University of Reading

Accounting for model error in strong-constraint 4DVar data assimilation

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Howes, K. E., Fowler, A. M. ORCID: https://orcid.org/0000-0003-3650-3948 and Lawless, A. S. ORCID: https://orcid.org/0000-0002-3016-6568 (2017) Accounting for model error in strong-constraint 4DVar data assimilation. Quarterly Journal of the Royal Meteorological Society, 143 (704). pp. 1227-1240. ISSN 1477-870X doi: 10.1002/qj.2996 Available at https://centaur.reading.ac.uk/68962/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://dx.doi.org/10.1002/qj.2996
To link to this article DOI: http://dx.doi.org/10.1002/qj.2996

Publisher: Royal Meteorological Society

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

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RMetS
Royal Meteorological Society

Accounting for model error in strong-constraint 4D-Var data assimilation

K. E. Howes,[a] A. M. Fowler[a,b]* and A. S. Lawless[a,b]
[a]*School of Mathematical, Physical and Computational Sciences, University of Reading, UK*
[b]*National Centre for Earth Observation, University of Reading, UK*

*Correspondence to: A. M. Fowler, Department of Meteorology, University of Reading, RG6 6BB, UK.
E-mail: a.m.fowler@reading.ac.uk

The strong-constraint formulation of four-dimensional variational data assimilation (4D-Var) assumes that the model used in the process perfectly describes the true dynamics of the system. However, this assumption often does not hold and the use of an erroneous model in strong-constraint 4D-Var can lead to a sub-optimal estimation of the initial conditions. We show how the presence of model error can be correctly accounted for in strong constraint 4D-Var by allowing for errors in both the observations and the model when considering the statistics of the innovation vector. We demonstrate that, when these combined model error and observation-error statistics are used in place of the standard observation error statistics in the strong-constraint formulation of 4D-Var, a statistically more accurate estimate of the initial state is obtained.

The calculation of the combined model error and observation-error statistics requires the specification of model error covariances, which in practice are often unknown. We present a method to estimate the combined statistics from innovation data that does not require explicit specification of the model error covariances. Numerical experiments using the linear advection equation and a simple nonlinear coupled model demonstrate the success of the new methods in reducing the error in the estimate of the initial state, even in the case when only the uncorrelated part of the model error is accounted for.

Key Words: strong constraint; weak constraint; variational data assimilation; model-error covariances

Received 19 August 2016; Revised 21 December 2016; Accepted 10 January 2017; Published online in Wiley Online Library

1. Introduction

Four-dimensional variational data assimilation (4D-Var) is a method for combining a time window of observations with a model. The output, known as the analysis, should provide a more accurate estimate of the state of the system at the start of the time window than either the observations or model can provide alone. 4D-Var is the data assimilation method of choice at many operational meteorological centres, for example to initialize numerical weather prediction (NWP) forecasts at the Met Office (Rawlins *et al.*, 2007) and ECMWF (Rabier *et al.*, 2000) and for reanalysis (Dee *et al.*, 2011).

The model equations are often implemented as a strong constraint within the 4D-Var algorithm (SC4DVar). This assumes that the observations over the time window (accounting for the observation-error statistics) are consistent with the model if initialized with the true state. However, the error in the model is often non-negligible and so a model trajectory initialized from the truth may not be consistent with the observations. This has the effect that the 4D-Var scheme is unable to find an analysis in good agreement with the truth. The problem of model error is greater as the assimilation window is increased and the error at each time step accumulates.

There are multiple sources of model error. These include:

- finite model resolution, which means that subgrid-scale phenomena cannot be modelled explicitly;
- errors in subgrid-scale parametrizations;
- errors in boundary conditions;
- numerical errors arising from temporal discretization.

Weak-constraint 4D-Var (WC4DVar) relaxes the perfect model assumption by assuming model error is present at each time step (Sasaki, 1970). This allows WC4DVar to provide an estimate of both the initial state and the model error over the time window. However, in its implementation, it requires an estimate of the model-error statistics throughout the assimilation window (Trémolet, 2006). It is not simple to estimate the model-error covariances and in practice many estimates are only available in the space of the observations rather than the full model space, as required by WC4DVar (Todling, 2015).

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

An additional difficulty with WC4DVar is how to best use the estimate of the model error to correct the model in the resulting forecast. For example, how long is the estimate of model error valid for beyond the assimilation window? If the model is not corrected, then it is possible that the forecast will be worse than using strong-constraint 4D-Var (Trémolet, 2007). In practice, there also exists a third difficulty in implementing WC4DVar related to separating out the effects of model and observation bias (Lea *et al.*, 2008).

This article takes an alternative approach to account for model error in 4D-Var. Instead of trying to correct the model, a method is developed to account for the model error when comparing the model evolution to the observations by combining the model-error and observation-error statistics. In practice, this is seen to effectively inflate the observation errors over time, giving the observations nearer the end of the time window less weight in the analysis. The combined model-error and observation-error statistics can be estimated using samples of observation minus model differences. This is a similar technique to the use of covariance matching proposed by Dee (1995) and Desroziers *et al.* (2005) and has been used extensively to estimate the observation-error statistics only (e.g. Weston *et al.*, 2014; Waller *et al.*, 2016b). This new approach has many potential practical benefits over WC4DVar:

- There is no need to explicitly define an estimate of the model-error covariance matrix. Instead an estimate of the matrix in observation space is required and can be obtained on-line.
- The initial state alone is estimated, therefore the size and complexity of the problem is smaller than WC4DVar.
- Although this method does not try to correct the model, the analysis of the initial state is consistent with the assumed (erroneous) model dynamics. Therefore, the chance of an initialization shock will be reduced.

We develop this new technique using linear theory, consistent with the assumptions regularly made in 4D-Var. However, we illustrate this method using two simple dynamical models; firstly a linear model and then a nonlinear model. We show the novel ideas developed within the linear model framework can be successfully applied to both linear and nonlinear dynamical models.

The paper is organized as follows. In section 2 we outline SC4DVar with a perfect model and then show the impact that the breakdown of the perfect model assumption has on the analysis and its error. In section 3, we show that combining the model-error and observation-error statistics does indeed restore the optimality of the analysis estimated with SC4DVar in the presence of model error. This can be shown to give a statistically more accurate analysis than using the sub-optimal SC4DVar which does not account for model error. The expression derived for the combined model-error and observation-error covariances requires an estimate of the elusive model-error statistics. In section 4 we formulate a method to estimate the combined model-error and observation-error statistics using a sample of observation−model misfits. In section 5 we introduce two dynamical systems, the linear advection equation and a nonlinear idealized coupled atmosphere−ocean model. Section 6 involves use of these numerical models to demonstrate the successful implementation of the theory developed in this article and investigate sensitivities of the results. Note that in these numerical experiments we only consider the effect of the model-error variances as this is more feasible for larger systems. Finally, we summarize and conclude our findings in section 7.

2. Strong-constraint 4D-Var

The aim of SC4DVar is to find the most probable state of a system given a background estimate and a time window of observations. A key underlying assumption of SC4DVar is that the model equations perfectly describe the evolution of the system. In this section we begin by introducing the notation to be used throughout the article.

2.1. Observations and background

Let the column vector $\mathbf{x}_i \in \mathbb{R}^m$ represent the model state vector at time t_i. Observations, $\mathbf{y}_i \in \mathbb{R}^{p_i}$, at time t_i are (linearly) related to the true model state, $\mathbf{x}_i^{\mathrm{t}}$, at time t_i by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i^{\mathrm{t}} + \boldsymbol{\epsilon}_i^{\mathrm{o}}, \qquad i = 0, 1, ..., N, \tag{1}$$

where $\boldsymbol{\epsilon}_i^{\mathrm{o}} \in \mathbb{R}^{p_i}$ represents the observation errors at time t_i. These are assumed to be drawn from a normal distribution with mean zero and error covariance matrix $\mathbf{R}_i \in \mathbb{R}^{p_i \times p_i}$. The linear observation operator $\mathbf{H}_i : \mathbb{R}^m \longrightarrow \mathbb{R}^{p_i}$ maps the state from model space to observation space.

The background, $\mathbf{x}^{\mathrm{b}}$, is related to the true state $\mathbf{x}_0^{\mathrm{t}}$ at time t_0 such that

$$\mathbf{x}^{\mathrm{b}} = \mathbf{x}_0^{\mathrm{t}} + \boldsymbol{\epsilon}^{\mathrm{b}}, \tag{2}$$

where the column vector $\mathbf{x}^{\mathrm{b}} \in \mathbb{R}^m$ contains a prior estimate of each of the model state variables. The column vector $\boldsymbol{\epsilon}^{\mathrm{b}} \in \mathbb{R}^m$ contains errors in the background model state and is assumed to be drawn from a normal distribution with mean zero and error covariance matrix $\mathbf{B} \in \mathbb{R}^{m \times m}$.

2.2. Strong-constraint 4D-Var with perfect model equations

We assume use of a linear model. The evolution of the state at time $i-1$ to time i is given by

$$\mathbf{x}_i^{\mathrm{t}} = \mathbf{M}_{\{i-1\}\rightarrow i} \mathbf{x}_{i-1}^{\mathrm{t}}. \tag{3}$$

Let us assume we have observations at $N+1$ times throughout the assimilation window and write the 4D-Var cost function as

$$\begin{aligned} J(\mathbf{x}_0) =& \mathcal{J}^{\mathrm{b}} + \mathcal{J}^{\mathrm{o}} \\ =& \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}})^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}) \\ &+ \frac{1}{2}(\widehat{\mathbf{y}} - \widehat{\mathbf{H}}\mathbf{x}_0)^{\mathrm{T}} \widehat{\mathbf{R}}^{-1} (\widehat{\mathbf{y}} - \widehat{\mathbf{H}}\mathbf{x}_0), \end{aligned} \tag{4}$$

where

$$\widehat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \qquad \widehat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}_{0\rightarrow 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N \mathbf{M}_{0\rightarrow N} \end{pmatrix},$$

and

$$\widehat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{R}_1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{R}_N \end{pmatrix}.$$

The hat notation represents vectors and matrices that are valid over the entire assimilation window, not just at individual time steps. For simplicity of presentation, we assume that the observation errors are not correlated in time, but the theory we develop is also applicable to the more general case. The minimum of (4) should provide the minimum variance estimate of the true model state at the initial time, $\mathbf{x}_0^{\mathrm{t}}$, if the Gaussian and linear assumptions are correct, and $\mathbf{B}$ and $\widehat{\mathbf{R}}$ correctly represent the

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

covariances of $\mathbf{x}_0^{\mathrm{t}} - \mathbf{x}^{\mathrm{b}}$ and $\widehat{\mathbf{y}} - \widehat{\mathbf{H}}\mathbf{x}_0^{\mathrm{t}}$ respectively (Lewis *et al.*, 2006).

The analysis which minimizes (4) is given by

$$\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{K}}\widehat{\mathbf{d}}_{\mathrm{b}}, \tag{5}$$

where $\widehat{\mathbf{K}}$ is the gain matrix given by $\widehat{\mathbf{K}} = \mathbf{B}\widehat{\mathbf{H}}^{\mathrm{T}}(\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{T}} + \widehat{\mathbf{R}})^{-1}$ and $\widehat{\mathbf{d}}_{\mathrm{b}}$ is the innovation vector given by $\widehat{\mathbf{d}}_{\mathrm{b}} = \widehat{\mathbf{y}} - \widehat{\mathbf{H}}\mathbf{x}^{\mathrm{b}}$. Therefore the gain matrix $\widehat{\mathbf{K}}$ determines the appropriate weighting of the observational information to be added to the background model state. The error covariance of the analysis in this case is given by

$$\mathbf{A} = (\mathbf{I} - \widehat{\mathbf{K}}\widehat{\mathbf{H}})\mathbf{B}, \tag{6}$$

(Lewis *et al.*, 2006).

2.3. *The effect of an erroneous model in strong-constraint 4D-Var*

Let us now consider the case when the model is no longer perfect. In this article the erroneous model equations take the form

$$\mathbf{x}_i = \mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i}\mathbf{x}_{i-1}, \qquad i = 1, 2, \ldots \tag{7}$$

The model matrix $\mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i} \in \mathbb{R}^{m\times m}$ is the best known description of the evolution of the state vector from the previous time t_{i-1} to time t_i, and is used to assimilate the data and to produce the subsequent forecast. We assume additive model error of a random nature is present at each time step, where to acquire the true model state $\mathbf{x}_i^{\mathrm{t}} \in \mathbb{R}^m$ at time t_i,

$$\begin{aligned}\mathbf{x}_i^{\mathrm{t}} &= \mathbf{M}_{\{i-1\}\to i}\mathbf{x}_{i-1}^{\mathrm{t}} \\ &= \mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i}\mathbf{x}_{i-1}^{\mathrm{t}} + \boldsymbol{\eta}_i, \qquad i = 1, 2, \ldots\end{aligned} \tag{8}$$

where $\boldsymbol{\eta}_i \in \mathbb{R}^m$ is a column vector containing the random model error at time t_i. The vector of random model error is assumed to be unbiased and to have a Gaussian distribution such that $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$, where $\mathbf{Q}_i$ is the model-error covariance matrix at time t_i.

We can extend (8) to relate the true state at time i to the initial true state evolved using the erroneous model

$$\begin{aligned}\mathbf{x}_i^{\mathrm{t}} &= \mathbf{M}_{0\to i}\mathbf{x}_0^{\mathrm{t}} \\ &= \mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i}\left(\mathbf{M}_{0\to\{i-1\}}\mathbf{x}_0^{\mathrm{t}}\right) + \boldsymbol{\eta}_i \\ &= \mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i}\left(\mathbf{M}^{\mathrm{e}}_{\{i-2\}\to\{i-1\}}\left(\mathbf{M}_{0\to\{i-2\}}\mathbf{x}_0^{\mathrm{t}}\right) + \boldsymbol{\eta}_{i-1}\right) + \boldsymbol{\eta}_i \\ &\vdots \\ &= \mathbf{M}^{\mathrm{e}}_{0\to i}\mathbf{x}_0^{\mathrm{t}} + \sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\to i}\boldsymbol{\eta}_j.\end{aligned} \tag{9}$$

The linearity of the model (7) allows us to define the erroneous operator, $\widehat{\mathbf{H}}^{\mathrm{e}} \in \mathbb{R}^{(\sum_{i=0}^{N} p_i)\times m}$, as

$$\widehat{\mathbf{H}}^{\mathrm{e}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}^{\mathrm{e}}_{0\to 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N\mathbf{M}^{\mathrm{e}}_{0\to N} \end{pmatrix}.$$

When erroneous model matrices of the form $\mathbf{M}^{\mathrm{e}}_{\{i-1\}\to i}$, (7), are used to evolve the model state, the operator $\widehat{\mathbf{H}}$ which contains the perfect model dynamics is replaced with $\widehat{\mathbf{H}}^{\mathrm{e}}$ containing the erroneous model matrices in the 4D-Var cost function, (4). The difference between the true state and the observations evaluated using the erroneous model, $\widehat{\mathbf{y}} - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{x}_0^{\mathrm{t}}$, no longer has a variance consistent with $\widehat{\mathbf{R}}$.

Let $\widehat{\boldsymbol{\epsilon}}^{\mathrm{o}*} = \widehat{\mathbf{y}} - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{x}_0^{\mathrm{t}}$. This can be evaluated, at time i, in terms of the observation error, $\boldsymbol{\epsilon}^{\mathrm{o}}$, and the model error, $\boldsymbol{\eta}$, using (1), (3), (7) and (9):

$$\begin{aligned}\boldsymbol{\epsilon}_i^{\mathrm{o}*} &= \mathbf{y}_i - \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\to i}\mathbf{x}_0^{\mathrm{t}} \\ &= \mathbf{y}_i - \mathbf{H}_i\mathbf{M}_{0\to i}\mathbf{x}_0^{\mathrm{t}} + \mathbf{H}_i\mathbf{M}_{0\to i}\mathbf{x}_0^{\mathrm{t}} - \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\to i}\mathbf{x}_0^{\mathrm{t}} \\ &= \boldsymbol{\epsilon}_i^{\mathrm{o}} + \mathbf{H}_i(\mathbf{x}_i^{\mathrm{t}} - \mathbf{M}^{\mathrm{e}}_{0\to i}\mathbf{x}_0^{\mathrm{t}}) \\ &= \boldsymbol{\epsilon}_i^{\mathrm{o}} + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\to i}\boldsymbol{\eta}_j.\end{aligned} \tag{10}$$

Equation (10) explicitly states that, at each observation time t_i, the error in the comparison of observations and the model state initial conditions evolved with an erroneous model is a combination of observation error and model error. We next seek to obtain the statistics of this error term $\boldsymbol{\epsilon}_i^{\mathrm{o}*}$.

By taking the statistical expectation of the error term $\boldsymbol{\epsilon}^{*}_{\mathrm{o}i}$ (10), and using the assumptions that both the observation errors and model errors are distributed with a zero mean, it is clear that $<\boldsymbol{\epsilon}_i^{\mathrm{o}*}>= 0$. The covariance of the combined error terms at observation times t_i and t_k respectively can be evaluated by taking the statistical expectation

$$\begin{aligned}\mathbf{R}^{*}{}_{(i,k)} &= <\boldsymbol{\epsilon}_i^{\mathrm{o}*}(\boldsymbol{\epsilon}_k^{\mathrm{o}*})^{\mathrm{T}}> \\ &= <\left(\boldsymbol{\epsilon}_i^{\mathrm{o}} + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\to i}\boldsymbol{\eta}_j\right)\left(\boldsymbol{\epsilon}_k^{\mathrm{o}} + \mathbf{H}_k\sum_{j=1}^{k}\mathbf{M}^{\mathrm{e}}_{j\to k}\boldsymbol{\eta}_j\right)^{\mathrm{T}}> \\ &= <\boldsymbol{\epsilon}_i^{\mathrm{o}}(\boldsymbol{\epsilon}_k^{\mathrm{o}})^{\mathrm{T}}> \\ &\quad + \mathbf{H}_i <\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\to i}\boldsymbol{\eta}_j\left(\sum_{j=1}^{k}\mathbf{M}^{e}_{j\to k}\boldsymbol{\eta}_j\right)^{\mathrm{T}}> \mathbf{H}_k^{\mathrm{T}},\end{aligned} \tag{11}$$

using (10) and assuming that both the model errors and observation errors are uncorrelated in time and the model and observation errors are uncorrelated between them. Evaluating the expectations in (11) leads to

$$\mathbf{R}^{*}{}_{(i,k)} = \begin{cases} \mathbf{R}_i & \text{for i=k=0,} \\ \mathbf{R}_i + \mathbf{Q}^{*}{}_{(i,k)} & \text{for i=k} \neq 0, \\ \mathbf{Q}^{*}{}_{(i,k)} & \text{otherwise,} \end{cases} \tag{12}$$

where $\mathbf{Q}^{*}_{(i,k)}$ are the terms present due to the errors in the model,

$$\mathbf{Q}^{*}_{(i,k)} = \mathbf{H}_i\left[\sum_{j=1}^{\min(i,k)}\mathbf{M}^{\mathrm{e}}_{j\to i}\mathbf{Q}_j\,\mathbf{M}^{\mathrm{e}}_{j\to k}{}^{\mathrm{T}}\right]\mathbf{H}_k^{\mathrm{T}}. \tag{13}$$

A more detailed derivation of (12) and (13) is given in Appendix A. We note that no model evolution is required in the comparison of observations and the model state at time t_0, hence the observation-error covariance matrix $\mathbf{R}_0$ fully describes the error statistics in the innovation at time t_0. We can present the full combined model-error and observation-error covariance matrix $\widehat{\mathbf{R}}^{*}$, containing the covariance submatrices $\mathbf{R}^{*}{}_{(i,k)}$ (12) as

$$\widehat{\mathbf{R}}^{*} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_1 + \mathbf{Q}^{*}{}_{(1,1)} & \cdots & \mathbf{Q}^{*}{}_{(1,N)} \\ \vdots & \mathbf{Q}^{*}{}_{(2,1)} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{Q}^{*}{}_{(N,1)} & \cdots & \mathbf{R}_N + \mathbf{Q}^{*}{}_{(N,N)} \end{pmatrix}. \tag{14}$$

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Note that the inclusion of time-correlated observation errors could also be included in the off block diagonal elements of (14), but for convenience we will continue to assume that the time correlations are only due to the accumulation of model error.

If the analysis scheme were to use $\widehat{\mathbf{R}}^*$ to weight the observation term in (4), we would expect the assimilation system to be optimal as the correct error statistics are used. (This will be explored further in section 3.) However, if this model error, $\mathbf{Q}^*$, is not accounted for, the analysis estimated using SC4DVar is given by

$$\mathbf{x}_0^{\mathrm{ae}} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{d}}_{\mathrm{b}}^{\mathrm{e}}, \tag{15}$$

with the gain matrix given by

$$\widehat{\mathbf{K}}^{\mathrm{e}} = \mathbf{B}\widehat{\mathbf{H}^{\mathrm{e}}}^{\mathrm{T}}(\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}^{\mathrm{e}}}^{\mathrm{T}} + \widehat{\mathbf{R}})^{-1}, \tag{16}$$

and the innovation vector $\widehat{\mathbf{d}}_{\mathrm{b}}^{\mathrm{e}} = \widehat{\mathbf{y}} - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{x}^{\mathrm{b}}$, analogous to (5) in the perfect model case. The corresponding error in the analysis is evaluated as

$$\boldsymbol{\epsilon}_{\mathrm{a}}^{\mathrm{e}} = \mathbf{x}_0^{\mathrm{ae}} - \mathbf{x}_0^{\mathrm{t}} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{d}}_{\mathrm{b}}^{\mathrm{e}} - \mathbf{x}_0^{\mathrm{t}} = \boldsymbol{\epsilon}_b + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{d}}_{\mathrm{b}}^{\mathrm{e}}. \tag{17}$$

We take the statistical expectations of $\boldsymbol{\epsilon}_a^{\mathrm{e}}$ to calculate the mean of the error in the analysis giving $<\boldsymbol{\epsilon}_a^{\mathrm{e}}>=\mathbf{0}$, using the assumptions that the background error, observation errors and model errors all have a zero mean. The corresponding analysis-error covariance matrix is

$$\mathbf{A}^{\mathrm{e}} = <\boldsymbol{\epsilon}_{\mathrm{a}}^{\mathrm{e}}(\boldsymbol{\epsilon}_{\mathrm{a}}^{\mathrm{e}})^{\mathrm{T}}> = (\mathbf{I} - \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{H}^{\mathrm{e}}})\mathbf{B} + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{Q}}^*(\widehat{\mathbf{K}}^{\mathrm{e}})^{\mathrm{T}}, \tag{18}$$

where $\widehat{\mathbf{Q}}^* = \widehat{\mathbf{R}}^* - \widehat{\mathbf{R}}$, using the assumption that the background errors, observation errors and model errors are uncorrelated (Appendix B gives a more detailed derivation of (18)).

3. An alternative 4D-Var approach

In this section we derive an alternative 4D-Var approach that amends the 4D-Var cost function (4) in order to correctly account for the error statistics when unbiased random error is present in a model. We demonstrate that, by using a combined model-error and observation-error covariance matrix to replace the observation-error covariance matrix in the 4D-Var cost function (4), we improve the accuracy of the analysis when error of a random nature is present in a model. Subsequently we use a simple scalar example to investigate the sensitivity of the improvement to analysis accuracy with respect to the size of the model error, background error and observation error.

To recapitulate, when a perfect model is used in the SC4DVar cost function (4), the error to be accounted for in the $\mathcal{J}^{\mathrm{o}}$ term is simply the observation-error covariance matrix $\widehat{\mathbf{R}}$. However, when an erroneous model of the form (7) is used in the strong-constraint 4D-Var cost function (4), the error to be accounted for in the $\mathcal{J}^{\mathrm{o}}$ term is the combined model-error and observation-error covariance matrix $\widehat{\mathbf{R}}^*$ given by (14). Compared with $\widehat{\mathbf{R}}$, the matrix $\widehat{\mathbf{R}}^*$

- contains an additional term, $\mathbf{Q}^*{}_{(i,i)}$, as described by (13), on the block diagonal submatrices $\mathbf{R}^*{}_{(i,i)}$. From (13) we see that this is an accumulation of the model error over the assimilation time window;
- contains off-diagonal block covariance submatrices, $\mathbf{Q}^*{}_{(i,k)}$, as described by (13). This is the presence of time correlations caused by the accumulation of error in the model trajectory.

We propose that use of the combined model-error and observation-error covariance matrix $\widehat{\mathbf{R}}^*$ (14), as opposed to $\widehat{\mathbf{R}}$, in the cost function (4), will statistically improve the analysis accuracy when random error is present in the model.

3.1. Improvements to analysis accuracy

The analysis when the erroneous model (7) is used within the SC4DVar cost function (4) *and* the error in the model is accounted for, by using $\widehat{\mathbf{R}}^*$ as opposed to $\widehat{\mathbf{R}}$, is

$$\mathbf{x}_0^{\mathrm{a}*} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{K}}^*\widehat{\mathbf{d}}_{\mathrm{b}}^*, \tag{19}$$

with the gain matrix $\widehat{\mathbf{K}}^* = \mathbf{B}\widehat{\mathbf{H}^{\mathrm{e}}}^{\mathrm{T}}(\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}^{\mathrm{e}}}^{\mathrm{T}} + \widehat{\mathbf{R}}^*)^{-1}$ and the innovation vector $\widehat{\mathbf{d}}_{\mathrm{b}}^* = \widehat{\mathbf{y}} - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{x}^{\mathrm{b}}$, which is the same as $\widehat{\mathbf{d}}_{\mathrm{b}}^{\mathrm{e}}$ defined previously. The corresponding error in the analysis is evaluated as

$$\boldsymbol{\epsilon}_{\mathrm{a}}^* = \mathbf{x}_0^{\mathrm{a}*} - \mathbf{x}_0^{\mathrm{t}} = \mathbf{x}^{\mathrm{b}} + \widehat{\mathbf{K}}^*\widehat{\mathbf{d}}_{\mathrm{b}}^* - \mathbf{x}_0^{\mathrm{t}} = \boldsymbol{\epsilon}_{\mathrm{b}} + \widehat{\mathbf{K}}^*\widehat{\mathbf{d}}_{\mathrm{b}}^*. \tag{20}$$

We take the statistical expectation of ϵ_a^* to calculate the mean of the error in the analysis giving $<\epsilon_a^*>=\mathbf{0}$, using the assumption that the background errors, observation errors and model errors all have a zero mean. We next evaluate the analysis-error covariance matrix,

$$\mathbf{A}^* = <\boldsymbol{\epsilon}_a^*(\boldsymbol{\epsilon}_a^*)^{\mathrm{T}}> = (\mathbf{I} - \widehat{\mathbf{K}}^*\widehat{\mathbf{H}^{\mathrm{e}}})\mathbf{B}, \tag{21}$$

using the assumption that the background errors, observation errors and model errors are uncorrelated (Howes, 2016).

Replacing $\widehat{\mathbf{R}}$ by $\widehat{\mathbf{R}}^*$ leads to the analysis-error covariance matrix $\mathbf{A}^*$ (21) having the same form as the analysis-error covariance matrix of the SC4DVar estimate in the absence of model error, given by (6). This suggests that the analysis solution (19) is the optimal minimum variance estimate of the linear strong-constraint 4D-Var problem, when the model has error of a random nature at each time step, and therefore implies that the analysis $\mathbf{x}_0^{\mathrm{a}*}$ (19) is more statistically accurate than the analysis $\mathbf{x}_0^{\mathrm{ae}}$ (15).

An important point to be noted is that when model error is not accounted for, the analysis error covariance matrix $\mathbf{A}^{\mathrm{e}}$ (18) is unbounded. As the model error increases, the diagonal of the matrix term $\widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{Q}}^*(\widehat{\mathbf{K}}^{\mathrm{e}})^{\mathrm{T}}$ present in (18) increases and therefore the error variances of the analysis variables increase. If the model error is large, it is feasible that the analysis is less accurate than the background. However the analysis-error covariance matrix $\mathbf{A}^*$ (21) is bounded by the background-error covariance matrix $\mathbf{B}$. As the model error increases, the analysis-error covariance matrix $\mathbf{A}^*$ tends to the background-error covariance matrix $\mathbf{B}$. Note that, although $\mathbf{A}^*$ can tend to the background-error covariance matrix $\mathbf{B}$, it cannot be equal $\mathbf{A}^* \neq \mathbf{B}$ for non-zero $\mathbf{B}$ and $\widehat{\mathbf{R}}^*$.

Model error can be regarded as a representivity-like error, which can be defined loosely as an error in the generalized operator $\widehat{\mathbf{H}}$. The method we have developed can therefore be related to those developed to account for other sources of representivity error, such as errors that result from the discretized model state not being able to represent the fine scales that the observation measures (Hodyss and Nichols, 2015; Li *et al.*, 2015). These two different types of representivity error can affect the comparison of observations with the model state mapped to observation space, although the source and structure of the two errors are very different, and so need to be treated separately.

3.2. Simple scalar example

We now investigate how changes in the accuracy of the model, observations and background respectively affect the increase in accuracy obtained when using combined model-error and observation-error statistics as opposed to observation-error statistics alone. Consider an erroneous linear scalar model over one time step of the form $x_i = \beta_i^{\mathrm{e}} x_{i-1}$, with the erroneous model constant $\beta_i^{\mathrm{e}} \in \mathbb{R}$ describing the evolution of the model state from time t_{i-1} to time t_i. To obtain the true model state at time t_i,

$$x_i^{\mathrm{t}} = \beta_i^{\mathrm{e}} x_{i-1}^{\mathrm{t}} + \eta_i, \tag{22}$$

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

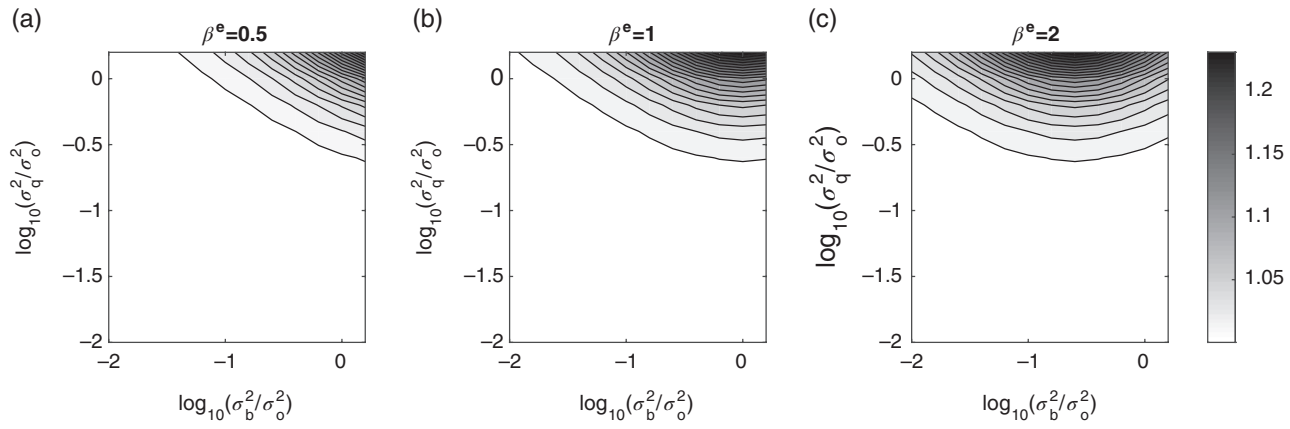

Figure 1. Ratio of analysis error variance, $\sigma_a^{e2}/\sigma_a^{*2}$ (27), for three different values of the erroneous model, $x_1 = \beta^e x_0$, with (a) $\beta^e = 0.5$, (b) $\beta^e = 1.0$ and (c) $\beta^e = 2.0$.

where η_i is random error normally distributed around a zero mean with variance σ_{qi}^2. Consider a direct observation y_i of the model state (observation operator $h_i = 1$) with error normally distributed around a zero mean with variance σ_{oi}^2 and a background model state x^b at time t_{i-1} with error normally distributed around a zero mean with variance σ_b^2. We consider only two times; the background time t_{i-1} and the observation time t_i, therefore to simplify notation we set $i = 1$. The analysis when the erroneous linear scalar model $x_1 = \beta^e x_0$ is used within the strong-constraint 4D-Var cost function (4) and the error in the model is not accounted for is

$$x_0^{ae} = x^b + k^e\, d_b, \tag{23}$$

where the gain constant $k^e = \sigma_b^2 \beta^e (\sigma_b^2 \beta^{e2} + \sigma_o^2)^{-1}$ and the innovation $d_b = y_1 - \beta^e x^b$. Using (14), we evaluate the combined error variance $\sigma_o^{*2} = \sigma_o^2 + \sigma_q^2$. When the erroneous linear scalar model $x_1 = \beta^e x_0$ is used within the strong-constraint 4D-Var cost function (4) and the error in the model is accounted for by replacing σ_o^2 with σ_o^{*2}, the analysis is

$$x_0^{a*} = x^b + k^* d_b, \tag{24}$$

where the gain constant $k^* = {\sigma_b}^2 \beta^e ({\sigma_b}^2 \beta^{e2} + {\sigma_o}^{*2})^{-1}$. Note $k^* \leq k^e$.

In this scalar framework, it is simple to show that the analysis-error variances when not accounting for model error, σ_a^{e2}, will always be greater than or equal to the analysis-error variances when the model error is correctly accounted for, σ_a^{*2}. The analysis-error variances when not accounting for model error can be evaluated using (18):

$$\sigma_a^{e2} = (1 - k^e \beta^e)\sigma_b^2 + (k^e)^2 \sigma_q^2. \tag{25}$$

The analysis-error variances when the model error is correctly accounted for can be evaluated using (21):

$$\sigma_a^{*2} = (1 - k^* \beta^e)\sigma_b^2. \tag{26}$$

The ratio of the error variance in the analysis outputs, $\sigma_a^{e2}/\sigma_a^{*2}$, is

$$\frac{\sigma_a^{e2}}{\sigma_a^{*2}} = \frac{r_b^2 - k^e \beta^e r_b^2 + (k^e)^2 r_q^2}{r_b^2 - k^* \beta^e r_b^2} \geq 1 \tag{27}$$

where $r_b^2 = \sigma_b^2/\sigma_o^2$, and $r_q^2 = \sigma_q^2/\sigma_o^2$. This ratio (27) is plotted in Figure 1 as a function of r_b^2 and r_q^2. We see that, as the magnitude of the model error variance increases, the difference between σ_a^{e2} and σ_a^{*2} also increases as expected. We also see that the difference between σ_a^{e2} and σ_a^{*2} increases initially as r_b^2 increases. This is because, as the observation error becomes smaller than the background error, the difference between the two schemes and the weight they give the observations is more apparent. The peak in the value of $\sigma_a^{e2}/\sigma_a^{*2}$ as a function of r_b^2 is seen to depend on the value of β^e, because the observation is at a different time to the background, e.g. a dispersive model ($\beta^e < 1$, Figure 1(a)) reduces the apparent background error at the observation time. However, as r_b^2 tends to infinity, we see from (27) that $\sigma_a^{e2}/\sigma_a^{*2}$ tends to 1 as in this case the Kalman gains for methods k^e and k^* both tend to $1/\beta$ and so the analyses are identical.

The proposed method relies on the specification of the model-error covariance, which is assumed to change throughout time. In practice model-error statistics are difficult to estimate. However, a key advantage of the proposed method is that an explicit estimate of $\mathbf{Q}_i$ is not needed, as the algorithm only needs the combined observation- and model-error covariances in the (often reduced) observation space. Therefore, we next present a method for estimation of the combined model-error and observation-error statistics required for specification in $\widehat{\mathbf{R}}^*$ (14).

4. Estimation of the combined model-error and observation-error statistics

We now return to the general multi-dimensional problem. The definitions of observation error (1) and background error (2) along with (9) enables the innovation involving the erroneous model at each time t_i to be defined as

$$\begin{aligned}(\mathbf{d}_b^*)_i &= \mathbf{y}_i - \mathbf{H}_i \mathbf{M}^e_{0\to i} \mathbf{x}^b \\ &= \epsilon_{oi} - \mathbf{H}_i \mathbf{M}^e_{0\to i} \epsilon_b + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e_{j\to i} \boldsymbol{\eta}_j. \end{aligned} \tag{28}$$

We calculate the expectation of the following innovation product,

$$\begin{aligned} &E[(\mathbf{d}_b^*)_i (\mathbf{d}_b^*)_k{}^{\mathrm{T}}] \\ &= < (\mathbf{y}_i - \mathbf{H}_i \mathbf{M}^e_{0\to i} \mathbf{x}^b)(\mathbf{y}_k - \mathbf{H}_k \mathbf{M}^e{}_{0\to k} \mathbf{x}^b)^{\mathrm{T}} > \\ &= < (\epsilon_{oi} - \mathbf{H}_i \mathbf{M}^e_{0\to i} \epsilon_b + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e_{j\to i} \boldsymbol{\eta}_j) \\ &\qquad (\epsilon_{ok} - \mathbf{H}_k \mathbf{M}^e_{0\to k} \epsilon_b + \mathbf{H}_k \sum_{j=1}^{k} \mathbf{M}^e_{j\to k} \boldsymbol{\eta}_j)^{\mathrm{T}} >, \end{aligned} \tag{29}$$

for times t_i and t_k using (28). The definition of $\mathbf{R}^*{}_{(i,k)}$ as defined in (11) allows us to deduce

$$E[(\mathbf{d}_b^*)_i (\mathbf{d}_b^*)_k^{\mathrm{T}}] = \mathbf{R}^*{}_{(i,k)} + \mathbf{H}_i \mathbf{M}^e{}_{0\to i} \mathbf{B} \mathbf{M}^e_{0\to k}{}^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}}. \tag{30}$$

Assembling the entries at all observation times into the vector $\widehat{\mathbf{d}}_b^*$ we have

$$E[\widehat{\mathbf{d}}_b^* (\widehat{\mathbf{d}}_b^*)^{\mathrm{T}}] = \widehat{\mathbf{R}}^* + \widehat{\mathbf{H}^e} \mathbf{B} \widehat{\mathbf{H}^e}^{\mathrm{T}}, \tag{31}$$

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

where $\widehat{\mathbf{R}}^*$ is defined by (14). Therefore if the evolution of the background-error covariance matrix throughout the assimilation time window can be computed, we have a way to explicitly estimate $\mathbf{R}^*$:

$$\widetilde{\mathbf{R}}^* = E[\widehat{\mathbf{d}}_{\mathrm{b}}^*(\widehat{\mathbf{d}}_{\mathrm{b}}^*)^{\mathrm{T}}] - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{e}^{\mathrm{T}}}, \tag{32}$$

where the tilde indicates that this is an estimate.

For a more detailed derivation of the estimated combined model-error and observation-error covariance matrix $\widetilde{\mathbf{R}}^*$ (32), we refer to Howes (2016). Note that this estimation $\widetilde{\mathbf{R}}^*$ (32) assumes an accurate specification of the background-error statistics in $\mathbf{B}$. To gain a sample of innovation vectors $\mathbf{d}_{\mathrm{b}}^*$ we require a sample from the observation error and background error (distributions of both can be assumed to be known) and from the model error (distribution unknown). One possible way to gain a sample of model error is to make use of ensemble prediction systems, such at those that have been developed at the Met Office and ECMWF, which represent random error in the model forecast using stochastic physics (Buizza *et al.*, 1999; Shutts, 2015; Brankart *et al.*, 2015). However, this alone may not be enough to represent the full model error.

To evolve the B-matrix to the time of the observations, we propose using the randomization technique to estimate the diagonal elements of $\mathbf{HMBM}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}$, as conducted previously in experiments at ECMWF using the atmospheric forecasting model (Andersson, 2003). This same methodology can be extended and applied to evaluate the diagonal elements of $\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{e}^{\mathrm{T}}}$ required in (32). Therefore use of (32) is potentially operationally feasible in areas of high observational frequency, as long as we restrict ourselves to estimating the diagonal elements of the combined error matrix $\widehat{\mathbf{R}}^*$ only. In section 6, it is shown that neglecting the effect of model-error covariances still leads to more accurate analysis than not accounting for model error at all.

5. Dynamical models

In this section we introduce two dynamical models which will be used to demonstrate the methods we have developed in this article. Specifically we present the linear advection equation, followed by a nonlinear idealized coupled atmosphere–ocean model.

5.1. Linear advection equation

The one-dimensional linear advection equation is used to model the transportation of a scalar quantity $u(x,t)$ carried along by a flow with constant speed v in one direction,

$$\frac{\partial u}{\partial t} + v\frac{\partial u}{\partial x} = 0, \tag{33}$$

where x and t are independent variables representing space and the time respectively. Although this model is far more simplistic than any operational model representing the dynamics of the atmosphere or ocean, the model dynamics (33) can be interpreted to represent an idealized description of a passive tracer transported in the atmosphere in one direction by a constant flow, e.g. water vapour carried along by a constant light breeze. The partial differential equation (PDE) (33) does not describe the shape of the passive tracer, only its movement in space and time. We use the spatial domain $x \in [0, 10)$ and describe the shape of the passive tracer $u(x, t_0 = 0)$ at the initial time as (Causon and Mingham, 2010),

$$f(x) = \begin{cases} \exp\left\{-(x - x_0)^2\right\} & \text{for } 2.5 \le x \le 7.5, \\ 0 & \text{otherwise,} \end{cases} \tag{34}$$

which is an exponential function centred around $x_0 = 5$. The PDE (33) can be solved analytically, subject to the initial conditions (34). Equation (33) is solved numerically over the time domain $t \in [0, 1]$ with periodic boundary conditions. The Crank and Nicolson (1947) scheme is an implicit method that computes a second-order approximation of the spatial derivative and a first-order approximation of the time derivative. We use this unconditionally stable method, with the spatial step $\Delta x = 0.1$ and time step $\Delta t = 0.1$ resulting in a time-stepping approximation of the PDE (33).

5.2. Idealized coupled atmosphere–ocean model

The idealized coupled atmosphere–ocean model we use was developed by Molteni *et al.* (1993) and is given by

$$\left.\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= -\sigma x + \sigma y + \alpha v, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -xz + rx - y + \alpha w, \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= xy - bz, \\ \frac{\mathrm{d}w}{\mathrm{d}t} &= -\Omega v - k(w - w^*) - \alpha y, \\ \frac{\mathrm{d}v}{\mathrm{d}t} &= \Omega(w - w^*) - kv - \alpha x, \end{aligned}\right\} \tag{35}$$

with parameter values $\sigma = 10$, $r = 30$, $b = 8/3$, $k = 0.1$, $\Omega = \pi/10$, $w^* = 2$ and coupling parameter $\alpha = 1$ as specified by Dubois and Yiou (1999). With the coupling parameter $\alpha = 0$, the atmosphere is represented by the Lorenz 63 equations (Lorenz, 1963) with the state variables x, y and z and the ocean is represented by two linear equations with the state variables w and v. However, with the coupling parameter $\alpha = 1$, the atmosphere variables x and y are coupled with the ocean variables v and w. The coupled model (35) describes the relationship between convection in the atmosphere and the influences that the sea surface temperature (SST) has on the convection in the atmosphere, as interpreted by Dubois and Yiou (1999). The atmospheric variables describe properties in a layer of fluid of uniform depth in the atmosphere. The atmospheric state variable x is proportional to the intensity of convective motion. The temperature difference between ascending and descending currents is proportional to the atmospheric state variable y. The atmospheric state variable z is proportional to the distortion of the vertical temperature profile from linearity. The model variables w and v represent equatorial SST anomalies' influence on the global system. The coupling parameter α can be interpreted as heat flux at the atmosphere–ocean interface. We apply the Runge–Kutta second-order scheme (Stuart and Humphries, 1998) to provide a time-stepping approximate solution to the coupled ordinary differential equations (ODEs) (35) with the fixed time step $\Delta t = 0.01$.

6. Numerical experiments

6.1. Linear advection equation

We first perform experiments using the linear advection equation to demonstrate the success of the sampling method to estimate the combined observation- and model-error covariances. In this linear system, this matrix can also be calculated explicitly using (12) and (13).

Let the erroneous linear model be the time-stepping solution of the linear advection equation over the spatial domain $x \in [0, 10)$, as described in section 5.1. The true model state at each time t_i differs from the erroneous model state by random error $\boldsymbol{\eta}_i$, as defined by (8). For simplicity, we define the structure of the model-error covariance matrix $\mathbf{Q}_i$ to be diagonal with variance $\sigma_{\mathrm{q}}{}^2 = 0.01$ at each model time step. The background-error correlations are given by the second-order auto-regressive

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

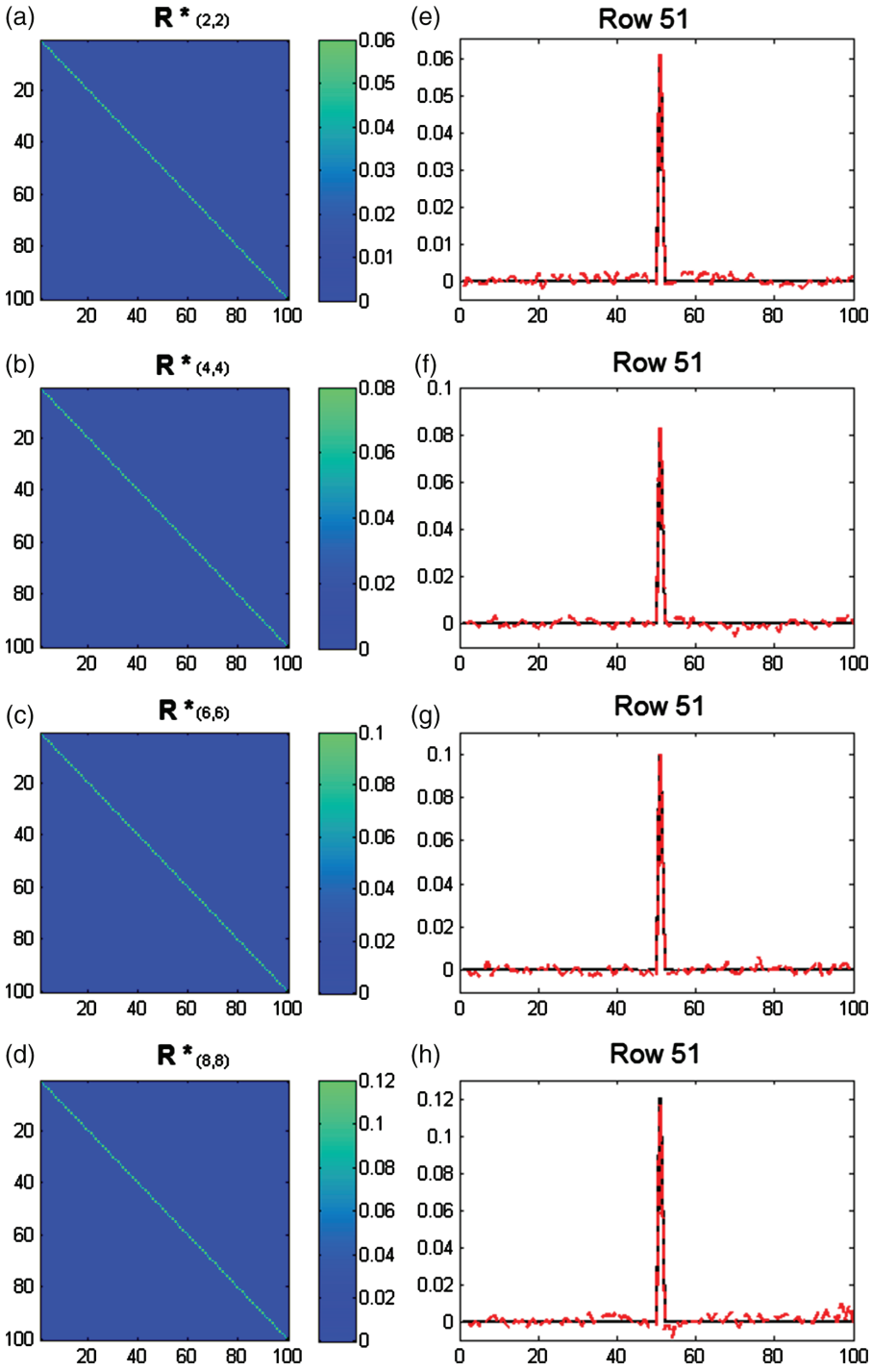

Figure 2. Combined model-error and observation-error covariance matrix $\mathbf{R}^*_{(i,i)} \in \mathbb{R}^{100\times100}$ at each observation time t_i: (a) $i = 2$, (b) $i = 4$, (c) $i = 6$, and (d) $i = 8$, calculated explicitly using (14). (e)–(h) show the central (51st) row of $\mathbf{R}^*_{(i,i)}$ calculated explicitly using (14) –black line – and the central row of $\widetilde{\mathbf{R}}^*_{(i,i)}$ estimated from (32) using sample data – red dotted line – both for the corresponding observation times t_i ($i = 2, 4, 6, 8$).

(SOAR) function with correlation length-scale $\mathcal{L} = 0.4$, where the SOAR correlation function is given by

$$c_k = (1 + r_k/\mathcal{L})\mathrm{e}^{-r_k/\mathcal{L}}, \tag{36}$$

where r_k is the separation of two points. The background-error variances are set to $\sigma_\mathrm{b}^2 = 0.04$. We assume we have direct observations $\mathbf{y}_i$ of all spatial points every two time steps over an assimilation window length of eight time steps. The observation-error covariance matrices are specified to be $\mathbf{R}_i = \sigma_\mathrm{o}^2\mathbf{I}$ (for $i = 2, 4, 6, 8$) with variance $\sigma_\mathrm{o}^2 = 0.04$.

6.1.1. *Estimation of the combined error statistics*

In this example, the combined model-error and observation-error statistics $\mathbf{R}^*_{(i,i)} \in \mathbb{R}^{100\times100}$ at observation times t_i ($i = 2, 4, 6, 8$) are calculated using the explicit equation (14), with the results shown in Figures 2(a)–(d)). These matrices maintain a diagonal structure at all observation times, due to the diagonal structure of $\mathbf{Q}$ and $\mathbf{R}$ and the linear nature of the model. At each subsequent observation time, the variances in the combined error matrix increase, in this case $\sigma_\mathrm{o}^{*2} = 0.06$ at t_2, $\sigma_\mathrm{o}^{*2} = 0.08$ at t_4, $\sigma_\mathrm{o}^{*2} = 0.10$ at t_6 and $\sigma_\mathrm{o}^{*2} = 0.12$ at t_8. This represents the increase in uncertainty of the model trajectory throughout the assimilation window, due to the errors present in the model.

We next estimate the same error covariance matrices $\widetilde{\mathbf{R}}^*_{(i,i)}$ ($i = 2, 4, 6, 8$) with (32). To do this we need to set up a numerical environment where the required sample innovation data are available. We take the true initial conditions of the model state as defined in (34) and from this compute the true trajectory which contains an additive stochastic term as in (8). We repeat

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Table 1. The RMSE of the matrix elements between $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each observation time t_i ($i = 2, 4, 6, 8$).

Time step	RMSE
t_2	0.0014
t_4	0.0017
t_6	0.0020
t_8	0.0023

the following steps 5000 times to produce a sample size of 5000 background values, model error vectors and observations:

1. Produce a background vector $\mathbf{x}^{\mathrm{b}}$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^{\mathrm{t}}_0$.
2. Produce a random model error vector $\boldsymbol{\eta}_i$ for each time $t_1, \ldots, t_8$ using the statistics specified in the model error covariance matrices $\mathbf{Q}_i$.
3. Produce a vector of observations $\mathbf{y}_i$ at each observation time t_i ($i = 2, 4, 6, 8$) by evaluating the true model state vector $\mathbf{x}^{\mathrm{t}}_i$ with (8) and adding noise consistent with the error statistics specified in $\mathbf{R}_i$.

This process gives us a sample size of 5000 innovation vectors $(\mathbf{d}^*_{\mathrm{b}})_i$ for the estimation of $\widetilde{\mathbf{R}}^*_{(i,i)} \in \mathbb{R}^{100\times 100}$ at each observation time t_i ($i = 2, 4, 6, 8$) in (32). The estimation is found to be very successful in representing the structure and values of the combined error covariance matrix. For example, the difference between the central (51st) row of $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i ($i = 2, 4, 6, 8$) can clearly be seen in Figures 2(e)–(h). The overall root mean square errors (RMSEs) between $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i ($i = 2, 4, 6, 8$) are shown in Table 1. These RMSEs are of an order of magnitude less than the specified error variances in $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$ and are only non-zero due to sampling error. Note that the larger the sample size, the smaller sampling error. Although the $\widetilde{\mathbf{R}}^*$ matrix we are estimating at each observation time only contains diagonal elements, we are not making use of this information and so a sample size of 5000 is relatively small in comparison to the 10 000 entries we are estimating.

6.1.2. *Use of combined error statistics in 4D-Var*

We now demonstrate that the replacement of the observation-error covariance matrix $\widehat{\mathbf{R}}$, with combined error statistics in $\widehat{\mathbf{R}}^*$, leads to an analysis of greater statistical accuracy when random error is present in the linear advection equation. Note that in these experiments the time correlations in the accumulated model error are not taken into account. We use the numerical set-up as described previously at the start of section 6.1. We

Table 2. List of data assimilation conditions for error covariance matrices used in strong-constraint 4D-Var with the linear advection equation with random error. Both $\mathbf{Q}_i$ and $\mathbf{R}_i$ are diagonal matrices, whereas $\mathbf{B}$ is constructed using the SOAR function with correlation length-scale $\mathcal{L} = 0.4$.

Conditions	Covariance matrix	Variance
A	$\mathbf{Q}_i$	0.01
	$\mathbf{B}$	0.04
	$\mathbf{R}_i$	0.04
B	$\mathbf{Q}_i$	0.01
	$\mathbf{B}$	0.04
	$\mathbf{R}_i$	0.0016
C	$\mathbf{Q}_i$	0.04
	$\mathbf{B}$	0.04
	$\mathbf{R}_i$	0.04

define two data assimilation methods which will be performed and compared.

- Method 1: the evaluation of the analysis $\mathbf{x}^{\mathrm{ae}}_0$ using (15), i.e. not accounting for model error,
- Method 2: the evaluation of the analysis $\mathbf{x}^{\mathrm{a}*}_0$ by replacing $\widehat{\mathbf{R}}$ in the 4D-Var cost function (4) with the block diagonal combined error covariance matrices $\mathbf{R}^*_{(i,i)}$ (14) for observation times $i = 2, 4, 6, 8$.

We conduct the following steps 100 times to produce a sample size of 100 analysis values $\mathbf{x}^{\mathrm{a}}_0$ for Methods 1 and 2:

1. Produce a background vector $\mathbf{x}^{\mathrm{b}}$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^{\mathrm{t}}_0$.
2. Produce a random model error vector $\boldsymbol{\eta}_i$ for each time $t_1, \ldots, t_8$ using the statistics specified in the model error covariance matrix $\mathbf{Q}_i$ (i=1,...,8).
3. Produce a vector of observations $\mathbf{y}_i$ at each observation time t_i, (i =2, 4, 6, 8) by evaluating the true model state vector $\mathbf{x}^{\mathrm{t}}_i$ with (8) and adding noise consistent with the error statistics specified in $\mathbf{R}_i$ (i=2, 4, 6, 8).
4. Compute both the analysis $\mathbf{x}^{\mathrm{a}*}_0$ and the analysis $\mathbf{x}^{\mathrm{ae}}_0$ as described by Method 1 and Method 2 respectively.

The RMSE for the sample of analysis outputs from Method 1 and Method 2 are calculated and compared. The resulting analysis RMSEs, using the three conditions as specified in Table 2, are shown in Figure 3(a). When accounting for the model error, in Method 2 with $\mathbf{R}^*_{(i,i)}$ ($i = 2, 4, 6, 8$), a more accurate analysis is obtained in each case. The combined error covariance matrix at each observation time is diagonal with the increasing variances, as can be seen in Figures 2(a)–(d) for conditions A. This represents

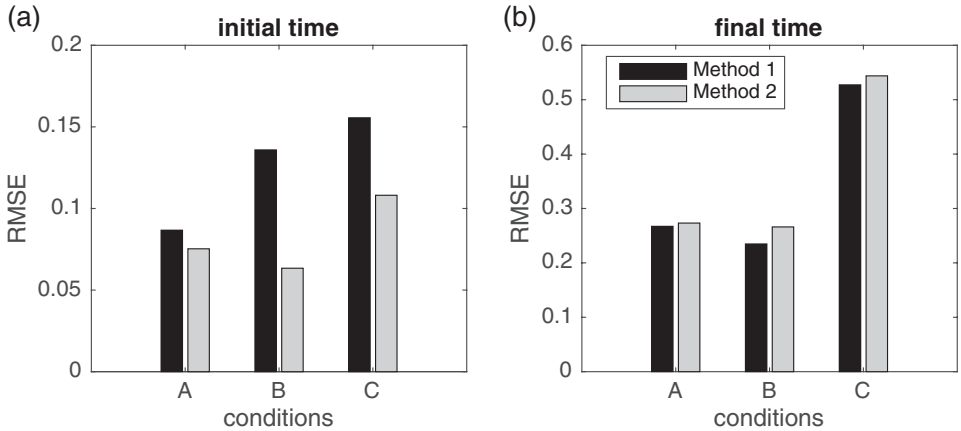

Figure 3. Analysis RMSE from a sample of 100 data assimilation runs, for both Method 1 (not accounting for the model error) and Method 2 (accounting for model error with the combined error covariance matrix $\widehat{\mathbf{R}}^*$ (14)): (a) at the beginning of the assimilation window and (b) at the end of the assimilation window.

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

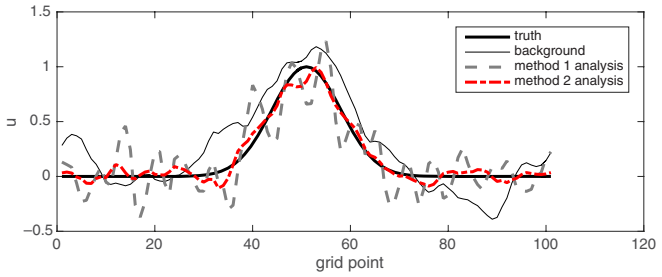

Figure 4. Comparison of the analysis from both Method 1 (not accounting for the model error) and Method 2 (accounting for the model error with the combined error covariance matrix $\widehat{\mathbf{R}}^*$ (14)) for Conditions B. [Colour figure can be viewed at wileyonlinelibrary.com].

the increase in uncertainty of the model trajectory throughout the assimilation window, due to the errors present in the model. Therefore the data assimilation scheme puts less weight on the comparison of the observations with the model evolved state as time increases and hence allows a more accurate estimation of the initial conditions. We illustrate the analysis outputs (Method 1 and Method 2) from one data assimilation cycle (from the sample) in Figure 4 for conditions B. From this figure we see that by accounting for model error (comparing dashed and dash-dot lines) the shape of the initial function is smoother.

6.1.3. *Sensitivity of results*

Work in section 3.2 showed that, for a scalar model, there is a more significant increase in analysis accuracy (of x_0^{a*} compared with x_0^{ae}) when the observations increase in accuracy (in comparison with the background accuracy) and when the size of the model error increases. These conclusions are not limited to models of a scalar nature and we demonstrate that these properties also hold when using the linear advection equation with random error in 4D-Var. Firstly, if we reduce the standard deviation of the observation errors by a factor of five so that $\mathbf{R}_i = 0.0016\mathbf{I}$ (i =2, 4, 6, 8), as shown by conditions B in Table 2, the corresponding combined error covariance matrix at each observation time is diagonal with the variances in the combined error matrix increasing with time. In this case $\sigma_o^{*2} = 0.0216$ at t_2, $\sigma_o^{*2} = 0.0416$ at t_4, $\sigma_o^{*2} = 0.0616$ at t_6 and $\sigma_o{}^{*2} = 0.0816$ at t_8. The resulting analysis RMSE for Method 1 and analysis RMSE for Method 2 are shown in the central two bars of Figure 3(a). These results demonstrate that an increase in observation accuracy leads to a more significant increase in analysis accuracy, when accounting for the model error as opposed to not (in comparison to conditions A). An interesting remark is that, when the model error is not accounted for, this increase in observation accuracy causes a degradation in the analysis $\mathbf{x}_0^{ae}$, as can be seen comparing Method 1 results from experiment B with Method 1 results from A. When Method 1 is used, a decrease in observation variance results in an analysis in which the trajectory of the erroneous model is confined to be closer to the observations, even at the end of the assimilation window, and so the effect of the model error is magnified. However, when accounting for the model error in the analysis $\mathbf{x}_0^{a*}$ in Method 2, an improvement is made to the analysis when the observations are of increased accuracy. With Method 2 the analysis is able to benefit from the increased accuracy of the observations.

Next we increase the standard deviation of the model error variance by a factor of two so that $\mathbf{Q}_i = 0.04\mathbf{I}$ ($i = 1, ..., 8$), as stated by conditions C in Table 2. The corresponding combined error covariance matrix at each observation time is diagonal with the variances in the combined error matrix increasing over time. In this case $\sigma_o^{*2} = 0.12$ at t_2, $\sigma_o^{*2} = 0.20$ at t_4, $\sigma_o^{*2} = 0.28$ at t_6 and $\sigma_o^{*2} = 0.36$ at t_8. The resulting analysis RMSEs are shown in Figure 3(a) (right two bars), where the analysis $\mathbf{x}_0^{a*}$ from Method 2 is again more accurate than when not accounting for the model error in the Method 1 analysis $\mathbf{x}_0^{ae}$. When not accounting for this significant increase in model error, there is a notable decrease in analysis accuracy. However, when accounting for this significant increase in model error, using the combined error covariance matrix, this has minimal effect on the accuracy of the analysis $\mathbf{x}_0^{a*}$, as can be seen comparing the analysis RMSE from experiment A with the analysis RMSE from experiment C.

In Figure 3(b) it is seen that, despite the increased accuracy of the analysis, the RMSE at the end of the assimilation window is slightly worse for Method 2 than for Method 1. This is because the analysis derived using Method 1 is more constrained to the observations at the end of the assimilation window than the analysis derived using Method 2. In the next section we will show the implications of this for the accuracy of a forecast beyond the assimilation window using the nonlinear model.

We have demonstrated the work in section 3.2 for a scalar model, showing that there is a significant increase in analysis accuracy when the observations increase in accuracy (in comparison with the background accuracy) and when the size of the model error increases; this also holds when using a non-scalar erroneous linear model. We next extend our investigation to use an erroneous model of a nonlinear nature.

6.2. *Idealized coupled atmosphere–ocean model*

We now demonstrate that use of combined model-error and observation-error statistics can account for random error in a model of a nonlinear nature and hence improve the analysis accuracy. We use the discretized solution to the idealized coupled atmosphere–ocean model, as described in section 5.2, as our nonlinear erroneous model of the form

$$\mathbf{x}_i = \mathcal{M}^e{}_{\{i-1\}\rightarrow i}(\mathbf{x}_{i-1}), \qquad i = 1, 2, .., 50, \tag{37}$$

where $\mathbf{x}_i = (x_i\ y_i\ z_i\ w_i\ v_i)^{\mathrm{T}}$ is the model state vector. The true model state can be obtained at time t_i as

$$\mathbf{x}^t{}_i = \mathcal{M}^e{}_{\{i-1\}\rightarrow i}(\mathbf{x}^t{}_{i-1}) + \boldsymbol{\eta}_i, \tag{38}$$

where the vector of model error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$. The forms of the erroneous model (37) and true model (38) are the same as (7) and (8) used in derivation of both the combined model-error and observation-error covariance matrix $\widehat{\mathbf{R}}^*$ (14) and the estimated combined error covariance matrix $\widetilde{\mathbf{R}}^*$ (32), with the exception that here nonlinear model equations are used, as opposed to linear model matrices. One of the key objectives in this section is to show that the theory developed with linear models is also successfully applicable using models of a nonlinear nature. We have discussed the capabilities of NWP centres, such as ECWMF, to estimate the diagonal entries of the background model-error covariance matrix evolved using the model matrix and subsequently mapped to observation space using the randomization method. If our developed method to compute $\widetilde{\mathbf{R}}^*$ (32) were to be implemented operationally using similar randomization techniques, then only the diagonal elements would be specified in the combined error covariance matrix. Therefore, we wish to show in our numerical experiments with the idealized erroneous model (37) that, even when only the diagonal elements of the combined error covariance matrix are calculated and used within the data assimilation process, improvements to the analysis accuracy can be obtained. Specifically, this ignores the presence of time and multivariate cross-correlations in both the observation error and model error.

For experiments in this section we use an assimilation window length of 50 time steps. We define the true initial conditions:
$x^t{}_0 = -3.4866$, $y^t{}_0 = -5.7699$, $z^t{}_0 = 18.341$,
$w^t{}_0 = -10.7175$ and $v^t{}_0 = -7.1902$,
which are on the coupled model attractor. We initially specify the model-error covariance matrix $\mathbf{Q}_i$ at each time t_i ($i = 1, 2, ..., 50$) to be diagonal, with variances along the diagonal set to 0.02,

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

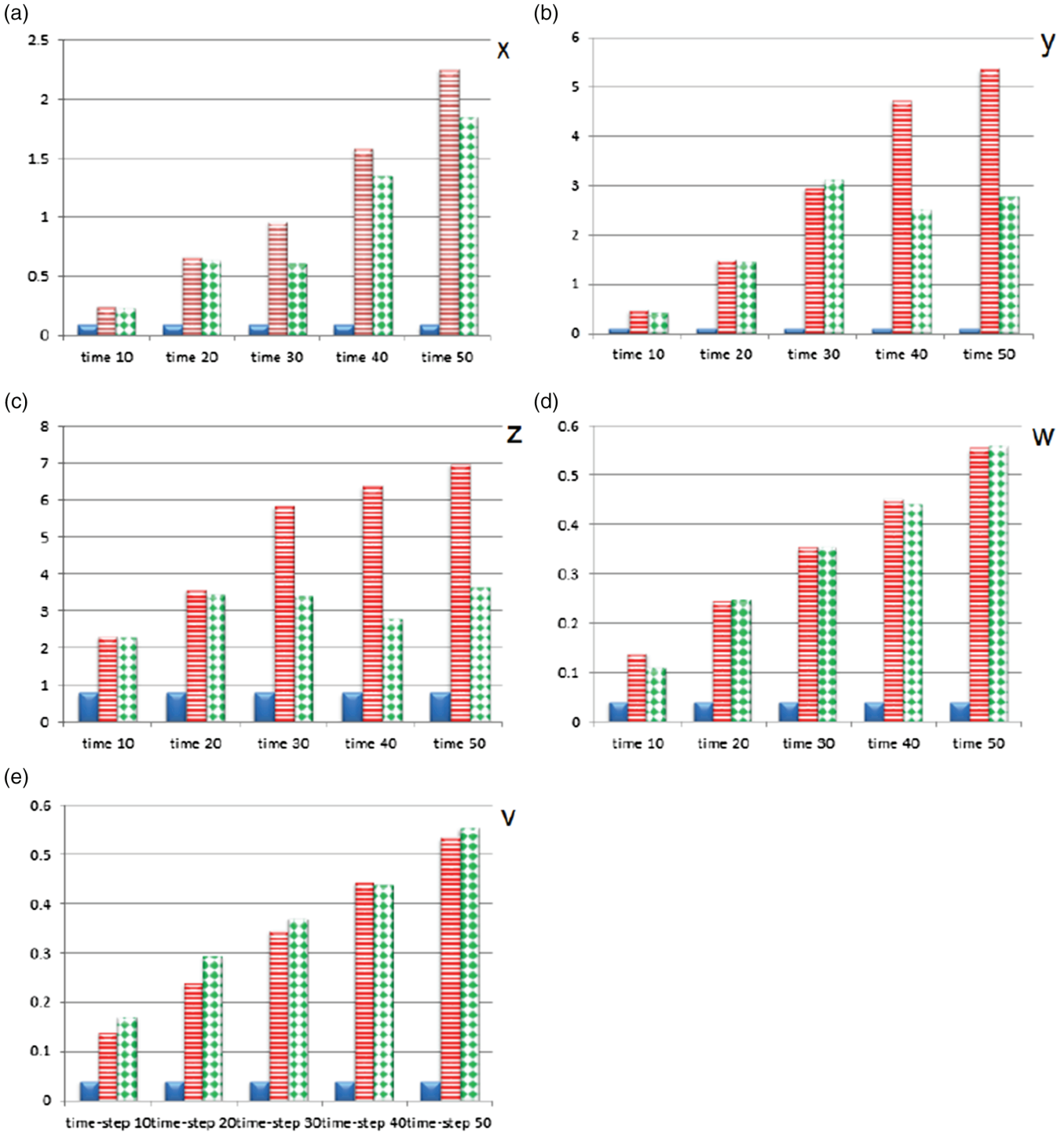

Figure 5. Diagonal elements of $\widehat{\mathbf{R}}^*$ (red stripes) and $\widetilde{\mathbf{R}}^*$ (green diamonds), compared with the observation-error covariance matrix $\mathbf{R}_i$ (blue), at all observation time steps t_{10}, t_{20}, t_{30}, t_{40}, t_{50}, for each of the variables in the idealized coupled nonlinear model: (a) x, (b) y, (c) z, (d) w, and (e) v.

0.02, 0.2, 0.01, 0.01. With this level of model error variance, the general behaviour of the coupled model is maintained, but there are significant variations in the model trajectories of both the atmospheric and oceanic model state variables throughout the time window. We initially specify a diagonal background-error covariance matrix $\mathbf{B}$ with standard deviations approximately 10% of the true initial conditions (specifically variances 0.1, 0.3, 3.4, 1.1 and 0.52 along the diagonal). We have direct observations of each model state variable present every 10 time steps with observation operator $\mathbf{H}_i = \mathbf{I}$. The observation-error covariance matrix $\mathbf{R}_i$ is diagonal at each observation time t_i in the assimilation window, initially set with standard deviations approximately 2% of the maximum absolute value of each respective variable. Specifically $\mathbf{R}_i$ contains the values 0.09, 0.09, 0.81, 0.04, 0.04 along the diagonal.

6.2.1. *Estimation of the combined error statistics*

We compare the diagonal entries of the combined error covariance matrix $\widehat{\mathbf{R}}^*$ computed using (14) with the diagonal entries of $\widetilde{\mathbf{R}}^*$ estimated using (32) with sample innovation data. The elements of $\widehat{\mathbf{R}}^*$ calculated directly with (14) require linearization of the nonlinear model (37) around the true model state trajectory, while the diagonal entries of $\widetilde{\mathbf{R}}^*$ are estimated with (32), which requires linearization of the nonlinear model (37) which is conducted around the background model state trajectory. The estimation of $\widetilde{\mathbf{R}}_i^*$ uses a sample size of 1000 innovations at each observation time t_i ($i = 10, 20, 30, 40, 50$). Figure 5 compares the diagonal entries (variances) of the three matrices $\widehat{\mathbf{R}}$, $\widehat{\mathbf{R}}^*$ and $\widetilde{\mathbf{R}}^*$ at the observation times t_i ($i = 10, 20, 30, 40, 50$). The observation error covariance matrix $\mathbf{R}_i$ is static throughout the window, whereas the variances

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

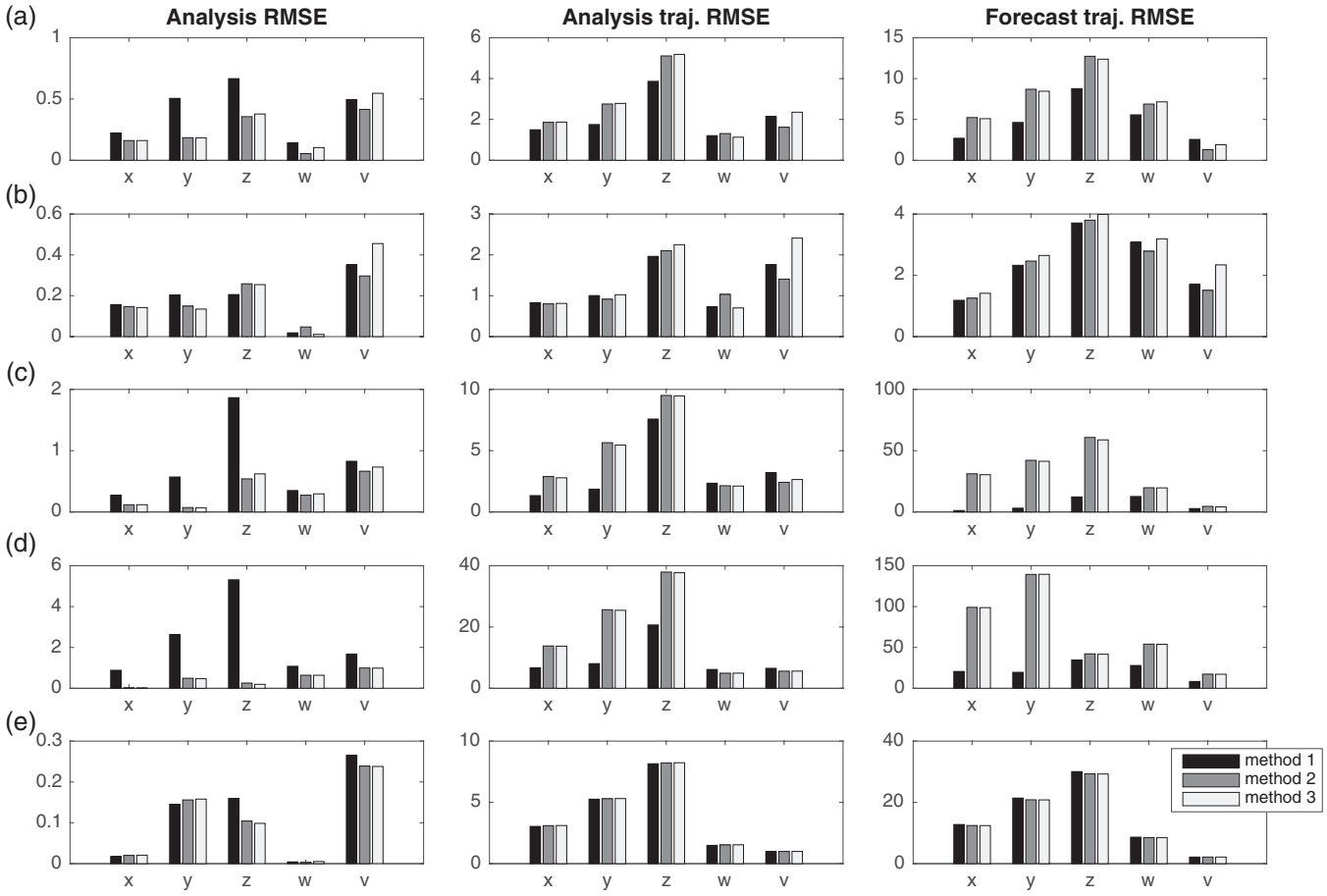

Figure 6. Analysis RMSE (left column), the subsequent RMSE of the analysis trajectories over the assimilation window (middle column) and the RMSE of the forecast trajectories beyond the assimilation window over 50 time steps (right column). Results are from applying Methods 1, 2 and 3 over a sample of 100 data assimilation runs using the covariance specified in Table 3 for (a) Conditions I, (b) Conditions II, (c) Conditions III, (d) Conditions IV, and (e) Conditions V.

in the combined error covariance matrix $\widehat{\mathbf{R}}^*$ computed directly with the formulae (14) are significantly larger as time increases, accounting for the uncertainty in the model state trajectories caused by the errors in the model. Note that the variances estimated in the combined model-error and observation-error covariance matrix $\widetilde{\mathbf{R}}^*$ do not always increase with time, as can be seen for the atmospheric variables x, y and z in Figure 5. Whether $\widehat{\mathbf{R}}^*$ or $\widetilde{\mathbf{R}}^*$ contain the most appropriate combined model-error and observation-error statistics to use in the 4D-Var cost function is an area of future work. The estimate $\widetilde{\mathbf{R}}^*$ uses a sample of innovations and therefore better recognises the spread of the model error at each specific observation time. However, the estimate $\widetilde{\mathbf{R}}^*$ also uses the linearization of the nonlinear model around the background model state trajectory (as opposed to the true model trajectory) and this could have a detrimental effect on the estimation of the combined statistics. Another key point to note is that the variances evaluated in $\widehat{\mathbf{R}}^*$ and estimated in $\widetilde{\mathbf{R}}^*$ are much closer in value for the ocean variables than the atmospheric variables. The reason for this is that the theory for the derivation of $\widehat{\mathbf{R}}^*$ (14) and estimation of $\widetilde{\mathbf{R}}^*$ (32) is based on the use of linear models and the ocean variables w and v are of a less chaotic nature than the atmospheric variables x, y and z, which are of a highly nonlinear chaotic nature. In the next section we perform strong-constraint 4D-Var, firstly not accounting for the error in the erroneous idealized coupled model, and compare the analysis outputs with strong-constraint 4D-Var performed when accounting for the error in the model.

6.2.2. *Use of combined error statistics in 4D-Var*

We define three data assimilation methods that vary depending on the error covariance matrix used in the $\mathcal{J}^{\mathrm{o}}$ term of the nonlinear 4D-Var cost function.

- Method 1: the evaluation of the analysis $\mathbf{x}_0^{\mathrm{ae}}$ using the nonlinear 4D-Var cost function with no changes;
- Method 2: the evaluation of the analysis $\mathbf{x}_0^{\mathrm{a}*}$ by replacing $\widehat{\mathbf{R}}$ in the nonlinear 4D-Var cost function with the diagonal entries of the combined error covariance matrix $\widehat{\mathbf{R}}^*$ (14);
- Method 3: the evaluation of the analysis $\mathbf{x}_0^{\mathrm{a}*}$ by replacing $\widehat{\mathbf{R}}$ in the nonlinear 4D-Var cost function with the diagonal entries of the estimated matrix $\widetilde{\mathbf{R}}^*$ (32) from sample innovation data.

In the left column of Figure 6, the analysis RMSE from a sample of 100 analyses is plotted when using the three different methods for each state variable. The five different covariance conditions are specified in Table 3 (note that Conditions I correspond to the covariances used in the previous subsection). The results show, in general, a significant reduction in analysis error for all model state variables when combined model-error and observation-error statistics are used in the strong-constraint 4D-Var cost function, compared with observation-error statistics only. When compared with the observation-error variances (used in Method 1), both the error variances in $\widehat{\mathbf{R}}^*$ (Method 2) and the estimated error variances in $\widetilde{\mathbf{R}}^*$ (Method 3) give less weight to the comparison of the model evolved state with observations, due to the uncertainty in the model trajectory. Let us focus first on Figure 6(a), Conditions I. The reduction in analysis RMSE, when accounting for the model error, is largest for the atmospheric variables y and z. We hypothesise that this is down to two factors: the first is the ratio between the background-error and observation-error variances and the second is the size of the model error. We will investigate this further in section 6.2.3. We find that accounting for errors in the model with the combined error statistics reduces the number of minimization iterations required to reach the tolerance level. This is demonstrated in this

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Table 3. List of error covariance matrices used in strong-constraint 4D-Var with the erroneous idealized coupled atmosphere–ocean model.

	Covariance matrix	Variance for x	Variance for y	Variance for z	Variance for w	Variance for v
Conditions I	$\mathbf{Q}_i$	0.02	0.02	0.2	0.01	0.01
	$\mathbf{B}$	0.1	0.3	3.4	1.1	0.52
	$\mathbf{R}_i$	0.09	0.09	0.81	0.04	0.04
Conditions II:	$\mathbf{Q}_i$	0.005	0.005	0.05	0.0025	0.0025
decrease in model error	$\mathbf{B}$	0.1	0.3	3.4	1.1	0.52
	$\mathbf{R}_i$	0.09	0.09	0.81	0.04	0.04
Conditions III:	$\mathbf{Q}_i$	0.08	0.08	0.8	0.04	0.04
increase in model error	$\mathbf{B}$	0.1	0.3	3.4	1.1	0.52
	$\mathbf{R}_i$	0.09	0.09	0.81	0.04	0.04
Conditions IV:	$\mathbf{Q}_i$	0.5	0.5	5.0	0.25	0.25
further increase in model error	$\mathbf{B}$	0.1	0.3	3.4	1.1	0.52
	$\mathbf{R}_i$	0.09	0.09	0.81	0.04	0.04
Conditions V:	$\mathbf{Q}_i$	0.02	0.02	0.2	0.01	0.01
decrease in observation accuracy	$\mathbf{B}$	0.004	0.012	0.136	0.044	0.0208
and increase in background accuracy	$\mathbf{R}_i$	2.25	2.25	20.25	1.0	1.0

case with the number of minimization iterations reducing from 12 for Method 1, to 10 for both Method 2 and Method 3. This is because the increase in variances in the 4D-Var cost function leads to the erroneous model trajectory no longer being so tightly constrained to the observations, and therefore a solution is more easily found. This work can be related to that of Haben *et al.* (2011), where it was shown that the 4D-Var problem becomes better conditioned, and therefore involves fewer minimization iterations, with increased observation error variances.

An increase in analysis accuracy does not necessarily lead to an increase in forecast accuracy. This is demonstrated in Figure 6, which shows the RMSE of the analysis trajectories throughout the assimilation window (middle column) and the RMSE of the forecast trajectories beyond the assimilation window over 50 time steps (right column). When we include the model-error statistics in the combined error covariance matrix, we are letting the analysis model trajectory depart further from the observations (which are measuring the truth). Even if the true initial conditions were used, we would expect a poor forecast due to the use of the erroneous model. In contrast, SC4DVar when not accounting for model error tries to fit all observations throughout the window, and so the poor analysis at the initial time compensates for the model error to give an improved forecast at a later time; this effect has also been noted by Fowler and Lawless (2016) and Griffith and Nichols (2000).

6.2.3. *Sensitivity of results*

Work with an erroneous scalar model in section 3.2 showed that there was the most improvement to the analysis accuracy when replacing the observation-error statistics in the cost function with combined error statistics, in the presence of large model-error variance and a large background-error variance in comparison with the observation-error variance. Next we investigate whether these characteristic features hold with use of the erroneous nonlinear model (37) in SC4DVar.

We investigate the effect of increasing and decreasing the size of the model error on the analysis accuracy. We previously defined $\mathbf{Q}_i$ at each time t_i using Conditions I in Table 3. We reduce the standard deviations of the model error for each of the model state variables by a factor of two and label these Conditions II, whereas in Conditions III we increase the model error standard deviations by a factor of two. Conditions IV are the initial model error standard deviations increased by a factor of five. For each of the Conditions II, III and IV we recalculate the diagonal entries in $\widetilde{\mathbf{R}}_i^*$ at observation times ($i = 10, 20, 30, 40, 50$) with a sample size of 1000 and subsequently conduct 100 strong-constraint 4D-Var runs. Each individual data assimilation run uses independent model error vectors, vectors of observations and background vectors and minimizes the nonlinear 4D-Var cost function using Methods 1, 2 and 3. This produces 100 analysis vectors for each method. The results are shown in Figures 6(b)–(d).

Comparing the levels of model error within the data assimilation process, where Conditions II is lowest and is Conditions IV highest, we demonstrate that the larger the model error, the more significant the increase in accuracy is when accounting for the model error (Methods 2 and 3), as opposed to when not (Method 1). The number of iterations that the minimization algorithm performs is lower when accounting for model error than when not accounting for model error. We also note, similarly to the previous experiment, that an increase in analysis accuracy does not necessarily lead to an increase in forecast accuracy. In Figures 6(b)–(d), it is seen that the larger the model error, the greater the degradation in the forecast when using Methods 2 and 3 than with Method 1.

The last of these experiments was conducted with very large model error variances (Conditions IV in Table 3). The decrease in analysis RMSE, when using the estimated combined error variances in Method 3, is very significant (Figure 6(d)). We re-emphasise that this remarkable improvement to the analysis accuracy is due to the specification of the diagonal entries in the estimated matrix $\widetilde{\mathbf{R}}^*$ (32) which did not require explicit knowledge of the model-error statistics in $\mathbf{Q}_i$. We also have disregarded any cross-covariance information in $\widetilde{\mathbf{R}}^*$, as we are only using the diagonal entries in these experiments. The impact of including this information in $\widetilde{\mathbf{R}}^*$ could produce even further improvement in the analysis accuracy and is an area of further work.

The ratio between the accuracy of the background error and observation error is also important when investigating the difference between the accuracy of the analysis obtained when accounting for model error, as opposed to not. We demonstrate this property by both increasing the accuracy of the background model state (by reducing the standard deviations in $\mathbf{B}$ by a factor of five) and decreasing the accuracy of the observations (by increasing the standard deviations in $\mathbf{R}_i$ by a factor of five; Conditions V in Table 3). The observations are now far less accurate than the background. We set $\mathbf{Q}_i$ back to Conditions I as stated in Table 3. We recalculate the diagonal entries of $\widetilde{\mathbf{R}}_i^*$, with the error covariance matrices as specified in Conditions V using a sample size of 1000. We subsequently conduct 100 strong-constraint 4D-Var runs for Methods 1, 2 and 3. Each individual data assimilation run uses independent model error vectors, vectors of observations and background vectors. This produces 100 analysis vectors for each method. Figure 6(e) shows that, in general, accounting for the error in the model can still produce an analysis of higher statistical accuracy, however the significance of this increase is minimized when the ratio $r = \sigma_b^2/\sigma_o^2$ is significantly reduced for each of the model state

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

variables. For Conditions I, the ratio $r = \sigma_b^2/\sigma_o^2$ for each of the model state variables x, y, z, w and v is 1.11, 3.33, 4.20, 27.5 and 13 respectively. However, for Conditions V the ratios r are far smaller at 0.002, 0.005, 0.007, 0.044 and 0.0208 for each of the respective model state variables x, y, z, w and v.

7. Conclusion

In this article we have developed a SC4DVar method which is able to account for model error over the assimilation window. Unlike WC4DVar, this method does not need an explicit estimate of the model-error covariances and does not try to estimate the model error but instead allows the analysis to be less constrained by observations at the end of the assimilation window. In doing so, this new method is able to find a more accurate analysis than SC4DVar when model error is not accounted for.

The new method accounts for model error by weighting the mismatch between the observations and the evolved model state using a combined model- and observation-error covariance, $\widehat{\mathbf{R}}^*$ rather that the observation-error covariances alone, $\widehat{\mathbf{R}}$. In theory this provides an optimal estimate of the true initial conditions, satisfying the assumption that this is the minimum variance estimate for a linear system.

Use of these combined error covariance statistics requires specification of the model-error statistics, which are often unknown. Therefore, in section 4 we formulated a method to estimate the combined model-error and observation-error covariance matrix using a sample of backgrounds and observations. This follows a similar methodology to the use of covariance innovation consistency diagnostics (Dee, 1995; Desroziers *et al.*, 2005; Ménard, 2016) for estimating observation-error variances alone. The difference here is that we need a window of observations to allow also for an estimate of model error. There are many potential difficulties to using such estimates of the error covariances in practice. To use the sampled estimate of the combined model- and observation-error covariance it is necessary to subtract an estimate of the evolved background-error covariances. This could potentially be estimated using the randomization method in which a sample from a standard normal distribution is transformed using the square root of the background-error covariances matrix and than propagated forward to the time of interest using the (potentially nonlinear) model. The sensitivity of the new method to this estimate of $\widehat{\mathbf{H}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{T}}$ should be investigated. Another potential issue is that the estimate of $\widehat{\mathbf{R}}^*$ relies on the accuracy of $\mathbf{B}$ which itself is difficult to quantify. This is a known issue when attempting to estimate $\mathbf{R}$ alone using samples of background minus observations (Ménard, 2016; Waller *et al.*, 2016a). The sampling method also relies on a sample of the model error. We anticipate that the use of stochastic physics alone may not be enough to represent all sources of model error. Assuming some degree of homogeneity and isotropy in the model error, an alternative approach could be to estimate $\widehat{\mathbf{R}}^*$ off-line allowing for averaging over time and space. This will be investigated in future work.

This new method was demonstrated on two dynamical systems, the linear advection equation and a nonlinear idealized coupled atmosphere–ocean model. In these numerical experiments, only the accumulated model-error variance was accounted for, as this is more feasible. However, it was shown that, when model error of a random nature is present at each time step, replacing the observation-error covariance matrix with combined model-error and observation-error variances statistically improves the analysis accuracy over a number of sample experiments. It was also demonstrated that an improved analysis does not necessarily lead to improvements in the forecast accuracy, due to the erroneous model. Therefore, this methodology could be best suited to the application of reanalysis, as opposed to forecasting.

In the experimental set-up, only the diagonal elements of $\widehat{\mathbf{R}}^*$ were estimated and used. Future work should look at the sensitivity of the method to the use of spatial/multivariate cross-covariances, and to temporal cross-correlations.

Acknowledgements

This work has been funded by a grant from the Natural Environmental Research Council (NERC), grant number NE/J005835/1. Lawless and Fowler have also received funding from NERC's support for the National Centre for Earth Observation (NCEO).

Appendices

Appendix A: Derivation of $\mathbf{Q}^*$

In (12) and (13) the covariance between the accumulation of model error at time step i and k is defined as

$$\mathbf{Q}^*{}_{(i,k)} = \mathbf{H}_i\left\langle\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\rightarrow i}\boldsymbol{\eta}_j\left(\sum_{j=1}^{k}\mathbf{M}^{e}_{j\rightarrow k}\boldsymbol{\eta}_j\right)^{\mathrm{T}}\right\rangle\mathbf{H}_k^{\mathrm{T}}. \tag{A1}$$

It is assumed that the model error $\boldsymbol{\eta}_j \sim N(\mathbf{0},\mathbf{Q}_j)$ is uncorrelated in time. Using this assumption we can write $\mathbf{Q}^*{}_{(i,k)}$ in terms of the model error covariance at each time step i, $\mathbf{Q}_i$.

$$\begin{aligned}
\mathbf{Q}^*{}_{(i,k)} &= \mathbf{H}_i\left\langle\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\rightarrow i}\boldsymbol{\eta}_j\left(\sum_{j=1}^{k}\mathbf{M}^{e}_{j\rightarrow k}\boldsymbol{\eta}_j\right)^{\mathrm{T}}\right\rangle\mathbf{H}_k^{\mathrm{T}}\\
&= \mathbf{H}_i\left\langle(\mathbf{M}^{\mathrm{e}}_{1\rightarrow i}\boldsymbol{\eta}_1+\ldots+\boldsymbol{\eta}_i)(\mathbf{M}^{\mathrm{e}}_{1\rightarrow k}\boldsymbol{\eta}_1+\ldots+\boldsymbol{\eta}_k)^{\mathrm{T}}\right\rangle\mathbf{H}_k^{\mathrm{T}}\\
&= \mathbf{H}_i\Big[\mathbf{M}^{\mathrm{e}}_{1\rightarrow i}\mathbf{Q}_1\mathbf{M}^{\mathrm{e}}{}^{\mathrm{T}}_{1\rightarrow k}+\ldots+\mathbf{M}^{\mathrm{e}}_{j\rightarrow i}\mathbf{Q}_j\mathbf{M}^{\mathrm{e}}{}^{\mathrm{T}}_{j\rightarrow k}+\ldots\\
&\qquad+\mathbf{M}^{\mathrm{e}}_{\min(i,k)\rightarrow i}\mathbf{Q}_{\min(i,k)}\mathbf{M}^{\mathrm{e}}{}^{\mathrm{T}}_{\min(i,k)\rightarrow k}\Big]\mathbf{H}_k^{\mathrm{T}}\\
&= \mathbf{H}_i\left[\sum_{j=1}^{\min(i,k)}\mathbf{M}^{\mathrm{e}}{}_{j\rightarrow i}\mathbf{Q}_j\mathbf{M}^{\mathrm{e}}{}^{\mathrm{T}}_{j\rightarrow k}\right]\mathbf{H}_k^{\mathrm{T}}.
\end{aligned}$$

Appendix B: Derivation of $\mathbf{A}^{\mathrm{e}}$, (18)

The error covariance of the analysis when the model error is not accounted for is given by

$$\mathbf{A}^{\mathrm{e}} = \left\langle\boldsymbol{\epsilon}^{\mathrm{e}}_{\mathrm{a}}(\boldsymbol{\epsilon}^{\mathrm{e}}_{\mathrm{a}})^{\mathrm{T}}\right\rangle. \tag{B1}$$

Equation (17) defined $\boldsymbol{\epsilon}^{\mathrm{e}}_{\mathrm{a}}$ as

$$\boldsymbol{\epsilon}^{\mathrm{e}}_{\mathrm{a}} = \boldsymbol{\epsilon}_{\mathrm{b}} + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{d}}^{\mathrm{e}}_{\mathrm{b}}. \tag{B2}$$

Recall $\widehat{\mathbf{d}}^{\mathrm{e}}_{\mathrm{b}} = \widehat{\mathbf{y}} - \widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{x}^{\mathrm{b}}$. At time i this can be written in terms of the observation error, $\boldsymbol{\epsilon}^{\mathrm{o}}_i$, the model error, $\boldsymbol{\eta}$, and the background error, $\boldsymbol{\epsilon}^{\mathrm{b}}_i$, using (1), (2) and (8).

$$\begin{aligned}
\widehat{\mathbf{d}}^{\mathrm{e}}_{\mathrm{b}i} &= \mathbf{y}_i - \mathbf{H}_i\mathbf{x}^{\mathrm{t}}_i + \mathbf{H}_i\mathbf{x}^{\mathrm{t}}_i - \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\rightarrow i}\mathbf{x}^{\mathrm{b}}\\
&= \boldsymbol{\epsilon}^{\mathrm{o}}_i + \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\rightarrow i}\mathbf{x}^{\mathrm{t}} + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\rightarrow i}\boldsymbol{\eta}_j - \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\rightarrow i}\mathbf{x}^{\mathrm{b}}\\
&= \boldsymbol{\epsilon}^{\mathrm{o}}_i - \mathbf{H}_i\mathbf{M}^{\mathrm{e}}_{0\rightarrow i}\boldsymbol{\epsilon}^{\mathrm{b}}_i + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^{\mathrm{e}}_{j\rightarrow i}\boldsymbol{\eta}_j\,.
\end{aligned} \tag{B3}$$

Therefore $\boldsymbol{\epsilon}^{\mathrm{e}}_{\mathrm{a}} = \boldsymbol{\epsilon}_{\mathrm{b}} + \widehat{\mathbf{K}}^{\mathrm{e}}\left(\widehat{\boldsymbol{\epsilon}}^{\mathrm{o}} + \widehat{\boldsymbol{\eta}} - \widehat{\mathbf{H}}^{\mathrm{e}}\boldsymbol{\epsilon}_{\mathrm{b}}\right)$, where

$$\widehat{\boldsymbol{\eta}} = \begin{pmatrix}\mathbf{0}\\ \mathbf{H}_1\boldsymbol{\eta}_1\\ \vdots\\ \mathbf{H}_N\sum_{j=0}^{N}\mathbf{M}^{\mathrm{e}}_{j\rightarrow N}\boldsymbol{\eta}_N\end{pmatrix}. \tag{B4}$$

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

From Appendix A we see that $\mathbf{Q}^* = <\widehat{\boldsymbol{\eta}}\widehat{\boldsymbol{\eta}}^{\mathrm{T}}>$.

We can now evaluate $\mathbf{A}^{\mathrm{e}}$. If we substitute (B2) into (B1), and use the definition of $\mathbf{Q}^*$, it can be shown that

$$\mathbf{A}^{\mathrm{e}} = \mathbf{B} + \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{eT}}\widehat{\mathbf{K}}^{\mathrm{eT}} + \widehat{\mathbf{K}}^{\mathrm{e}}\mathbf{R}\widehat{\mathbf{K}}^{\mathrm{eT}} + \widehat{\mathbf{K}}^{\mathrm{e}}\mathbf{Q}^*\widehat{\mathbf{K}}^{\mathrm{eT}} - \mathbf{B}\widehat{\mathbf{H}}^{\mathrm{eT}}\widehat{\mathbf{K}}^{\mathrm{eT}} - \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}. \tag{B5}$$

The last two terms in the above equation can be shown to be equal (by substituting $\widehat{\mathbf{K}}^{\mathrm{e}}$ from (16)).

$$\mathbf{A}^{\mathrm{e}} = \widehat{\mathbf{K}}^{\mathrm{e}}(\widehat{\mathbf{H}}^{\mathrm{e}}\mathbf{B}\widehat{\mathbf{H}}^{\mathrm{eT}} + \mathbf{R})\widehat{\mathbf{K}}^{\mathrm{eT}} + \widehat{\mathbf{K}}^{\mathrm{e}}\mathbf{Q}^*\widehat{\mathbf{K}}^{\mathrm{eT}} + (\mathbf{I} - 2\widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{H}}^{\mathrm{e}})\mathbf{B}. \tag{B6}$$

Substituting (16) for the second occurrence of $\widehat{\mathbf{K}}^{\mathrm{e}}$ only it can be shown

$$\mathbf{A}^{\mathrm{e}} = (\mathbf{I} - \widehat{\mathbf{K}}^{\mathrm{e}}\widehat{\mathbf{H}}^{\mathrm{e}})\mathbf{B} + \widehat{\mathbf{K}}^{\mathrm{e}}\mathbf{Q}^*\widehat{\mathbf{K}}^{\mathrm{eT}}. \tag{B7}$$

References

Andersson E. 2003. 'Modelling the temporal evolution of innovation statistics'. In *Recent Developments in Data Assimilation for Atmosphere and Ocean. ECMWF: Reading, UK.*

Brankart J, Candille G, Garnier F, Calone C, Melet A, Bouttier P, Brasseur P, Vernon J. 2015. A generic approach to explicit simulation of uncertainty in the NEMO model. *Geosci. Model Dev.* **8**: 1285–1297.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* **125**: 2887–2908.

Causon D, Mingham C. 2010. *Introductory Finite Difference Methods for PDEs*. Bookboon: London.

Crank J, Nicolson P. 1947. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Math. Proc. Cambridge Phil. Soc.* **43**: 50–67.

Dee DP. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.* **123**: 1128–1145, doi: 10.1175/1520-0493(1995)123<1128:oleoec>2.0.CO;2.

Dee DP, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda M, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.

Desroziers G, Berre L, Chapnik B, Poli P. 2005. Diagnosis of observation, background- and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.* **131**: 3385–3396.

Dubois M, Yiou P. 1999. Testing asynchronous coupling on simple 'ocean–atmosphere' dynamic systems. *Clim. Dyn.* **15**: 1–7.

Fowler AM, Lawless AS. 2016. An idealized study of coupled atmosphere–ocean 4D-Var in the presence of model error. *Mon. Weather Rev.* **144**: 4007–4030.

Griffith A, Nichols NK. 2000. Adjoint methods in data assimilation for estimating model error. *Flow Turbul. Combust.* **65**: 469–488.

Haben S, Lawless AS, Nichols NK. 2011. Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus A* **63**: 782–792.

Hodyss D, Nichols NK. 2015. The error of representation: Basic understanding. *Tellus A* **67**: 24822, doi: 10.3402/tellusa. v67.24822.

Howes KE. 2016. 'Accounting for model error in four-dimensional variational data assimilation', PhD thesis. University of Reading: Reading, UK.

Lea D, Drecourt J, Haines K, Martin M. 2008. Ocean altimeter assimilation with observational and model bias correction. *Q. J. R. Meteorol. Soc.* **134**: 1761–1774.

Lewis J, Lakshmivarahan S, Dhall S. 2006. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press: Cambridge, UK.

Li Z, Feng S, Liu Y, Lin W, Zhang M, Toto T, Vogelmann A, Endo S. 2015. Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation. *J. Geophys. Res. Atmos.* **120**: 654–666, doi: 10.1002/2014JD022245.

Lorenz EN. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141.

Ménard R. 2016. Error covariance estimation methods based on analysis residuals: Theoretical foundation and convergence properties derived from simplified observation networks. *Q. J. R. Meteorol. Soc.* **142**: 257–273.

Molteni F, Ferranti L, Palmer TN, Viterbo P. 1993. A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *J. Clim.* **6**: 777–795.

Rabier F, Järvinen H, Klinker E, Mahfouf J-F, Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **126**: 1143–1170.

Rawlins F, Ballard S, Bovis K, Clayton AM, Li D, Inverarity G, Lorenc AC, Payne T. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**: 347–362.

Sasaki Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Weather Rev.* **98**: 875–883.

Shutts G. 2015. A stochastic convective backscatter scheme for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **141**: 2602–2616.

Stuart A, Humphries A. 1998. *Dynamical Systems and Numerical Analysis* **2**. Cambridge University Press Cambridge, UK.

Todling R. 2015. A complementary note to 'A lag-1 smoother approach to system-error estimation': The intrinsic limitations of residual diagnostics. *Q. J. R. Meteorol. Soc.* **141**: 2917–2922.

Trémolet Y. 2006. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.* **132**: 2483–2504.

Trémolet Y. 2007. Model-error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.* **133**: 1267–1280.

Waller JA, Dance SL, Nichols NK. 2016a. Theoretical insight into diagnosing observation error correlations using observation-minus-background and observation-minus-analysis statistics. *Q. J. R. Meteorol. Soc.* **142**: 418–431.

Waller JA, Simonin D, Dance SL, Nichols NK, Ballard SP. 2016b. Diagnosing observation error correlations for Doppler radar radial winds in the Met Office UKV model using observation-minus-background and observation-minus-analysis statistics. *Mon. Weather Rev.* **144**: 3533–3551.

Weston P, Bell W, Eyre J. 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. *Q. J. R. Meteorol. Soc.* **140**: 2420–2429.

© 2017 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.