Tellus (2004), 56A, 16–28
Printed in UK. All rights reserved

TELLUS

Wavelet approximation of error covariance propagation in data assimilation

By ANDREW TANGBORN*, *Data Assimilation Office NASA-GSFC, Code 910.3, Greenbelt, MD, USA and JCET, University of Maryland-Baltimore County, Baltimore, MD, USA*

(Manuscript received 2 May 2002; in final form 19 May 2003)

ABSTRACT

Estimation of the state of the atmosphere with the Kalman filter remains a distant goal in part because of high computational cost of evolving the error covariance for both linear and non-linear systems (in this case, the extended Kalman filter). Wavelet approximation is presented here as a possible solution that efficiently compresses both global and local covariance information. We demonstrate the compression characteristics by implementing a wavelet approximation scheme on the assimilation of the one-dimensional Burgers' equation. The discrete linearized equations (tangent linear model) and analysis covariance are projected onto a wavelet basis and truncated to just 6% of the coefficients. A nearly optimal forecast is achieved and we show that errors due to truncation of the dynamics are no greater than the errors due to covariance truncation.

1. Introduction

Data Assimilation is in general the estimation of the state of a system, including state variables and model parameters, achieved through the combination of observations and a physical/dynamical model of the system. Most data assimilation systems use Bayesian estimation theory to obtain an optimal estimate of the state. Achieving this optimal estimate requires the minimization of a *cost function* which in turn requires a statistical knowledge of the forecast and observation errors. In schemes where the error statistics are not evolved in time errors are generally assumed homogeneous and isotropic, in spite of the numerous physical sources of inhomogeneity (e.g. fronts, Desroziers and Lafore, 1993). In the Kalman filter, the forecast errors consist of propagated initial errors and errors created by the model (which are in turn propagated forward). Error statistics are also altered by the observation network, which in reality is non-uniform and temporally changing. Localized features like fronts affect the error covariance to the extent that the computational grid can resolve them. Thus the propagated error fields produced by the Kalman filter contain the inhomogeneities which represent both the physical and data driven variations in accuracy. However, propagation of forecast errors is perhaps one of the most computationally expensive component of a data assimilation system. Approximation of the propagation step therefore has become an important area of investigation. It can also be argued that because so little is known about error covariances to begin with, it is meaningless to try to evolve them. However, as methods to estimate error statistics improve, covariance evolution will become more important (Dee, 1995).

Numerous techniques which approximate the forecast errors and their propagation have been proposed and tested. The goal of the approximation is to represent the error covariance with fewer degrees of freedom while retaining the most significant part of the forecast errors. The many approximation schemes differ in what part is retained (and therefore also differ as to which part of the error covariance is assumed to be the most significant). These include those that evolve just the error variance (Dee, 1991), SVD and eigen-decomposition (Cohn and Todling, 1996; Tippet et al., 2000), which retain either the leading singular values or eigenmodes, and flow-dependent covariances (Riishøjgaard, 1998) that assume a correlation between the state forecast and its error covariance (and therefore retain just this correlated part of the covariance). The ensemble Kalman filter (Evensen, 1994; Houtekamer and Mitchell, 1998) evolves error statistics using an ensemble of forecasts rather than applying the dynamical model (or its tangent linear model) to the full error covariance. In this way, a subspace estimate of the error covariance is evolved by the full dynamical system, avoiding any concerns about truncation of dynamics. Error subspace estimation (ESSE) (Lermusiaux and Robinson, 1999), retains the "dominant" eigenmodes statistical properties of an ensemble forecast. Farrell and Ioannou (2001; 2002) adapted the "balanced truncation" method from engineering control theory (Moore, 1981), which retains both leading terms in both the covariance and the dynamics in terms of Hankel singular values.

*e-mail: tangborn@dao.gsfc.nasa.gov

Wavelet representation has been studied as an alternative approach to covariance approximation (Chin and Mariano, 1994; Chin et al., 1999; Tangborn and Zhang, 2000). Chin and Mariano used an orthonormal wavelet transform to approximate the error covariance for the Kalman filtering of linearized shallow water equations. Tangborn and Zhang applied wavelet approximation to the linear conduction diffusion equation. Both of these showed that accurate representation of error covariances for linear dynamics could be achieved using a small fraction of the wavelet coefficients.

In the present work, we apply the wavelet approximation scheme to a more realistic non-linear system, the one-dimensional Burger's equation. We focus in particular on the scale features of error covariances and how they influence the success of the wavelet approximation for a non-linear dynamical system. Evolution of error covariance by a Tangent Linear Model (TLM) in terms of wavelet expansion coefficients, requires the same truncation of the TLM coefficients as for the error covariance coefficients. This could potentially cause a loss of some of the growing coefficients which become important in later stages of the assimilation. Therefore we investigate the impact of both error covariance and propagator truncation.

The paper is organized in the following way: In Section 2 we discuss representation of the covariance dynamics in wavelet space, and its implications for reducing the order of the system. Section 3 describes the extended Kalman filter (EKF) for Burger's equation and Section 4 presents the EKF with error covariances propagated in wavelet space. Results for two different observation networks are given in Section 5 followed by the conclusions in Section 6.

2. Representation of linear error covariance propagation in wavelet space

The accuracy and efficiency of covariance propagation depends on the approximation of the linearized model dynamics (which propagates the covariance) and on the approximation of the error covariance itself. Error covariance propagation in a non-linear system is carried out by the TLM, so we only need consider a linear system here. This will be combined with the non-linear state evolution in the next section. As an example of linear propagation, consider the solution to a linear convection/diffusion type problem by a finite difference method with implicit time-stepping. The system will have the form

$$\mathbf{A}\mathbf{u}^{\mathrm{f}}_{k+1} = \mathbf{B}\mathbf{u}^{\mathrm{f}}_{k}, \tag{1}$$

where $\mathbf{u}^{\mathrm{f}}_{k}$ is the state of the system at time t_k and $\mathbf{A}$ and $\mathbf{B}$ are matrix representations of the discretization (See Appendix B for more details on a typical discretization). $\mathbf{A}$ is generally invertible when boundary conditions are properly imposed, and we can define the propagator as $\mathbf{\Psi} = \mathbf{A}^{-1}\mathbf{B}$ so that the system becomes

$$\mathbf{u}^{\mathrm{f}}_{k+1} = \mathbf{\Psi}\mathbf{u}^{\mathrm{f}}_{k}. \tag{2}$$

In most NWP models, $\mathbf{A}$ has some banded structure, which is taken advantage of when choosing a solution algorithm. The actual computation is based on eq. (1) while eq. (2) is a convenient notation for presenting the propagation portion of a data assimilation system. If we define u^{t} as the true state of the system then the forecast error covariance is defined as $\mathbf{P}^{\mathrm{f}} = \langle(\mathbf{u}^{\mathrm{t}} - \mathbf{u}^{\mathrm{f}})(\mathbf{u}^{\mathrm{t}} - \mathbf{u}^{\mathrm{f}})^T\rangle$ and, assuming the model is perfect, the evolution of $\mathbf{P}^{\mathrm{f}}$ can be calculated by

$$\mathbf{P}^{\mathrm{f}}_{k+1} = \mathbf{\Psi}\mathbf{P}^{\mathrm{a}}_{k}\mathbf{\Psi}^T. \tag{3}$$

The numerical algorithm is actually solving for $\mathbf{P}^{\mathrm{f}}$ from

$$\mathbf{A}\mathbf{P}^{\mathrm{f}}_{k+1}\mathbf{A}^T = \mathbf{B}\mathbf{P}^{\mathrm{a}}_{k}\mathbf{B}^T, \tag{4}$$

so we analyze the wavelet representation of the matrix $\mathbf{A}$.

The particular wavelet transform used in this work comes from the family of compactly supported orthonormal wavelets introduced by Daubechies (1988). Our approach involves starting with the discretized equations from another numerical scheme (e.g. finite differences) and projecting them onto a wavelet basis (Tangborn and Zhang, 2000). We have not considered other wavelets because our focus is on the algorithm itself. As algorithms to approximate error covariance from the continuous equations are developed, issues related to spherical geometry and edge effects will become more important. Basis functions suitable for the two-dimensional sphere have been developed by Gottelmann (1999), for example, and may provide a solution to these problems.

We represent the discrete wavelet transform of a vector $\mathbf{b}$ by the $n \times n$ matrix, $\mathbf{W}$, so that

$$\hat{\mathbf{b}} = \mathbf{W}\mathbf{b}. \tag{5}$$

The matrix $\mathbf{W}$ need not be formed explicitly, and it consists of a sequence of linear operations on the elements of the vector. Details of on discrete wavelet transforms are given in Appendix A. Projection of the analysis error covariance matrix onto a wavelet basis requires the same matrix operations on both rows and columns, which can be written as

$$\hat{\mathbf{P}} = \mathbf{W}\mathbf{P}\mathbf{W}^T. \tag{6}$$

In order to carry out the covariance propagation, we need to project the matrices $\mathbf{A}$ and $\mathbf{B}$ onto the wavelet basis as well:

$$\hat{\mathbf{A}} = \mathbf{W}\mathbf{A}\mathbf{W}^T, \tag{7}$$

$$\hat{\mathbf{A}}^T = \mathbf{W}\mathbf{A}^T\mathbf{W}^T, \tag{8}$$

$$\hat{\mathbf{B}} = \mathbf{W}\mathbf{B}\mathbf{W}^T, \tag{9}$$

$$\hat{\mathbf{B}}^T = \mathbf{W}\mathbf{B}^T\mathbf{W}^T. \tag{10}$$

The structure of coefficient matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{b}}$ depend on the structures of the matrices $\mathbf{A}$ and $\mathbf{B}$. One might be concerned that the

sparse banded structure that occurs in finite difference discretization of convection-diffusion type problems might be destroyed by the wavelet transform. Quite fortunately, the reverse occurs: Most of the entries become very small, so that the matrices can be made more sparse by applying thresholding to the smallest entries. Ground breaking work by Beylkin et al. (1991) showed that diagonally dominant matrices have fewer terms with significant magnitude when projected onto a wavelet basis. This is another sort of compression since fewer terms are needed to represent the operator, thereby reducing the operation count. The meaning of this *near* diagonalization of the propagator is that there is much less interaction between the wavelet coefficients representing the error covariance than there is in physical space. This is a natural effect of the orthogonality and quasi localizability of the basis.

We can illustrate this on a relatively simple system, the finite difference approximation to the two- dimensional univariate convection diffusion equation. The discretization of this equation on an $n \times n$ grid (details are given in Appendix B) results in a sparse pentadiagonal matrix structure ($n^2 \times n^2$) for **A** where the bandwidth is $2n$. Solution of this system generally requires $O(n^2)$ operations. Figure 1 shows a 40×40 pentadiagonal matrix (a) and its projection onto the Daub12 wavelet basis (b). In wavelet space the matrix is no longer pentadiagonal, but rather has a still sparse *hierarchically banded diagonal* form. If we can truncate the very small off diagonal terms (about 1/100th the size of the diagonal terms), we then end up with a diagonal system. Other options include using solvers created specifically for solving systems of this form (Beylkin et al., 1991) which require only $O(n)$ operations. In addition, the covariance and propagator are truncated from $n \times n$ to $L \times L$, so this smaller system can be solved by an $O(L)$ algorithm.

Truncation of the error covariance will retain the greatest information if truncation is carried out first on the diagonal elements (and their associated rows and columns) that have the smallest magnitude. This requires reordering of both the error covariance matrix and the propagator. The selective truncation of wavelet coefficients could introduce loss of potentially important dynamics from the propagator when the largest coefficients of the TLM differ from the largest coefficients for the covariance. Farrell and Ioannou (2001; 2002) dealt with this issue by using a balancing transformation to obtain a representation in coordinates where the stochastic optimals (preferred structures of excitation) and the empirical orthogonal functions (EOFs) coincide. The simple system of Fig. 1 is normal and no special treatment is needed. However, the TLM of the Burgers' equation in the next section is non-normal and dynamical truncation may result in eventual covariance loss. While there is not yet a general method for ensuring that the important parts of the dynamics are retained, from Fig. 1 and the work of Beylkin et al., one can see that diagonally dominant matrices tend to have fewer important elements (i.e. with large magnitude) in wavelet space.

By examining the structures of the error covariance propagation matrices in wavelet space, we have shown that a wavelet basis is an efficient representation of the propagator for an implicit convection/diffusion type solver. Wavelet representation should therefore greatly reduce the computational cost of error covariance propagation in data assimilation without significant loss in accuracy. In the next section we apply the wavelet compact representation method to a simple non-linear system using the extended Kalman filter (EKF).

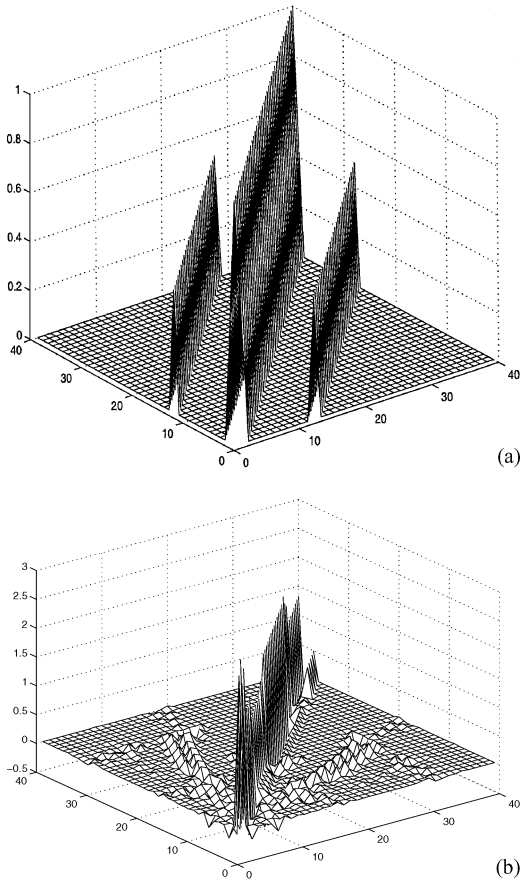

Fig 1. (a) Plot of the 40×40 pentadiagonal matrix that results from the finite difference discretization of the convection/diffusion equations in two-dimensions. Note that the two off-diagonals are so close to the main diagonal that they appear as one. (b) Projection of the above matrix onto Daub12 wavelet basis. The finest scale is represented by half the indices in each direction.

3. The extended Kalman filter for Burgers' equation

We have chosen the one-dimensional Burgers' equation to demonstrate the wavelet approximation scheme for several reasons. Solutions to this non-linear equation tend to contain sharp frontal features (particularly when small viscosity is used). Localized structures of this sort are generally difficult to approximate by some traditional expansions (i.e. spectral, eigendecomposition). For details of solutions to Burgers' equation, see

Benton and Platzman (1972), and for Kalman filtering of Burgers' equation, see Ménard (1994). We wish to compare the success of the assimilation scheme with the full extended Kalman filter. Burgers' equation with a viscosity term, ν, is used as the dynamics of interest, i.e.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{11}$$

with periodic boundary conditions:

$$u(0, t) = u(1, t);$$

and initial conditions:

$$u(x, 0) = f(x).$$

Discretization is carried out using a second order centered differencing scheme in space. Temporal discretization is the implicit Crank–Nicolson scheme for viscous terms and the explicit Adams–Bashforth scheme for advective terms. Details of the discretization are given in Appendix C. The resulting equations can be written in matrix-vector notation as

$$\mathbf{A}\mathbf{u}_{k+1} = \mathbf{B}(\mathbf{u}_\mathbf{k}). \tag{12}$$

The $n \times n$ matrix $\mathbf{A}$ is constant while $\mathbf{B}$ is state dependent because of non-linearity.

The full extended Kalman filter, with analysis carried out every m timesteps, consists of the following steps:

update forecast:

$$\mathbf{u}^{\mathrm{f}}_{k+m} = \mathbf{\Psi}_m\left(\mathbf{u}^{\mathrm{a}}_k\right) \tag{13}$$

$\mathbf{\Psi}_m = \mathbf{\Psi}^m$ represents m applications of the non-linear operator. Update true state (as a substitute for a real physical process to observe)

$$\mathbf{u}^{\mathrm{t}}_{k+m} = \mathbf{\Psi}\left(\cdots\left\{\mathbf{\Psi}\left[\mathbf{\Psi}\left(\mathbf{u}^{\mathrm{t}}_k\right) + \mathbf{w}_k\right] + \mathbf{w}_{k+1}\right\}\cdots\right) + \mathbf{w}_{k+m}, \tag{14}$$

by applying the $\mathbf{\Psi}$ operator m times. with the Gaussian distributed random noise vectors, $\mathbf{w}_k$, $\mathbf{w}_{k+1}$, etc, which act as model error. An observation process is defined by

$$\mathbf{u}_{\mathrm{o}} = \mathbf{H}\mathbf{u}_t + \mathbf{w}_{\mathrm{o}} \tag{15}$$

where where $\mathbf{u}_{\mathrm{o}}$ is the observed value, $\mathbf{H}$ is the observation matrix, and $\mathbf{w}_{\mathrm{o}}$ is the observational error, which is also Gaussian distributed.

Update forecast covariance using the TLM of $\mathbf{\Psi}$

$$\mathbf{P}^{\mathrm{f}}_{k+m} = \mathbf{\Psi}^{\mathrm{TLM}}_m\mathbf{P}^{\mathrm{a}}_k(\mathbf{\Psi}^{\mathrm{TLM}})^T_m + \mathbf{Q}_m. \tag{16}$$

Calculate Kalman gain:

$$\mathbf{K}_{k+m} = \mathbf{P}^{\mathrm{f}}_{k+m}\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^{\mathrm{f}}_{k+m}\mathbf{H}^T + \mathbf{R}_{k+m}\right)^{-1}. \tag{17}$$

Update analysis variable

$$\mathbf{u}^{\mathrm{a}}_{k+m} = \mathbf{u}^{\mathrm{f}}_{k+m} + \mathbf{K}_{k+m}\left(\mathbf{u}^{\mathrm{o}}_{k+m} - \mathbf{H}\mathbf{u}^{\mathrm{f}}_{k+m}\right). \tag{18}$$

Update analysis covariance

$$\mathbf{P}^{\mathrm{a}}_{k+m} = (\mathbf{I} - \mathbf{K}_{k+m}\mathbf{H})\mathbf{P}^{\mathrm{f}}_{k+m}, \tag{19}$$

where $\mathbf{P}^{\mathrm{f}}$ is the forecast covariance matrix, $\mathbf{P}^{\mathrm{a}}$ is the analysis covariance matrix, $\mathbf{Q}$ is the model error covariance and $\mathbf{R}$ is the observational error covariance. $\mathbf{K}$ is the Kalman gain matrix.

The operation count for the covariance propagation step (16) is $O(n^2)$ (where n is the number of grid points). If the number of observations at each analysis time is significantly less than n, then equation (16) is the most computationally intensive step in the assimilation system. The next section outlines a scheme for approximating the error covariance propagation by a truncated wavelet expansion.

4. A Wavelet approximation to the EKF

The wavelet approximation scheme involves projecting both the tangent linear propagator $\mathbf{\Psi}^{\mathrm{TLM}}$ and the analysis covariance $\mathbf{P}^{\mathrm{a}}$ onto the wavelet basis and retaining L wavelet coefficients, where $L \ll n$. The error covariance propagation is then carried out in the small space of the wavelet coefficients. We repeat the Kalman filter algorithm here with the wavelet space approximation of the error covariance and propagator. The state update equation for $\mathbf{u}^{\mathrm{f}}$ and $\mathbf{u}^{\mathrm{t}}$ are carried out in physical space as

$$\mathbf{u}^{\mathrm{f}}_{k+m} = \mathbf{\Psi}_m\mathbf{u}^{\mathrm{a}}_k \tag{20}$$

$$\mathbf{u}^{\mathrm{t}}_{k+m} = \mathbf{\Psi}_m\left\{\cdots\mathbf{\Psi}_m\left[\mathbf{\Psi}_m\left(\mathbf{u}^{\mathrm{t}}_k\right) + \mathbf{w}_k\right]\cdots\right\} + \mathbf{w}_{k+m} \tag{21}$$

The projection of the analysis error covariance onto a wavelet basis follows from eq. (6):

$$\hat{\mathbf{P}}^{\mathrm{a}}_k = \mathbf{W}^T\mathbf{P}^{\mathrm{a}}_k\mathbf{W} \tag{22}$$

and the projection of the error covariance propagator is:

$$\hat{\mathbf{\Psi}}^{\mathrm{TLM}} = \mathbf{W}^T\mathbf{\Psi}^{\mathrm{TLM}}\mathbf{W}. \tag{23}$$

The important question here is how to decide which part of $\hat{\mathbf{\Psi}}$ and $\hat{\mathbf{P}}^{\mathrm{a}}_k$ to retain. Simply using the first L terms is equivalent to removing all small-scales represented by the remaining coefficients. In the case of Burgers' equation, small-scale features are particularly significant as viscosity decreases. In a steep frontal system, forecast errors tend to be large and very localized because any velocity (or dispersion) errors are amplified. It is also important that the resulting error forecast covariance be positive definite. Any truncation of the wavelet space representation of $\hat{\mathbf{P}}^{\mathrm{a}}$ is carried out on entire rows and columns, rather than on individual matrix entries.

The matrix equation (16) can be projected onto the wavelet basis following eqs. (4)–(10), and therefore can be written as

$$\hat{\mathbf{P}}^{\mathrm{f}}_{k+m} = (\hat{\mathbf{\Psi}}_m)\left(\hat{\mathbf{P}}^{\mathrm{a}}_k\right)\left(\hat{\mathbf{\Psi}}^T_m\right) + (\hat{\mathbf{Q}}_m). \tag{24}$$

Table 1. Operation count

Ref.	Equation	Flops
20	$\mathbf{u}^{\mathrm{f}}_{k+m} = \mathbf{\Psi}_m \mathbf{u}^{\mathrm{a}}_k$	$O(n)$
22	$\hat{\mathbf{P}}^{\mathrm{a}}_k = \mathbf{W}^T \mathbf{P}^{\mathrm{a}}_k \mathbf{W}$	$O(n^2)$
23	$\hat{\Psi}^{\mathrm{TLM}} = \mathbf{W}^T \mathbf{\Psi}^{\mathrm{TLM}} \mathbf{W}.$	$O(n^2)$
25	$(\hat{\mathbf{P}}^{\mathrm{f}}_{k+m})_L = (\hat{\Psi}_m)_L (\hat{\mathbf{P}}^{\mathrm{a}}_k)_L (\hat{\Psi}^T_m)_L + (\hat{\mathbf{Q}}_m)_L.$	$O(L^2)$
26	$\mathbf{P}^{\mathrm{f}}_{k+m} = \mathbf{W} \hat{\mathbf{P}}^{\mathrm{f}}_{k+m} \mathbf{W}^T$	$O(n^2)$
27	$\mathbf{K}_{k+m} = \mathbf{P}^{\mathrm{f}}_{k+m} \mathbf{H}^T (\mathbf{H} \mathbf{P}^{\mathrm{f}}_{k+m} \mathbf{H}^T + \mathbf{R}_{k+m})^{-1}$	$O(nM)$
28	$\mathbf{u}^{\mathrm{a}}_{k+m} = \mathbf{u}^{\mathrm{f}}_{k+m} + \mathbf{K}_{k+m} (\mathbf{u}^{\mathrm{o}}_{k+m} - \mathbf{H}\, \mathbf{u}^{\mathrm{a}}_{k+m})$	$O(nM)$
29	$\mathbf{P}^{\mathrm{a}}{}_{k+m} = (\mathbf{I} - \mathbf{K}_{k+m} \mathbf{H}) \mathbf{P}^{\mathrm{f}}{}_{k+m}$	$O(nM^2)$

In order to retain the most important components of the covariance, the rows and columns of $\hat{\mathbf{P}}^{\mathrm{a}}$ are reordered using the magnitude of the main diagonal as the ordering criterion. Note that in wavelet space, this diagonal term contains information on both the variance and correlation fields. Simultaneously, we must reorder the covariance propagator $\hat{\Psi}^{\mathrm{TLM}}$ and its transpose as well.

After reordering, the covariance propagation equations are truncated to L terms and covariance propagation is carried out in wavelet space:

$$\left(\hat{\mathbf{P}}^{\mathrm{f}}_{k+m}\right)_L = (\hat{\Psi}_m)_L \left(\hat{\mathbf{P}}^{\mathrm{a}}_k\right)_L \left(\hat{\Psi}^T_m\right)_L + (\hat{\mathbf{Q}}_m)_L. \tag{25}$$

The forecast error covariance is transformed back to physical space:

$$\mathbf{P}^{\mathrm{f}}_{k+m} = \mathbf{W} \hat{\mathbf{P}}^{\mathrm{f}}_{k+m} \mathbf{W}^T, \tag{26}$$

and the Kalman gain is computed in physical space:

$$\mathbf{K}_{k+m} = \mathbf{P}^{\mathrm{f}}_{k+m} \mathbf{H}^T \left(\mathbf{H} \mathbf{P}^{\mathrm{f}}_{k+m} \mathbf{H}^T + \mathbf{R}_{k+m}\right)^{-1}. \tag{27}$$

The analysis is updated as in eq. (18):

$$\mathbf{u}^{\mathrm{a}}_{k+m} = \mathbf{u}^{\mathrm{f}}_{k+m} + \mathbf{K}_{k+m} \left(\mathbf{u}^{\mathrm{o}}_{k+m} - \mathbf{H} \mathbf{u}^{\mathrm{f}}_{k+m}\right), \tag{28}$$

and the analysis covariance is also updated in physical space:

$$\mathbf{P}^{\mathrm{a}}_{k+m} = (\mathbf{I} - \mathbf{K}_{k+m} \mathbf{H}) \mathbf{P}^{\mathrm{f}}_{k+m}. \tag{29}$$

The operation count of this scheme is summarized in Table 1, where the significant savings are shown to be in the computation of the forecast error covariance [eq. (25)], which is reduced from n^2 to L^2 for the one-dimensional Burgers' equation. The added expenses of transforming error covariance to and from wavelet space is required only at analysis times. The reduction in computational cost will depend on the number of timesteps between analysis times, k_{anal}, and the change in computational expense would then be

$$O\left[k_{\mathrm{anal}} \times (n^2 - L^2) - n^2\right]$$

per analysis time. If $L \ll n$ then the saving is

$$O\left[(k_{\mathrm{anal}} - 1) \times n^2\right]$$

5. Assimilation experiments

We have carried out assimilation experiments for each of four wavelet expansion sizes, $L = 4$, 8, 16 and 128 coefficients. For each case, 15 twin experiments are run in order to obtain meaningful statistics of the assimilations. The total number of grid points is $n = 128$ so that the full covariance evolution is included in the experiments. Each assimilation differs in initial condition and model error perturbations, but all have the same error statistics. We assume that both the initial, model error and observation statistics are known perfectly.

The full (untruncated) and approximate EKF are compared for Burgers' equation on the domain

$$0 \leq x \leq 1$$

using initial conditions on $\mathbf{u}$

$$u_{\mathrm{o}}(x) = \sin(2\pi x) \quad x \leq 0.1 \tag{30}$$

$$u_{\mathrm{o}}(x) = 0, 0.1 \leq x \leq 1.0 \tag{31}$$

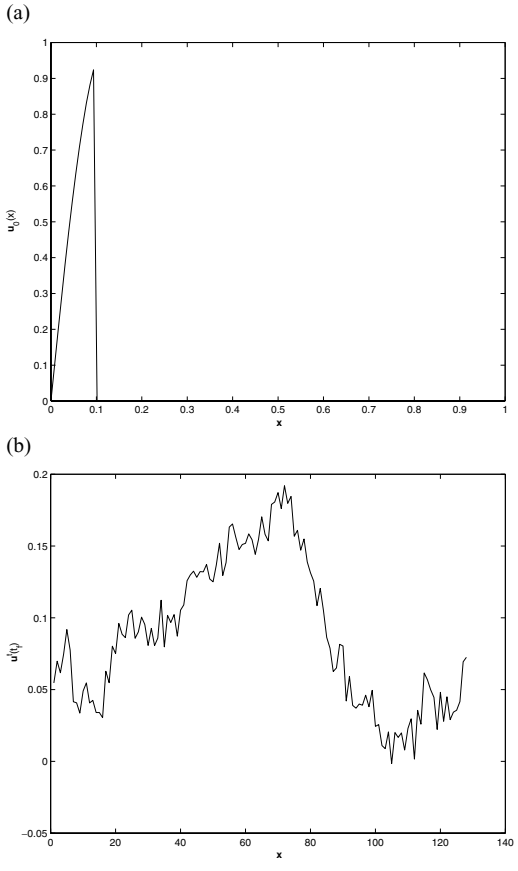

Fig 2. (a) Initial value of $u(x, t)$, (b) final value of $u^{\mathrm{t}}(x, t)$. The peak has moved to the right and the stochastic component is the result of the model error term added to eq. (21).

and run parameters: $\delta t = 0.01$, timestep for the discretization; $m = 40$, number of timesteps between analyses; and $\nu = 0.005$, fluid kinematic viscosity. The model error and observational error covariances have correlation lengths of $L_c = 0.02$ and variance of 0.0001. The simulations ran for 360 timesteps with observations taken every 40 timesteps. In the experiments, we separate our analyses with respect to error covariance and propagator truncation.

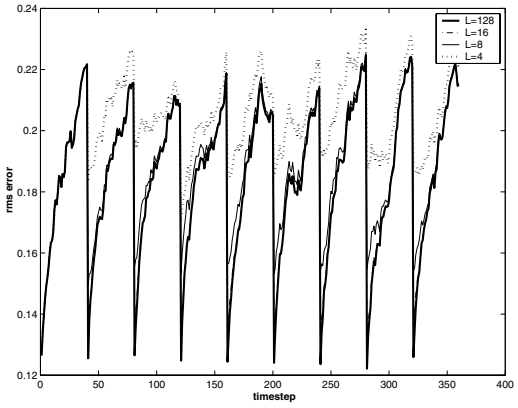

Fig 3. Ensemble mean RMS error for the four assimilation experiments of the one-dimensional Burgers' equation using the uniform 42 observation network (UO). The wavelet expansions have sizes $L = 128$ (thick solid line), $L = 16$ (dash–dot), $L = 8$ (thin solid) and $L = 4$ (dotted). The assimilations that result from $L = 8$, 16 approximations show only a slight increase in RMS error, while the $L = 4$ case shows a more noticeable increase.

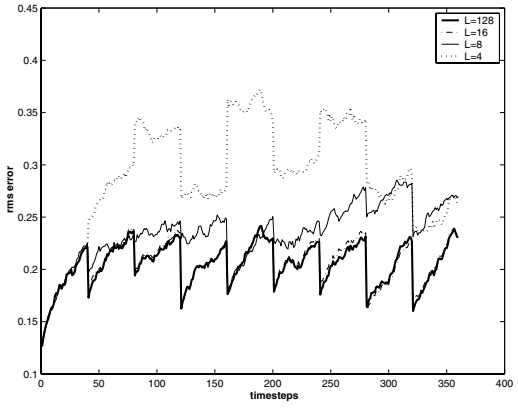

Fig 4. Ensemble mean RMS error for the different four experiments of the one-dimensional Burgers' equation using the non-uniform 80 observation network (NUO). The wavelet expansions have sizes $L = 128$ (thick solid line), $L = 16$ (dash–dot), $L = 8$ (thin solid), and $L = 4$ (dotted) for each set of 15 twin experiments. The 16 coefficient expansion experiment follows the full system closely for the entire time period, while the eight coefficient experiments show significantly larger errors. The $L = 4$ system shows filter divergence due to errors in the correlation field.

5.1. *Effect of truncated error covariance*

The important components of error covariance that need to be maintained are the error variance and the correlation length. The error variance determines the relative weight given to the forecast and observations, while the correlation length determines the spread of the impact of the observations. The effect that the accuracy of these two components has on the assimilation will depend on the nature of the observation network. By examining whether the analysis increment, $u_a - u_f$, is spreading information from the observations into data void regions where their impact

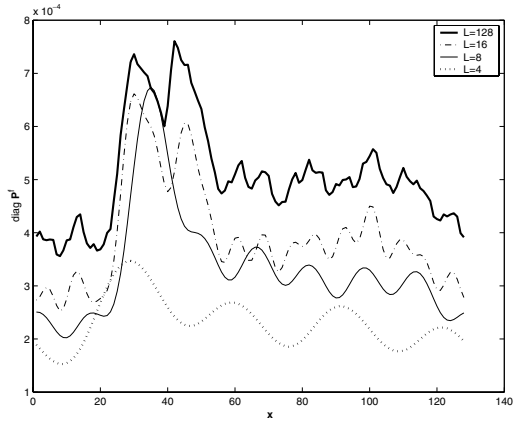

Fig 5. Forecast error variance obtained from the diagonal of $\mathbf{P}$f for the uniformly distributed 42 observation experiment just before the second analysis time. The expansions are $L = 128$ (thick solid line), $L = 16$ (dash–dot), $L = 8$ (thin solid) and $L = 4$ (dotted). The error variances are almost identical to the NUO experiments at this point in the assimilation.

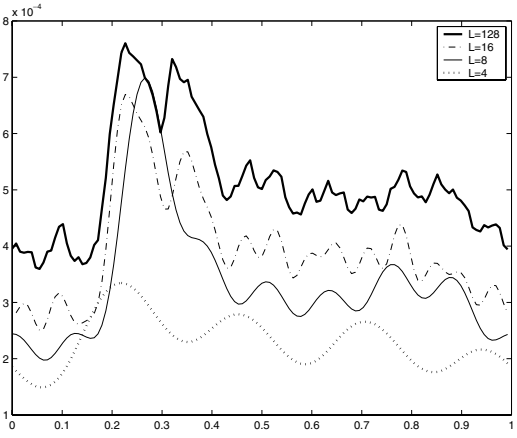

Fig 6. Forecast error variance obtained from the diagonal of $\mathbf{P}$f for the non-uniformly distributed 80 observation experiment (NUO) just before the second analysis time. The expansion sizes are $L = 128$ (thick solid line), $L = 16$ (dash–dot), $L = 8$ (thin solid) and $L = 4$ (dotted). All of the approximations lose some variance information, but only the $L = 4$ case does not resolve the variance peak occurring around $x = 0.3$.

may be greater, one can determine the effect of truncation on error variance and particularly on correlation length.

In the first set of experiments the observations are uniformly spaced (henceforth called the UO experiment) and are made at every third grid point ($n_{\text{obs}} = 42$). The second set of experiments the observations are non-uniformly spaced (called UO experiments), with observations at every grid point for $0.38 \leq x \leq 1$ and none elsewhere ($n_{\text{obs}} = 80$).

The initial $u_{\text{o}}(x)$ field used in both sets of experiments, from eqs. (30) and (31), is shown in Fig. 2(a). Advection will cause the spike in $u(x)$ to move to the right, while viscosity will result in a decay in the peak. The non-linearity of Burgers' equation maintains a steep front on the right side of the peak. Figure 2(b) shows a single realization of the true state $\mathbf{u}^{\text{t}}$ at a later time. The initial spike has shifted to the right and decreased in magnitude, while the stochastic component is due to the addition of the random noise vector in eq. (21).

The ensemble mean of the RMS errors ($u - u^t$) for the UO experiments using each of the wavelet expansion sizes is shown in Fig. 3. experiments. Reductions of 87% ($L = 16$) and 94% ($L = 8$) in the number of coefficients do not increase the RMS errors, and even when $L = 4$ (97% reduction), the RMS error increases only about 5%.

Figure 4 shows the ensemble mean of the RMS errors for the NUO experiments. Here the errors in $\mathbf{u}$ due to approximations in the error covariance field are substantially larger. While the $L = 16$ case has nearly identical RMS error to the full covariance evolution case, further reduction to $L = 8$ increases the error by about 50% and the $L = 4$ case results in filter divergence.

In order to understand how approximating the error covariance by truncating its wavelet expansion [eq. (25)] affects the analysis errors of Figs. 3 and 4, we need to consider representation errors in the error covariance itself. Figures 5 and 6 show the diagonal of the forecast covariance matrix, $\mathbf{P}^{\text{f}}$, (a prediction of the forecast error variance) after being propagated for one analysis interval for the UO and NUO experiments, respectively. The two experiments have remarkably similar error variances, and therefore this can not explain the differences in the forecasts. In both cases the approximate error variances have some loss that

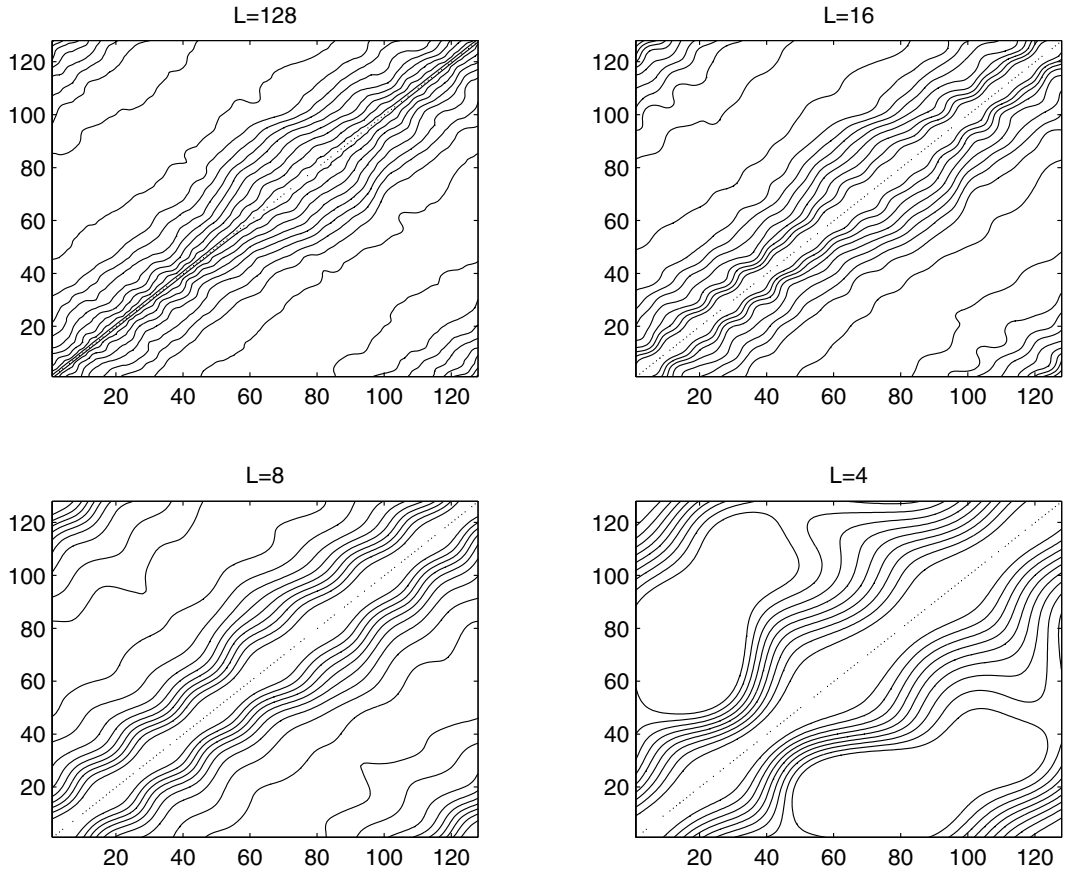

Fig 7. Forecast error correlations extracted from $\mathbf{P}^{\text{f}}$ for the UO (uniform observing network) experiments just before the second analysis time, for $L = 128$, 16, 8 and 4.

increase as more coefficients are truncated. More significantly, both the $L = 8$ and $L = 16$ cases mostly retain the peak in error variance around $x = 0.25$ that is associated with the frontal structure in u. The accurate approximation of the variance spike in the observation void region appears to be one important component of the covariance approximation.

We consider error correlations to help explain the difference in the relative success of the two observation networks. The forecast error correlation, $\mathbf{C}^{\mathrm{f}}$ is extracted from the covariance fields using

$$C^{\mathrm{f}}_{i,j} = \frac{P^{\mathrm{f}}_{i,j}}{\sigma_i \sigma_j} \tag{32}$$

where $C^{\mathrm{f}}_{i,j}$ is the (i, j) entry in $\mathbf{C}^{\mathrm{f}}$ and σ_i, σ_j are the forecast error standard deviations at the ith and jth grid points. Figures 7 and 8 are the forecast error correlations just after propagation, for NUO and UO experiments, with coefficient truncations of $L = 128$, $L = 16$, $L = 8$ and $L = 4$. Both experiments show an increase in correlation length as L is reduced. This is the result of the elimination of some of the smaller-scale wavelets which are required to represent the short correlation length structure. Because the correlations in Figs. 7 and 8 are so similar, they do not explain the differences between the analyses of the UO and NUO experiments. To understand this we need to examine how the increase in correlation length affects the Kalman gain in eq. (27). The Kalman gain represents both the relative weight of the forecast and observations and the spreading of the increments due to the observations. When the correlation length scale is overestimated, as it is in the $L = 4$ truncation for both experiments, the effect of observations will be spread too far. In the UO experiments, the impact of incorrect correlation lengths does not cause filter divergence because there are nearby observations which act to constrain any errors. In the NUO experiments, the large data void region contains significant analysis increments from distant observations because the correlations are too large. Figure 9 shows the difference between the Kalman gain for full error covariance propagation and the Kalman gain at different els of truncation. In the NUO experiments, (a–c), there is a significant difference between the Kalman gains for the full and truncated systems. As L is reduced from 16 to 4, the difference can be seen to stretch across the entire data void (where off-diagonal values are non-zero). From the UO experiments we have plotted the difference between the full system Kalman gain and the $L = 4$

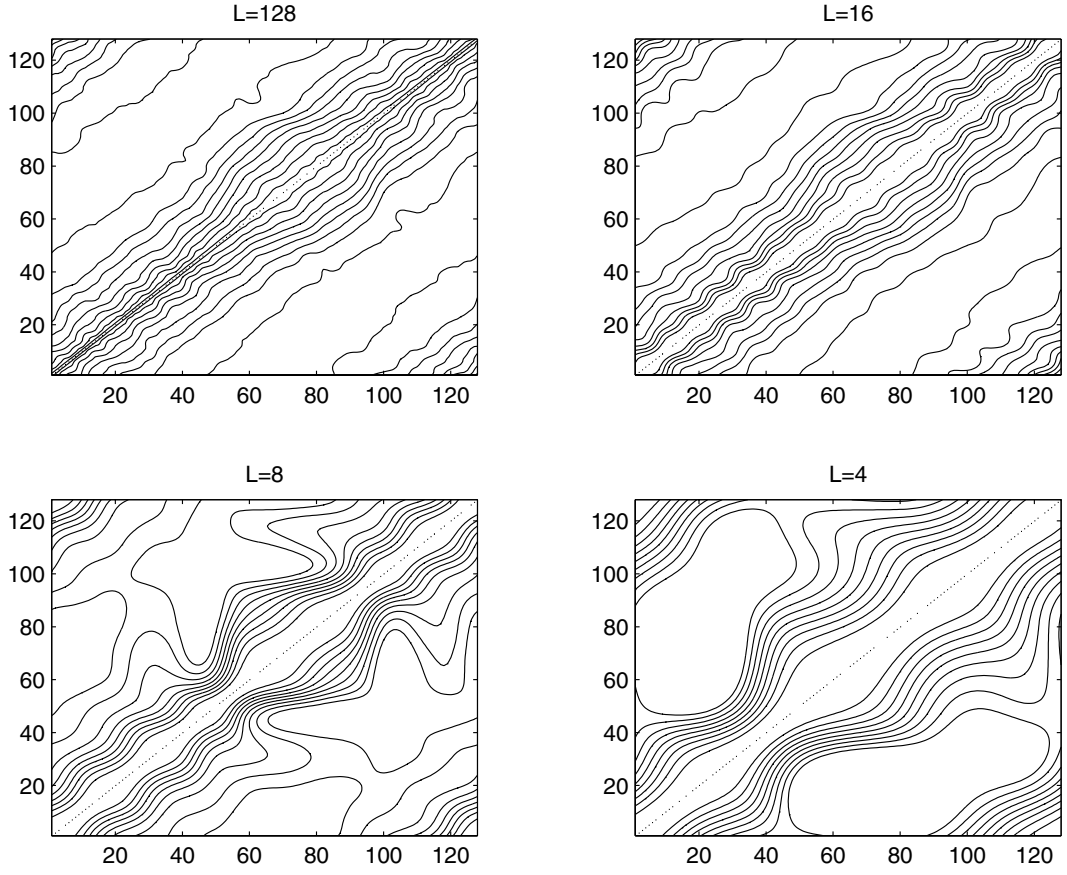

Fig 8. Forecast error correlations extracted from $\mathbf{P}^{\mathrm{f}}$ for the NUO (non-uniform observing network) experiments, just before the second analysis time, for $L = 128$, 16, 8 and 4.

case (d), and see there is no significant influence from distant observations. Thus, severe truncation of the wavelet expansion is found to affect the assimilation when there are large regions without observations. With a uniformly distributed observation network, the RMS errors are significantly lower, even though only about half the number of observations are used.

Another measure of the information contained in the wavelet expansions is the energy retained in the approximated covariance matrices. We define the fraction of "energy" in a reconstruction as

$$\frac{E_L}{E_{\text{total}}} = \left(\frac{\sum_{k=1}^{L} (\hat{u}_k)^2}{\sum_{k=1}^{n} (\hat{u}_k)^2} \right)^{1/2}. \tag{33}$$

This comparison is made after one propagation interval and just before the first analysis time, so that the observation network has no impact and it is valid for either observation network. The 4-, 8-and 16-term expansions represent 3, 6 and 12% of the coefficients in one dimension (or about 0.1, 0.4 and 1.6% if extended to two dimensions). The energy retained in each case is found to be [by eq. (33)] 54, 72 and 81%, respectively. Therefore, in our one-dimensional Burgers' equation assimilation, we need to retain about 74% of the energy by using 6% of the coefficients in order to achieve assimilation results close to the full system for the uniform observation network, but need 81% of the energy using 12% of the coefficients for the non-uniform observation network.

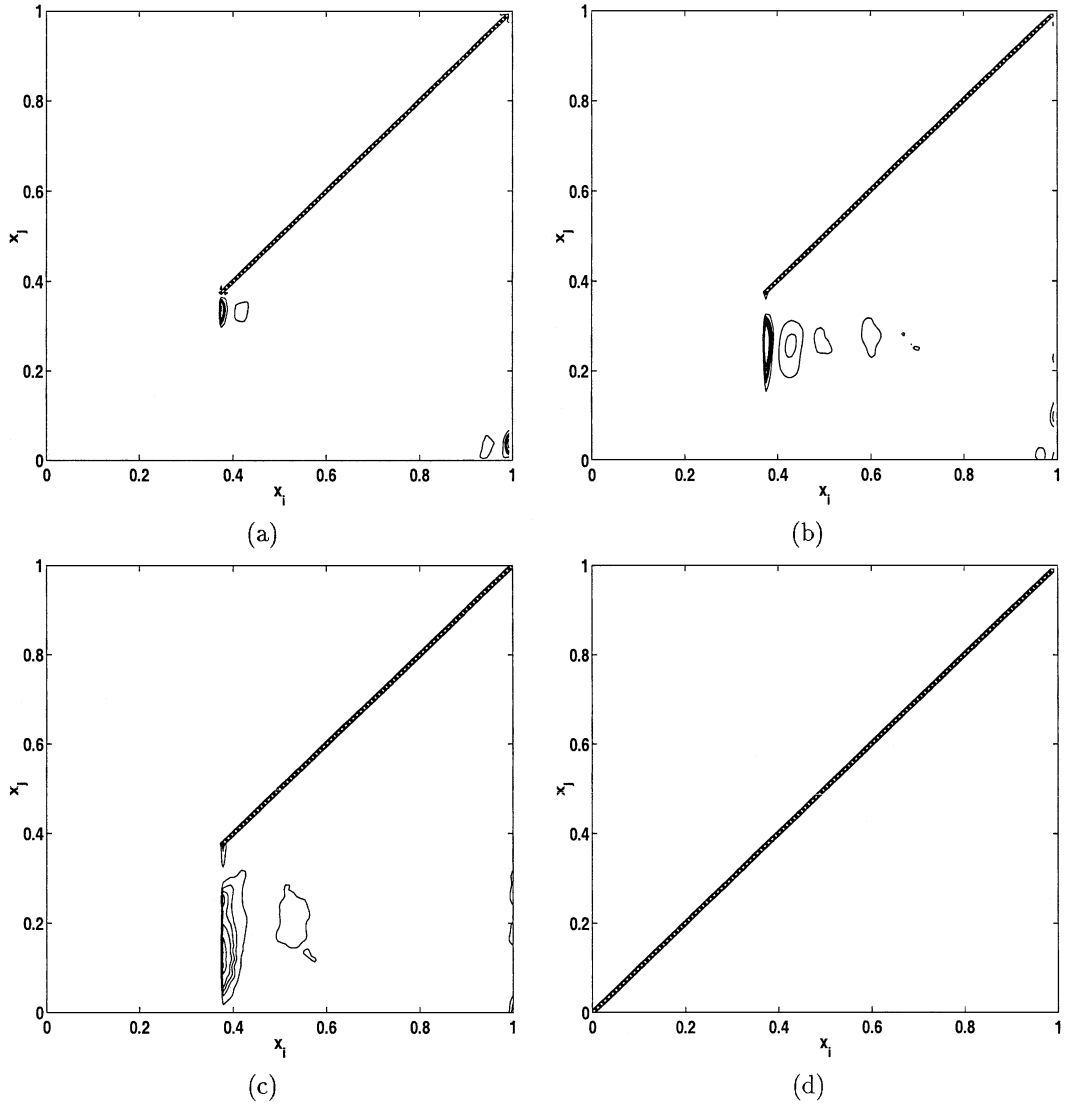

Fig 9. contour plots of the difference between the Kalman gain for the full and approximated error covariances ($K_128 - K_L$) where $L = 16$ (a), $L = 8$ (b) and $L = 4$ (c) using the non-uniform observing network (NUO). The uniform observing network (UO) is used in (d) for the case $L = 4$.

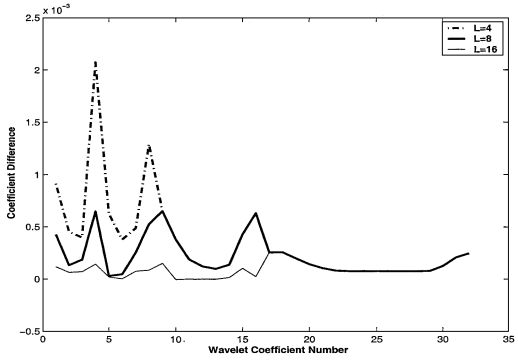

Fig 10. Difference between between full and partial covariance matrix diagonals in wavelet wavelet space. The Diagonal terms of $(\hat{P}^{\mathrm{f}})_{128} - (\hat{P}^{\mathrm{f}})_{16}$ is represented by the thin solid line, while the thick solid line is $(\hat{P}^{\mathrm{f}})_{128} - (\hat{P}^{\mathrm{f}})_{8}$, and the dash–dot line is $(\hat{P}^{\mathrm{f}})_{128} - (\hat{P}^{\mathrm{f}})_{4}$.

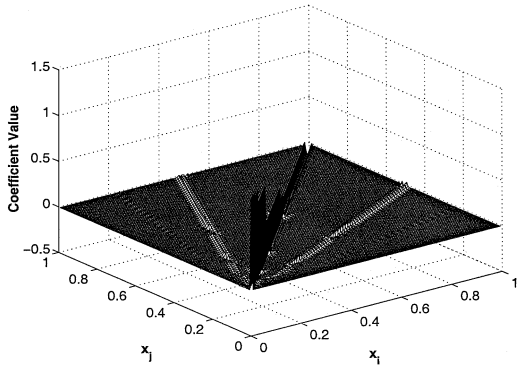

Fig 11. Projection of discretized tangent linear model onto wavelet basis. The wavelet representation of the dynamics is nearly diagonal and shows a exponential decay in diagonal elements with increasing scale j.

5.2. *Effect of truncated propagator dynamics*

We next consider whether the loss of accuracy is due to the truncation of the covariance itself or to the truncation of the propagator. Figure 10 shows the difference between the full and truncated wavelet representation along the main diagonal $(\hat{\mathbf{P}}^{\mathrm{f}} - \hat{\mathbf{P}}^{\mathrm{f}}_{L})$ for the cases $L = 8$, 16. For $L = 16$, the difference in the diagonal are very small up to (and including) the 16th and then become large. This jump at the 17th coefficient is due entirely to the truncation of the covariance, and the propagator truncation has no impact. For the $L = 8$ case, we see that the differences in the first eight coefficients are on the same order as the larger coefficients, indicating that propagation errors have become important. When the expansion is truncated at four coefficients, the propagator truncation errors become more important than covariance truncation errors, since the 4th coefficient difference is the largest. Thus we see that it is only when the system becomes most severely truncated that the errors incurred by truncating the dynamics $\mathbf{\Psi}$ dominate. This confirms the qualitative argument made about the reduction of significant modes in the projection of Fig. 1(b).

Finally, we return to the question of the impact of truncation of the TLM on the dynamics. While the TLM used here for Burgers' equation is non-normal, the coefficient ordering has been such that, except for the first eight terms, the coefficient magnitudes decrease monotonically. Thus, as long as the first few coarsest scales are retained (up to $j = 3$ in this case), the remaining decay exponentially. In this TLM, wavelet representation results in a nearly diagonal matrix with entries that decay rapidly towards zero as the coefficient index increases, as shown in Fig. 11. Thus truncation of finer scales past some cutoff will not result in a loss of the most significant dynamics in this particular case.

6. Conclusions

We have developed a reduced rank extended Kalman filter which uses truncated wavelet expansions to represent error covariances and covariance dynamics. For Burgers' equation assimilation, we demonstrate that the representation is accurate enough to obtain nearly optimal assimilation results using an extended Kalman filter when about 6% of the coefficients and 74% of the energy are retained with a uniform observation network. The assimilation results in sharply higher RMS errors when the number of retained coefficients is dropped below this level (3% of coefficients). At this point the forecast error covariance matrix does not resolve the spike in variance occurring at the front and results in a correlation length that is too long. The effect of errors in the correlation length scale become more pronounced when the observation network is non-uniform. In this case, when the correlation length becomes to long, analysis increments are incorrectly spread into data void regions, generating larger errors. Significant loss of the dynamics is seen for the most severe truncation ($L = 4$). The important conclusion from these results is that these important covariance characteristics can be attained with just 6% of the coefficients resulting in a near optimal assimilation. Better results can be achieved when the observation network is uniform.

The extension of these techniques to global three-dimensional models will naturally involve substantial further work. For example, the task of storing the entire error covariance field becomes unwieldy at these scales. Existing schemes that store error covariances in spectral (Parish and Derber, 1992) and physical (Cohn et al., 1998) space calculate the covariances only when and where they are needed. This approach can easily be adapted to wavelet representation because of their inherent localization in space. Thus the wavelet transform need not be carried out everywhere at one time, but in spatially localized pieces that will access only a fraction of the stored coefficients at a given time. This contrasts

sharply with existing spectral methods in which every coefficient is needed to calculate the error covariance in some sub-region.

7. Acknowledgments

I thank Steve Cohn for his encouragement during the course of this work, and Ricardo Todling and Richard Menard for their comments on an early version of the manuscript. The anonymous reviewers a careful reading of the paper have helped to improve it greatly.

8. Appendices

A. *Wavelet transforms*

Wavelets are sets of basis functions with local support. The Daubechies family of wavelets are generated from scaling functions which satisfy the recursion relationship

$$\phi(x) = \sum_k c_k \phi(2x - k), \tag{A.1}$$

where the choice of filter coefficients c_k determines the properties of the resulting scaling function and wavelet. A small number of non-zero c_k will give rise to more localized wavelets and a larger number of c_k will produce smoother (i.e. continuous higher derivatives) wavelets. The basic Wavelet function, ψ, for a given set of filter coefficients can be derived from scaling functions by taking differences

$$\psi(x) = \sum_k (-1)^k c_{1-k} \phi(2x - k). \tag{A.2}$$

The representations of different scales and locations are then obtained by dilating and translating $\psi(x)$

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \tag{A.3}$$

where j refers to dilation (scale) and k refers to translation (location).

The pyramidal wavelet transform matrix (Mallat, 1989) is constructed from the filter coefficients which acts on a vector as a series of "averages" and "differences" differences at increasingly coarser scales. For example, given a vector $\mathbf{b}$ of length $n = 2^J$, the first operation in the decomposition of $\mathbf{b}$ into its multi-resolution representation would be to calculate the "averages" at the next coarser scale $(J - 1)$, which is followed by a series of differences that determine the wavelet coefficients, $\hat{\mathbf{b}}^{J-1,k}$. This process is continued through increasingly coarse scales, $J - 2$, $J - 3$, $\ldots, j, \ldots, 1, 0$, until all of the coefficients are determined. The operation count for this algorithm is $o(n)$.

We represent this discrete wavelet transform by the matrix $\mathbf{W}$ (which is not actually constructed), so that the wavelet transform of a one-dimensional (discretized) data set, $\mathbf{b}$, is

$$\hat{\mathbf{b}} = \mathbf{W}\mathbf{b} \tag{A.4}$$

where $\hat{\mathbf{b}}$ is the vector of wavelet coefficients of $\mathbf{b}$. The the gridded values in one dimension are then represented by the expansion

$$b(x_i) = \sum_j \sum_k \hat{b}_{j,k} \psi_{j,k}(x_i). \tag{A.5}$$

For a two-dimensional data set, $\mathbf{P}$ for example, the wavelet transform must be applied to both the rows and columns, so we have

$$\hat{\mathbf{P}} = \mathbf{W}\mathbf{P}\mathbf{W}^T \tag{A.6}$$

where $\hat{\mathbf{P}}$ is the matrix of wavelet coefficients of $\mathbf{P}$. Then the gridded values in two-dimensions are represented by the tensor product basis

$$P(x_i, y_{i'}) = \sum_j \sum_k \sum_{j'} \sum_{k'} \hat{P}_{i,j,i',k'} \psi_{j,k}(x_i) \psi_{j',k'}(y_{i'}). \tag{A.7}$$

Interested readers are directed to either short (Press et al., 1992) or longer (Strang and Nguyen, 1996) introductions to wavelet transforms.

B. *Discretization of the convection–diffusion equation*

An example of discretization of the two-dimensional convection–diffusion is centered differencing in space and implicit Euler-backward in time. The linear differential equation is

$$\frac{\partial C}{\partial t} + U(x, y)\frac{\partial C}{\partial x} + V(x, y)\frac{\partial C}{\partial y} = \alpha \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial^2 y^2} \right), \tag{B.1}$$

where $C(x, y)$ is a two-dimensional conserved scaler field, $U(x, y)$ and $V(x, y)$ are known velocity components and α is a diffusivity coefficient. The second order spatial approximations are

$$\left(\frac{\partial C}{\partial x} \right)^{(i,j)} = \frac{C^{(i+1,j)} - C^{(i-1,j)}}{2\Delta x} + o(\Delta x^2), \tag{B.2}$$

$$\left(\frac{\partial C}{\partial y} \right)^{(i,j)} = \frac{C^{(i,j+1)} - C^{(i,j-1)}}{2\Delta y} + o(\Delta y^2), \tag{B.3}$$

$$\left(\frac{\partial^2 C}{\partial x^2} \right)^{(i,j)} = \frac{C^{(i+1,j)} - 2C^{(i,j)} + C^{(i-1,j)}}{2\Delta x^2} + o(\Delta x^2), \tag{B.4}$$

$$\left(\frac{\partial^2 C}{\partial y^2} \right)^{(i,j)} = \frac{C^{(i,j+1)} - 2C^{(i,j)} + C^{(i,j-1)}}{2\Delta y^2} + o(\Delta y^2), \tag{B.5}$$

where Δx and Δy are the x and y direction grid spacing, (i, j) are their node numbers on a uniform grid. The first order implicit

time stepping is

$$\left(\frac{\partial C}{\partial t}\right)_{k+1}^{(i,j)} = \frac{C_{k+1}^{(i,j)} - C_k^{(i,j)}}{\Delta t} + o(\Delta t), \tag{B.6}$$

where Δt is the timestep size and k is the timestep number. Inserting eqs. (B.2)–(B.6) into eq. (B.1) and rearranging yields

$$\begin{aligned}
&-\left(\frac{\Delta t}{2\Delta x}U^{(i,j)} + \frac{\Delta t\alpha}{2\Delta x^2}\right)C_{k+1}^{(i-1,j)} \\
&-\left(\frac{\Delta t}{2\Delta y}V^{(i,j)} + \frac{\Delta t\alpha}{2\Delta y^2}\right)C_{k+1}^{(i,j-1)} \\
&+\left(1 + \frac{\Delta t\alpha}{\Delta x^2} + \frac{\Delta t\alpha}{\Delta y^2}\right)C_{k+1}^{(i,j)} \\
&+\left(\frac{\Delta t}{2\Delta x}U^{(i,j)} - \frac{\Delta t\alpha}{2\Delta x^2}\right)C_{k+1}^{(i+1,j)} \\
&+\left(\frac{\Delta t}{2\Delta y}V^{(i,j)} - \frac{\Delta t}{2\Delta y}V^{(i,j)} - \frac{\Delta t\alpha}{2\Delta x^2}\right)C_{k+1}^{(i,j+1)} = C_k^{(i,j)}.
\end{aligned} \tag{B.7}$$

Equation (B.7) results in the pentadiagonal matrix from Fig. 1 when the $C^{(i,j)}$ are put into a single column vector form. Since we are not solving this system we have not defined any boundary conditions.

C. *Discretization of Burgers' equation*

The semi-implicit scheme used to solve the one-dimensional Burgers' equation employs centered difference approximation in x [eqs. (B.2) and (B.4)], Euler backward in time [eq. (B.6)] for the linear (diffusion) term and second order Adams–Bashforth for the non-linear term. This results in the discretized equations

$$-\frac{\nu\Delta t}{\Delta x^2}u_{k+1}^{(i+1)} + \left(1 + \frac{\nu\Delta t}{\Delta x^2}\right)u_{k+1}^{i} - \frac{\nu\Delta t}{\Delta x^2}u_{k+1}^{(i-1)} \tag{C.1}$$

$$= -\frac{\Delta t}{2\Delta x}u_k^{(i+1)} + \left(1 - \frac{\nu\Delta t}{\Delta x^2}\right)u_k^{i} + \frac{\Delta t}{2\Delta x}u_k^{(i-1)} \tag{C.2}$$

where Δt is the timestep size, Δx is the spatial discretization and i is the spatial index. In this case, both **A** and **B** are tridiagonal matrices. Periodic boundary conditions will add the equation

$$u_{k+1}^{(n)} = u_{k+1}^{(1)} \tag{C.3}$$

where n is the number of grid points.

References

Benton, E. R. and Platzman, G. W. 1972. A table of solutions of the one-dimensional Burgers' equation. *Quart. Appl. Math.* **30**, 195–212.

Beylkin, G., Coifman, R. and Rokhlin, V. 1991. Fast wavelet transforms and numerical algorithms. *Commun. Pure Appl. Math.* **41**, 141–183.

Chin, T. M. and Mariano, A. J. 1994. Wavelet-based compression of covariances in Kalman filtering of geophysical flows. *Proc. SPIE*, **2242**.

Chin, T. M., Mariano, A. J. and Chassignet, E. P. 1999. Spatial regression and multiscale approximations for sequential data assimilation in ocean models. *J. Geophys. Res.* **104**, 7991–8014.

Cohn, S. E., da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.* **126**, 2913–2926.

Cohn, S. E. and Todling, R. 1996. Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteorol. Soc. Jpn.* **74**, 63–75.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **41**, 909–996.

Dee, D. P. 1991. Simplification of the Kalman filter for meteorological data assimilation. *Q. J. R. Meteorol. Soc.* **117**, 365–384.

Dee, D. P. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.* **123**, 1128–1145.

Desroziers, G. and Lafore, J-P. 1993. A coordinate transformation for objective frontal analysis. *Mon. Wea. Rev.* **121**, 1531–1553.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.-Oceans* **99**, 10 143–10 162.

Farrell, B. F. and Ioannou, P. J. 2002. Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.* **58**, 2771–2789.

Farrell, B. F. and Ioannou, P. J. 2001. State estimation using a reduced-order Kalman filter. *J. Atmos. Sci.* **58**, 3666– 3680.

Gottelmann, J. 1999. Locally supported wavelets on manifolds with Applications to the 2D sphere. *Appl. Computa. Harmonic Anal.* **7**, 1–33.

Kalman, R. E. 1960. A new approach to linear filter and prediction problems. *Trans. ASME, Ser. D*, **98**, 35– 45.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Lermusiaux, P. F. J. and Robinson, A. R. 1999. Data assimilation via error subspace statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.* **127**, 1385–1407.

Lermusiaux, P. F. J. 1999. Data assimilation via error subspace statistical estimation. *Mon. Wea. Rev.* **127**, 1408–1432.

Mallat, S. G. 1989. A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence* **11**, 675– 693.

Ménard, R. 1994. Kalman filtering of Burgers' equation and its application to atmospheric data assimilation. Ph.D. Thesis, Mcgill University, Montreal.

Moore, B. C. 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Autom. Control* **AC-26**, 17– 31.

Parish, D. F. and 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. and Flannery, B. P. 1992. *Numerical recipes*. Cambridge Univ. Press, Cambridge, UK.

Riishøjgaard, L. P. 1998. A direct way of specifying flow-dependent background error Correlations for meteorological analysis systems. *Tellus* **50A**, 43–57.

Strang, G. and Nguyen, T. 1996. *Wavelets and filter banks*. Wellesley-Cambridge Press, Wellesley, MA.

Tangborn, A. and Zhang, S. 2000. Wavelet transform adapted to an approximate Kalman filter system. *Appl. Num. Math.* **33**, 307–316.

Tippet, M. K., Cohn, S. E., Todling, R. and Marchesin, D. 2000. Low-dimensional representation of error covariance. *Tellus* **52**, 533–553.