University of Reading

A systematic method of parameterisation estimation using data assimilation

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Lang, M. ORCID: https://orcid.org/0000-0002-1904-3700, Van Leeuwen, P. J. and Browne, P. (2016) A systematic method of parameterisation estimation using data assimilation. Tellus A: Dynamic Meteorology and Oceanography, 68 (1). 29012. ISSN 1600-0870 doi: 10.3402/tellusa.v68.29012 Available at https://centaur.reading.ac.uk/55822/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.3402/tellusa.v68.29012

Publisher: Taylor & Francis

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

University of Reading

Central Archive at the University of Reading

Reading's research outputs online

Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

A systematic method of parameterisation estimation using data assimilation

By MATTHEW LANG[1]*, PETER JAN VAN LEEUWEN[1,2] and PHILIP BROWNE[1], [1]*Department of Meteorology, University of Reading, Reading, UK*; [2]*National Centre for Earth Observation, Reading, UK*

(Manuscript received 30 June 2015; in final form 11 January 2016)

ABSTRACT

In numerical weather prediction, parameterisations are used to simulate missing physics in the model. These can be due to a lack of scientific understanding or a lack of computing power available to address all the known physical processes. Parameterisations are sources of large uncertainty in a model as parameter values used in these parameterisations cannot be measured directly and hence are often not well known; and the parameterisations themselves are also approximations of the processes present in the true atmosphere. Whilst there are many efficient and effective methods for combined state/parameter estimation in data assimilation (DA), such as state augmentation, these are not effective at estimating the structure of parameterisations. A new method of parameterisation estimation is proposed that uses sequential DA methods to estimate errors in the numerical models at each space-time point for each model equation. These errors are then fitted to pre-determined functional forms of missing physics or parameterisations that are based upon prior information. We applied the method to a one-dimensional advection model with additive model error, and it is shown that the method can accurately estimate parameterisations, with consistent error estimates. Furthermore, it is shown how the method depends on the quality of the DA results. The results indicate that this new method is a powerful tool in systematic model improvement.

Keywords: *data assimilation, parameterisation estimation, parameter estimation*

1. Introduction

Data assimilation (DA) is the process by which observational data are incorporated into numerical models to improve knowledge of the trajectory, and uncertainties, of the state. In meteorological models, DA is used to generate an optimal initial state to be used in subsequent forecasts through the use of prior knowledge of the state (e.g. a previous forecast) and observational data obtained from satellites/weather stations, etc. This combination of prior information and observational data is commonly referred to as state estimation.

Models have parameters that are often unobservable quantities, such as latent heat flux, entrainment rate and surface albedo. Using observations to infer the values that these parameters take is known as *parameter estimation*. Parameter estimation is a more complex problem than state estimation, as even for simple, linear models, the parameter estimation problem becomes non-linear (Evensen et al., 1998). The majority of parameter estimation methods are based on some form of state augmentation, where the state vector is augmented, or extended, with a vector of parameter values. This allows the parameters to be estimated via the DA analysis equations in the same manner that any unobserved prognostic variables would be updated. In a sequential DA scheme, this would involve updating the parameters at every observation time during the model run.

State augmentation is an effective method of parameter estimation and is used to estimate optimal parameters with varying degrees of success (e.g. Annan et al., 2005; Kondrashov et al., 2008; Smith, 2010). However, the process of changing the parameters during a model run may lead to parameters being modified to a region of dynamical instability which may cause the model run/state estimation to fail (Yang and Delsole, 2009; Williamson et al., 2014). Another common problem with state augmentation is that the parameter ensemble tends to collapse before

*Corresponding author.
email: m.s.lang@pgr.reading.ac.uk

Tellus A 2016. © 2016 M. Lang et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

Citation: Tellus A 2016, **68**, 29012, http://dx.doi.org/10.3402/tellusa.v68.29012

an optimal parameter value is found (Santitissadeekorn and Jones, 2014). This can be mitigated by adding a small perturbation to the parameter in order to stop the ensemble from collapsing before the optimal parameter is found. However, the method of generating the perturbation to add to the parameter value is not well understood, with suggestions for improving the parameter ensemble in Liu and West (2001). State augmentation methods, however, cannot be used to estimate systematic functional errors in the parameterisations.

Parameterisations are functions that represent simplified forms of processes in the models that are either sub-grid scale processes or are too complex (either due to lack of scientific understanding or not enough computational power) to represent explicitly. For example, in numerical weather prediction (NWP) models, parameterisations are used to approximate processes such as convective processes, cloud microphysics and radiative processes.

Parameterisation estimation is even more complex than parameter estimation, as it requires not only the estimate of correct parameters, but also the estimate of the functional form of the parameterisation itself. Furthermore, as parameters/parameterisations do not necessarily represent true physical qualities that occur within a system, rarely are they observed quantities.

Parameterisation estimation is often done in an ad-hoc way by running models using the current parameterisations and compensating for errors in the functional forms by adding new parameters or functions when they arise. However, these new additions may not improve a fundamental error with the parameterisation. This is because forecasts are often verified against reanalysis data every 1–6 hours, hence errors have time to accumulate making it difficult to infer the source of the errors.

In this study, DA is used to obtain information about errors in the numerical models, specifically errors that come from incorrectly specified parameterisations. A new method of parameterisation estimation is proposed that makes use of the model equations that will allow for an estimate of the structure in the model errors to be found. This new method uses the differences between a DA analysis and an analysis forecast to estimate the errors in the parameterisations each space-time point. In this study, this parameterisation estimation method is applied to an advection model with additive model error as it is a simple model where it is well known how changes in the functional form affect the dynamics of the system.

This paper shall describe the DA methods used in Section 2. Section 3 outlines the new parameterisation estimation method that is proposed by this study, followed by numerical results and conclusions in Sections 4 and 5, respectively.

2. Data assimilation methods

In this section, the data assimilation method used in this study, the ensemble Kalman smoother (EnKS), is described, and the notation used throughout this paper is introduced.

2.1. Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) was developed by Evensen (1994) and has many different variations that are mainly split into two groups. There are the stochastic EnKFs, such as developed by Burgers et al. (1998) and Houtekamer and Mitchell (1998); and the Ensemble Square Root Filters [EnSRF, Tippett et al. (2003)], including the ensemble adjustment Kalman filter [EAKF, Anderson (2001)] and the ensemble transform Kalman filter [ETKF, Bishop et al. (2001) and Majumdar et al. (2002)]. This study shall use the stochastic EnKF [from Burgers et al. (1998)] but any of the other methods can be used.

The EnKF is a sequential DA method that uses Monte Carlo methods to approximate the error statistics of the state. The evolution of the state is given by the following equation:

$$\mathbf{x}(t+1) = f(\mathbf{x}(t)) + \xi(t) \tag{1}$$

where $\mathbf{x}(t)$ is the N-dimensional state vector at time t, $f : \mathbb{R}^N \to \mathbb{R}^N$ represents the deterministic model and ξ is an unknown model error (assumed to be random).

The EnKF approximates the forecast error covariance matrix, $\mathbf{P_t^f}$, by using an ensemble of M distinct state vectors at time t, $\mathbf{X}(t)$, such that:

$$\mathbf{P_t^f} = \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\mathbf{x_m^f}(t) - \overline{\mathbf{x^f}(t)}\right)^T\left(\mathbf{x_m^f}(t) - \overline{\mathbf{x^f}(t)}\right)\right] \tag{2}$$

where $\mathbf{x_m^f}(t)$ represents an N-dimensional ensemble member of the forecast state and $\bar{\cdot}$ is the mean over the ensemble.

The k^{th} set of observations (occurring at timestep t_k) denoted $\mathbf{y_k}$ with an observation error $\boldsymbol{\epsilon_k}$ is assumed to be Gaussian, with zero mean and a covariance matrix given by $\mathbf{R_k} \in \mathbb{R}^{p_k \times p_k}$ (where p_k is the dimension of $\mathbf{y_k}$). The observations are related to the state by:

$$\mathbf{y_k} = \mathcal{H}_k(\mathbf{x^t}(t_k)) + \boldsymbol{\epsilon_k} \tag{3}$$

where $\mathcal{H}_k : \mathbb{R}^N \to \mathbb{R}^{p_k}$ is the k^{th} observation operator that maps the state vector to observation space and $\mathbf{x^t}$ is the true state.

The initial ensemble is created by sampling ensemble members from a (normal) distribution centred on the initial background state with covariance matrix, $\mathbf{P_0^f}$, generated from prior information. The ensemble is then propagated

forward using the model equations to the observation time, at which point the state is updated by the formula:

$$\mathbf{x}^{\mathbf{a}}_{\mathbf{m}}(t_k) = \mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_k) + \mathbf{K}_{\mathbf{k}}\left[\mathbf{y}_{\mathbf{k}} - \left(\mathcal{H}_k\left(\mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_k)\right) + \delta_{\mathbf{m}}\right)\right] \quad (4)$$

where $\mathbf{x}^{\mathbf{a}}_{\mathbf{m}}(t_k)$ is the m^{th} ensemble member of the updated analysis at timestep t_k, δ_{m} is a stochastic perturbation added to the observation operator that has the same distribution as $\boldsymbol{\epsilon}_{\boldsymbol{k}}$ and $\mathbf{K}_{\mathbf{k}}$ is the Kalman gain matrix for the k^{th} observation with the form

$$\mathbf{K}_{\mathbf{k}} = \mathbf{P}^{\mathbf{f}}_{\mathbf{k}}\mathbf{H}^T_{\mathbf{k}}\left(\mathbf{H}_{\mathbf{k}}\mathbf{P}^{\mathbf{f}}_{\mathbf{k}}\mathbf{H}^T_{\mathbf{k}} + \mathbf{R}_{\mathbf{k}}\right)^{-1} \quad (5)$$

where $\mathbf{H}_{\mathbf{k}}$ is the linearisation of $\mathcal{H}_k$.

The result is a new *analysis* ensemble of state vectors which are then propagated forward by the model. The process is repeated every time there is an observation.

When using the ensemble to estimate the forecast error covariance matrix, spurious correlations can arise in the ensemble covariance matrix that would not occur in the full forecast error covariance matrix. The method used to reduce the impact of these spurious correlations is localisation [see, for example Houtekamer and Mitchell (2001) and Janjic et al. (2011)]. To do this, a localisation function, $\mathcal{L}$ is defined, typically a function of distance in the spatial domain. The localisation matrix, $L_{i,j} = \mathcal{L}(i,j)$ is defined, where i and j are points in the domain and is applied to $\mathbf{P}^{\mathbf{f}}_{\mathbf{k}}$ via the Schur, or Hadamard, product such that:

$$\mathbf{P}^{\mathbf{f}}_{\mathbf{k}_{\mathbf{loc}}} = \mathbf{L} \circ \mathbf{P}^{\mathbf{f}}_{\mathbf{k}} \quad (6)$$

where $\mathbf{P}^{\mathbf{f}}_{\mathbf{k}_{\mathbf{loc}}}$ is the localised forecast error covariance matrix and $\circ$ represents the Schur product of the two $\mathbb{R}^{N\times N}$ matrices.

2.2. Ensemble Kalman Smoother

The EnKS was developed by Evensen and van Leeuwen (2000) and is an extension of the EnKF that aims to update the state vector between observations. This is done by considering the temporal cross-covariances between the model state at the analysis timestep and the observation timestep to update the state accordingly. The first step of the EnKS is to run the EnKF to the observation time, t_k, to obtain the analysed state vector, $\mathbf{x}^{\mathbf{a}}_{\mathbf{m}}(t_k)$. The ensemble forecast cross-covariance matrices between timesteps t_l and t_k are then calculated as:

$$\mathbf{P}^{\mathbf{f}}_{\mathbf{kl}} = \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_k) - \overline{\mathbf{x}^{\mathbf{f}}(t_k)}\right)^T\left(\mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_l) - \overline{\mathbf{x}^{\mathbf{f}}(t_l)}\right)\right] \quad (7)$$

where $0 < t_l < t_k$.

The m^{th} ensemble member of the state vector at t_l is then updated using the following equation:

$$\mathbf{x}^{\mathbf{a}}_{\mathbf{m}}(t_l) = \mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_l) + \mathbf{K}_{\mathbf{kl}}\left[\mathbf{y}_{\mathbf{k}} - \left(\mathcal{H}_k\left(\mathbf{x}^{\mathbf{f}}_{\mathbf{m}}(t_k) + \delta_{\mathbf{m}}\right)\right)\right] \quad (8)$$

where the Kalman gain matrix, $\mathbf{K}_{\mathbf{kl}}$, between t_k and t_l is defined as:

$$\mathbf{K}_{\mathbf{kl}} = \mathbf{P}^{\mathbf{f}}_{\mathbf{kl}}\mathbf{H}^T_{\mathbf{k}}\left(\mathbf{H}_{\mathbf{k}}\mathbf{P}^{\mathbf{f}}_{\mathbf{kk}}\mathbf{H}^T_{\mathbf{k}} + \mathbf{R}_{\mathbf{k}}\right)^{-1}. \quad (9)$$

The difference between the EnKS and the EnKF is now the observations influence the trajectory between observation timesteps as opposed to only at the observation timestep.

3. The parameterisation estimation method

In this section, we discuss a new algorithm for parameterisation estimation using a DA scheme. Traditionally, new parameterisations are chosen by making a large number of forecasts with different parameterisation versions in the model and testing which verifies the best. The method that we propose identifies errors in the functional form of the model (i.e. the model equations) based upon results obtained from a sequence of DA steps. The methodology is general and will apply to any numerical model of any dimension. Let the true state vector at timestep t be denoted by $\mathbf{x}(t)$. Hence, the evolution of this state over one timestep can be written as:

$$\mathbf{x}(\mathbf{t}+\mathbf{1}) = f(\mathbf{x}(\mathbf{t}), G) + \xi(t) \quad (10)$$

where f is the deterministic part of the model, G is the parameterisation we wish to estimate with input parameters given by α, and $\xi(t)$ is a stochastic forcing term representing model error.

The method consists of two parts: (1) finding an estimate of the model error at every space-time point and every variable and (2) extract the structural part of the model error and generate a new parameterisation.

Part I: Estimation of model errors via state estimation

(1) Given forecast state $\mathbf{x}^{\mathbf{f}}_{\mathbf{0}}$, background parameterisation G^b, with background parameters $\boldsymbol{\alpha}^{\mathbf{b}}$, and $\mathbf{y}$, perform state estimation to produce an analysis trajectory, $\mathbf{x}^{\mathrm{a}}(t)$ for all timesteps.
(2) Define an analysis forecast variable, $\widetilde{\mathbf{x}}$ by running the model forward one timestep using the prior parameterisation, such that $\widetilde{\mathbf{x}}(t) = f(\mathbf{x}^{\mathbf{a}}(t-1), G^b)$, for all timesteps. Note that each starting point is from the analysis trajectory, in Fig. 1b.
(3) Compute $\mathbf{e}(t) = \mathbf{x}^{\mathbf{a}}(t) - \widetilde{\mathbf{x}}(t)$ over the whole spatial domain at every timestep. This is the difference between a model forecast of one timestep and the

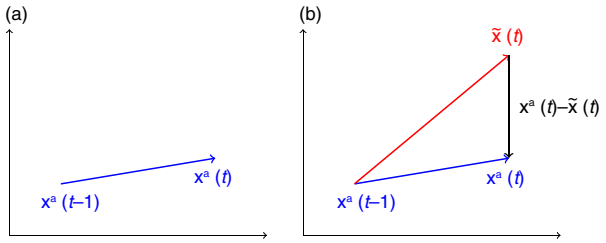

Fig. 1. A single timestep from the data assimilation trajectory from timestep $t-1$ to t. Step 3 defines $\tilde{\mathbf{x}}$ as a single model timestep from the data assimilation trajectory at timestep $t-1$ using the model f with the prior parameterisation G^b.

analysis, and contains information about structural errors in the model equations.

Part II: Estimation of parameterisation errors that give rise to model error

(4) Using any available prior information, propose a collection of n potentially missing parameterisation terms which may contribute to the parameterisation error.
(5) Group these terms into sets of test functions, g_i, $i=0,\ldots n-1$ that contain the first $i+1$ terms, such that $g_j \subset g_{j+1}$.
(6) Use an optimisation scheme to fit the differences between the analysis and analysis forecast, $\mathbf{e}(t)$, to all of the functions, g_i.
(7) Use Bayesian Information Criterion (BIC) to find the terms that contain the most information about the model error.
(8) Reorder the terms in ascending order of BIC differences and define a new set of functions, h_i, such that each contain the first $i+1$ reordered terms.
(9) Repeat Steps 5 to 7 for the new test functions, h_i, to determine the dominant terms that are missing from the model.
(10) The function corresponding to the minimum BIC value corresponds to the optimal functional form of the model error, h^* given the test function, g.
(11) The new estimate for the true state is $\mathrm{G}^a = \mathrm{G}^b + h^*$.
(12) Fit the $\boldsymbol{e}(t)$ for each ensemble member using the terms in the optimal functional form and h^* as a first guess to obtain an ensemble of functional forms for the true model error. The error variance for each functional form from the fit is averaged over the ensemble members to obtain a final error estimate of the parameterisation.

3.1. *Explanation of algorithm*

The most important ingredient of this new method is that concentrates on estimating parameterisations at the level of the model equations directly, and not after errors have accumulated after several timesteps. In the latter case, it will be very difficult to unravel the cause of the accumulated model error in non-linear models.

The first step of the parameterisation estimation method is to run a DA scheme with a prior estimate of the parameterisation, G^b, generating a time series of analysis state vectors, $\mathbf{x}^{\mathbf{a}}(t)$ which represent the best estimate of the true state through time.

Using the time series of $\mathbf{x}^{\mathbf{a}}(t)$ values generated, a new time series representing a forecast of the analysis state, $\tilde{\mathbf{x}}(t)$, is generated such that:

$$\tilde{\mathbf{x}}(t) = f\left(\mathbf{x}^{\mathbf{a}}(t-1), G^b\right) \tag{11}$$

The difference between the analysis and the analysis forecast at all analysis timesteps for each variable at each gridpoint is given by $\mathbf{x}^{\mathbf{a}} - \tilde{\mathbf{x}}$, illustrated for a single timestep in Fig. 1, which represents the best estimate that is currently available regarding the model error at each timestep in each model equation. In order to extract the functional form of this term, a test function needs to be defined to compare the structure in $\mathbf{x}^{\mathbf{a}}(t) - \tilde{\mathbf{x}}(t)$.

The terms chosen in Step 3 can be arranged into a function of the form:

$$g(\mathbf{x}, \beta) = \beta_0 f_0(\mathbf{x}) + \beta_1 f_1(\mathbf{x}) + \ldots + \beta_n f_n(\mathbf{x}) + \xi \tag{12}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_n)$ are the coefficients to be computed in an optimisation scheme. The $f_i(\mathbf{x})$ are functions of the state variable that may include derivatives of the state. ξ represents the stochastic model error assumed to be Gaussian with zero mean and covariance determined by the fit. It is important that the terms are specified well, as this parameterisation estimation method will only look for possible parameterisations in the span of the function $g(\mathbf{x}, \beta)$. Without any prior information, finding a true parameterisation is an ill-posed problem. The prior information and therefore the terms to be considered are generally obtained with help from an expert with knowledge of likely errors in the model. These may include terms assumed negligible such as higher order terms.

Using this form, the parameterisation estimation problem has been translated into a parameter estimation problem, also known as regression analysis. By using an optimisation scheme to calculate the optimum coefficients, $\widehat{\beta}$, for the test function, a functional estimate for the errors in the parameterisation will be found. The choice of optimisation scheme is free; for simplicity, we have chosen linear regression, but non-linear regression or sparse regression techniques, etc. may be used.

This is done sequentially by using the optimisation scheme to initially estimate the coefficient for the first

term in g, $\beta_0 f_0(\mathbf{x})$. In other words, the first test function is defined as:

$$g_0(\mathbf{x}) = \beta_0 f_0(\mathbf{x}) + \xi \tag{13}$$

and β_0 is calculated by using the optimisation scheme to fit to $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$.

The test function is then updated to include the first and second terms of $g(\mathbf{x}, \beta)$ and the optimisation scheme is used again to fit $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$ to:

$$g_1(\mathbf{x}) = \beta_0 f_0(\mathbf{x}) + \beta_1 f_1(\mathbf{x}) + \xi. \tag{14}$$

This is repeated until functional estimates are obtained for all g_i, $i = 0, \ldots, n$.

To verify the quality of the functional estimates calculated in the above step, the BIC (Schwarz, 1978) is used. The Akaike Information Criterion AIC, Akaike (1974) can also be used and produces similar analysis. The BIC is given by:

$$BIC = k \log(N) - 2\log(\mathcal{L}) \tag{15}$$

where k is the number of parameters used in the test function, N is the number of points in $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$ and $\mathcal{L}$ is the maximised likelihood function, $p(\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}|g(\widehat{\beta}))$. The BIC represents the trade-off between how well the functional estimate fits $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$ against the complexity of the model. The BIC is a quantification of Occam's razor states that the optimal form of the parameterisation is the simplest form of the parameterisation that fits the data well. As the regression produces a better fit to $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$, the log-likelihood term ($\log(\mathcal{L})$) increases, resulting in a decrease in BIC. However, as the number of terms in the functional estimate increases, the BIC increases, hence punishing overly complex models. This implies that the optimal form of the parameterisation is given when the BIC is minimal.

When the BIC has been calculated for all g_i functions in Step 6, the terms whose addition to the test function coincide with the biggest decreases in BIC from g_{i-1} to g_i (where the decrease from g_{-1} to g_0 is defined as 0) will contain the most information about the structure of $\mathbf{x}^{\mathbf{a}}(t) - \widetilde{\mathbf{x}}(t)$.

In Step 8, the terms in g are then reordered in descending order of terms with the greatest decrease in BIC to create a new functional form for the model error, $h(\mathbf{x})$. $h(\mathbf{x})$ is then split into sub-functions in the same way as $g(\mathbf{x})$ and the optimisation scheme is again run for each $h_i(\mathbf{x})$ and the coefficients are calculated as before. As the terms are now ordered such that the terms with most information in them are first in the test functions, the BIC will decrease until it reaches the optimal function and then increase for all functional forms after that. Therefore, the minimum in the BIC calculated for each $h_i(\mathbf{x})$ will correspond to the optimal functional form of the model error, within the set that is being considered.

By calculating the coefficients for all $h_i(\mathbf{x})$ individually, we obtain a set of functions with various degrees of confidence based upon the BIC (i.e. the lower the BIC is for an estimate, the higher the confidence in that functional estimate).

To calculate the uncertainty in the coefficients obtained from the method, the uncertainties in the coefficients obtained from each ensemble member are averaged over the ensemble.

For example, for least-squares (used in the results in this paper), the error covariance matrix for the parameters is obtained as (Friedman et al., 2001):

$$\mathrm{Cov}(\widehat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma_y^2 \tag{16}$$

where $\mathbf{X} \in \mathbb{R}^{(n+1)\times(NT)}$ is a matrix of all the functions, f_i, in the test function at all N gridpoints and all T timesteps and σ_y^2 is the variance of the resulting random errors from the regression, such that:

$$\sigma_y^2 = \frac{1}{NT-1}\sum_{i=1}^{NT}\left[\left((\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}) - \widehat{(\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}})}\right)^2\right] \tag{17}$$

where $\widehat{(\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}})} = \mathbf{X}\widehat{\beta}$ is the function obtained from the least squares fit of the $\mathbf{x}^{\mathbf{a}} - \widetilde{\mathbf{x}}$ over all space-time points.

4. Numerical experiments

In this section, experiments are performed on an imperfect 1D advection model described by the following continuous equation:

$$\frac{\partial x}{\partial t} + \alpha\frac{\partial x}{\partial s} + G = \xi_s(t) \tag{18}$$

where s is the space dimension, t is time, α is the linear component of the advection velocity and $\xi_s(t)$ is a stochastic term representing the model error present in the system. $x(t)$ is the quantity being advected, and is a scalar at each point in space-time. In the notation outlined in the previous section, $\mathbf{x}(t)$ is the collection of x at all gridpoints at time-step, t. G is an additive parameterisation of the – possibly non-linear – model error present in the advection equation and is the parameterisation that we shall estimate in the following experiments.

The advection model is applied to a sine wave advected for 100 timesteps over a periodic domain on the interval [0,1] with 101 gridpoints spread uniformly across it, such that $\Delta s = 0.01$.

The initial true state is defined as:

$$x^t(0) = \sin(2\pi s) + \xi_s^0 \tag{19}$$

where the initial model errors are distributed normally with covariance given by a Second Order Autoregressive (SOAR) function, that is:

$$\xi_s^0 \sim N(0, B) \text{ where } B_{k,l} = \left(1 + \frac{d(k,l)}{L}\right) e^{-\frac{d(k,l)}{L}} \tag{20}$$

where k and l are two gridpoints, $d(k,l)$ is the shortest distance between k and l, and $\mathrm{L} = 0.06$ is a correlation length scale.

For subsequent timesteps, the model error, $\xi_s(t)$ is distributed by an uncorrelated Gaussian distribution with zero mean and error covariance, $\mathbf{Q} = 0.01^2\mathbf{I_{101}}\forall t$.

The initial state covariance of the ensemble is also given by a SOAR function, but with a smaller correlation length scale of 0.05. SOAR functions are used in order to introduce cross-covariances between gridpoints.

Observations at timestep, k, are taken of the true state directly such that:

$$\mathbf{y_k} = \mathbf{x^t}(k) + \epsilon_\mathbf{k} \tag{21}$$

where observation errors, $\epsilon_\mathbf{k}$, are distributed normally with zero mean and error covariance, $\mathbf{R_k}$, where $\mathbf{R_k} = 0.005^2\mathbf{I_{p_k}}$ (p_k is the number of observations at timestep k). It is assumed that each observation is independent, hence $\mathbf{R_K}$ is diagonal. Unless otherwise stated, observations are taken at all timesteps and gridpoints in the following experiments.

The EnKS is used with 40 ensemble members and localisation applied to the $\mathbf{P^f}$ matrix in the space dimension, using the localisation function:

$$\mathcal{L}(i,j) = \exp\left(\frac{-d(i,j)^2}{2R_{loc}^2}\right) \tag{22}$$

where the localisation radius, R_{loc}, is defined as 0.1 for these experiments.

The initial test function for use in the parameterisation method is defined as:

$$\begin{aligned} g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 \frac{\partial x}{\partial s} + \beta_4 x \frac{\partial x}{\partial s} + \beta_5 x^2 \frac{\partial x}{\partial s} \\ + \beta_6 \frac{\partial^2 x}{\partial s^2} + \beta_7 x \frac{\partial^2 x}{\partial s^2} + \beta_8 x^2 \frac{\partial^2 x}{\partial s^2}. \end{aligned} \tag{23}$$

The terms were chosen as they may appear in an advection model, and each has a physical meaning. The x terms represent the quantity being advected, $\frac{\partial x}{\partial s}$ is the velocity of the advection and $\frac{\partial^2 x}{\partial s^2}$ represents diffusion.

The optimal functional difference, as calculated by our parameterisation estimation method for each experiment is presented with uncertainty specified on each coefficient calculated as the standard deviation of the parameter errors averaged over the ensemble, as described in Section 3.1.

4.1. Linear advection model

In this section, the method outlined in Section 3 will be applied to a simple parameter estimation problem where the only structural/deterministic error in the model comes from an error in a scalar parameter.

A true state trajectory, $\mathbf{x}^t$ is generated using:

$$\alpha = 1 \tag{24}$$

and a true parameterisation value of the form:

$$G^t = 1\frac{\partial x}{\partial s} \tag{25}$$

which we wish to estimate.

The background state/initial estimate of the true state trajectory, $\mathbf{x}^b$ is generated using the parameterisation of the form:

$$G^b = 0 \tag{26}$$

Substituting these values into the true and background models given in eq. (18) leads to a true model given by:

$$\frac{\partial x}{\partial t} + 2\frac{\partial x}{\partial s} + G = \xi_s(t) \tag{27}$$

and a background model

$$\frac{\partial x}{\partial t} + 1\frac{\partial x}{\partial s} + G = \xi_s(t) \tag{28}$$

Each model described in eqs. (27) and (28) is integrated using a upwind Euler scheme with timestep:

$$\Delta t = \frac{\Delta s}{2.2} \tag{29}$$

in order to maintain the Courant–Friedrichs–Lewy (CFL) criterion. This value of Δt is chosen to minimise spurious numerical diffusion in the experiments.

Using the method outlined in Section 3, the optimal functional form of the parameterisation, G, is given by:

$$G^a = (0.9917 \pm 0.0051)\frac{\partial x}{\partial s} + \left(4.0714 \times 10^{-4} \pm 0.0064\right)\frac{\partial^2 x}{\partial s^2} \tag{30}$$

This is very close to desired solution given in eq. (25). The true parameterisation falls within the error, generated from the ensemble, of the function obtained by this method.

In addition to the simple adjustment of parameter, it can be seen that the parameterisation estimation method also calculates the numerical diffusion that occurs in the system. Analytical calculations show that the expected difference in numerical diffusion produced by the upwind scheme between the true and background models (with no model

error) is $0.0019\frac{\partial^2 x}{\partial s^2}$, which is within the uncertainty estimated by the parameterisation estimation method.

4.2. A non-linear advection equation

In this experiment, the true state is no longer generated by the linear advection equation, instead it is generated with:

$$\alpha = 4.5 \tag{31}$$

and truth parameterisation:

$$G^t = -1.5\frac{\partial x}{\partial s} + x\frac{\partial x}{\partial s}. \tag{32}$$

This leads to a true model given by:

$$\frac{\partial x}{\partial t} + 3\frac{\partial x}{\partial s} + x\frac{\partial x}{\partial s} = \xi_s(t) \tag{33}$$

The introduction of the state variable into the advection equation leads to the model becoming non-linear. The non-linear term, $x\frac{\partial x}{\partial s}$, has the effect of steepening the gradient of the sine wave between the peak and trough of the sine wave, by making the peak move faster and the trough slower.

The background model shall continue to assume that the truth is generated by a linear advection model such that the background parameterisation is given by:

$$G^b = 0 \tag{34}$$

Therefore, the background model becomes:

$$\frac{\partial x}{\partial t} + 4.5\frac{\partial x}{\partial s} + G = \xi_s(t) \tag{35}$$

The advection speed, $\alpha = 4.5$, is chosen such that the background advection velocity is approximate to the advection velocity of the peak of the non-linear advection model in equation.

The initial conditions are the same as the previous experiments on the linear advection model in Section 4.1. The value of Δt is changed to fulfil the CFL criterion of the non-linear advection model and is now given by

$$\Delta t = \frac{\Delta x}{22.5} \tag{36}$$

where 22.5 is chosen such that the system is comfortably within the stable regime of the upwind Euler scheme.

The parameterisation estimation procedure gives an estimate of the true parameterisation in the form

$$G^a = (-1.5101 \pm 0.0408)\frac{\partial x}{\partial s} + (1.0313 \pm 0.0741)x\frac{\partial x}{\partial s}. \tag{37}$$

Comparing this to eq. (33) indicates that this is a consistent estimate of the true parameterisation given the estimated errors.

4.2.1. State augmentation.

In this section, the new method of parameterisation estimation is compared with an existing method of parameter estimation, state augmentation, when applied to the estimation of parameterisations. The true and background models and observation density are the same as in Section 4.2. It is unknown how to define a localisation matrix such that the augmented forecast error covariance matrix is positive definite. This is due to the bordering entries in the augmented localisation matrix, corresponding to the cross-covariances between the state and parameters; setting these to be one can destroy the definiteness of the matrix. This results in non-positive eigenvalues occurring in the 'localised' augmented forecast covariance matrix. Therefore, a 2000-member ensemble is used in the EnKF such that localisation is not required to account for spurious correlations in the forecast covariance matrix.To extend state augmentation to the estimation of parameterisations, we again assume that the true model error terms are unknown and add the full test function defined in (23) to the background model. So that the augmented background model for state augmentation becomes:

$$\begin{aligned}\frac{\partial x}{\partial t} + 4.5\frac{\partial x}{\partial s} + a_0 + a_1 x + a_2 x^2 + a_3\frac{\partial x}{\partial s} + a_4 x\frac{\partial x}{\partial s} + a_5 x^2\frac{\partial x}{\partial s} \\ + a_6\frac{\partial^2 x}{\partial s^2} + a_7 x\frac{\partial^2 x}{\partial s^2} + a_8 x^2\frac{\partial^2 x}{\partial s^2} = \xi_q.\end{aligned} \tag{38}$$

The a_i's are the parameters to be estimated using state augmentation, and the results obtained are summarised in Fig. 2. The parameterisation error found through using state augmentation is:

$$\begin{aligned}h^*(x) = 0.1074 - 1.2748x - 1.1892x^2 + 1.1048\frac{\partial x}{\partial s} \\ - 0.9224x\frac{\partial x}{\partial s} + 0.6589x^2\frac{\partial x}{\partial s} - 0.0012\frac{\partial^2 x}{\partial s^2} \\ - 0.0027x\frac{\partial^2 x}{\partial s^2} + 0.0012x^2\frac{\partial^2 x}{\partial s^2}.\end{aligned} \tag{39}$$

Therefore, the analysis model obtained from state augmentation is:

$$\begin{aligned}\frac{\partial x}{\partial t} + 0.1074 - 1.2748x - 1.1892x^2 + 5.5048\frac{\partial x}{\partial s} \\ - 0.9224x\frac{\partial x}{\partial s} + 0.6589x^2\frac{\partial x}{\partial s} - 0.0012\frac{\partial^2 x}{\partial s^2} \\ - 0.0027x\frac{\partial^2 x}{\partial s^2} + 0.0012x^2\frac{\partial^2 x}{\partial s^2}.\end{aligned} \tag{40}$$

It can be seen that this form of the parameterisation error does not recreate the true model error $-1.5\frac{\partial x}{\partial s} + x\frac{\partial x}{\partial s}$, from eq. (32). This is not the simplest parameterisation error

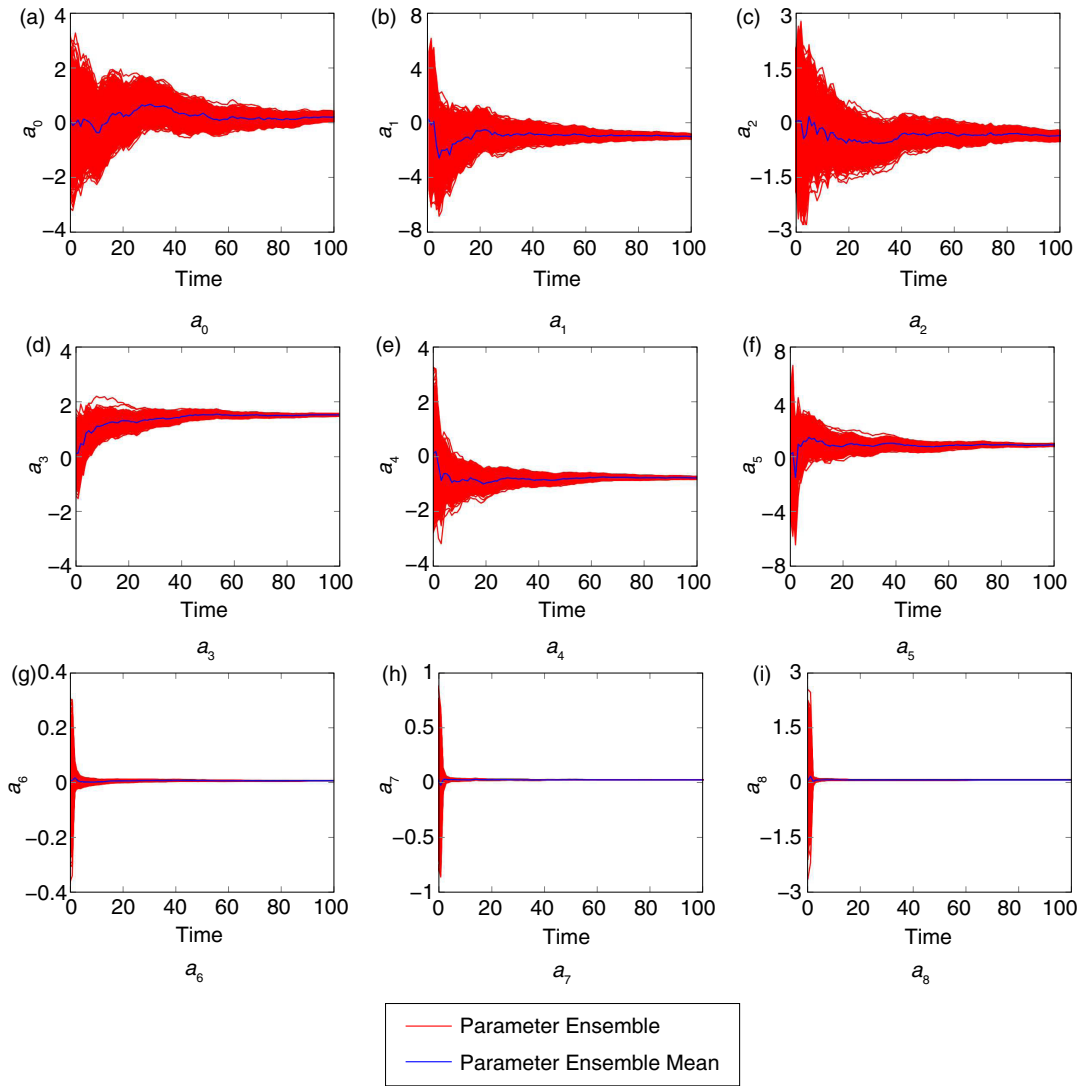

Fig. 2. Estimated parameter values from state augmentation, for the parameters described in eq. (38).

possible, and no information is obtained from state augmentation that suggests how important the terms are relative to one another so that the number of terms in the test function can be reduced.

4.3. *Experiments with different observation densities*

The aim of this section is to show the sensitivity the parameterisation estimation method to the quality of the state estimation. This is done by reducing the observation density in space and/or time resulting in a less accurate DA trajectory.

The experiments in this section, shown in Table 1, will have the same true and background parameterisations as in Section 4.2, such that the true state is generated from the non-linear advection eq. (33) and the background state is generated from the linear advection eq. (18). The observation error covariance matrix and stochastic model error covariance matrix and the localisation function and radius are the same as used in Section 4.1.

It can be seen in Table 1 that when the observation density decreases, the parameterisation estimation method produces less accurate estimations of the true parameter values. This is due to the observations not being able to fully constrain the DA analysis to the true state. When observations are taken every three gridpoints, the diffusion terms become more prominent. However, they are always within the error bounds specified by our methods. When observations are taken every gridpoint and every three timesteps, the parameterisation estimate falls outside the one-standard deviation error bound from the truth, but only slightly. In all other cases the estimates are consistent.

Table 1. Analysis parameterisations from parameterisation estimation method for different observation densities

Observations at all space-time points $G^a = (-1.5101 \pm 0.0408)\frac{\partial x}{\partial s} + (1.0313 \pm 0.0741)x\frac{\partial x}{\partial s}$
Observations at every three gridpoints and every timestep $G^a = (-1.5216 \pm 0.0555)\frac{\partial x}{\partial s} + (0.9783 \pm 0.0902)x\frac{\partial x}{\partial s} + (0.0074 \pm 0.0967)\frac{\partial^2 x}{\partial s^2} + (-0.0048 \pm 0.1695)x^2\frac{\partial^2 x}{\partial s^2}$
Observations at every gridpoint and every three timesteps $G^a = (-1.4146 \pm 0.0484)\frac{\partial x}{\partial s} + (0.8492 \pm 0.0872)x\frac{\partial x}{\partial s}$
Observations at every three gridpoints and every three timesteps $G^a = (-1.4683 \pm 0.0541)\frac{\partial x}{\partial s} + (1.0149 \pm 0.1129)x\frac{\partial x}{\partial s} + (0.0035 \pm 0.1148)x\frac{\partial^2 x}{\partial s^2}$

4.4. Relation between magnitudes of stochastic model error and stochastic observation error

In this section, we study the sensitivity of the method by varying the relative magnitude of model error covariance matrix, $\mathbf{Q}$, and observation error covariance matrix, $\mathbf{R}$.

The true model and background models are both generated using eqs. (33) and (18), from Section 4.2, respectively and $\Delta t = \frac{\Delta x}{22.5}$ to ensure that the CFL criterion is conserved. The experiments are performed in the well-observed case where observations occur at all space-time points and the same localisation function is used as in Section 4.1, given by eq. (22) with localisation radius, $R_{loc} = 0.1$.

Recall that in the previous experiments, $\mathbf{Q} = 0.01^2\mathbf{I}$ and $\mathbf{R} = 0.05^2\mathbf{I}$ (i.e. $\mathbf{Q} = 4\mathbf{R}$) and that the experiments in this section should be compared directly with the experiment in Section 4.2.

Table 2 shows that the parameterisation terms become less accurate when $\mathbf{R}$ increases compared to $\mathbf{Q}$, but estimated errors also increase in a consistent manner. Additionally, the diffusive terms become more prominent as $\mathbf{R}$ increases.

This is because as the observation error covariances increase relative to the model error covariances, the state estimation will have more confidence in the prior model with respect to the observations meaning that the analysis state will be closer to the prior model equations, hence the signal in the observations will be much weaker. As a result, the diffusion terms become more prominent in the analysis state. This is to be expected, as the observations do not contain enough information to get the functional estimate closer to the desired solution.

Table 2. Analysis parameterisations from parameterisation estimation method for different magnitudes of observation error relative to model error

$\mathbf{R} = 0.005^2\ \mathbf{I}$ such that $\mathbf{Q} = 4\mathbf{R}$ $G^a = (-1.5101 \pm 0.0408)\frac{\partial x}{\partial s} + (1.0313 \pm 0.0741)x\frac{\partial x}{\partial s}$
$\mathbf{R} = 0.01^2\ \mathbf{I}$ such that $\mathbf{Q} = \mathbf{R}$ $G^a = (-1.4651 \pm 0.0423)\frac{\partial x}{\partial s} + (1.0217 \pm 0.0778)x\frac{\partial x}{\partial s}$
$\mathbf{R} = 0.02^2\ \mathbf{I}$ such that $\mathbf{Q} = 0.25\mathbf{R}$ $G^a = (-1.4472 \pm 0.0481)\frac{\partial x}{\partial s} + (1.0055 \pm 0.0847)x\frac{\partial x}{\partial s} + (-0.0023 \pm 0.0742)\frac{\partial^2 x}{\partial s^2}$

5. Conclusions

This study has presented a new method of parameterisation estimation that uses the model equations directly to estimate the underlying errors in the model. This is done by comparing the DA analysis with an analysis forecast at every space-time point for every model variable to retrieve errors in the model and to use this information to improve model parameterisations. It is important to realise that the method uses the errors at the level of the model equations directly, instead of accumulating errors over a number of model timesteps. In the latter case, it is difficult to find the cause of the errors in a non-linear model, as model equation errors will interfere with each other during the accumulation. As the method uses the DA analysis as the best estimate of the truth, this method is dependent on the quality of the DA analysis.

In this paper, experiments have been shown using the linear advection equation to estimate parameters. Existing methods of parameter estimation are mostly based on state augmentation, which is an online method where the parameters are updated along with the state. While this means that state augmentation is potentially more computationally efficient, there is a risk that the updated parameter values could cause the numerical model to become unstable (e.g. for the advection models, an advection velocity could be updated such that the CFL criterion is violated) in the middle of computation. For simple parameter estimation, state augmentation may give a better estimate of the true parameter values. However, our new method has been developed to find overall structural errors in the model where state augmentation cannot. There is no risk of numerical instability with our method, as it is an offline method. Model runs are only computed with background parameterisations.

In the numerical experiments conducted in this paper, when the DA analysis was poor, due to less accurate

observations or due to a less dense observations network, poorer estimates of the model error were obtained. This meant that $\mathbf{x}^a - \tilde{\mathbf{x}}$ was less representative of the true model error present in the system, hence the optimal equations obtained from the regression were less representative of the true model error. It is worth noting that the diffusion terms become more prominent when there are fewer observations in space, and that our method is also able to estimate the scale of numerical diffusion present in system. Our method does provide an error estimate on the new parameterisation which is related to the spread in the ensemble. So a less accurate DA result will lead to a larger error in the parameterisation estimate, leading to consistent estimates even if the DA is poor.

The method introduced in this study will only find parameterisations within the pre-determined functional form. Our method appeals to the BIC to select the most influential terms. The optimal equation for the true model error corresponds to the equation with the lowest value of the BIC.

Work still needs to be done to reduce the sensitivity of the method to the quality of the analysis state in order to produce better/more consistent estimates of the model error. However, the new method of parameterisation estimation presented in this study has been shown to estimate simple parameterisations and has the potential to estimate functional forms of model errors and hence parameterisations in more complex models. A natural application of this method would be to infer parameterisations for a low-resolution model when a higher resolution model is available. In this case, all variables can be observed which is favourable for this method. We are currently planning to test the method on NWP models.

6. Acknowledgements

This work is funded by the Natural Environment Research Council under NE/J005878/1 and via the National Centre for Earth Observation (PJvL).

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**(6), 716–723.

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**(12), 2884–2903.

Annan, J. D., Hargreaves, J. C., Edwards, N. R. and Marsh, R. 2005. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model.* **8**(12), 135–154.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **129**(3), 420–436.

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.* **126**(6), 1719–1724.

Evensen, G. 1994. Sequential data assimilation with a nonlinear Quasi-Geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* **99**(C5), 10143–10162.

Evensen, G., Dee, D. P. and Schröter, J. 1998. Parameter estimation in dynamical models. In: *Ocean Modeling and Parameterization* (eds. E. P. Chassignet and J. Verron). Springer, Netherlands, pp. 373–398.

Evensen, G. and van Leeuwen, P. J. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.* **128**(6), 1852–1867.

Friedman, J., Hastie, T. and Tibshirani, R. 2001. *The Elements of Statistical Learning*. 2nd ed. Springer, Berlin.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**(3), 796–811.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**(1), 123–137.

Janjic, T., Nerger, L., Albertella, A., Schröter, J. and Skachko, S. 2011. On domain localization in ensemble-based Kalman filter algorithms. *Mon. Wea. Rev.* **139**(7), 2046–2060.

Kondrashov, D., Sun, C. and Ghil, M. 2008. Data assimilation for a coupled ocean-atmosphere model. Part II: Parameter estimation. *Mon. Wea. Rev.* **136**(12), 5062–5076.

Liu, J. and West, M. 2001. Combined parameter and state estimation in simulation-based filtering. In: *Sequential Monte Carlo Methods in Practice* (eds. A. Doucet, N. de Freitas and N. Gordon). Statistics for Engineering and Information Science. Springer, New York. pp. 197–223.

Majumdar, S. J., Bishop, C. H., Etherton, B. J. and Toth, Z. 2002. Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation. *Mon. Wea. Rev.* **130**(5), 1356–1369.

Santitissadeekorn, N. and Jones, C. 2014. Two-state filtering for joint state-parameter estimation. *arXiv preprint arXiv:1403.5989*.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464.

Smith, P. 2010. *Joint state and Parameter Estimation using Data Assimilation with Application to Morphodynamic Modelling*. PhD Thesis, University of Reading, UK.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**(7), 1485–1490.

Williamson, D., Blaker, A. T., Hampton, C. and Salter, J. 2014. Identifying and removing structural biases in climate models with history matching. *Clim. Dyn.* **45**(5–6), 1299–1324.

Yang, X. and Delsole, T. 2009. Using the ensemble Kalman filter to estimate multiplicative model parameters. *Tellus A* **61**(5), 601–609.