School of Mathematical

Department of Mathematics and Statistics

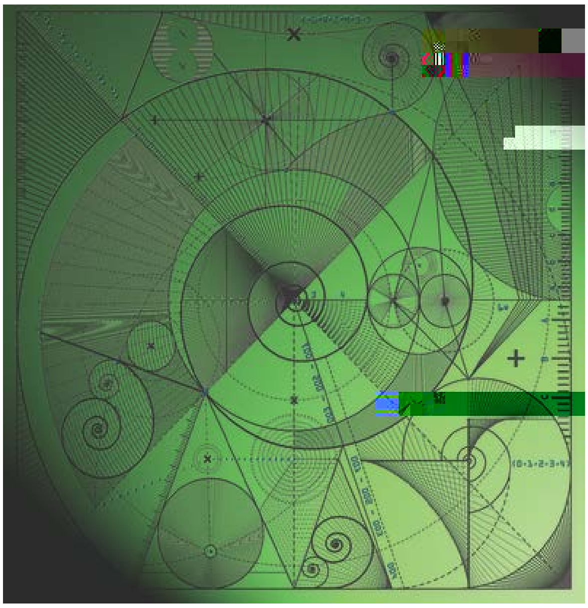
Preprint MPS-2014-26

10 December 2014

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 inexpainve-371.105(t)6.4802(o).356.309(r)-7.5753(u)2.80589(n-343.4397fr)-7238345(o)4.03117rolarge-3413.197(s)-2.57462(y)2.9177(s)-2.57462(t)6.48
[illegible]
seme-371.105(i)-6.12487(s)-24..596(a)4.03117(b)2.80589(l)-6.12487(e)35261035(t)6.4802(o).356.309(r)-7.5753(e)5.25492(c)5.25645(o346.1214(v)33.0
ue to be easily transferable to much larger models.

neter estimation, variational data assimilation, filtering,

are intrinsic to numerical modelling. Parameteri- f representing processes that are not fully known o computer power constrain the model resolution can be described. Numerical models will often onents derived from practical experience rather of this is that model parameters often do not ectly measurable quantity. Whilst these approx- uncertainties in the model parameters can lead dicted and actual states of the system.

is addressed as a separate issue to state esti- formed o-line in a separate calculation. Model ed theoretically or by adhoc calibration of the More recently, improvements in computational nt of many novel, and often complex, automated Generally, these methods involve data fitting function [16], [20], [35]. The main distinctions how the minimum is located, how the observed tions made about the error statistics. A useful tion techniques for parameter estimation applied s is given in [48]. Another approach is to use tainty in the parameter estimate[illegible]

CASE (Co-operative Awards in Science and Engineering) scheme and the for Earth Observation (NCEO).

l and Physical Sciences, University of Reading, UK
email: p.j.smith@reading.ac.uk

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.

where the matrix $\mathbf{M}_k \in \mathsf{R}^{n\times n}$ 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 where $\tilde{\mathbf{h}}_k : \mathsf{R}^{n+q} \longrightarrow \mathsf{R}^{r_k}$ 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 $\mathbf{w}_k^b$

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}

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

$$P_k = \begin{pmatrix} P_{xx_k} & P_{xp_k} \\ (P_{xp_k})^T & P_{pp_k} \end{pmatrix} \tag{16}$$

with the superscript f or a used to indicate forecast or analysis. Here $P_{xx_k} \in \mathbb{R}^{n \times n}$ is the forecast (analysis) error covariance matrix for the model state vector x_k at time t_k, $P_{pp_k} \in \mathbb{R}^{q \times q}$ is the covariance matrix describing the errors in the parameter vector p_k and $P_{xp_k} \in \mathbb{R}^{n \times q}$ 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

$$M_k = \left.\frac{\partial f(x;p)}{\partial x}\right|_{x_k^a;p_k^a} \quad \text{and} \quad N_k = \left.\frac{\partial f(x;p)}{\partial p}\right|_{x_k^a;p_k^a}; \tag{17}$$

where $M_k \in \mathbb{R}^{n \times n}$ and $N_k \in \mathbb{R}^{n \times q}$, and substitute into (15),(14) we obtain the following expressions for the blocks of P^f_{k+1}

$$P^f_{xx_{k+1}} = M_k P^a_{xx_k} M_k^T + N_k (P^a_{xp_k})^T M_k^T + M_k P^a_{xp_k} N_k^T + N_k P^a_{pp_k} N_k^T; \tag{18}$$

$$P^f_{xp_{k+1}} = M_k P^a_{xp_k} + N_k P^a_{pp_k}; \tag{19}$$

$$P^f_{pp_{k+1}} = P^a_{pp_k}. \tag{20}$$

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

$$P^f_{xx_k} = P^f_{xx} \quad \text{for all } k. \tag{21}$$

The choice for P^f_{xx} will depend on the particular model application; a simple and commonly adopted approach is to define P^f_{xx} 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

$$P^f_{pp_k} = P^f_{pp} \quad \text{for all } k. \tag{22}$$

Specification of P^f_{pp} requires some *a priori* knowledge of the parameter error statistics. The error variance of each parameter should reflect our unc 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-

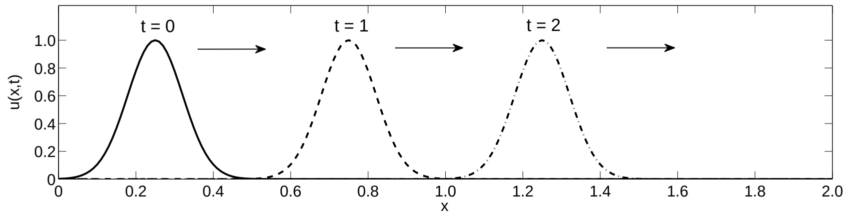


Fig. 1. Solutions $u(x;t)$ to the linear advection equation (27) for Gaussian initial data at model times $t = 0; 1; 2$:

assume that the observation errors are spatially and temporally uncorrelated and set the observation error covariance matrix R_k

$$R_k = R = \sigma_o^2 I; \qquad I \in \mathbb{R}^{r \times r}: \tag{26}$$

stable we set $\mu = 1$ 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 N_k.

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

$$x_{k+1} = M x_k; \tag{30}$$

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

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

Setting $w_k = (x_k, c_k)^T$, we combine (30) with the parameter evolution model (3) to give the augmented system model

$$w_{k+1} = \begin{pmatrix} M_k & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_k \\ c_k \end{pmatrix} \tag{32}$$

Note that the constant matrix M in (30) has been replaced by the time varying matrix $M_k = M(c_k)$. Although the exact system matrix 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 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 P^f_{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

$$P^f_{xp_{k+1}} = \sigma_c^2 N_k. \tag{33}$$

For the linear advection model, the matrix N_k is defined as

$$N_k = \left. \frac{\partial (M_k x_k)}{\partial c} \right|_{x_k^a, c_k^a}; \tag{34}$$

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

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

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 $\mu = 1$. The initial state of the reference solution is given by the Gaussian function

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

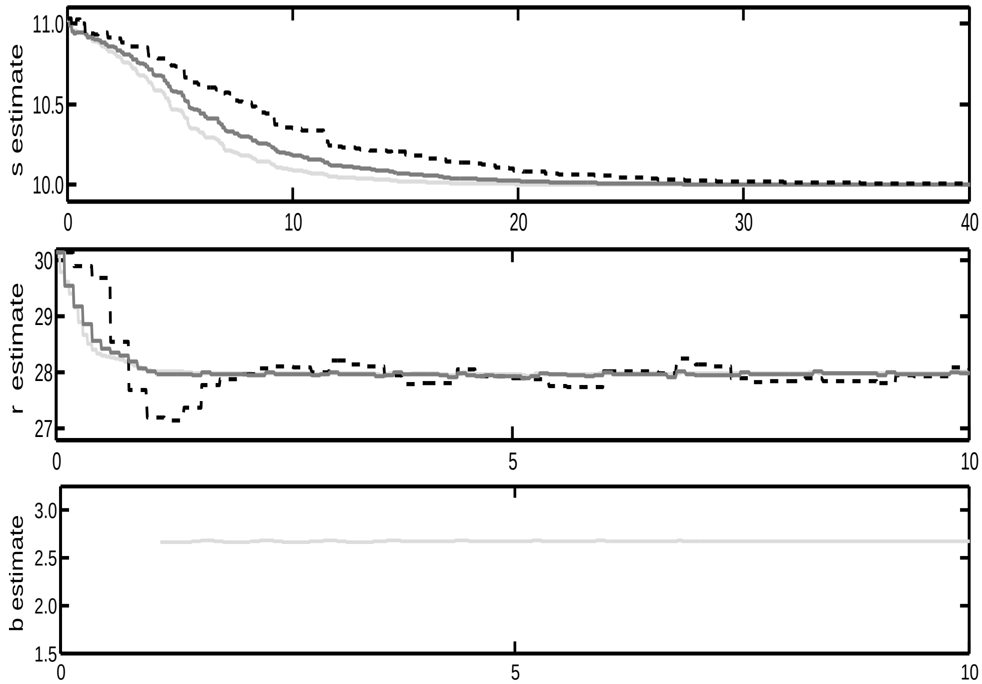
the case 25 t. When observations are taken every 50 t the estimates for c take longer to converge and as a result the model takes longer to stabilise. Once the model has settled the analysis for u

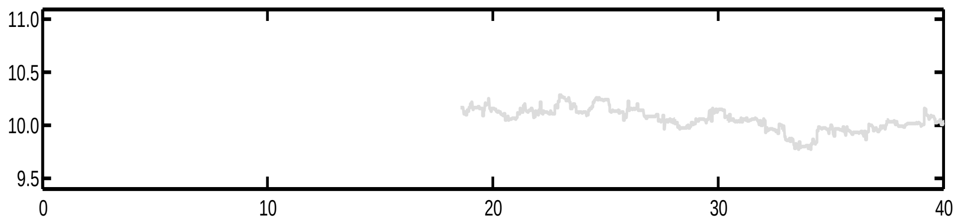
using di erent observation intervals. There was also variation across model runs. A notable result is the estimates of parameter d when observations are available at every timestep (solid grey line in gure 6(a)). The estimates initially appear to be moving towards the correct value but at around 40 timesteps they begin to increase away. The experiment was repeated with di erent noise simulations and di erent starting values for d but similar behaviour was found in every case. It is possible that the interval between updates is insu cient 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 gures 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 ^2_o is relatively small and so the
oc689988(is)-461(io)-5.88993(n)-324.294(b)-5(e)-1.666343(n-5.89054(t55839(i45-1.66516(s)-485.301(f)-449.159(a)]TJ 31

$x_0 = \; 5{:}4458$, $y_0 = \; 5{:}4841$ and $z_0 = 22{:}5606$. The solutions for x and z are illustrated in gure 8 for $t \; 2 \; [0; 30]$. The initial model background state vector 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 $P^f_{xx} = \; {}^2_b I \; 2 \; R^{3 \;\; 3}$ with error variance $\; {}^2_b = 1{:}0$. 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.


s estimate
11.0
10.5
10.0
0
10
20
30
40
r estimate
30
29
28
27
0
5
10
b estimate
3.0
2.5
2.0
1.5
0
5
10


11.0
10.5
10.0
9.5
0
10
20
30
40

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 identi -ability [2], [30]; whether the available observations contain su cient 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 de ning 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 identi ability 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 signi cant impact on the computational cost of the estimation scheme and the re-calculation of the matrix 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 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 nite di erence approach, as demonstrated in [43], [42], [41]. A further option is automatic di erentiation [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 matricespocgetrti-

[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]