Department of Mathematics and Statistics
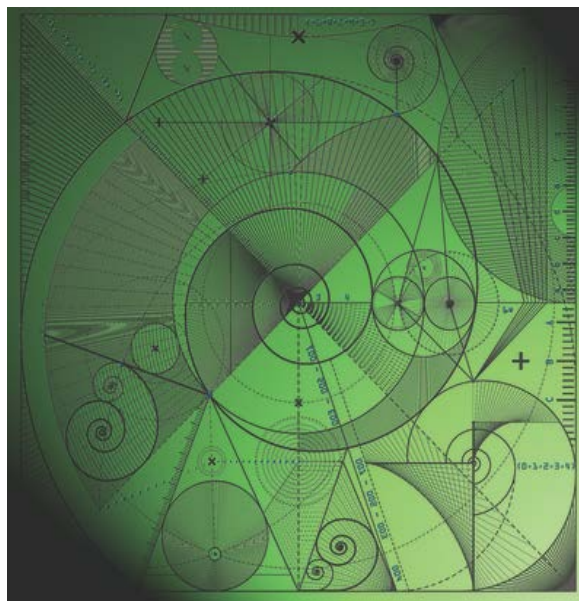
A joint state and parameter estimation scheme for nonlinear dynamical systems

by

Polly J. Smith, **Sarah L. Dance** and
Nancy K. Nichols

A JOINT STATE AND PARAMETER ESTIMATION SCHEME FOR NONLINEAR DYNAMICAL SYSTEMS*

POLLY J. SMITH[†‡], SARAH L. DANCE[†], AND NANCY K. NICHOLS[†]

Abstract. We present a novel algorithm for concurrent model state and parameter estimation in nonlinear dynamical systems. The new scheme uses ideas from three dimensional variational data assimilation (3D-Var) and the extended Kalman filter (EKF) together with the technique of state augmentation to estimate uncertain model parameters alongside the model state variables in a sequential filtering system. The method is relatively simple to implement and computationally inexpensive to run for large systems with relatively few parameters. We demonstrate the efficacy of the method via a series of identical twin experiments with three simple dynamical system models. The scheme is able to recover the parameter values to a good level of accuracy, even when observational data are noisy. We expect this new technique to be easily transferable to much larger models.

Key words. state estimation, parameter estimation, variational data assimilation, filtering, nonlinear dynamical systems.

AMS subject classifications.

1. Introduction. Parameters are intrinsic to numerical modelling. Parameterisations are typically used as a way of representing processes that are not fully known or understood, or where limitations to computer power constrain the model resolution and therefore the level of detail that can be described. Numerical models will often contain empirical or heuristic components derived from practical experience rather than physical laws. A consequence of this is that model parameters often do not represent an accurately known or directly measurable quantity. Whilst these approximations are generally well founded, uncertainties in the model parameters can lead to significant errors between the predicted and actual states of the system.

Typically, parameter estimation is addressed as a separate issue to state estimation and model calibration is performed off-line in a separate calculation. Model parameters are ordinarily determined theoretically or by adhoc calibration of the model against historical data sets. More recently, improvements in computational capabilities have seen the development of many novel, and often complex, automated parameter optimisation algorithms. Generally, these methods involve data fitting via the minimisation of an objective function [16], [20], [35]. The main distinctions between the different methods are how the minimum is located, how the observed data are processed and the assumptions made about the error statistics. A useful inter-comparison of several optimisation techniques for parameter estimation applied to terrestrial biogeochemical models is given in [48]. Another approach is to use Bayesian ideas which allow the uncertainty in the parameter estimates to be assessed [36], [37], [51].

Efforts to improve computational models of dynamical systems tend to concentrate on either improving methods for parameter estimation (without explicitly accounting for uncertainty in the model state estimate) or improving methods for state estimation (estimating model variables whilst keeping the model parameters fixed).

[†]School of Mathematical and Physical Sciences, University of Reading, UK
[‡]Corresponding author email: p.j.smith@reading.ac.uk

However, in some scenarios it can make sense to perform joint estimation of both the model state and parameters simultaneously. One method for doing this is to use data assimilation techniques based on filtering methods.

State estimation, or data assimilation, is a technique for combining partial observations of a dynamical system with a mathematical model of that system in order to accurately estimate the current system state and in turn produce better predictions of future states. In filtering based estimation methods the observational data are received sequentially in time and the system state is updated in an on-line fashion as part of an analysis-forecast loop.

Data assimilation methods are most commonly used in the earth sciences to provide initial conditions for model forecasting in applications such as meteorology and oceanography. However, they can also be used to estimate uncertain model parameters concurrently with the dynamical system state. Here, we do this by employing the method of *state augmentation* [19]. State augmentation is conceptually very simple and can, in theory, be applied with any standard state estimation method. The parameters are regarded as variables in the dynamical system and are appended to the model state vector to give an augmented state vector; the equations governing the evolution of the model state are combined with the equations describing the evolution of these parameters and the chosen state estimation algorithm is simply applied to this new augmented system. This enables us to estimate the model parameters and update the predicted model state simultaneously, rather than treating the problem as two individual processes, and means that observational data can be used much more efficiently.

The state augmentation approach has previously been successfully employed in the context of model error or bias estimation (see e.g. [3], [7], [14], [27]). The review article [30] discusses state augmentation for parameter estimation in relation to four dimensional variational data assimilation (4D-Var) techniques and surveys the literature relating to parameter estimation in meteorology and oceanography. The technique has also been applied with the Kalman filter. In [25] the method is used for model bias estimation in ocean modelling, [47] employs the technique with the extended and ensemble Kalman filters for parameter estimation in a simplified biogeochemical model, and [11] uses the approach with an Ensemble Transform Kalman Filter (ETKF) for parameter and bias estimation in river flood modelling. Here, we implement the method within a sequential three dimensional variational (3D-Var) data assimilation scheme (e.g. [5], [23]). Variational data assimilation is a popular choice for state estimation in large problems. 3D-Var is a well established method that has many practical advantages over other estimation techniques, such as ease of implementation, computational efficiency and robustness.

A key challenge in the construction of a state estimation algorithm is specification of the statistics of the errors in the *a priori* system state estimate (referred to as the *background* errors). These statistics, in the form of error covariances, play an important role in the filtering and spreading of the observational data and are therefore fundamental in determining the nature of the solution. Ideally the background error covariance should be evolved with the model, but this is computationally very expensive, and infeasible when the system of interest is of high dimension. Conventional 3D-Var algorithms assume that the background error covariances are statistically stationary; the structure of the covariance matrix is specified at the start of the assimilation and kept fixed throughout.

A particular issue highlighted by [39] and [40] is the role of the state-parameter

cross-covariances in joint state-parameter estimation. It is these cross-covariances that transfer information from the observations to the parameter estimates and determine the nature of the parameter updating. In [39] it was found that whilst the assumption of static error covariances was sufficient for state estimation, it was insufficient for joint state-parameter estimation. In order to yield reliable estimates of the exact parameters, a flow dependent representation of the state-parameter cross-covariances is required. Crucially, however, it is not necessary to evolve the full augmented system covariance matrix. This result led to the development of a novel algorithm that uses ideas from 3D-Var and the extended Kalman filter (EKF) to construct a hybrid error covariance matrix. The new approach enables us to capture the flow dependent nature of the state-parameter error cross-covariances whilst avoiding the explicit propagation of the full background error covariance matrix. As we demonstrate here, the method has proved to be applicable to a range of dynamical system models. An additional example of its application is given in [43].

In this paper we give details of the formulation of our new method and demonstrate its efficacy using three simple models: a single parameter 1D linear advection model, a two parameter nonlinear damped oscillating system, and a three parameter nonlinear chaotic system. The scheme has been tested by running a series of identical twin experiments. The results are positive and confirm that our new scheme can indeed be a valuable tool in identifying uncertain model parameters. We are able to recover the model parameter values to a good level of accuracy, even when observations are noisy. We expect that there is potential for successful application of our new methodology to larger, more realistic models with more complex parameterisations.

This paper is organised as follows. In §2 we introduce the augmented system model and give a brief overview of the sequential 3D-Var and Kalman filter algorithms upon which our new method is based. Details of the formulation of our new method are given in §3. In §4 we derive estimates of the state-parameter cross covariances for each of our three test models and present results from our experiments. Finally, in §5 we summarise the conclusions from this work.

2. State estimation and state augmentation. In this section we present the background material necessary for understanding the formulation and implementation of our new scheme as described in §3. We start by introducing the equations for a general dynamical system model and explaining the notation and terminology that we will be using throughout this paper.

2.1. The model system equations. We consider the discrete nonlinear time invariant dynamical system model

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{p}), \qquad k = 0, 1, \dots \tag{1}$$

The column vector $\mathbf{x}_k \in \mathbb{R}^n$ is known as the state vector; it contains the model variables we wish to estimate and represents the system state at time t_k. The vector $\mathbf{p} \in \mathbb{R}^q$ is a vector of q (uncertain) model parameters, and the operator $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a nonlinear function describing the evolution of the state from time t_k to t_{k+1}. We assume that specification of $\mathbf{x}$ and $\mathbf{p}$ at time t_k uniquely determines the model state at all future times. We also assume that $\mathbf{f}(\mathbf{x}, \mathbf{p})$ is differentiable with respect to $\mathbf{x}$ and $\mathbf{p}$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{p} \in \mathbb{R}^q$.

In §4.1 we consider a model for which the operator $\mathbf{f}$ is a linear function of the model state. In this case we can re-write the model (1) in the form

$$\mathbf{x}_{k+1} = \mathbf{M}_k(\mathbf{p})\mathbf{x}_k, \qquad k = 0, 1, \dots, \tag{2}$$

where the matrix $\mathbf{M}_k \in \mathbb{R}^{n \times n}$ depends nonlinearly on the parameters $\mathbf{p}$.

In this paper we use the 'perfect model assumption' [32]; for any given initial state, the model equations (1), together with the known exact parameter values, give a unique description of the behaviour of the underlying exact dynamical system. We also assume that the model parameters remain constant over time, that is, they are not altered by the forecast model from one time step to the next. The equation for the evolution of the parameters therefore takes the simple form

$$\mathbf{p}_{k+1} = \mathbf{p}_k, \qquad k = 0, 1, \ldots . \tag{3}$$

2.2. The augmented system. We define the augmented state vector $\mathbf{w}$ by appending the parameter vector $\mathbf{p}$ to the model state vector $\mathbf{x}$,

$$\mathbf{w} = \left(\begin{array}{c} \mathbf{x} \\ \mathbf{p} \end{array} \right) \in \mathbb{R}^{n+q}. \tag{4}$$

Combining the equation for the evolution of the parameters (3) with the state forecast model (1) we can write the equivalent augmented system model as

$$\mathbf{w}_{k+1} = \widetilde{\mathbf{f}}(\mathbf{w}_k), \tag{5}$$

where

$$\widetilde{\mathbf{f}}(\mathbf{w}_k) = \left(\begin{array}{c} \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k) \\ \mathbf{p}_k \end{array} \right), \tag{6}$$

with $\widetilde{\mathbf{f}} : \mathbb{R}^{n+q} \longrightarrow \mathbb{R}^{n+q}$.

In most cases the parameters will enter the governing equations nonlinearly so that even if the original dynamical model is a linear function of the state variables the resulting augmented system model will be nonlinear. However, since the number of parameters is typically small relative to the dimension of the state vector and the dynamics of the parameters are simple, this does not have a significant impact on computational cost.

2.3. Sequential state and parameter estimation. For sequential estimation, we start with a background state $\mathbf{w}_k^b = (\mathbf{x}_k^b, \mathbf{p}_k^b)^T$ that represents *a priori* estimates of the exact system state $\mathbf{x}_k^t$ and model parameters $\mathbf{p}^t$ at time t_k, with error $\boldsymbol{\varepsilon}_k^b \in \mathbb{R}^{(n+q)}$. This should be the best available approximation of the current exact system state and parameters and is typically obtained from a previous model forecast.

We suppose that we have a set of r_k observations to assimilate and that these are related to the model state by the equations

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k^t) + \boldsymbol{\delta}_k, \qquad k = 0, 1, \ldots \tag{7}$$

Here $\mathbf{y}_k \in \mathbb{R}^{r_k}$ is a vector of r_k observations valid at time t_k. Note that the number of available observations r_k may vary with time. The operator $\mathbf{h}_k : \mathbb{R}^n \longrightarrow \mathbb{R}^{r_k}$ is a nonlinear observation operator that maps from model state space to observation space and the vector $\boldsymbol{\delta}_k \in \mathbb{R}^{r_k}$ represents the observational errors. These errors are commonly assumed to be unbiased, serially uncorrelated, stochastic variables, with a given probability distribution [22]. For the augmented problem, the equation for the observations (7) becomes

$$\mathbf{y}_k = \widetilde{\mathbf{h}}_k(\mathbf{w}_k^t) + \boldsymbol{\delta}_k = \widetilde{\mathbf{h}}_k \left(\begin{array}{c} \mathbf{x}_k^t \\ \mathbf{p}_k^t \end{array} \right) + \boldsymbol{\delta}_k \stackrel{\text{def}}{=} \mathbf{h}_k(\mathbf{x}_k^t) + \boldsymbol{\delta}_k, \qquad k = 0, 1, \ldots, \tag{8}$$

where $\tilde{\mathbf{h}}_k : \mathbb{R}^{n+q} \longrightarrow \mathbb{R}^{r_k}$ maps from augmented model space to observation space. The equivalence of equations (8) and (7) comes from the fact that the parameters cannot be observed.

The aim of state-parameter estimation is to combine the measured observations $\mathbf{y}_k$ with the prior estimates $\mathbf{w}_k^b$ to produce an updated augmented model state that gives the best estimate of the expected value of the exact system state $\mathbf{w}_k^t$ at time t_k. This optimal estimate is called the *analysis* and is denoted $\mathbf{w}_k^a$. For sequential estimation, this updating procedure is carried out repeatedly in a series of forecast-analyis steps; the current analysis state is used to forecast the background state for the next analysis time by taking it as the initial condition and evolving the model forward to the time at which a new set of observations becomes available. Note that although the exact model parameters are assumed to be constant, the value of the estimated parameters used in the state forecast model will vary with time as they are updated by the estimation.

If we assume that the background and observation errors are unbiased and uncorrelated with Gaussian probability density functions, the optimal analysis is equal to the maximum *a posteriori* Bayesian estimate of the system states [32]. The estimation problem reduces to minimising a cost function measuring the misfit between the model state $\mathbf{w}_k$, and the background state $\mathbf{w}_k^b$ and the model and observations $\mathbf{y}_k$, weighted by the inverse of their respective error covariances,

$$J(\mathbf{w}_k) = \frac{1}{2}\left(\mathbf{w}_k - \mathbf{w}_k^b\right)^T \left(\mathbf{P}_k^f\right)^{-1}\left(\mathbf{w}_k - \mathbf{w}_k^b\right) + \frac{1}{2}\left(\mathbf{y}_k - \tilde{\mathbf{h}}_k(\mathbf{w}_k)\right)^T \mathbf{R}_k^{-1}\left(\mathbf{y}_k - \tilde{\mathbf{h}}_k(\mathbf{w}_k)\right).$$
(9)

The covariance matrices $\mathbf{P}_k^f \in \mathbb{R}^{(n+q)\times(n+q)}$ and $\mathbf{R}_k \in \mathbb{R}^{r_k \times r_k}$ are taken to be symmetric and positive definite. These matrices characterise the uncertainties in the background state and the observations and determine the relative contribution of the background and observations in the analysis. They are defined as

$$\mathbf{P}_k^f = E\left[\boldsymbol{\varepsilon}_k^b \, \boldsymbol{\varepsilon}_k^{b\,T}\right] \quad \text{and} \quad \mathbf{R}_k = E\left[\boldsymbol{\delta}_k \, \boldsymbol{\delta}_k^{\,T}\right].$$
(10)

There are a variety of state estimation methods that can be used to solve the nonlinear optimisation problem (9), each varying in formulation, complexity, computational burden, optimality and suitability for practical application. Our novel scheme utilises ideas from the 3D-Var and extended Kalman filter methods. In 3D-Var (e.g. [5], [23], [31]) the minimising solution at each observation time t_k is found iteratively via a gradient descent algorithm. Generally, the assumption made in 3D-Var schemes is that the statistics of the model state background errors are homogenous, isotropic and independent of the flow. The background error covariances are then approximated by a fixed matrix (i.e. $\mathbf{P}_k^f = \mathbf{P}^f$ for all k in (9)), thus making 3D-Var an efficient approach for large scale problems. If a 3D-Var scheme is applied cyclically as described above it can be regarded as a sequential estimation method, thus allowing us to utilise time series of observations.

In the EKF (e.g. [12], [19]), the analysis is taken to be the best linear unbiased estimate (BLUE) [22] of the solution to the optimisation problem and is calculated directly as

$$\mathbf{w}_k^a = \mathbf{w}_k^b + \mathbf{K}_k(\mathbf{y}_k - \tilde{\mathbf{H}}_k \mathbf{w}_k^b).$$
(11)

The Kalman gain $\mathbf{K}_k$ is given by

$$\mathbf{K}_k = \mathbf{P}_k^f \tilde{\mathbf{H}}_k^T (\tilde{\mathbf{H}}_k \mathbf{P}_k^f \tilde{\mathbf{H}}_k^T + \mathbf{R}_k)^{-1}.$$
(12)

where the matrix $\tilde{\mathbf{H}}_k \in \mathbb{R}^{r_k \times (n+q)}$ represents the linearisation (or Jacobian) of the augmented observation operator $\tilde{\mathbf{h}}_k$ evaluated at the background state $\mathbf{w}_k^b$.

Unlike in 3D-Var, the EKF algorithm forecasts the error covariance matrix $\mathbf{P}_k^f$ forward, using the quality of the current analysis to specify the covariances for the next update step. If the Kalman gain (12) has been computed exactly, the analysis (posterior) error covariance $\mathbf{P}_k^a$ is given by

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \tilde{\mathbf{H}}_k)\mathbf{P}_k^f. \tag{13}$$

The background (forecast) state at t_{k+1} is again found by evolving the model forward from the current analysis state. The background (or forecast) error covariance at t_{k+1} is determined by propagating the analysis error covariance forward from time t_k using a linearisation of the model dynamics

$$\mathbf{P}_{k+1}^f = \mathbf{F}_k \mathbf{P}_k^a \mathbf{F}_k^T, \tag{14}$$

where

$$\mathbf{F}_k = \left. \frac{\partial \tilde{\mathbf{f}}}{\partial \mathbf{w}} \right|_{\mathbf{w}_k^a} = \left. \begin{pmatrix} \frac{\partial \mathbf{f}(\mathbf{x},\mathbf{p})}{\partial \mathbf{x}} & \frac{\partial \mathbf{f}(\mathbf{x},\mathbf{p})}{\partial \mathbf{p}} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \right|_{\mathbf{x}_k^a, \mathbf{p}_k^a} \tag{15}$$

is the Jacobian of the augmented system forecast model evaluated at the current analysis state $\mathbf{w}_k^a$.

If the system is linear, equation (11) gives the exact solution to the minimisation problem (9), but otherwise the EKF solution it is not optimal due to linearisation of the model dynamics and observation operators. The high computational cost of the EKF means that it becomes prohibitively expensive for large scale systems. In practice the matrix $\mathbf{P}_k^f$ is kept constant or a much simpler updating is performed. However, the equations provide a useful basis for the development of approximate algorithms.

3. A hybrid approach. Although state augmentation appears straightforward in theory, practical implementation of the approach relies strongly on the relationships between the parameters and model state components being well defined and assumes that we have sufficient knowledge to reliably describe them. Since it is not possible to observe the parameters themselves, the parameter estimates will depend on the observations of the state variables. For joint state-parameter estimation, it is the state-parameter cross-covariances that govern how information in the observed variables is translated into updates in the estimates of the unobserved parameters. A good *a priori* specification of these covariances is therefore fundamental to reliable joint state and parameter estimation. Since by the nature of the problem the true error statistics of the system are unknown we have to approximate them in some manner; this can offer a significant challenge.

Previous work [39], [40] indicated that in order to be able to reliably update the model parameters the state-parameter cross-covariances in $\mathbf{P}_k^f$ need to have a flow-dependent structure. However, an important part of our findings was that, provided the state-parameter cross-covariances were well defined, it was not necessary to propagate the model state background error covariances; here a static representation was sufficient. Following this result, we developed a novel hybrid approximation that captures the flow dependence of the state-parameter errors without the computational expense and complexity of explicitly propagating the full system covariance matrix.

The state-parameter cross-covariances are estimated based on a simplified version of the EKF error covariance forecast step (14) and this is then combined with an empirical, fixed approximation of the model state background error covariances and a fixed parameter error covariance matrix. We give details of the formulation of this new approach in the following section.

3.1. Formulation. The augmented EKF forecast and analysis error covariance matrices can be partitioned as follows

$$\mathbf{P}_k = \left(\begin{array}{cc} \mathbf{P}_{\mathbf{xx}_k} & \mathbf{P}_{\mathbf{xp}_k} \\ (\mathbf{P}_{\mathbf{xp}_k})^T & \mathbf{P}_{\mathbf{pp}_k} \end{array} \right) \tag{16}$$

with the superscript f or a used to indicate forecast or analysis. Here $\mathbf{P}_{\mathbf{xx}_k} \in \mathbb{R}^{n \times n}$ is the forecast (analysis) error covariance matrix for the model state vector $\mathbf{x}_k$ at time t_k, $\mathbf{P}_{\mathbf{pp}_k} \in \mathbb{R}^{q \times q}$ is the covariance matrix describing the errors in the parameter vector $\mathbf{p}_k$ and $\mathbf{P}_{\mathbf{xp}_k} \in \mathbb{R}^{n \times q}$ is the covariance matrix for the cross correlations between the forecast (analysis) errors in the state and parameter vectors.

Starting at time t_k, we consider the form of the EKF forecast error covariance matrix (14) for a single step of the filter. If we denote the Jacobian of the state forecast model with respect to the model state and model parameters respectively as

$$\mathbf{M}_k = \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{p})}{\partial \mathbf{x}} \right|_{\mathbf{x}_k^a, \mathbf{p}_k^a} \qquad \text{and} \qquad \mathbf{N}_k = \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{p})}{\partial \mathbf{p}} \right|_{\mathbf{x}_k^a, \mathbf{p}_k^a}, \tag{17}$$

where $\mathbf{M}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{N}_k \in \mathbb{R}^{n \times q}$, and substitute into (15),(14) we obtain the following expressions for the blocks of $\mathbf{P}_{k+1}^f$

$$\mathbf{P}_{\mathbf{xx}_{k+1}}^f = \mathbf{M}_k \mathbf{P}_{\mathbf{xx}_k}^a \mathbf{M}_k^T + \mathbf{N}_k (\mathbf{P}_{\mathbf{xp}_k}^a)^T \mathbf{M}_k^T + \mathbf{M}_k \mathbf{P}_{\mathbf{xp}_k}^a \mathbf{N}_k^T + \mathbf{N}_k \mathbf{P}_{\mathbf{pp}_k}^a \mathbf{N}_k^T, \tag{18}$$

$$\mathbf{P}_{\mathbf{xp}_{k+1}}^f = \mathbf{M}_k \mathbf{P}_{\mathbf{xp}_k}^a + \mathbf{N}_k \mathbf{P}_{\mathbf{pp}_k}^a, \tag{19}$$

$$\mathbf{P}_{\mathbf{pp}_{k+1}}^f = \mathbf{P}_{\mathbf{pp}_k}^a. \tag{20}$$

We do not want to recompute the full augmented matrix (18–20) at every time step. Guided by the results of our previous work [39], [40] we simplify as follows. We substitute the EKF model state forecast error covariance matrix (18) with a conventional 3D-Var fixed approximation

$$\mathbf{P}_{\mathbf{xx}_k}^f = \mathbf{P}_{\mathbf{xx}}^f \qquad \text{for all } k. \tag{21}$$

The choice for $\mathbf{P}_{\mathbf{xx}}^f$ will depend on the particular model application; a simple and commonly adopted approach is to define $\mathbf{P}_{\mathbf{xx}}^f$ using an analytic correlation function [6]. Alternatively, a more sophisticated covariance representation can be obtained using one of the various empirical techniques discussed in the literature (see e.g. [1], [10]). We make the same assumption for the parameter error covariances and set

$$\mathbf{P}_{\mathbf{pp}_k}^f = \mathbf{P}_{\mathbf{pp}}^f \qquad \text{for all } k. \tag{22}$$

Specification of $\mathbf{P}_{\mathbf{pp}}^f$ requires some *a priori* knowledge of the parameter error statistics. The error variance of each parameter should reflect our uncertainty in its initial estimate and should be chosen so as to ensure that the parameter updates are realistic and consistent with the scales of the model. When the number of model parameters

we wish to estimate, q, is greater than one, we also need to consider the relationships between individual parameters.

For our new method we focus on the state-parameter background error cross-covariances given by the off-diagonal blocks (19). Even with assumptions (21) and (22), evaluating (19) at every time step would introduce unwanted complexity into our scheme. We want a simple approximation that provides an adequate characterisation of the state-parameter cross-covariances whilst being straightforward to compute.

It is not unreasonable to assume that at the initial time the background state and parameter errors are uncorrelated. Setting $\mathbf{P}^a_{\mathbf{xp}_k} = \mathbf{0}$ and using (22) gives us the following approximation to (19)

$$\mathbf{P}^f_{\mathbf{xp}_{k+1}} = \mathbf{N}_k \mathbf{P}^f_{\mathbf{pp}}. \tag{23}$$

Our augmented background error covariance matrix now takes the following form

$$\mathbf{P}^f_{k+1} = \begin{pmatrix} \mathbf{P}^f_{\mathbf{xx}} & \mathbf{N}_k \mathbf{P}^f_{\mathbf{pp}} \\ \mathbf{P}^f_{\mathbf{pp}} \mathbf{N}_k^T & \mathbf{P}^f_{\mathbf{pp}} \end{pmatrix}. \tag{24}$$

To summarise, all elements of the augmented background error covariance matrix (24) are kept fixed except the off-diagonal blocks which are updated by recomputing the matrix $\mathbf{N}_k$ at each new analysis time, where $\mathbf{N}_k$ is given by (17). This approximation enables us to capture the flow dependent nature of the state-parameter cross-covariances without having to explicitly evolve the full augmented system matrix.

4. Models and experiments. In this section we demonstrate our new method using three simple models: (i) a single parameter 1D linear advection model [49]; (ii) a two parameter nonlinear damped oscillating system [15]; and (iii) the three parameter nonlinear chaotic Lorenz 63 equations [24], [45]. We give details of each model and its discretisation before using the method formulated in the previous section to derive an approximation of the state-parameter error cross-covariances. A brief description of the experimental design is given, followed by selected results.

The scheme has been tested via a series of twin experiments using pseudo observations with a range of temporal frequencies. For the linear advection model we also vary the spatial frequency of the observations. In each case we specify an exact or reference solution; this solution is used to provide the observations and also to evaluate the performance of the scheme in terms of estimating the state variables. It is generated by running the model from a given initial condition with pre-specified parameter values.

Observations are assimilated sequentially and are generated by sampling the reference solution at regular intervals. The temporal and (where applicable) spatial frequency of the observations remains fixed for the duration of each individual experiment but is varied between experiments as described in §4.1.3, §4.2.3 and §4.3.3. In all cases, the observation operator is linear and takes the form

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{H}} \equiv \begin{pmatrix} \mathbf{H} & \mathbf{0} \end{pmatrix} \qquad \text{for all } k, \tag{25}$$

where $\mathbf{H} \in \mathbb{R}^{r \times n}$, with the number of observations $r_k = r$ the same for all k. In cases (ii) and (iii) we have $\mathbf{H} = \mathbf{I} \in \mathbb{R}^{n \times n}$.

Experiments have been run with both perfect and imperfect observations. The imperfect observations are created by adding random noise from a Gaussian distribution with mean zero and variance σ_o^2, where σ_o^2 is the observation error variance. We
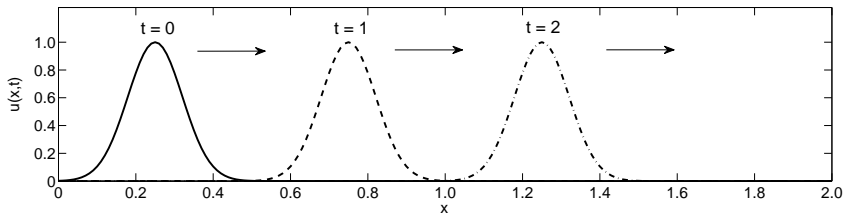
FIG. 1. *Solutions $u(x,t)$ to the linear advection equation (27) for Gaussian initial data at model times $t = 0$, 1, 2.*

assume that the observation errors are spatially and temporally uncorrelated and set the observation error covariance matrix $\mathbf{R}_k$

$$\mathbf{R}_k = \mathbf{R} = \sigma_o^2 \mathbf{I}, \qquad \mathbf{I} \in \mathbb{R}^{r \times r}. \tag{26}$$

In the perfect observations experiments, the observation error variance is set at $\sigma_o^2 = 0.01$. The use of a non-zero observation error variance allows us to investigate the impact of the accuracy of the observations without actually adding noise. This is common practice in the preliminary testing of state estimation schemes with pseudo data (e.g. [27]).

The initial background state is generated by perturbing the initial state of the reference solution as described in §4.1.2, §4.2.2 and §4.3.2. The initial parameter estimates are generated by adding Gaussian random noise with zero mean and given variance to the exact parameter values.

4.1. Linear advection. We consider the one-dimensional linear advection equation [49]

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0, \tag{27}$$

where $u(x,t)$ is the model state variable, t is time, and the parameter we wish to estimate, c, represents the advection velocity or wave speed in the x direction.

The model (27) has the benefit of an analytical solution [21]. Given initial data, $u(x,0)$, and known, non-zero, constant velocity, the solution has the property that it preserves its initial shape. As figure 1 illustrates, as time evolves, the initial data propagates undistorted at constant speed c to the right (for $c > 0$).

For the experiments, we solve (27) numerically on a finite spatial domain with periodic boundary conditions. We assume that the parameter c is positive and discretise using the upwind method (e.g. [21], [29])

$$u_{j,k+1} = u_{j,k} + c\lambda(u_{j-1,k} - u_{j,k}), \qquad k = 0, 1, \dots, \ j = 1, \dots n \tag{28}$$

with boundary conditions

$$u_{1,k} = u_{n,k} \tag{29}$$

where $u_{j,k} \approx u(x_j, t_k)$, with $x_j = (j-1)\Delta x$, $t_k = k\Delta t$, and $\lambda = \frac{\Delta t}{\Delta x}$ where Δx is the spatial grid spacing and Δt is the model time step.

The upwind scheme is first order accurate in space and time and stable provided that the CFL condition $|\lambda c| \leq 1$ is satisfied [8]. To ensure that the model remains

stable we set $\lambda = 1$ and assume that c is known to be somewhere in the interval $[0,1]$. The upwind scheme is numerically diffusive; this results mainly in amplitude errors in the solution when the forecast model is run with the correct c value and can be reduced by choosing a small Δx. The scheme (28) can be written as a linear matrix system enabling us to obtain an explicit expression for the elements of the Jacobian matrix $\mathbf{N}_k$.

The state forecast model (28), with known constant advection speed c, can be written as

$$\mathbf{x}_{k+1} = \mathbf{M}\mathbf{x}_k, \tag{30}$$

where $\mathbf{x}_k = (u_{1,k}, u_{2,k}, \ldots u_{n,k})^T \in \mathbb{R}^n$ is the model state at time t_k and $\mathbf{M}$ is a (constant) $n \times n$ matrix, that depends nonlinearly on the advection velocity c,

$$M_{i,j} = \begin{cases} (1 - \lambda c) & i = j \\ \lambda c & i = j+1 \,, \text{ and } (i,j) = (1,n) \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

Setting $\mathbf{w}_k = (\mathbf{x}_k, c_k)^T$, we combine (30) with the parameter evolution model (3) to give the augmented system model

$$\mathbf{w}_{k+1} = \begin{pmatrix} \mathbf{M}_k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_k \\ c_k \end{pmatrix} \tag{32}$$

Note that the constant matrix $\mathbf{M}$ in (30) has been replaced by the time varying matrix $\mathbf{M}_k = \mathbf{M}(c_k)$. Although the exact system matrix $\mathbf{M}$ is constant, during the state-parameter estimation the forecast model at time t_k will depend on the current estimate, c_k, of the exact advection velocity, c. The matrix $\mathbf{M}_k$ will therefore vary as c_k is updated.

4.1.1. State-parameter cross-covariance. In this case, we only have a single unknown parameter; the parameter vector is scalar and the parameter background error covariance matrix $\mathbf{P}^f_{\mathbf{pp}}$ is simply the parameter error variance, σ_c^2. The approximation of the cross-covariances between the errors in the model state and the parameter c at time t_{k+1} is therefore given by

$$\mathbf{P}^f_{\mathbf{x}\mathbf{p}_{k+1}} = \sigma_c^2 \mathbf{N}_k \,. \tag{33}$$

For the linear advection model, the matrix $\mathbf{N}_k$ is defined as

$$\mathbf{N}_k = \left. \frac{\partial (\mathbf{M}_k \mathbf{x}_k)}{\partial c} \right|_{\mathbf{x}_k^a, c_k^a}, \tag{34}$$

which is a vector in $\mathbb{R}^n$ with elements

$$N_{j,k} = \lambda(u_{j-1,k} - u_{j,k}), \qquad j = 1, \ldots, n; \;\; k = 0, 1, \ldots \tag{35}$$

4.1.2. Experiments. We run the linear advection model on the domain $x \in [0,3]$ with grid spacing $\Delta x = 0.01$ and time step $\Delta t = 0.01$, giving $\lambda = 1$. The initial state of the reference solution is given by the Gaussian function

$$u(x,0) = \begin{cases} 0 & x < 0.01 \\ e^{-\frac{(x-0.25)^2}{0.01}} & 0.01 < x < 0.5 \\ 0 & x \geq 0.5 \end{cases} . \tag{36}$$

The exact advection velocity is set as $c = 0.5$. The parameter error variance is set to be $\sigma_c^2 = 0.1$, which corresponds to an error variance of 20%. The initial background state for u is also specified as a Gaussian function but is rescaled so that it is a different height, width and centred around a slightly different point to the reference state. The model state background error covariance matrix is defined using a isotropic correlation function [34]

$$\mathbf{P}_{\mathbf{xx}}^f = \sigma_b^2 \boldsymbol{\rho}, \tag{37}$$

where σ_b^2 is the model state background error variance and $\boldsymbol{\rho}$ is the Markov matrix

$$\rho_{i,j} = e^{-\Delta x |i-j|/L} \qquad i,j = 1,\ldots,n. \tag{38}$$

The element $\rho_{i,j}$ defines the correlation between components i and j of the model state background error vector $\boldsymbol{\varepsilon}_x^b = (\mathbf{x}^b - \mathbf{x}^t)$ and L is a correlation length scale that is adjusted empirically. For these experiments L is set at twice the current observation spacing and $\sigma_b^2 = 0.05$.

The analysis $\mathbf{w}_{k+1}^a$ at time t_{k+1} is found by solving the nonlinear optimisation problem (9) numerically, using a quasi-Newton descent algorithm [13] to iterate to the minimising solution. At the end of each update cycle the state analysis is integrated forward, using the new value of c, to become the background state at the next analysis time.

4.1.3. Results. *Perfect observations.* Experiments were carried out using a range of both over- and under-estimated initial c values and different initial background guesses for $\mathbf{x}^b$. We found that the quality of the state analysis and the convergence and accuracy of the parameter estimates depends on a number of factors including the initial background guess, the location and spatial frequency of the observations, the time between successive updates, and the presence of observational noise. Here, we discuss the results from one example case where the advection velocity c is initially over estimated as $c_0 = 0.87116$.

Figure 2(a) shows the effect of varying the spatial frequency of the observations. Observations are assumed to be available every $10\Delta t$ with the grid spacing ranging from every Δx to every $50\Delta x$. The hybrid approach works extremely well. For observations taken at intervals between Δx and $25\Delta x$ the value of c is found, at worst, to within 2 decimal places. The speed of convergence of the estimate decreases as the number of observations decreases. For intervals up to $10\Delta x$ the analysis for u is consistently good and closely tracks the reference solution. When observations are taken every $25\Delta x$ there is initially some variation in the quality of the analysis for u but once the c estimate has converged it is also very good (see [38] for further discussion and figures). If the observation spacing is further increased to $50\Delta x$, the estimates of c are too low. However, if the time window is extended beyond that shown, they do slowly begin to converge towards the exact value. The state analysis also consequently improves.

Figure 2(b) shows the effect of varying the temporal frequency of the observations between $5\Delta t$ and $50\Delta t$. For these experiments, the spatial frequency is fixed at $10\Delta x$. The results are similar to the previous experiment; the speed of convergence decreases as the frequency of the observations decreases but the final estimated c values are very close to the exact value. There are only small differences in convergence when the time between successive updates is increased from every $5\Delta t$ to $10\Delta t$ to $25\Delta t$. The analysis for u is also very good, with only slight fluctuations in predicted height for
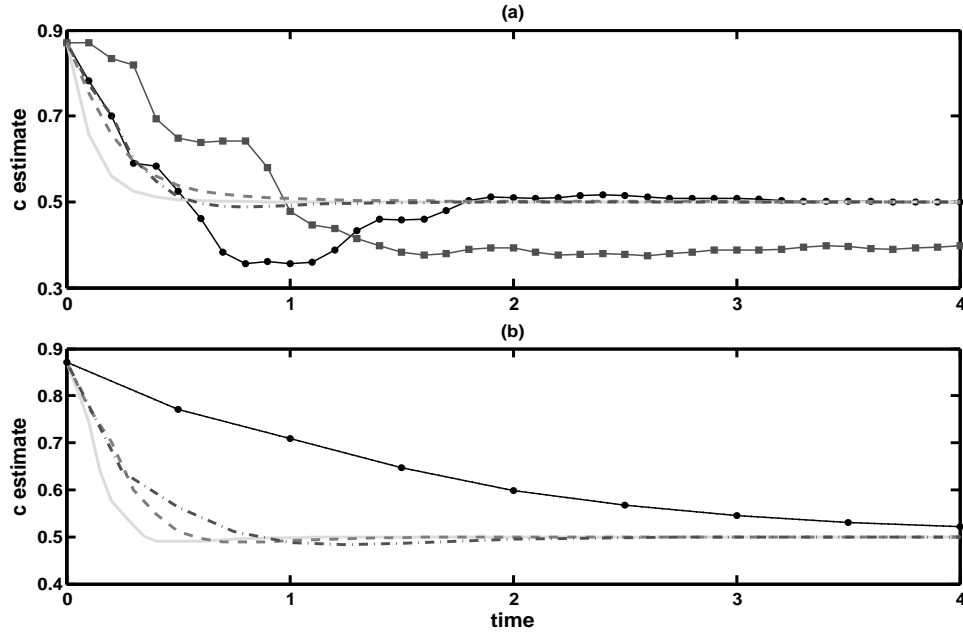
Fig. 2. *Linear advection model: parameter updates for initial estimate $c_0 = 0.87116$, exact $c = 0.5$.* **(a)** *Varying the spatial frequency of observations: solid grey line - observations at $1\Delta x$ intervals; dashed line - observations at $5\Delta x$ intervals; dot-dash line - observations at $10\Delta x$ intervals; solid line with round markers - observations at $25\Delta x$ intervals; solid line with square markers - observations at $50\Delta x$ intervals.* **(b)** *Varying the temporal frequency of observations: solid grey line - observations every $5\Delta t$; dashed line - observations every $10\Delta t$; dot-dash line - observations every $25\Delta t$; solid line with round markers - observations every $50\Delta t$.*
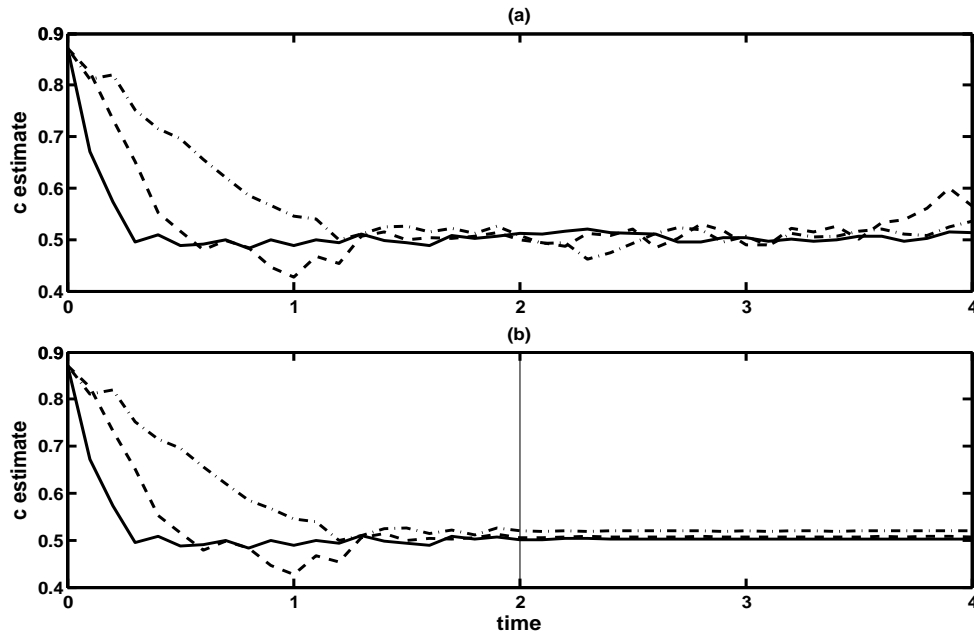

Fig. 3. *Linear advection model: imperfect observations. Parameter updates for initial estimate $c_0 = 0.87116$, exact $c = 0.5$.* **(a)** *unaveraged estimates.* **(b)** *time averaged estimates. Solid line - $\sigma_o^2 = 0.001$; dashed line - $\sigma_o^2 = 0.01$; dot-dash line - $\sigma_o^2 = 0.1$. Vertical line indicates the start of time averaging.*

the case $25\Delta t$. When observations are taken every $50\Delta t$ the estimates for c take longer to converge and as a result the model takes longer to stablise. Once the model has settled the analysis for u is relatively good. If the time between observations is doubled to $100\Delta t$ the scheme completely fails to recover c.

Noisy observations. For these experiments, noisy observations were taken with temporal frequency $10\Delta t$ at spatial intervals of $10\Delta x$. Figure 3 (a) shows the parameter estimates produced for observation error variance increasing from $\sigma_o^2 = 0.001$ to $\sigma_o^2 = 0.1$. This represents errors with variance of up to 10% of the maximum curve height.

As we would expect, when the observations are noisy the resulting state analysis and parameter estimates are also noisy. The amplitude of the oscillations in the estimated c values increases as σ_o^2 is increased. The oscillations are, however, approximately centered around the reference c value and lie within the bounds of uncertainty placed on the observations. We found that smoother and more accurate parameter estimates could be obtained by averaging over a moving time window, as is shown figure 3(b). Here, the c estimates are averaged over a moving time window of 50 time steps. Note that to allow time for the scheme to settle we omit the early estimates and begin the averaging at $t = 2.0$.

4.2. Nonlinear oscillator. Our second test model is a two parameter, unforced, damped non-linear oscillator given by the second order ordinary differential equation

$$\ddot{x} + d\dot{x} + mx + x^3 = 0\,, \tag{39}$$

where the parameters we wish to estimate d and m are real and constant, and $x = x(t)$.

Often referred to as the Duffing equation or Duffing oscillator, equation (39) arises in number of forms and has a variety of applications. See e.g. [15], [46], [50] for more detailed discussions. For $d, m > 0$, the form (39) describes the motion of a single mass attached to a spring with nonlinear elasticity and linear damping. The parameter d is the damping coefficient and m is the square of the frequency of oscillation.

We can rewrite (39) as the first order system

$$
\begin{aligned}
\dot{x} &= y, \\
\dot{y} &= -(mx + x^3 + dy).
\end{aligned}
\tag{40}
$$

The nature of the solution of this system varies greatly depending on the values of the parameters. For $d, m > 0$ the system (40) has a single stable equilibrium at $(x, \dot{x}) = (0, 0)$. We solve (40) numerically using a second order Runge-Kutta method (see e.g. [4]). The discrete system is given by the following set of equations

$$x_{k+1} = \left(\Delta t - d\frac{\Delta t^2}{2}\right) y_k + \left(1 - m\frac{\Delta t^2}{2} - \frac{\Delta t^2}{2}x_k^2\right) x_k \tag{41}$$

$$y_{k+1} = \left(1 - d\Delta t - m\frac{\Delta t^2}{2} + d^2\frac{\Delta t^2}{2}\right) y_k + \left(-m\Delta t + dm\frac{\Delta t^2}{2} + (d\frac{\Delta t^2}{2} - \frac{\Delta t}{2})x_k^2\right) x_k \ldots$$

$$-\frac{\Delta t}{2}\left(x_k + \Delta t y_k\right)^3\,, \qquad\qquad k = 0, 1, \ldots \tag{42}$$

where $x_k \approx x(t_k)$ and $y_k \approx y(t_k)$, $t_k = k\Delta t$.

We combine the parameters d and m in the vector $\mathbf{p}_k = (d_k, m_k)^T \in \mathbb{R}^2$ with the parameter evolution model given by (3). Adding the parameter vector to the state
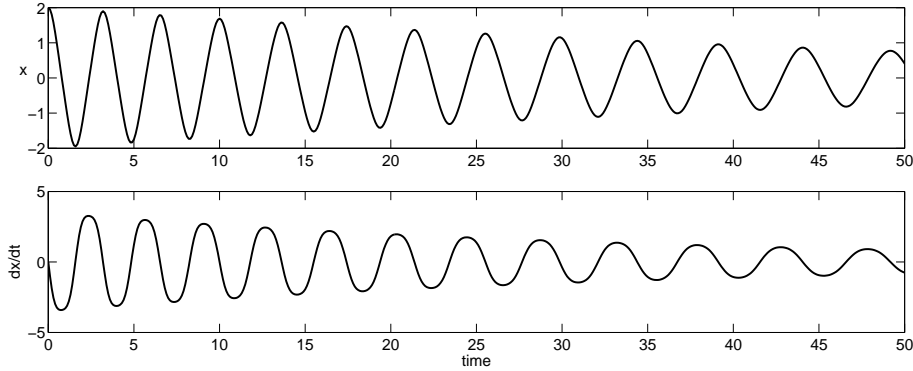
FIG. 4. *Damped, unforced nonlinear oscillator: computed numerical solution for x and y.*

vector $\mathbf{x}_k = (x_k,\, y_k)^T$ gives the augmented state vector

$$\mathbf{w}_k = (x_k,\, y_k,\, d_k,\, m_k)^T. \tag{43}$$

The augmented system model takes the form (5), (6) with $\mathbf{f} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by (41)–(42).

4.2.1. State-parameter cross-covariance. For the oscillating system, we assume that the parameters d and m are uncorrelated and set the parameter error covariance matrix $\mathbf{P}^f_{\mathbf{pp}} = \mathrm{diag}(\sigma_d^2, \sigma_m^2)$ where σ_d^2 and σ_m^2 are the error variances of d and m respectively. The Jacobian of the state forecast model with respect to the parameters is a 2×2 matrix defined as

$$\mathbf{N}_k = \left(\left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{p})}{\partial d} \quad \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{p})}{\partial m} \right) \right|_{\mathbf{x}_k^a, \mathbf{p}_k^a}, \tag{44}$$

so that

$$\mathbf{P}^f_{\mathbf{xp}_{k+1}} = \left(\left. \sigma_d^2 \frac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k)}{\partial d} \quad \sigma_m^2 \frac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k)}{\partial m} \right) \right|_{\mathbf{x}_k^a, \mathbf{p}_k^a}. \tag{45}$$

The elements of (45) can be computed directly from the discrete equations (41)–(42). Details of this calculation are given in [38].

4.2.2. Experiments. We define the reference solution to be that given by the discretised equations (41)–(42) with model time step $\Delta t = 0.1$, initial displacement $x_0 = 2.0$, initial velocity $y_0 = 0.0$ and parameter values $d = 0.05$ and $m = 1.0$. It is assumed that we know the values of d, m are positive. The evolution of x and y is shown in figure 4 for $t \in [0, 50]$. The initial background estimate for the state $\mathbf{x}_0^b$ is generated by adding random noise to the initial conditions of the reference solution. This noise is taken from a Gaussian distribution with zero mean and variance $\sigma_b^2 = 0.01$. The state background error covariance matrix is assumed to be a diagonal matrix, $\mathbf{P}^f_{\mathbf{xx}} = \sigma_b^2 \mathbf{I} \in \mathbb{R}^{2 \times 2}$. The error variance of parameters d and m are set at $\sigma_d^2 = 0.005$ and $\sigma_m^2 = 0.1$ respectively. Since the dimension of the augmented state is small in this case, and the observation operator is linear, we compute the analysis directly from the BLUE equation (11).
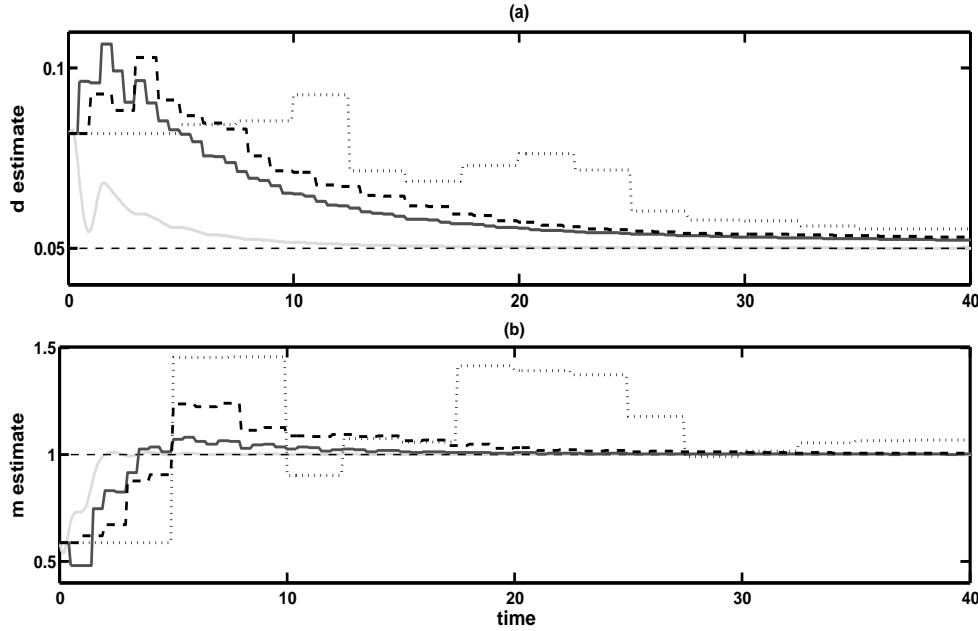
FIG. 5. *Nonlinear oscillator:* **(a)** *Parameter d updates for initial estimate* $d_0 = 0.081877$, **(b)** *Parameter m updates for initial estimate* $m_0 = 0.58617$: *solid light grey line - observations every* Δt; *solid dark grey line - observations every* $5\Delta t$; *dashed line - observations every* $10\Delta t$; *dotted line - observations every* $25\Delta t$. *The horizontal dashed line indicates the exact parameter value.*

4.2.3. Results. *Perfect observations.* Figures 5(a) and (b) show the parameter d and m updates for an example model run with initial parameter estimates $d_0 = 0.081877$ (over estimated), $m_0 = 0.58617$ (under estimated) and observations of x and y at varying temporal frequencies. The scheme manages to retrieve the values of both d and m to a good level of accuracy for observation intervals up to every $25\Delta t$. We see a large increase in error in the estimated parameter values when the observation frequency is decreased to $50\Delta t$ (not shown). The x and y analyses are not shown; for observation intervals of less than $25\Delta t$ they are qualitatively indistinguishable from the reference solution. When observations are taken every $25\Delta t$ we start to see very small differences, and when the observation frequency is halved to $50\Delta t$ there is a marked deterioration in the analysis, with the over estimation of both d and m causing significant phase and amplitude errors. In this case, too much weight is being given to the background state which suggests that a background error variance value of $\sigma_b^2 = 0.01$ is too low. Indeed, we would expect the errors in the model forecast to grow as the time between successive state and parameter updates is increased. We found that both the state and parameter estimates could be improved by inflating the background error variance. Further examples and discussion can be found in [41] and [38].

Noisy observations. Figures 6(a) and (b) show the results produced when random noise with variance $\sigma_o^2 = 0.01$ was added to the observations. Note that the length of the time window has been increased to 80 time units for these experiments. Overall the scheme performs well, although there is some curious behaviour. Unlike the perfect observation case, there is no clear relationship between the observation frequency and the accuracy of the parameter estimates; the best estimates of d and m were obtained

using different observation intervals. There was also variation across model runs. A notable result is the estimates of parameter d when observations are available at every timestep (solid grey line in figure 6(a)). The estimates initially appear to be moving towards the correct value but at around 40 timesteps they begin to increase away. The experiment was repeated with different noise simulations and different starting values for d but similar behaviour was found in every case. It is possible that the interval between updates is insufficient for the model to adjust to the new value of d before the next input of data. A further hypothesis is that this behaviour is related to the role of d in the model equations. The parameter d determines how quickly the solution becomes damped. As we move forward in time the amplitude of the solution decreases, the relative size of the observational noise therefore increases causing greater misrepresentation of the true amplitude and making it harder to identify the exact value of d.

We found that this behaviour could be remedied by averaging the estimates as is illustrated in figures 7(a) and (b). The parameter estimates were averaged over a moving time window of 50 time steps starting at $t = 30$. This produces more stable estimates for the parameters which in turn gives greater stability to the forecast model. In this example, the value prescribed for σ_o^2 is relatively small and so the observational noise has very little impact on the overall quality of the state analysis. We see a much greater effect on the state analysis if the observation error variance is increased to $\sigma_o^2 = 0.1$ but we are still able to estimate the parameters to a reasonable level of accuracy for observation intervals up to $10\Delta t$.

4.3. The Lorenz 63 equations. The Lorenz equations is the name given to a system of first order differential equations describing a simple nonlinear dynamical system that exhibits chaotic behaviour. The system was originally derived from a model of fluid convection and consists of the three coupled, nonlinear ordinary differential equations [24]

$$\dot{x} = -s(x - y), \quad \dot{y} = \rho x - y - xz, \quad \dot{z} = xy - \beta z. \tag{46}$$

where $x = x(t)$, $y = y(t)$ and $z = z(t)$ and s, ρ and β are real, positive parameters.

The strong nonlinearity of these equations means that the model solution is extremely sensitive to perturbations in the initial conditions and parameters; they are often used as a framework for examining the properties of state estimation methods applied to highly nonlinear dynamical systems [9], [28]. The origin is a stationary point for all parameter values. When $\rho > 1$ there are two other stationary points

$$\left(\pm\sqrt{\beta(\rho - 1)}, \pm\sqrt{\beta(\rho - 1)}, \rho - 1 \right).$$

For these experiments we set the reference parameters as $s = 10$, $\rho = 28$ and $\beta = 8/3$. These are the classic values first used by Lorenz. At these values all three equilibrium points are unstable and give rise to chaotic solutions [45].

To investigate the applicability of our new method to this system we adapt a pre-existing Matlab routine written by M.J Martin [27], [25]. The code was developed as a data assimilation training tool to illustrate sequential state estimation in simplified models; a copy of the original, unmodified code can be obtained from [26]. Equations (46) are solved numerically using the same second order Runge–Kutta scheme as in
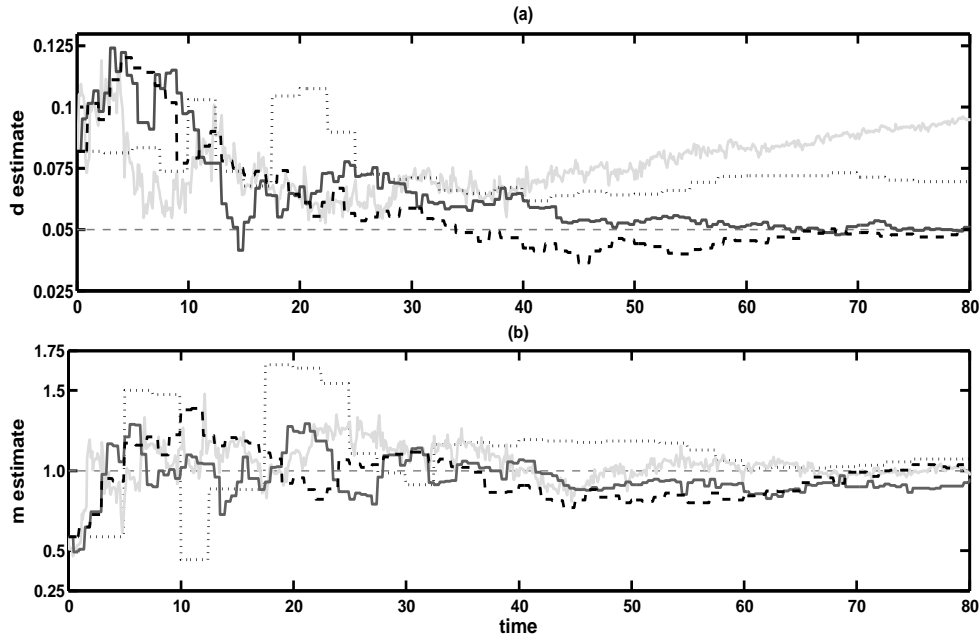
FIG. 6. *Nonlinear oscillator: imperfect observations with observation error variance $\sigma_o^2 = 0.01$. Parameter updates for initial estimate (a) $d_0 = 0.081877$ and (b) $m_0 = 0.58617$. Lines as figure 5.*
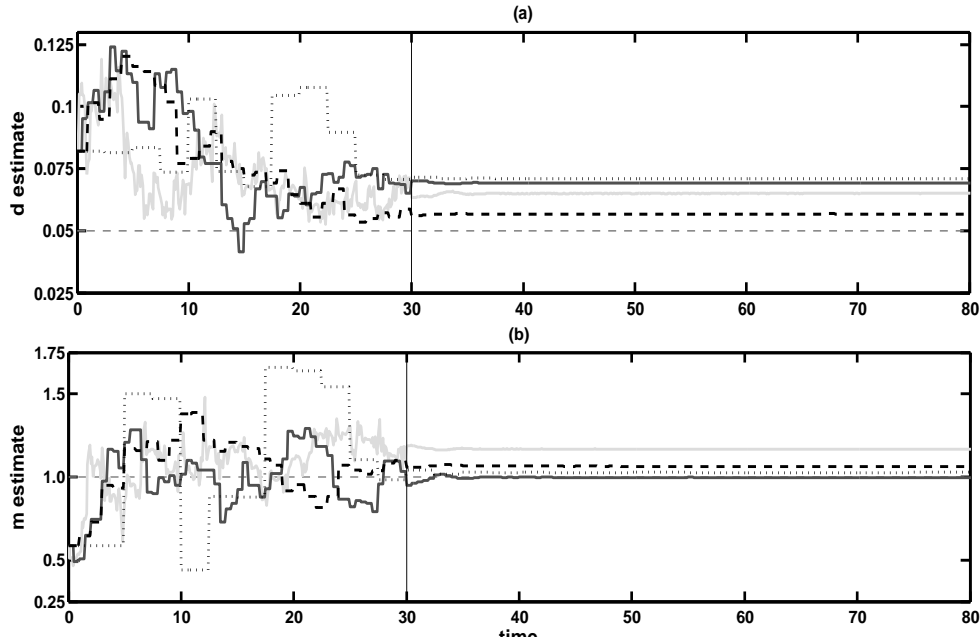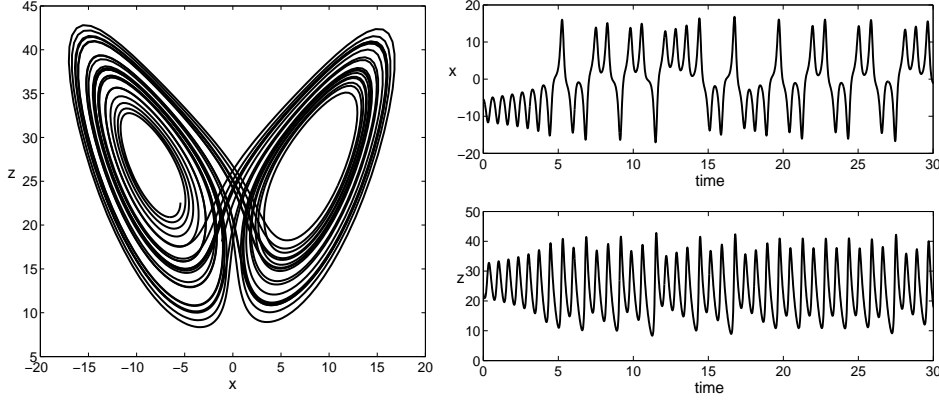

FIG. 7. *Nonlinear oscillator: imperfect observations with observation error variance $\sigma_o^2 = 0.01$. Time averaged parameter updates for initial estimate (a) $d_0 = 0.081877$ and (b) $m_0 = 0.58617$. Lines as figure 5. Vertical line indicates the start of time averaging.*

FIG. 8. *Lorenz equations: reference solution for x and z.*

§4.2. The discrete system is given by

$$x_{k+1} = x_k + s\frac{\Delta t}{2}\Big[2(y_k - x_k) + \Delta t(\rho x_k - y_k - x_k z_k) - s\Delta t(y_k - x_k)\Big], \tag{47}$$

$$y_{k+1} = y_k + \frac{\Delta t}{2}\Big[\rho x_k - y_k - x_k z_k + \rho(x_k + s\Delta t(y_k - x_k)) - y_k - \Delta t(\rho x_k - y_k - x_k z_k)$$
$$- (x_k + s\Delta t(y_k - x_k))(z_k + \Delta t(x_k y_k - \beta z_k))\Big] \tag{48}$$

$$z_{k+1} = z_k + \frac{\Delta t}{2}\Big[x_k y_k - \beta z_k + (x_k + \Delta t s(y_k - x_k))(y_k + \Delta t(\rho x_k - y_k - x_k z_k))$$
$$- \beta(z_k + \Delta t(x_k y_k - \beta z_k))\Big] \qquad\qquad k = 0, 1, \dots \tag{49}$$

where $x_k \approx x(t_k)$, $y_k \approx y(t_k)$ and $z_k \approx z(t_k)$, $t_k = k\Delta t$.

Combining the state and parameter vectors $\mathbf{x}_k = (x_k, \, y_k, \, z_k)^T$ and $\mathbf{p}_k = (s_k, \, \rho_k, \, \beta_k)^T$ gives the augmented state vector

$$\mathbf{w}_k = (x_k, \, y_k, \, z_k, \, s_k, \, \rho_k, \, \beta_k)^T. \tag{50}$$

Our augmented system model (5),(6) is then given by the state evolution model (47)–(49) together with the parameter model (3).

4.3.1. State-parameter cross-covariance. We assume that the parameters s, ρ and β are uncorrelated and set $\mathbf{P}_{\mathbf{pp}}^f = \mathrm{diag}(\sigma_s^2, \, \sigma_\rho^2, \, \sigma_\beta^2)$, where σ_s^2, σ_ρ^2 and σ_β^2 are the error variances of parameters s, ρ and β respectively. The Jacobian matrix $\mathbf{N}_k$ can be computed directly by differentiating the discrete equations (47)–(49) with respect to each parameter as described in [38]. The state-parameter cross-covariance matrix at t_{k+1} is then given by

$$\mathbf{P}_{\mathbf{xp}k+1} = \left(\begin{array}{ccc} \sigma_s^2 \dfrac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k)}{\partial s} & \sigma_\rho^2 \dfrac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k)}{\partial \rho} & \sigma_\beta^2 \dfrac{\partial \mathbf{f}(\mathbf{x}_k, \mathbf{p}_k)}{\partial \beta} \end{array} \right) \Bigg|_{\mathbf{x}_k^a, \mathbf{p}_k^a}. \tag{51}$$

4.3.2. Experiments. The reference solution is taken to be that given by the discrete equations (47)–(49) with model timestep $\Delta t = 0.01$ and initial conditions

$x_0 = -5.4458$, $y_0 = -5.4841$ and $z_0 = 22.5606$. The solutions for x and z are illustrated in figure 8 for $t \in [0, 30]$. The initial model background state vector $\mathbf{x}_0^b$ is generated by adding Gaussian random noise with zero mean and variance 0.1 to the reference state at t_0. The state background error covariance matrix is given by $\mathbf{P}_{\mathbf{xx}}^f = \sigma_b^2 \mathbf{I} \in \mathbb{R}^{3 \times 3}$ with error variance $\sigma_b^2 = 1.0$. The error variances of the parameters are set equal to 20% of their reference value. As in §4.2, we use the BLUE equation (11) to compute the analysis directly.

4.3.3. Results. *Perfect observations.* Once again we find that our method performs extremely well. Figure 9 shows the parameter updates for model runs with observations of x, y and z at decreasing temporal frequency. In this example, the initial parameter values are $s_0 = 11.0311$, $\rho_0 = 30.1316$ and $\beta_0 = 1.6986$. For observation frequencies $5\Delta t$, $10\Delta t$ and $20\Delta t$ the estimates of ρ and β converge to their exact values very rapidly. The updating of s is much slower but the correct value is eventually recovered to an accuracy of 3 decimal places. There are no distinguishable differences between the model and reference solution for x, y and z at these frequencies and so we do not show the results here. When the observation interval is increased to $30\Delta t$ (not shown due to scale) the scheme takes longer to stabilise; there are large initial deviations in the estimated parameters but all three do eventually converge to their reference value. This in turn effects the updating of the state variables; initially there are clear differences between the predicted and reference solutions. However, once the model has stabilised the reference solution is reproduced almost perfectly (see [38]). If the time period between updates is further increased to $40\Delta t$ the scheme completely fails to find the correct parameter values and the model state analysis is poor across the entire time window.

Noisy observations. The experiments were re-run using noisy observations with observation error variance ranging from $\sigma_o^2 = 0.01$ to $\sigma_o^2 = 0.25$. Figure 10 shows the parameter updates obtained with observation error variance $\sigma_o^2 = 0.1$. When $\sigma_o^2 = 0.01$ the convergence and quality of the parameter and state estimates is very similar to the perfect observation case. With $\sigma_o^2 = 0.1$ and $\sigma_o^2 = 0.25$ the parameter estimates are very noisy but the oscillations are centered around their reference values. The size of the parameter errors increase as σ_o^2 increases and also as the frequency of the observations decreases. When $\sigma_o^2 = 0.1$ there is a significant growth in the size of the oscillations when the time between updates is increased to $30\Delta t$. Again these are not shown due to scale. If we try to extend the observation interval any further, the model fails to produce any meaningful results.

Figure 11 shows the effect of averaging the parameter estimates over a moving time window of 50 time steps, starting at $t = 10$. With $\sigma_o^2 = 0.1$ the parameters are predicted to a good level of accuracy for observation intervals up to $20\Delta t$. With $\sigma_o^2 = 0.25$ (not shown) we get similarly good results for ρ and β, but the parameter s estimates are less accurate for observation intervals greater than $5\Delta t$. In the perfect observation experiments, we found that this parameter converged much slower than the other two. This perhaps suggests that the model is relatively insensitive to small deviations in its value. The averaged s estimate could potentially be improved by starting the averaging at a later time.

For observation intervals of $5\Delta t$ to $20\Delta t$, the quality of the analysis for the model state variables is good for both $\sigma_o^2 = 0.1$ and $\sigma_o^2 = 0.25$ even without the averaging of the parameter estimates. There is a noticeable deterioration in the analysis when the observation frequency is decreased to $30\Delta t$, particularly in the $\sigma_o^2 = 0.25$ case. When $\sigma_o^2 = 0.1$, we found that the state analysis improved when the model was re-run using
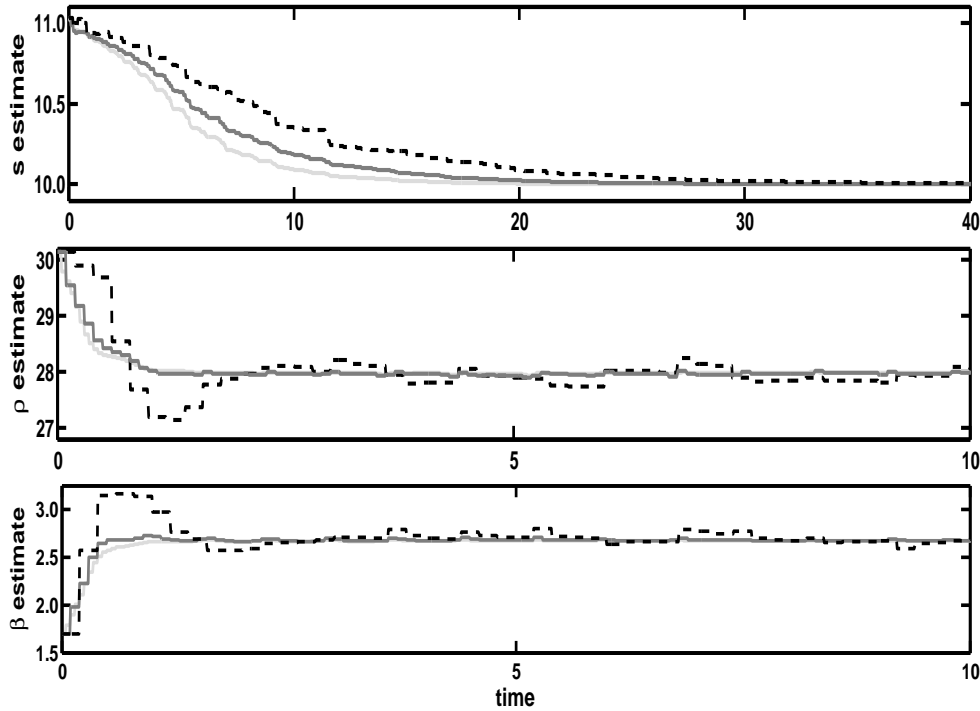
Fig. 9. *Lorenz equations: Parameter updates for initial estimates $s_0 = 11.0311$, $\rho_0 = 30.1316$ and $\beta_0 = 1.6986$. Solid light grey line - observations every $5\Delta t$; solid dark grey line - observations every $10\Delta t$; dashed line - observations every $20\Delta t$. Note the different timescale for parameter s.*

the time averaging of the parameter estimates. The same approach failed to produce any substantial improvement for the case $\sigma_o^2 = 0.25$.

5. Conclusions. A new method for concurrent model parameter and state estimation, employing a joint sequential estimation algorithm, has been proposed and demonstrated via a series of identical twin experiments with three simple numerical models. The approach combines ideas from 3D-Var and the extended Kalman filter methods to produce a flow dependent approximation of the state-parameter cross-covariances whilst avoiding the computational complexities associated with implementation of the full Kalman filter equations. This allows us to use the state augmentation technique with the 3D-Var algorithm which traditionally makes the assumption of a static background error covariance matrix.

In this paper we have presented details of this new methodology and illustrated its versatility by applying it to a range of simple dynamical system models in which the use of incorrect parameters has a direct impact on the model solution. Although each system has different characteristics the technique performed well in all three cases. As the results show, the scheme was successful in recovering the reference parameter values we had specified to a good level of accuracy, even when the observational data were noisy. This had a positive impact on the forecast model and enabled more accurate estimation of the reference model state.

As we would expect, there are limits to the success of the method; when observational data become too infrequent or too noisy, or if the initial state and parameter
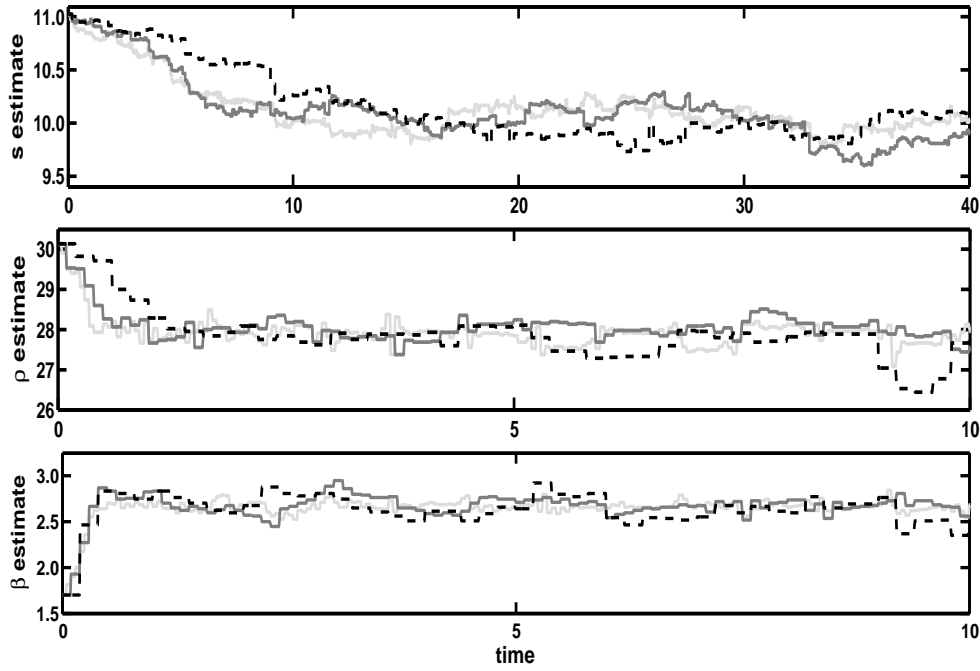
FIG. 10. *Lorenz equations: imperfect observations with observation error variance $\sigma_o^2 = 0.1$. Parameter updates for initial estimates $\sigma_0 = 11.0311$, $\rho_0 = 30.1316$ and $\beta_0 = 1.6986$. Lines as figure 9. Note the different timescale for parameter s.*
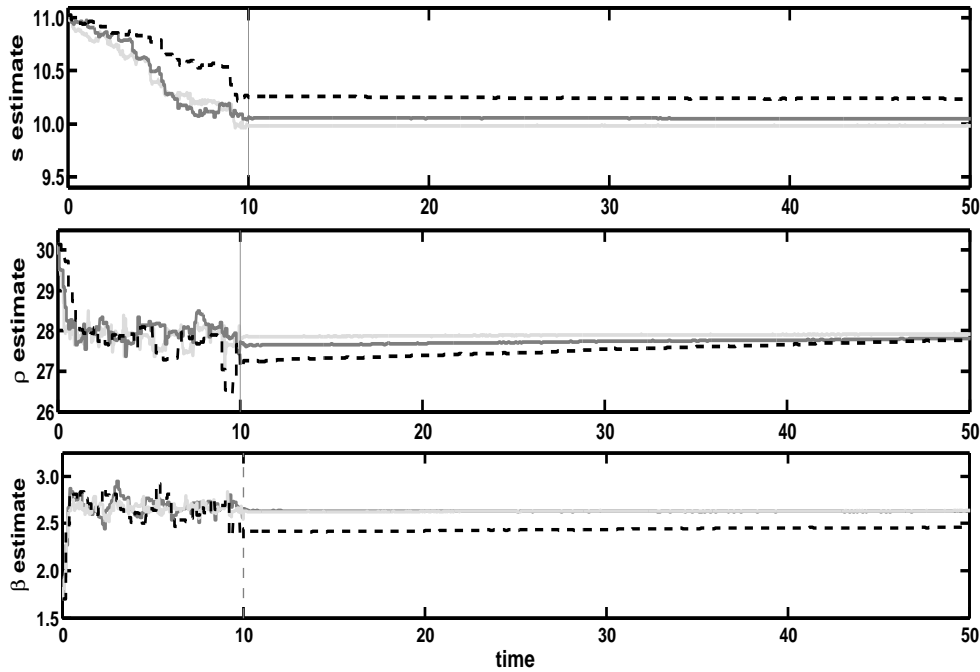

FIG. 11. *As figure 10 but showing time averaged parameter updates. Vertical line indicates the start of time averaging.*

background estimates are particularly poor then we are unable to yield reliable results. The threshold for each model varied depending on properties of the model structure and the underlying dynamics, but was not overly restrictive.

The scheme is inevitably less successful in situations where the model is relatively insensitive to a particular parameter, as was the case for certain settings in the nonlinear oscillating system. This is not surprising as we cannot expect to be able to correct parameters that cause errors in the model solution that are on smaller scales than can be reliably observed. Other parameter estimation techniques would also be likely to fail in such a scenario. This is linked to the concepts of observability and identifiability [2], [30]; whether the available observations contain sufficient information for us to be able to determine the parameters of interest and whether these parameters have a unique deterministic set of values. A method can only be expected to work reliably when both these properties hold. Future work will consider these issues in more depth and examine how they formally relate to our new algorithm.

For models with more than one parameter, consideration must be given to the relationship between individual parameters. In this work we assumed that the parameters in the oscillating and Lorenz models were uncorrelated and set the cross-covariances between the parameters equal to zero. Whilst this assumption worked for these particular models it may not adequate for models in which the parameters exhibit strong correlation. A model sensitivity analysis can be used to help identify the interdependence of parameters and ascertain whether cross-correlations are needed. In this case, more attention will need to be given to the parameter error covariance matrix and methods for defining the cross-correlations will need to be considered [43]. In some situations, it may be prudent to consider a re-parameterisation of the model equations to improve the identifiability of the parameters or even to transform the parameters to a set of uncorrelated variables [44].

To date, our new technique has only been tested in models of relatively low dimension, where the number of parameters is small and, since the required parameters are constants, the dynamics of the parameter model are simple. The increase in the dimension of the problem caused by the addition of the parameters to the state vector does not have a significant impact on the computational cost of the estimation scheme and the re-calculation of the matrix $\mathbf{N}_k$ at each new observation time is not infeasible. Here we chose model discretisations that allowed us to obtain explicit expressions for the matrix $\mathbf{N}_k$ thereby avoiding any additional computational complexity. However, an explicit computational form for the Jacobian is not necessarily required; it can, for example, be approximated using a simple local finite difference approach, as demonstrated in [43], [42], [41]. A further option is automatic differentiation [33].

Important advantages of our new approach are that the background error covariance matrix only needs to be updated at each new analysis time rather than at every time step and it does not require the previous cross-covariance matrices to be stored. It also avoids many of the potential problems associated with implementation of fully flow dependent algorithms like the EKF and ensemble Kalman filters such as filter divergence, rank deficiency [17], spurious correlations and imbalance [18].

This study has provided a valuable insight into how our new method is likely to perform in a range of dynamical systems. The scheme has proven to be largely successful and we believe that it offers an efficient and versatile solution to the problem of sequential joint state-parameter estimation across a variety of applications.

REFERENCES

[1] R. N. BANNISTER, *A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances*, Quarterly Journal of Royal Meteorological Society, 134 (2008), pp. 1955–1970.

[2] S. BARNETT AND R. G. CAMERON, *Introduction to Mathematical Control Theory*, Oxford Applied Mathematics and Computing Science Series, Oxford Clarendon Press, second ed., 1990.

[3] M. J. BELL, M. J. MARTIN, AND N. K. NICHOLS, *Assimilation of data into an ocean model with systematic errors near the equator*, Quarterly Journal of the Royal Meteorological Society, 130 (2004), pp. 873–894.

[4] R. L. BURDEN AND D. J. FAIRES, *Numerical Analysis*, Brooks-Cole Publishing Company, 6th ed., 1997.

[5] P. COURTIER, E. ANDERSSON, W. HECKLEY, J. PAILLEUX, D. VASILJEVIC, M. HAMRUD, A. HOLLINGSWORTH, F. RABIER, AND M. FISHER, *The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation*, Quarterly Journal of the Royal Meteorological Society, 124 (1998), pp. 1783–1807.

[6] R. DALEY, *Atmospheric Data Analysis*, Cambridge University Press, 1991.

[7] D. P. DEE, *Bias and data assimilation*, Quarterly Journal of the Royal Meteorological Society, 131 (2005), pp. 3323–2243.

[8] D. R. DURRAN, *Numerical Methods for Wave Equations in Geophysical Fluid Dynamics*, Texts in Applied Mathematics 32, Springer, 1999.

[9] G. EVENSEN, *Advanced data assimilation for strongly nonlinear dynamics*, Monthly Weather Review, 125 (1997), pp. 1342–1354.

[10] M. FISHER, *Background error covariance modelling*, in Seminar on recent developments in data assimilation for atmosphere and ocean, ECMWF, 2003, pp. 49–64.

[11] J. GARCÍA-PINTADO, J. C. NEAL, D. C. MASON, S. L. DANCE, AND PA. D. BATES, *Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling*, Journal of Hydrology, 495 (2013), pp. 252–266.

[12] A. GELB, *Applied Optimal Estimation*, M.I.T Press, 1974.

[13] P. E. GILL, W. MURRAY, AND M. H. WRIGHT, *Practical Optimization*, Academic Press, 1981.

[14] A. K. GRIFFITH AND N. K. NICHOLS, *Adjoint techniques in data assimilation for treating systematic model error*, Journal of Flow, Turbulence and Combustion, 65 (2000), pp. 469–488.

[15] J. GUCKENHEIMER AND P. HOLMES, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Applied Mathematical Sciences 42, Springer-Verlag, 1986.

[16] D. C. HILL, S. E. JONES, AND D. PRANDLE, *Derivation of sediment resuspension rates from accoustic backscatter time-series in tidal waters*, Continental Shelf Research, 23 (2003), pp. 19–40.

[17] P. L. HOUTEKAMER AND H. L. MITCHELL, *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Review, 126 (1998), pp. 796–811.

[18] ———, *Ensemble Kalman filtering*, Quarterly Journal of the Royal Meteorological Society, 131 (2005), pp. 3269–3289.

[19] A. H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[20] M. A. F. KNAAPEN AND S. J. M. H. HULSCHER, *Use of genetic algorithm to improve prediction of alternate bar dynamics*, Water Resources Research, 39 (2003), p. 1231.

[21] RANDALL J. LEVEQUE, *Numerical Methods for Conservation Laws*, Birkhäuser-Verlag, 1992.

[22] JOHN M. LEWIS, S. LAKSHMIVARAHAN, AND S. K. DHALL, *Dynamic Data Assimilation: A Least Squares Approach*, vol. 104 of Encyclopedia of Mathematics and its applications, Cambridge University Press, 2006.

[23] A. C. LORENC, S. P. BALLARD, R. S. BELL, N. B. INGLEBY, P. L. F. ANDREWS, DvM. BARKER, J. R. BRAY, A. M. CLAYTON, T. DALBY, D. LI, T. J. PAYNE, AND F. W. SAUNDERS, *The Met Office global three-dimensional variational data assimilation scheme*, Quarterly Journal of the Royal Meteorological Society, 126 (2000), pp. 2991–3012.

[24] E. N. LORENZ, *Deterministic nonperiodic flow*, Journal of the Atmospheric Sciences, 20 (1963), pp. 130–141.

[25] M. J. MARTIN, *Data Assimilation in Ocean Circulation Models with Systematic Errors*, PhD thesis, University of Reading, 2000.

[26] M. J MARTIN AND A. S LAWLESS, *Oscillating system with sequential data assimilation scheme*, NERC National Centre for Earth Observation. Available at http://www.nceo.ac.uk/training.php. (last accessed June 2010).

[27] M. J. MARTIN, N. K. NICHOLS, AND M. J. BELL, *Treatment of systematic errors in sequential data assimilation*, Technical Note No. 21, Meteorological Office, Ocean Applications Division, 1999.

[28] R. N. MILLER, M. GHIL, AND F. GAUTHIEZ, *Advanced data assimilation in strongly nonlinear dynamical systems*, Journal of the Atmospheric Sciences, 51 (1994), pp. 1037–1056.

[29] K. W. MORTON AND D. F. MAYERS, *Numerical solution of partial differential equations*, Cambridge University Press, 2nd ed., 2005.

[30] I. M. NAVON, *Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography*, Dynamics of Atmosphere and Oceans, 27 (1997), pp. 55–79.

[31] N. K. NICHOLS, *Data assimilation: Aims and basic concepts*, in Data Assimilation for the Earth System, R. Swinbank, V. Shutyaev, and W A. Lahoz, eds., vol. 26 of Nato Science Series IV: Earth & Environmental Sciences, Kluwer Academic, 2003, pp. 9–20.

[32] ———, *Mathematical concepts of data assimilation*, in Data Assimilation: Making Sense of Observations, W A. Lahoz, R. Swinbank, and B. Khattatov, eds., Springer, 2009.

[33] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.

[34] C. D. RODGERS, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, 2000.

[35] B. G. RUESSINK, *Calibration of nearshore process models - application of a hybrid genetic algorithm*, Journal of Hydroinformatics, 7 (2005), pp. 135–149.

[36] ———, *Predictive uncertainty of a nearshore bed evolution model*, Continental Shelf Research, 25 (2005), pp. 1053–1069.

[37] ———, *A Bayesian estimation of parameter-induced uncertainty in a nearshore alongshore current model*, Journal of Hydroinformatics, 7 (2006), pp. 37–49.

[38] P. J. SMITH, *Joint state and parameter estimation using data assimilation with application to morphodynamic modelling*, PhD thesis, University of Reading, 2010.

[39] P. J SMITH, M. J. BAINES, S. L. DANCE, N. K NICHOLS, AND T. R. SCOTT, *Data assimilation for parameter estimation with application to a simple morphodynamic model*, Mathematics Report 2/2008, Department of Mathematics, University of Reading, 2008. Available at `http://www.reading.ac.uk/maths/research/`.

[40] ———, *Variational data assimilation for parameter estimation: application to a simple morphodynamic model*, Ocean Dynamics, 59 (2009), pp. 697–708.

[41] P. J. SMITH, S. L. DANCE, AND N. K. NICHOLS, *A hybrid sequential data assimilation scheme for model state and parameter estimation*, Mathematics Report 2/2010, Department of Mathematics, University of Reading, 2010. Available at `http://www.reading.ac.uk/maths/research/`.

[42] ———, *A hybrid data assimilation scheme for model parameter estimation: application to morphodynamic modelling*, Computers and Fluids, 46 (2011), pp. 436–441.

[43] P. J. SMITH, G. D. THORNHILL, S. L. DANCE, A. S. LAWLESS, D. C. MASON, AND N. K. NICHOLS, *Data assimilation for state and parameter estimation: application to morphodynamic modelling*, Quarterly Journal of the Royal Meteorological Society, 139 (2013), pp. 314–327.

[44] S. SOROOSHIAN AND V. K. GUPTA, *Model calibration*, in Computer models of watershed hydrology, V J Singh, ed., Water Resources Publications, Colorado, 1995, ch. 2, pp. 23–68.

[45] C. SPARROW, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, Applied Mathematical Sciences 41, Springer-Verlag, 1982.

[46] J. M. T. THOMPSON AND H. B. STEWART, *Nonlinear Dynamics and Chaos: Geometrical Methods for Engineers and Scientists*, John Wiley and Sons Ltd., 1986.

[47] C. M. TRUDINGER, M. R. RAUPACH, P. J. RAYNER, AND I. G. ENTING, *Using the Kalman filter for parameter estimation in biogeochemical models*, Environmetrics, 19 (2008), pp. 849–870.

[48] C. M. TRUDINGER, M. R. RAUPACH, P. J. RAYNER, J. KATTGE, Q. LIU, B. PAK, M. REICHSTEIN, L. RENZULLO, A. D. RICHARDSON, S. H. ROXBURGH, J. STYLES, Y. P. WANG, P. BRIGGS, D. BARRETT, AND S. NIKOLOVA, *OptIC: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models*, Journal of Geophysical Research, 112 (2007).

[49] G. B. WHITHAM, *Linear and Nonlinear Waves*, John Wiley & Sons, 1974.

[50] S. WIGGINS, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Texts in Applied Mathematics 2, Springer-Verlag, 1990.

[51] J. C. WÜST, *Data-driven probablistic predictions of sand wave bathymetry*, in Marine Sandwave and River Dune Dynamics II, Hulsher S J M H., T. Garlan, and D. Idier, eds., 2004. Proceedings of International Workshop, University of Twente.