether the analysis state is closer to the observation or the background (e.g. [13]). The forecast error covariance matrix is particularly important because it spreads-out observational information spatially and to other variables by its off-diagonal elements [1]. Correct specification of the forecast error covariance matrix will aid the assimilation in production of a more realistic analysis.

For systems that evolve linearly, the Kalman Filter equations provide a framework in which the forecast error covariance matrix can be calculated [12]. Given a known error covariance matrix valid at time $k-1$, $\mathbf{P}^{\mathrm{a}}_{k-1}$, the forecast error covariance matrix at the forecast time, $\mathbf{P}^{\mathrm{f}}_{k}$ is found by evolving

$\mathbf{P}^{\mathrm{a}}_{k-1}$ with the linear model, $\mathbf{M}_{t_{k-1}\rightarrow t_k}$. Assuming a perfect model this gives

$$\mathbf{P}^{\mathrm{f}}_{k} = \mathbf{M}_{t_{k-1}\rightarrow t_k}\mathbf{P}^{\mathrm{a}}_{k-1}\mathbf{M}^{\mathrm{T}}_{t_{k-1}\rightarrow t_k}. \tag{1}$$

This equation yields a forecast error covariance matrix that is dependent upon the flow. This is often considered the 'gold standard' representation of the forecast error covariance matrix for linear systems with Gaussian statistics, which many data assimilation schemes try to emulate.

The state vector for operational NWP has $O(10^7)$ elements which leads to a $\mathbf{P}^{\mathrm{f}}$-matrix that has $O(10^{14})$ elements. This is too large to store or to propagate explicitly and therefore some approximations need to be made. There are a number of ways in which the $\mathbf{P}^{\mathrm{f}}$-matrix is represented in practice.

1. The Ensemble Kalman Filter (EnKF) [8] approximates the forecast state error covariance using an ensemble of forecasts. For an N-member ensemble the state vector (of dimension n) for member i is $\mathbf{x}_i$ for $i = 1, 2, \ldots, N$. The sample error covariance $\mathbf{P}^{\mathrm{f}}_{\mathrm{e}}$ is written as

$$\mathbf{P}^{\mathrm{f}}_{\mathrm{e}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_i - \langle\mathbf{x}\rangle\right)\left(\mathbf{x}_i - \langle\mathbf{x}\rangle\right)^{\mathrm{T}}, \tag{2}$$

   where $\langle x\rangle$ represents the ensemble mean.

2. In variational data assimilation a cost function is minimised to provide an optimal analysis e.g [14, 6]. The incremental 3D-VAR cost function is the sum of the background, $\mathcal{J}^{\mathrm{b}}$, and observation, $\mathcal{J}^{\mathrm{o}}$, penalties

$$\begin{aligned}\mathcal{J}(\delta\mathbf{x}) &= \underbrace{\frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x}}_{\mathcal{J}^{\mathrm{b}}} \\ &+ \underbrace{\frac{1}{2}\left[\mathbf{y} - \mathbf{H}(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y} - \mathbf{H}(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x})\right]}_{\mathcal{J}^{\mathrm{o}}}.\end{aligned} \tag{3}$$

Here $\delta\mathbf{x}$ is the increment, $\mathbf{x}^{\mathrm{b}}$ is the background state (where the full state vector is defined by $\mathbf{x} = \mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}$), the vector containing observational data is $\mathbf{y}$ and $\mathbf{H}$ is is the linear operator that maps from model to observational space. $\mathcal{J}^{\mathrm{b}}$ penalises the distance between $\mathbf{x}$ and $\mathbf{x}^{\mathrm{b}}$, and $\mathcal{J}^{\mathrm{o}}$ penalises the distance of the model observations to the measured observations. In variational data assimilation the forecast error covariances are crudely approximated by a matrix known as the background error covariance matrix, $\mathbf{B}$, which is usually a static, and thus suboptimal, estimate of $\mathbf{P}^{\mathrm{f}}$. For practical purposes, $\mathbf{B}$ can be represented in a compact way via use of a Control Variable Transform (CVT), denoted $\mathbf{L}$, as follows

$$\delta\mathbf{x} = \mathbf{L}\delta\boldsymbol{\chi}, \tag{4}$$

where $\delta\boldsymbol{\chi}$ is called the control vector. $\mathbf{L}$ is designed so that the forecast error covariance matrix, when in the represention of $\delta\boldsymbol{\chi}$, is the unit matrix. Thus $\left\langle \boldsymbol{\chi}\boldsymbol{\chi}^{\mathrm{T}} \right\rangle = \mathbf{I}$ where $\langle\rangle$ indicates an average over a population of forecast errors. Combining this property with (4) results in a simplified cost function

$$\begin{aligned} \mathcal{J}(\delta\boldsymbol{\chi}) &= \frac{1}{2}\delta\boldsymbol{\chi}^{\mathrm{T}}\delta\boldsymbol{\chi} \\ &+ \frac{1}{2}\left[\mathbf{y} - \mathbf{H}(\mathbf{x}^{\mathrm{b}} + \mathbf{L}\delta\boldsymbol{\chi})\right]^{\mathrm{T}} \mathbf{R}^{-1} \left[\mathbf{y} - \mathbf{H}(\mathbf{x}^{\mathrm{b}} + \mathbf{L}\delta\boldsymbol{\chi})\right]. \end{aligned} \tag{5}$$

Minimizing (5) with respect to $\delta\boldsymbol{\chi}$ is equivalent to minimizing (3) with respect to $\delta\mathbf{x}$ with the $\mathbf{B}$-matrix given by

$$\mathbf{B}_{\mathrm{imp}} = \left\langle \delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}} \right\rangle = \mathbf{L}\mathbf{L}^{\mathrm{T}}, \tag{6}$$

which is known as the implied $\mathbf{B}$-matrix. Part of the CVT introduces important large-scale meteorological balances into the system,

e.g. geostrophic and hydrostatic balances (e.g. [2]).

There are advantages and disadvantages of these two representations. The EnKF provides a flow-dependent $\mathbf{P}_{\mathrm{e}}^{\mathrm{f}}$-matrix, but the number of ensemble members, N, is in practice restricted by cost and is vastly less than n. This restriction on N leads to sampling errors, filter divergence, the introduction of spurioius correlations, and a rank deficient $\mathbf{P}_{\mathrm{e}}^{\mathrm{f}}$-matrix (e.g. [7]). The $\mathbf{B}$-matrix of standard variational assimilation is full rank, but is static and simplified. Variational data assimilation is currently the most common method for operational NWP for reasons such as computational efficiency [6]. As the resolution of weather forecast models increases so does their ability to resolve smaller-scale weather systems (e.g. thunderstorms) that can have a significant error growth. These systems would be particularly poorly represented by a static $\mathbf{B}$-matrix [11] and so the introduction of appropritate flow-dependent error statistics into the variational framework is an important area for investigation.

Hybrid methods, which combine ensemble and variational techniques, are currently a promising area of research. Various hybrid methods have been proposed which allow the introduction of flow-dependent error statistics into the variational framework. For example Liu et al. [15] proposed an ensemble based 4DVAR that allows flow-dependent forecast error statistics and avoids the need for the tangent linear and adjoint model. More recently Cheng et al. [5] proposed a hybrid method for determining the forecast error covariances by removing the most significant error directions in 4DVAR from a set of background ensemble perturbations and then adding back estimates of the forecast error in the same subspace. Other hybrid schemes exist though a

full review is beyond the scope of this paper (further details are found in [15] and [5]). In this paper a new mathematically rigorous hybrid method is proposed whereby a subset of flow-dependent covariance information from an ensemble is calculated for use within a variational setting. We call this method the Ensemble Reduced Rank Kalman Filter (EnRRKF) which builds on the Reduced Rank Kalman Filter (RRKF) introduced by Fisher [9].

In section 2 the RRKF and the EnRRKF are introduced, in section 3 the key differences between the schemes and the potential benefits of the EnRRKF over the RRKF are discussed, and in section 4 is a brief summary.

2. Mathematical Formulations

The RRKF [9] is employed in variational setting as a method of including a subset of explicitly evolving covariance information. A special subspace is identified in which the forecast error statistics are treated explicitly during the assimilation, the residual space is treated with the static **B**-matrix and the error covariances between the special and residual subspaces are also treated explicitly. In the standard RRKF [9] the special subspace is defined using the K dominant Hessian Singular Vectors (HSVs) [3, 9, 4], which are each evolved with the dynamic model to the background time and blended with the otherwise static **B**-matrix. Instead of defining the special subspace with HSVs, the proposed EnRRKF defines a subspace using a set of ensemble members. This merging of variational data assimilation and ensemble based techniques qualifies this scheme as a hybrid-method. The EnRRKF is a mathematically rigorous way of combining an ensemble estimated and a static representation of forecast error information.

The conventional RRKF has been dervied and analysed in [9, 10] and [4]. The implementation of the RRKF requires three stages: firstly the identification of the subspace in which the forecast error covariances are to gain explicit flow-dependence, secondly a modification to the otherwise standard control variable transform and thirdly devising a a means by which the flow-dependent information can be incorporated in variational data assimilation. The basic framework of the RRKF is given below, with some alterations for convienience, which is followed by the modifications made to form the EnRRKF.

2.1. Reduced Rank Kalman Filter

2.1.1. Partition the vector space and redefine the cost function

The first step is to partition a state vector increment, $\delta\mathbf{x}$, into a part that is to have flow-dependent error statistics, $\delta\mathbf{x}_\mathrm{s}$, and a residual $\delta\overline{\mathbf{x}}_\mathrm{s}$

$$\delta\mathbf{x} = \delta\mathbf{x}_\mathrm{s} + \delta\overline{\mathbf{x}}_\mathrm{s}. \tag{7}$$

In the RRKF the subspace is defined using the K leading HSVs held in a matrix $\mathbf{S}$ of dimensions $n \times K$, where generally $K \ll n$. It is convienient (though not necessary) to orthogonalise $\mathbf{S}$ to give $\tilde{\mathbf{S}}$. This provides an orthogonal basis to represent the part of the increment $\delta\mathbf{x}$ that lies in the subspace $\delta\mathbf{x}_\mathrm{s}$ as a projection from a vector, $\delta\mathbf{a}$, of K coefficients representing the special subspace

$$\delta\mathbf{x}_\mathrm{s} = \tilde{\mathbf{S}}\delta\mathbf{a}. \tag{8}$$

The orthogonal property of $\tilde{\mathbf{S}}$, $\tilde{\mathbf{S}}^\mathrm{T}\tilde{\mathbf{S}} = \mathbf{I}$, means that $\delta\mathbf{a}$ can be easily determined from $\delta\mathbf{x}$

$$\delta\mathbf{a} = \tilde{\mathbf{S}}^\mathrm{T}\delta\mathbf{x}, \tag{9}$$

noting that $\tilde{\mathbf{S}}^{\mathrm{T}}\delta\overline{\mathbf{x}}_{\mathrm{s}} = \mathbf{0}$.

The next step is to reformulate the background term of the cost function:

$$\mathcal{J}_{\mathrm{b}} = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\mathbf{x}, \tag{10}$$

here the formally correct error covariance matrix $\mathbf{P}^{\mathrm{f}}$ has been used to describe the forecast error covariances in contrast to (3) which uses the $\mathbf{B}$ matrix approximation. Substituting (7) into (10) we find

$$\mathcal{J}^{\mathrm{b}} = \frac{1}{2}\delta\mathbf{x}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\mathbf{x}_{\mathrm{s}} + \frac{1}{2}\delta\overline{\mathbf{x}}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\mathbf{x}_{\mathrm{s}} + \frac{1}{2}\delta\mathbf{x}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\overline{\mathbf{x}}_{\mathrm{s}} + \frac{1}{2}\delta\overline{\mathbf{x}}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\overline{\mathbf{x}}_{\mathrm{s}}. \tag{11}$$

The goal of the RRKF is to treat terms containing the increment $\delta\mathbf{x}_{\mathrm{s}}$ (which lies in the identifed subspace) with explicitly calculated flow dependent error statistics. The term containing only the increment $\delta\overline{\mathbf{x}}_{\mathrm{s}}$ (which lies in the residual space) is to be treated with the implicit and static $\mathbf{B}$-matrix approximation, this is achieved by replacing the forecast error covariance $\mathbf{P}^{\mathrm{f}}$ with $\mathbf{B}$ in the fourth term of (11). The second and third terms of (11) contain both $\delta\mathbf{x}_{\mathrm{s}}$ and $\delta\overline{\mathbf{x}}_{\mathrm{s}}$ which arise due to cross covariances between these spaces. As covariance matrices are symmetric we can write

$$\left(\delta\mathbf{x}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\overline{\mathbf{x}}_{\mathrm{s}}\right)^{\mathrm{T}} = \delta\overline{\mathbf{x}}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}^{-1}}\delta\mathbf{x}_{\mathrm{s}},$$

allowing the second and third terms to be summed to one cross term. This cross term is is treated with the explicitly evolved error covariances this term is anticipated to be important and non-trivial. An additional factor α added by Fisher is typically included in the cross term to ensure the cost function remains positive definite [9] but is omitted here for clarity. Combining these approximations allows the background term of the RRKF cost function to be written as

$$\mathcal{J}^{\mathrm{b}} = \frac{1}{2}\delta\mathbf{x}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}\,-1}\delta\mathbf{x}_{\mathrm{s}} + \delta\overline{\mathbf{x}}_{\mathrm{s}}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}\,-1}\delta\mathbf{x}_{\mathrm{s}} + \frac{1}{2}\delta\overline{\mathbf{x}}_{\mathrm{s}}^{\mathrm{T}}\mathbf{B}^{-1}\delta\overline{\mathbf{x}}_{\mathrm{s}}. \tag{12}$$

2.1.2. Control Variable Transform

The CVT in the RRKF is an adaptation of (4) used in standard variational assimilation, but there is an additional transform $\mathbf{X}$ that accounts for the partioning between $\delta\mathbf{x}_{\mathrm{s}}$ and $\delta\overline{\mathbf{x}}_{\mathrm{s}}$

$$\delta\mathbf{x} = \mathbf{L}\mathbf{X}\delta\boldsymbol{\chi}, \tag{13}$$

where $\delta\boldsymbol{\chi}$ is now the RRKF's control vector. The transform $\mathbf{X}$ is an $n \times n$ orthogonal matrix ($\mathbf{X}\mathbf{X}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{I}$). Additionally $\mathbf{X}^{\mathrm{T}}$ is designed to have the following properties. (i) When $\delta\mathbf{x}$ lies in the subspace spanned by $\delta\mathbf{x}_{\mathrm{s}}$, $\delta\boldsymbol{\chi} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\mathbf{x}$ produces a vector whose first K elements only are able to be non-zero. (ii) When $\delta\mathbf{x}$ lies in the subspace spanned by $\delta\overline{\mathbf{x}}_{\mathrm{s}}$, $\delta\boldsymbol{\chi} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\mathbf{x}$ produces a vector whose last $n-K$ elements only are able to be non-zero. Thus in the $\delta\boldsymbol{\chi}$-representation in (13), the first K (last $n-K$) elements are associated exclusively with $\delta\mathbf{x}_{\mathrm{s}}$ ($\delta\overline{\mathbf{x}}_{\mathrm{s}}$). This property can be achieved with a sequence of Householder transformations [9]. Substituting (13) into (12), noting that $(\mathbf{L}\mathbf{X})(\mathbf{L}\mathbf{X})^{\mathrm{T}} = \mathbf{B}$, and defining

$$\mathbf{P}_{\boldsymbol{\chi}}^{\mathrm{f}\,-1} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{P}^{\mathrm{f}\,-1}\mathbf{L}\mathbf{X}, \tag{14}$$

transforms (12) into the following form in control space

$$\begin{aligned}\mathcal{J}^{\mathrm{b}} \;=\; & \frac{1}{2}\delta\boldsymbol{\chi}_{\mathrm{s}}{}^{\mathrm{T}}\mathbf{P}_{\boldsymbol{\chi}}^{\mathrm{f}\,-1}\delta\boldsymbol{\chi}_{\mathrm{s}} + \delta\overline{\boldsymbol{\chi}}_{\mathrm{s}}{}^{\mathrm{T}}\mathbf{P}_{\boldsymbol{\chi}}^{\mathrm{f}\,-1}\delta\boldsymbol{\chi}_{\mathrm{s}} + \\ & \frac{1}{2}\delta\overline{\boldsymbol{\chi}}_{\mathrm{s}}{}^{\mathrm{T}}\delta\overline{\boldsymbol{\chi}}_{\mathrm{s}},\end{aligned} \tag{15}$$

where $\delta\boldsymbol{\chi}_{\mathrm{s}} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\mathbf{x}_{\mathrm{s}}$ and $\delta\overline{\boldsymbol{\chi}}_{\mathrm{s}} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\delta\overline{\mathbf{x}}_{\mathrm{s}}$.

In (15), and by the design of $\mathbf{X}^{\mathrm{T}}$ specified above, ${\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}}^{-1}$ always acts on vectors that are zero in all but the first K elements. This means that only the first K columns of ${\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}}^{-1}$ are required Fisher [9] shows how this matrix is found given the HSV calculation.

2.2. Ensemble Reduced Rank Kalman Filter (EnRRKF)

Most of the formulation is shared between the EnRRKF and the RRKF. The main difference is the way that the special flow-dependent subspace is identified. In the EnRRKF this subspace is defined by part of the space spanned by an N-member ensemble. Each member is forecast with the dynamic model and is valid at the same time as the background. The ensemble is not restricted to sampling initial condition error, it is feasible to use an ensemble that also describes the random sources of model error. By analogy with the HSVs used with the RRKF, let $\mathbf{S}_{n\times N}$ be the $n \times N$ matrix that holds the ensemble members and let $\tilde{\mathbf{S}}_{n\times N}$ be an orthogonal matrix that spans the same space (some matrices are now labelled with their dimensions to distingish between different representations). Each column of $\mathbf{S}_{n\times N}$ may be represented as a linear combination of the columns of $\tilde{\mathbf{S}}_{n\times N}$ with weights specified by the $N \times N$ matrix $\mathbf{S}_{N\times N}$ (c.f. (8,9)), i.e.

$$\mathbf{S}_{n\times N} = \tilde{\mathbf{S}}_{n\times N}\mathbf{S}_{N\times N} \quad \text{and} \quad \mathbf{S}_{N\times N} = \tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N}\mathbf{S}_{n\times N}. \tag{16}$$

The rank N sample forecast error covariance found from the N members calculated in the $\tilde{\mathbf{S}}_{n\times N}$-space, ${}^{N}\mathbf{P}^{\mathrm{f}}_{N\times N}$, is as follows (c.f. (2))

$${}^{N}\mathbf{P}^{\mathrm{f}}_{N\times N} = \frac{1}{N-1}\mathbf{S}_{N\times N}\mathbf{S}^{\mathrm{T}}_{N\times N}. \tag{17}$$

To minimize the effect of undersampling by the ensemble, a truncated space is constructed comprising the K eigenvectors of $\mathbf{P}^{\mathrm{f}}_{N\times N}$ that have the largest

eigenvalues ($K \leq N$). Let these be the columns of $\mathbf{U}_{N\times K}$. In n-space these eigenvectors are

$$\mathbf{U}_{n\times K} = \tilde{\mathbf{S}}_{n\times N}\mathbf{U}_{N\times K}, \tag{18}$$

which themselves define the special subspace and effectively take the role of the orthogonalised singular vectors in the RRKF. The truncated (to rank K) eigenvalue decomposition of ${}^{N}\mathbf{P}^{\mathrm{f}}_{N\times N}$ is

$${}^{K}\mathbf{P}^{\mathrm{f}}_{N\times N} = \mathbf{U}_{N\times K}\mathbf{\Lambda}_{K\times K}\mathbf{U}^{\mathrm{T}}_{N\times K}, \tag{19}$$

where $\mathbf{\Lambda}_{K\times K}$ is the diagonal matrix of the largest K eigenvalues. The $N-K$ discarded modes are associated with noise and are treated by the $\mathbf{B}$-matrix rather than the $\mathbf{P}^{\mathrm{f}}$-matrix, i.e. they are now in the residual space.

This information from the ensemble is used to construct the first K columns of ${\mathbf{P}^{\mathrm{f}}_{\chi}}^{-1}$ needed in (15). Act with ${\mathbf{P}^{\mathrm{f}}}^{-1}$ on $\mathbf{U}_{n\times K}$, call the result $\mathbf{Z}$, and insert the forward and inverse CVT from (13)

$$\mathbf{Z} = {\mathbf{P}^{\mathrm{f}}}^{-1}\mathbf{U}_{n\times K} = {\mathbf{P}^{\mathrm{f}}}^{-1}\mathbf{L}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{U}_{n\times K}. \tag{20}$$

Acting from the left with $\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}$ and using (14) gives an expression for ${\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}}^{-1}$

$${\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}}^{-1} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{Z}\left(\hat{\mathbf{I}}_{K\times n}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{U}_{n\times K}\right)^{-1}, \tag{21}$$

where $\hat{\mathbf{I}}_{K\times n}$ is a $K\times n$ quasi identity matrix which has the structure $\hat{\mathbf{I}}_{K\times n} = (\mathbf{I}_{K\times K}\mathbf{0}_{K\times(n-K)})$. This removes the last $n-K$ superfluous rows of $\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{U}_{n\times K}$ which contain only zero elements. Since interest is in the first K columns only, ${\mathbf{P}^{\mathrm{f}}_{\chi}}^{-1}$ in (21) (an $n\times K$ matrix), does not appear as a symmetric matrix.

In n-space the rank K error covariance from (19) is

$${}^{K}\mathbf{P}^{\mathrm{f}}_{n\times n} = \tilde{\mathbf{S}}_{n\times N}({}^{K}\mathbf{P}^{\mathrm{f}}_{N\times N})\tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N} = \tilde{\mathbf{S}}_{n\times N}\mathbf{U}_{N\times K}\mathbf{\Lambda}_{K\times K}\mathbf{U}^{\mathrm{T}}_{N\times K}\tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N}. \tag{22}$$

In the following, it is assumed that the operators act only on the space spanned by the special K vectors identified above. This means that the restricted inverse of ${}^{K}\mathbf{P}^{\mathrm{f}}_{n\times n}$ can be found. Inverting (22) by noting that under the same assumption, $\tilde{\mathbf{S}}_{n\times N}\tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N} = \mathbf{I}$ and $\mathbf{U}_{N\times K}\mathbf{U}^{\mathrm{T}}_{N\times K} = \mathbf{I}$, gives

$$ {}^{K}\mathbf{P}^{\mathrm{f}}_{n\times n}{}^{-1} = \tilde{\mathbf{S}}_{n\times N}\mathbf{U}_{N\times K}\mathbf{\Lambda}^{-1}_{K\times K}\mathbf{U}^{\mathrm{T}}_{N\times K}\tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N}. \tag{23} $$

Approximating $\mathbf{P}^{\mathrm{f}-1}$ in (20) by the inverse error covariance matrix in (23) found from the truncated sample allows $\mathbf{Z}$ to be found, which may be substituted into (21) for $\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}{}^{-1}$

$$ \mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}{}^{-1} = \mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\tilde{\mathbf{S}}_{n\times N}\mathbf{U}_{N\times K}\mathbf{\Lambda}^{-1}_{K\times K}\mathbf{U}^{\mathrm{T}}_{N\times K}\tilde{\mathbf{S}}^{\mathrm{T}}_{n\times N}\mathbf{U}_{n\times K}\left(\hat{\mathbf{I}}_{K\times n}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{U}_{n\times K}\right)^{-1}, \tag{24} $$

This expression for $\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}{}^{-1}$ is intentionally non-square, it contains calculable quantities for the typical size of n needed for useful models and has the minimum amount of information needed to evaluate (15).

2.3. The implied forecast error covariance matrix of the EnRRKF

In a standard variational data assimilation scheme the forecast error covariance implied by the CVT is given by (6). In the EnRRKF the implied forecast error covariance in model space could in theory be calculated from

$$ \mathbf{P}^{\mathrm{f}}_{\mathrm{imp}} = \mathbf{L}\mathbf{X}\mathbf{P}^{\mathrm{f}}_{\boldsymbol{\chi}}\mathbf{X}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}. \tag{25} $$

In practice this is not straightforward as it would require a difficult inversion of (24). An alternative approach to determining part of the implied forecast error covariance is to implement the EnRRKF and look at the structure of the analysis increments associated with single observation experiments which reveals chosen columns of $\mathbf{P}^{\mathrm{f}}_{\mathrm{imp}}$.

3. Discussion

The standard RRKF has been tested in operational 4DVAR at the European Centre for Medium Range Weather Forecasts, with the conclusion that the RRKF in its standard form does not have a significant impact on forecast scores [10]. Potential problems with the current RRKF are as follows. (i) Singular vectors are a linear construct and so may not adequately describe the error statistics of systems with a non-linear error growth, e.g. thunderstorms. The limited performance reported may not necessarily be attributed to the RRKF per se, rather just to the use of singular vectors which may not be identifying the most dynamically relevant subspace. (ii) The RRKF's knowledge about the Hessian in the previous cycle may be inadequate, particularly as the Hessian itself is approximate and depends upon the previous cycle's forecast error covariance matrix and so problems with the RRKF may cycle through the system. (iii) Small scale weather systems (thunderstorms or mesoscale convective systems) require high resolution models to capture their structure. To describe the error characteristics using the HSV calculation of these features at high resolution would be computationally demanding, possibly prohibitively so.

The potential benefits of the EnRRKF are as follows. (i) As in the RRKF the flow-dependent error statistics for part of the state space are calculated explicitly, which is done for each assimilation cycle. The remaining part of the state space is treated conventionally, the method also takes into account the covariances between each part. (ii) The description of this subset of error statistics applies even to structures with non-linear error growth as the ensemble trajectories are found with the non-linear forecast model. When

used in a data assimilation environment this will hopefully lead to an analysis state that is improved compared with either 4DVAR or an RRKF and ultimately improved forecast scores. (iii) Ensemble forecast systems are used at a number of operational centres and so ensemble information is readily available. An issue that will require close study in future tests of the EnRRKF is whether a small ensemble will give adequate information about the special K-dimensional subspace. Methods such as the EnKF that use sample covariances do suffer from the problems highlighted in section 1. It will be interesting to see how the EnRRKF can help to overcome these problems given the total effective error covariance matrix in the EnRRKF remains full-rank.

The generation of an ensemble is a crucial component of the EnRRKF. If ensemble information is not available one method of generating initial ensemble perturbations is to use a technique known as error breeding [13, 16]. This breeding method generates growing modes [16], known as bred vectors which have a larger growth rate than perturbations generated by a Monte Carlo technique [13]. Bred vectors may make an ideal set of initial perturbations for the EnRRKF as it is the set of most unstable modes (in a non-linear sense) that we seek to represent explicitly.

We plan to test the EnRRKF in a high-resolution toy model based on simplified equations equations of the atmosphere (Petrie et al., in preparation). This model has been designed to have multiscale behaviour, where large-scale balances (e.g. hydrostatic and geostrophic balance) coexist with unbalanced motion at the small-scales. This will provide an interesting testbed of the EnRRKF and allow us to investigate the covariance spectra for different vari-

ables with varying numbers of ensemble members and degrees of truncation.

4. Summary

The purpose of this paper is to present the methodology of a new hybrid data assimilation scheme. The background error statistics used in variational data assimilation are largely static due to the practial implications of the large dimensions of the $\mathbf{P}^{\mathrm{f}}$-matrix. The static representation is particularly poor in the case of extreme weather events where the uncertainty is likely to be large, fast growing and non-linear. The EnRRKF is an extension to the RRKF and is proposed as a hybrid method of data assimilation that has partially flow-dependent error statistics. A subspace defined using an ensemble (in contrast to the use of singular vectors in the RRKF) is explicitly represented and flow-dependent. It is hoped that this scheme will reduce analysis error and provide useful statistical information on the uncertainty in extreme weather events.

Acknowledgements

The authors would like to thank the Natural Environmental Research Council through the National Centre for Earth Observation and the UK Met. Office for financial support. The useful and considered comments of two anonymous reviewers of this paper are also gratefully acknowledged.

References

[1] Bannister, R.N. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and mea-

surements of forecast error covariances. *Q. J. R. Meteorol. Soc.* 2008; **134**:1951–1970.

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

[3] Barkmeijer, J., Gijzen, M., van Bouttier, F., Singular vectors and estimates of the analysis-error covariance metric. *Q. J. R. Meteorol. Soc.* 1998; **124**:1695–1713.

[4] Beck, A., Ehrendorfer, M. Singular-vector-based covariance propagation in a quasigeostrophic assimilation system. *Mon. Wea. Rev.* 2005; **133**:1295–1310.

[5] Cheng, H., Mohamed, J., Mihai, A., Sandu, A. A hybrid approach to estimating error covariances in variational data assimilation. *Tellus* 2010; **62A**:288–297

[6] Courtier, P., Thepautand, J.-N., Hollingsworth, A. A Strategy for operational implementation of 4D-Var, using an incremental approach *Q. J. R. Meteorol. Soc.* 1994; **120**:1367–1387.

[7] Ehrendorfer, M. A review of issues in ensemble-based Kalman filtering *Meteorol. Z.* 2007 **16**:795–818.

[8] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.* 1994; **99(C5)**:10143–10162.

[9] Fisher, M. Development of a simplified Kalman Filter *European Centre for Medium Range Weather Forecasts Technical Note* 1998; **260**

[10] Fisher, M., Andersson, A. Developments in 4DVAR and Kalman Filtering *European Centre for Medium Range Weather Forecasts Technical Note* 1998; **347**

[11] Isaksen, L. Use of analysis ensembles in estimating flow-dependent background error variances *ECMWF Workshop Proceedings: Flow dependent aspects of Data assimilation* 2007; 65–86

[12] Kalman, R., Bucy, K. New results in linear prediction filtering theory. *Trans. AMSE J. Basic Eng.* 1961 **83D**(ISS):95–108.

[13] Kalnay, E. Atmospheric Modelling: Data assimilation and predictability. *Cambridge University Press, Cambridge* 2003

[14] Le Dimet, F.X., Talagrand, O. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects *Tellus* 1986; **38A**:97–110.

[15] Liu, C., Qingnong, X., Wang, B. An ensemble-based four-dimensional variational data assimialtion scheme. Part 1: Technical formulation and Preliminary test. *Mon. Wea. Rev.* 2008; **136**:3363–3373.

[16] Toth, Z., Kalnay, E. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Meteorl. Soc.* 1993 **74**:2317–2330