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

Asynchronous data assimilation with the EnKF in presence of additive model error

By PAVEL SAKOV[1*] and MARC BOCQUET[2], [1]*Bureau of Meteorology, Melbourne, Australia;* [2]*CEREA, Joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France*

(Manuscript received 20 July 2017; in final form 20 November 2017)

ABSTRACT

The term 'asynchronous data assimilation' (ADA) refers to modifications of sequential data assimilation methods that take into consideration the observation time. In Sakov et al. [*Tellus A*, **62**, 24–29 (2010)], a simple rule has been formulated for the ADA with the ensemble Kalman filter (EnKF). To assimilate scattered in time observations, one needs to calculate ensemble forecast observations using the forecast ensemble at observation time. Using then these ensemble observations in the EnKF update matches the optimal analysis in the linear perfect model case. In this note, we generalise this rule for the case of additive model error.

Keywords: data assimilation, ensemble Kalman filter, asynchronous data assimilation, model error, additive model error

1. Introduction

The standard sequential data assimilation (DA) framework assumes that the time is discrete, and that observations assimilated at each time step are made at the analysis time. In practice, the time intervals between consecutive analyses can be substantial, and model states within the observation window can be quite different. It is often important in such cases for a DA method to take into account, in one way or another, the timing of observations. The method is then referred to as asynchronous, in contrast to its standard sequential formulation, which is referred to as synchronous.

The ensemble Kalman filter (EnKF, Evensen, 1994) is based on the Kalman filter (KF), which is a sequential method; however, in the linear perfect model case, the EnKF can be easily adapted to accommodate the asynchronous DA (ADA). Consider the associated minimisation problem:

$$\mathbf{x}_0^\star = \arg\min\ J(\mathbf{x}_0), \tag{1a}$$

$$J(\mathbf{x}_0) = \|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{a}}\|^2_{(\mathbf{P}_0^{\mathrm{a}})^{-1}} + \sum_{i=1}^{k} \|\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\|^2_{\mathbf{R}_i^{-1}}, \tag{1b}$$

$$\mathbf{x}_i = \mathcal{M}_{i-1\to i}(\mathbf{x}_{i-1}), \quad i = 1, \ldots, k. \tag{1c}$$

Here the lower index i refers to time t_i, $\mathbf{x}_i$ is the model state, $\mathbf{x}_0^{\mathrm{a}}$ – the initial state estimate, $\mathbf{P}_0^{\mathrm{a}}$ – the initial state error covariance estimate,[1] $\mathbf{y}_i$ – observations, $\mathcal{H}_i$ and $\mathbf{R}_i$ – the corresponding observation operator and observation error covariance matrix, $\mathcal{M}_{i\to j}$ is the model resolvent that maps the state from time t_i to t_j; the norm notation $\|\mathbf{x}\|^2_{\mathbf{B}} \equiv \mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x}$ is used.

In order to simplify the notations, we define blkdiag($\mathbf{A}_1, \ldots, \mathbf{A}_k$) as the block diagonal matrix (or operator) made of the square matrices (or square operators) $\{\mathbf{A}_i\}_{i=1,\ldots,k}$. We also define vec($\mathbf{A}_1, \ldots, \mathbf{A}_k$) as the matrix (or operator) that stacks the matrices (or operators) $\{\mathbf{A}_i\}_{i=1,\ldots,k}$ that are assumed to have the same column number. With these notations, the cost function (1b) can be written as follows:

$$J(\mathbf{x}_0) = \|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{a}}\|^2_{(\mathbf{P}_0^{\mathrm{a}})^{-1}} + \|\mathbf{y} - \mathcal{H} \circ \mathcal{M}(\mathbf{x}_0)\|^2_{\mathbf{R}^{-1}}, \tag{2}$$

where

$$\mathbf{y} \equiv \mathrm{vec}(\mathbf{y}_1, \ldots, \mathbf{y}_k), \tag{3}$$

$$\mathbf{R} \equiv \mathrm{blkdiag}(\mathbf{R}_1, \ldots, \mathbf{R}_k) \tag{4}$$

are the concatenated observation vector and corresponding observation error covariance matrix, respectively, and the forward map $\mathcal{H} \circ \mathcal{M}(\mathbf{x}_0)$ relates the initial state $\mathbf{x}_0$ to the observations:

$$\mathcal{H} \circ \mathcal{M}(\mathbf{x}_0) \equiv \mathrm{vec}\left(\{\mathcal{H}_i \circ \mathcal{M}_{0\to i}(\mathbf{x}_0)\}_{i=1,\ldots,k}\right). \tag{5}$$

*Corresponding author. e-mail: pavel.sakov@bom.gov.au

Citation: Tellus A: 2018, **70**, 1414545, https://doi.org/10.1080/16000870.2017.1414545

Apart from the forward map (5) that yields forecast observations at times $t_1, \ldots, t_k$, the cost function (2) has the same form as that for a single DA cycle with synchronous observations. Consequently, in the linear case, when the tangent linear forward operator $\mathbf{H} \circ \mathbf{M}$ (the sensitivity of observations to the initial state) does not depend on $\mathbf{x}_0$, the cost function (2) can be minimised using existing EnKF solutions, provided that the standard (synchronous) observation functions are replaced by the generic forward function (5):

$$\mathcal{H}(\mathbf{E}^{\mathrm{f}}) \leftarrow \mathrm{vec}\left(\{\mathcal{H}_i \circ \mathcal{M}_{0\rightarrow i}(\mathbf{E}_0^{\mathrm{a}})\}_{i=1,\ldots,k}\right), \tag{6}$$

where $\mathbf{E}$ is the EnKF ensemble; upper "f" refers to forecast; $\mathcal{H}(\mathbf{E}^{\mathrm{f}})$ denotes the forecast ensemble observations and $\mathbf{E}_0^{\mathrm{a}}$ is the initial ensemble. In practice, this means that to assimilate asynchronously with the EnKF one simply needs to calculate ensemble observations using forecast ensemble at observation time (Sakov et al., 2010). There are no specific restrictions on $\mathbf{R}$ in (2) – in particular, it does not have to be block-diagonal – therefore in theory observation errors can be correlated in time within a DA cycle.

Before formulated in Sakov et al. (2010), the recipe (6) has been used in a scheme called 4D-LETKF (Hunt et al., 2007), where it was empirically derived from a method called 4D-EnKF (Hunt et al., 2004).

In practical terms, the ADA makes it possible to avoid storing the full ensemble at each observation time, storing only ensemble observations instead. Because in large-scale geophysical systems the model size is typically much larger than the number of observations, this yields a significant economy of resources.

Formulation (1) of the minimisation problem implies assimilation time t_0; however, one interesting feature of the EnKF in the linear perfect model case is the time invariance of the optimising ensemble transform. If the update is conducted at time t_i by applying transform $\mathbf{X}$ to a forecast ensemble $\mathbf{E}_i^{\mathrm{f}}$,

$$\mathbf{E}_i^a = \mathbf{E}_i^f \mathbf{X}, \tag{7}$$

then, to the same effect, it can be applied at other time t_j:

$$\mathbf{E}_j^{\mathrm{a}} = \mathbf{M}_{i\rightarrow j}\mathbf{E}_i^{\mathrm{a}} = \mathbf{M}_{i\rightarrow j}(\mathbf{E}_i^{\mathrm{f}}\mathbf{X}) = (\mathbf{M}_{i\rightarrow j}\mathbf{E}_i^{\mathrm{f}})\mathbf{X} = \mathbf{E}_j^{\mathrm{f}}\mathbf{X}. \tag{8}$$

This time invariance of the ensemble transform has been used in the ensemble Kalman smoother (Evensen, 2003) and made it possible to interpret the ensemble transform as a linear combination of the ensemble trajectories in the 4D-EnKF (Hunt et al., 2004).

The above framework for the ADA with the EnKF is based on two important assumptions: linearity and perfect model. While the KF is a linear method, it can be (and indeed is) successfully applied to weakly nonlinear cases, when the Jacobians $\mathbf{M}(\mathbf{x}) = \nabla_x \mathcal{M}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x}) = \nabla_x \mathcal{H}(\mathbf{x})$ do not change much through the DA cycle and over the characteristic uncertainty range of the state. The same principle motivates the applicability of the approach (6) to the ADA with the EnKF. This is why we kept the nonlinear notation for the forward operator.[2] In strongly nonlinear cases, the method is, generally, no longer applicable, and iterative minimisation algorithms (e.g. Gu and Oliver, 2007; Bocquet and Sakov, 2014) must be used.

The situation in the case with the model error is somewhat similar. Various aspects of additive deterministic (possibly time-correlated) model error and how formulation (1) should be generalised have been considered in Lorenc (2003), Trémolet (2006) and Carrassi and Vannitsem (2010). However, because (1c) assumes a lossless transfer of information between DA system states in time by the model, it has not been clear whether the concept of the ADA can still be useful in cases with substantial model error. Recently, an iterative method for strongly nonlinear systems with additive stochastic model error called IEnKF-Q has been proposed (Sakov et al., in press) that successfully employs combining propagated ensemble anomalies and model noise ensemble anomalies into a single ensemble of anomalies to obtain an algebraically simple solution for the Gauss–Newton minimisation. The authors suggested that a similar approach can possibly be used to generalise the concept of the ADA with the EnKF in the case of additive model error. This study proposes such a generalisation.

2. The minimisation problem

Below we mainly follow the formulation and approach to the solution of the minimisation problem from Sakov et al. (in press), with the aim to put up an analogue of the recipe (6) that would permit a simultaneous assimilation of observations made at different times in the case of additive model error without loss of optimality. Sakov et al. (in press) considered a single DA cycle starting at time t_0 with observations at t_1, while we consider a cycle with an arbitrary number of model steps:

$$\{\mathbf{x}_0^\star, \ldots, \mathbf{x}_k^\star\} = \arg\min_{\{\mathbf{x}_0,\ldots,\mathbf{x}_k\}} J(\mathbf{x}_0, \ldots, \mathbf{x}_k), \tag{9a}$$

$$J(\mathbf{x}_0, \ldots, \mathbf{x}_k) = \left\|\mathbf{x}_0 - \mathbf{x}_0^{\mathrm{a}}\right\|^2_{(\mathbf{P}_0^{\mathrm{a}})^{-1}} + \sum_{i=1}^{k} \left\|\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\right\|^2_{\mathbf{R}_i^{-1}} + \sum_{i=1}^{k} \left\|\mathbf{x}_i - \mathcal{M}_{i-1\rightarrow i}(\mathbf{x}_{i-1})\right\|^2_{\mathbf{Q}_i^{-1}}. \tag{9b}$$

Here $\mathbf{x}_0^{\mathrm{a}}$ is the initial state (the analysis from the previous cycle), $\mathbf{P}_0^{\mathrm{a}}$ – the initial state error covariance, $\mathbf{Q}_i$ – the model error covariance at t_i, and upper star refers to the solution of the minimisation problem.

From a statistical perspective, (9b) relates to the joint probability density function on state and observation vectors since $J(\mathbf{x}_0, \ldots, \mathbf{x}_k)$ coincides with $-2 \ln p(\mathbf{x}_0, \ldots, \mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_k)$ up to a constant. This implies that the model noise, as random variable, is assumed Gaussian, unbiased, of covariance matrix $\mathbf{Q}_i$ at t_i and uncorrelated in time (see Carrassi and Vannitsem, 2010, for a generalisation).

Let the covariances $\mathbf{P}_0^{\mathrm{a}}$ and $\mathbf{Q}_i$ be factorised by the corresponding ensemble anomalies $\mathbf{A}_0^{\mathrm{a}}$ and $\mathbf{A}_i^{\mathrm{q}}$:

$$\mathbf{A}_0^{\mathrm{a}}(\mathbf{A}_0^{\mathrm{a}})^{\mathrm{T}} = \mathbf{P}_0^{\mathrm{a}}, \quad \mathbf{A}_0^{\mathrm{a}}\mathbf{1} = \mathbf{0}, \tag{10a}$$

$$\mathbf{A}_i^{\mathrm{q}}(\mathbf{A}_i^{\mathrm{q}})^{\mathrm{T}} = \mathbf{Q}_i, \quad \mathbf{A}_i^{\mathrm{q}}\mathbf{1} = \mathbf{0}, \quad i = 1, \ldots, k, \tag{10b}$$

where $\mathbf{1}$ is a vector of entries 1. In many applications, the representation of model errors through the anomalies $\mathbf{A}_i^{\mathrm{q}}$ is more natural and convenient than through the $\mathbf{Q}_i$. A discussion on this specific topic can be found in Section 4 of Sakov et al. (in press). Let us seek the solution of the minimisation problem (9) in the ensemble space of these anomalies:

$$\mathbf{x}_0 = \mathbf{x}_0^{\mathrm{a}} + \mathbf{A}_0^{\mathrm{a}}\mathbf{w}_0, \tag{11a}$$

$$\mathbf{x}_i = \mathcal{M}_{i-1\to i}(\mathbf{x}_{i-1}) + \mathbf{A}_i^{\mathrm{q}}\mathbf{w}_i, \quad i = 1, \ldots, k. \tag{11b}$$

The weights $\mathbf{w}_i$ are similar to the control variable $\boldsymbol{\eta}_i$ in the 'forcing' formulation of the weak-constraint 4D-Var (see Equation (2) in Trémolet, 2006), although using $\mathbf{A}_i^{\mathrm{q}}\mathbf{w}_i$ instead of $\boldsymbol{\eta}_i$ has a convenience of diagonalising $\mathbf{Q}_i$.

The problem (9) becomes equivalent to

$$\{\mathbf{w}_0^\star, \ldots, \mathbf{w}_k^\star\} = \arg\min_{\{\mathbf{w}_0,\ldots,\mathbf{w}_k\}} \tilde{J}(\mathbf{w}_0, \ldots, \mathbf{w}_k), \tag{12a}$$

$$\tilde{J}(\mathbf{w}_0, \ldots, \mathbf{w}_k) = \sum_{i=0}^{k} \mathbf{w}_i^{\mathrm{T}}\mathbf{w}_i + \sum_{i=1}^{k} \|\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\|^2_{\mathbf{R}_i^{-1}}. \tag{12b}$$

Let us concatenate $\mathbf{w}_0, \ldots, \mathbf{w}_k$ into $\mathbf{w}$:

$$\mathbf{w} \equiv \mathrm{vec}\,(\mathbf{w}_0, \ldots, \mathbf{w}_k)\,; \tag{13}$$

then (12b) becomes

$$\tilde{J}(\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{w} + \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|^2_{\mathbf{R}^{-1}}, \tag{14}$$

where $\mathbf{y}$ and $\mathbf{R}$ are given by (3) and (4), and

$$\mathbf{x} \equiv \mathrm{vec}(\mathbf{x}_0, \ldots, \mathbf{x}_k), \tag{15}$$

$$\mathcal{H}(\mathbf{x}) \equiv \mathrm{vec}\left(\{\mathcal{H}_i(\mathbf{x}_i)\}_{i=1,\ldots,k}\right). \tag{16}$$

The cost function (14) is similar to the formalism developed by Lorenc (2003) or to Equation (11) of Desroziers et al. (2014), but differs in that it is not incremental and can be obtained directly.

In (14), $\mathbf{x}$ is a function of $\mathbf{w}$ because of (11). Let us expand $\mathbf{x}$ in power series of $\mathbf{w}$ and retain the linear term only:

$$\mathbf{x} = \mathbf{x}^{\mathrm{f}} + \mathbf{A}\mathbf{w} + O(\|\mathbf{w}\|^2), \tag{17}$$

where[3]

$$\mathbf{x}^{\mathrm{f}} \equiv \mathrm{vec}\left(\{\mathcal{M}_{0\to i}(\mathbf{x}_0^{\mathrm{a}})\}_{i=0,\ldots,k}\right), \tag{18}$$

and

$$\mathbf{A} \equiv \mathrm{vec}(\tilde{\mathbf{A}}_0, \ldots, \tilde{\mathbf{A}}_k), \tag{19a}$$

$$\tilde{\mathbf{A}}_i \equiv [\mathbf{A}_i, \mathbf{0}], \quad i = 0, \ldots, k, \tag{19b}$$

$$\mathbf{A}_i \equiv \begin{cases} \mathbf{A}_0^{\mathrm{a}}, & i = 0 \\ [\mathbf{M}_{i-1\to i}\mathbf{A}_{i-1}, \mathbf{A}_i^{\mathrm{q}}], & i = 1, \ldots, k. \end{cases} \tag{19c}$$

Here the ensemble anomalies $\tilde{\mathbf{A}}_i$ are completed by zeros so that they all have the same number of columns as $\mathbf{A}_k$. Note that the augmentation of the propagated state error anomalies and model error anomalies has also been used in the reduced rank square root filter by Verlaan and Heemink (1997) in their Equation (28). Substituting (17) into (14) yields a form convenient for Gauss–Newton minimisation:

$$\tilde{J}(\mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{w} + \left\|\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{f}}) - \mathbf{Y}\mathbf{w} + O(\|\mathbf{w}\|^2)\right\|^2_{\mathbf{R}^{-1}}, \tag{20}$$

where

$$\mathbf{Y} \equiv \mathrm{vec}\left(\{\mathbf{H}_i\tilde{\mathbf{A}}_i\}_{i=1,\ldots,k}\right). \tag{21}$$

In the nonlinear case, (20) can be minimised by Gauss–Newton iterations. This involves using the approximate Hessian obtained by neglecting nonlinear terms $O(\|\mathbf{w}\|^2)$. In the linear case, the cost function becomes quadratic, and the solution is given by:

$$\mathbf{x}^\star = \mathbf{x}^{\mathrm{f}} + \mathbf{A}\mathbf{w}^\star, \tag{22a}$$

$$\mathbf{A}^\star = \mathbf{A}\mathbf{T}, \quad \mathbf{T} = \mathbf{D}^{-1/2}\mathbf{U}, \tag{22b}$$

$$\mathbf{w}^\star = \mathbf{D}^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y} - \mathcal{H}(\mathbf{x}^{\mathrm{f}})\right], \tag{22c}$$

$$\mathbf{D} \equiv \mathbf{I} + \mathbf{Y}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}, \tag{22d}$$

where $(\cdot)^{1/2}$ denotes the unique positive (semi)definite square root of a positive (semi)definite matrix, and $\mathbf{U}$ is an arbitrary mean preserving orthonormal matrix[4]: $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I},\ \mathbf{U}\mathbf{1} = \mathbf{1}$.

These equations provide the update of the ensemble at all time levels. Similarly to the perfect model case, the same coefficients $\mathbf{w}^\star$ and the same ensemble transform $\mathbf{T}$ are used over the whole assimilation window. However, due to the augmentation of the propagated ensemble anomalies with model error anomalies in (19c), the updated trajectories of the initial ensemble $\mathbf{E}_0^\star$ are no longer continuous in time.

3. Asynchronous DA

Based on the solution of the minimisation problem in the previous section, we can now formulate a framework (rules) for the ADA with the EnKF in presence of additive model error. According to equations (22), to update the system state, one needs to calculate the forecast observations $\mathcal{H}(\mathbf{x}^{\mathrm{f}})$ and forecast ensemble observation anomalies $\mathbf{Y}$; they in turn allow one to calculate the Hessian $2\mathbf{D}$ (approximate Hessian in the nonlinear case), coefficients $\mathbf{w}^\star$ and transform $\mathbf{T}$. Therefore, the essence of the ADA with the EnKF can be formulated as follows:

$$\mathcal{H}(\mathbf{x}^{\mathrm{f}}) = \mathrm{vec}\left(\{\mathcal{H}_i(\mathbf{x}_i^{\mathrm{f}})\}_{i=1,\ldots,k}\right), \tag{23}$$

$$\mathbf{Y} = \mathrm{vec}\left(\{\mathbf{H}_i\tilde{\mathbf{A}}_i\}_{i=1,\ldots,k}\right), \tag{21}$$

where $\tilde{\mathbf{A}}_i$ is defined by (19b) and (19c).

Subject to the new evolution equation for ensemble anomalies (19c), the rules (21) and (23) can be interpreted in the same way as in the perfect model case: calculate ensemble observations using the forecast ensemble at observation time.

The sequence of the complete linear analysis cycle is synthesised in Fig. 1 (upper panel).

4. Discussion

4.1. Increasing ensemble size

The update equations (22) have the same form as in the perfect model case; however, the propagation of the ensemble anomalies in (19c) is substantially different. According to (19c), in presence of model error, the forecast ensemble of anomalies is not only transformed with the tangent linear model, but is also augmented with the corresponding ensemble of model error anomalies. This procedure implicitly implements the forecast covariance equation of the KF

$$\mathbf{P}_i = \mathbf{M}_i\mathbf{P}_{i-1}\mathbf{M}_i^{\mathrm{T}} + \mathbf{Q}_i, \tag{24}$$

so that $\mathbf{A}_i\mathbf{A}_i^{\mathrm{T}} = \mathbf{P}_i$, and manifests an effective increase of the ensemble size at every time step:

$$m_i = m_{i-1} + m_i^{\mathrm{q}}, \quad i = 1, \ldots, k, \tag{25}$$

where m_i is the size of $\mathbf{A}_i$; m_i^{q} is the size of $\mathbf{A}_i^{\mathrm{q}}$ and $m_0 \equiv m$ is the size of $\mathbf{A}_0^{\mathrm{a}}$, i.e. the initial ensemble size. The practical choice for the m_i^{q} has been discussed in Section 4 of Sakov et al. (in press).

Note that this increase of the ensemble size may not be strictly necessary for implementing (24). For example, if the column space of $\mathbf{A}_i^{\mathrm{q}}$ is a subspace of the column space of $\mathbf{M}_i\mathbf{A}_{i-1}$, then adding model error covariance can be done without changing the ensemble size. Yet, the augmentation of the state error and model error ensemble anomalies in (19c) is an essential part of the described minimisation procedure, regardless of whether it is possible to factorise the forecast state error covariance with a smaller ensemble or not. It enables to concatenate the vectors of ensemble weights $\mathbf{w}_i$ into a single vector $\mathbf{w}$ in (13) so that each component $\mathbf{w}_i$ of this vector acts only on the columns of the augmented ensemble associated with the model error added at time t_i (except that $\mathbf{w}_0$ is indeed associated with the initial state error). In this way, the minimisation uses the same vector of coefficients $\mathbf{w}$ over the whole observation window, similarly to the perfect model case.

4.2. Nonlinearity and approximation of $\mathbf{M}$ and $\mathbf{H}$

Although the EnKF is a linear method targeting weakly nonlinear systems, it is still desirable to optimise it for strongly nonlinear situations, for two reasons. Firstly, it is quite common for a weakly nonlinear system to become strongly nonlinear during relatively short and infrequent transition periods. Secondly, there are a variety of EnKF-based iterative methods for strongly nonlinear systems that would benefit from such an optimisation.

The EnKF is a derivative-less method: it approximates the Jacobians $\mathbf{M}_i$ and $\mathbf{H}_i$ based on application of the forward function $\mathcal{H}_i \circ \mathcal{M}_i$ to the ensemble of model states $\mathbf{E}_{i-1}$. In the linear case, the results of these approximations do not depend on the ensemble spread; in the nonlinear case to achieve correct sampling of nonlinear effects, the magnitude of the ensemble anomalies should be of order of the standard deviation of the state error. For a system with additive model error, this means that the forecast error anomalies should have magnitude similar to that of samples from $\mathcal{N}(\mathbf{0}, \mathbf{P}_i)$, where the forecast covariance $\mathbf{P}_i$ is given by (24).

Such sampling can be associated with the statistical estimator of the covariance,

$$\mathbf{P} = \frac{1}{m_A - 1}\mathbf{A}\mathbf{A}^{\mathrm{T}}, \tag{26}$$

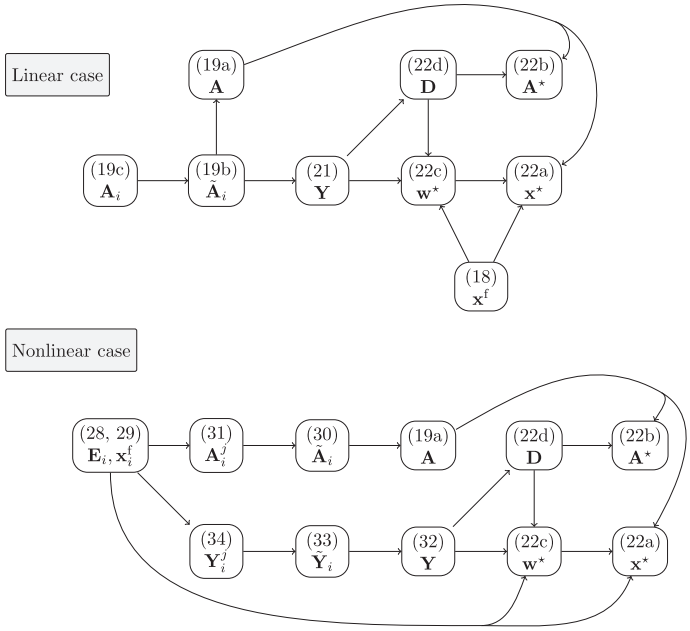

Fig. 1. The sequence of the complete analysis cycle in both the linear (upper) and the nonlinear (lower) case. From left to right, the nodes indicate the equation to use as referred to in the paper and the yielded quantities. The second, lower chain in the nonlinear case is needed if the observation operator is nonlinear. Arrows mean that the computation of the target node content requires the content of the origin node.

where $\mathbf{A}$ is an ensemble of m_A anomalies. Using this factorisation yields the characteristic magnitude of the anomalies of the order of its standard deviation, regardless of the ensemble size.

The minimisation procedure in Section 2 uses a different factorisation,

$$\mathbf{P} = \mathbf{A}\mathbf{A}^{\mathrm{T}}, \tag{27}$$

which simplifies the algebraic side. Using this factorisation reduces the characteristic magnitude of anomalies with increasing ensemble size. To obtain anomalies of physically 'right' magnitude, one needs to scale them up with a factor of $\sqrt{m_A - 1}$, but only if all samples in the ensemble are drawn from the same pool. If, for example, we introduce a new ensemble

$$\mathbf{B} = [\mathbf{A}, \mathbf{0}],$$

then we should still scale it with $\sqrt{m_A - 1}$ rather than with $\sqrt{m_B - 1}$, where m_B is the size of $\mathbf{B}$.

Because the magnitudes of the anomalies $\mathbf{M}_i\mathbf{A}_{i-1}$ and $\mathbf{A}_i^{\mathrm{q}}$ in (19c) cannot be assumed to be similar, it seems technically challenging to put up a generic re-scaling procedure for switching between factorisations (26) and (27) in the course of minimisation. This leaves us the following two main options.

(1) Using factorisation (27) over the whole DA cycle, which means no rescaling before or after application of nonlinear functions $\mathcal{M}$ and $\mathcal{H}$. This approach underestimates the ensemble spread in approximations and shifts the minimisation towards the Newton method from the more inherent for the EnKF (and generally more robust) secant method.

(2) Scaling components of $\mathbf{A}_i$ appended at different time steps according to the size of these components before application of $\mathcal{M}$ and $\mathcal{H}$, and scaling back after that. This approach still does not make all anomalies of $\mathbf{A}_i$ being sampled from $\mathcal{N}(\mathbf{0}, \mathbf{P}_i)$, but can be a good practical compromise. It is elaborated below.

First, we define the forecast ensembles $\mathbf{E}_i$ as

$$\mathbf{E}_i = \begin{cases} \mathbf{E}_0^{\mathrm{a}}, & i = 0, \\ \left[\mathcal{M}_{i-1\to i}(\mathbf{E}_{i-1}),\right. & \\ \quad \left. \mathbf{x}_i^{\mathrm{f}}\mathbf{1}^{\mathrm{T}} + \sqrt{m_i^{\mathrm{q}} - 1}\mathbf{A}_i^{\mathrm{q}}\right], & i = 1, \ldots, k, \end{cases} \tag{28}$$

where $\mathbf{E}_0^{\mathrm{a}}$ is the initial ensemble and

$$\mathbf{x}_i^{\mathrm{f}} \equiv \mathcal{M}_{i-1\to i}(\mathbf{E}_{i-1})\,\mathbf{1}/m_{i-1}. \tag{29}$$

We then define the new anomalies $\tilde{\mathbf{A}}_i$ that yield $\mathbf{A}$ via (19). They are composed of sub-blocks $\mathbf{A}_i^j$, each of them corresponding to the ensemble appended at time t_j:

$$\tilde{\mathbf{A}}_i \equiv [\mathbf{A}_i^0, \ldots, \mathbf{A}_i^k], \tag{30}$$

where

$$\mathbf{A}_i^j \equiv \begin{cases} (\mathbf{E}_i^0 - \mathbf{x}_i^{\mathrm{f}}\mathbf{1}^{\mathrm{T}})/\sqrt{m_0 - 1}, & j = 0 \\ (\mathbf{E}_i^j - \mathbf{x}_i^{\mathrm{f}}\mathbf{1}^{\mathrm{T}})/\sqrt{m_j^{\mathrm{q}} - 1}, & 1 \le j \le i \\ \mathbf{0}, & i < j \le k. \end{cases} \tag{31}$$

The submatrix $\mathbf{E}_i^0$ corresponds to the first m_0 columns of $\mathbf{E}_i$; $\mathbf{E}_i^j$, for $1 \le j \le k$, corresponds to the submatrix from column $m_{j-1} + 1$ to column m_j of $\mathbf{E}_i$.

The forecast observation anomalies (21) are then calculated as

$$\mathbf{Y} = \mathrm{vec}\,(\mathbf{Y}_1, \ldots, \mathbf{Y}_k)\,, \tag{32}$$

where

$$\mathbf{Y}_i \equiv [\mathbf{Y}_i^0, \ldots, \mathbf{Y}_i^k], \tag{33}$$

and

$$\mathbf{Y}_i^j \equiv \begin{cases} [\mathcal{H}_i(\mathbf{E}_i^0) - \mathcal{H}_i(\mathbf{x}_i^{\mathrm{f}})\mathbf{1}^{\mathrm{T}}]/\sqrt{m_0 - 1}, & j = 0 \\ [\mathcal{H}_i(\mathbf{E}_i^j) - \mathcal{H}_i(\mathbf{x}_i^{\mathrm{f}})\mathbf{1}^{\mathrm{T}}]/\sqrt{m_j^{\mathrm{q}} - 1}, & 1 \le j \le i \\ \mathbf{0}, & i < j \le k. \end{cases} \tag{34}$$

To summarise, the sequence of the complete nonlinear analysis cycle is synthesised in Fig. 1 (lower panel).

4.3. *Observations correlated in time*

The formulation of the minimisation problem (9) assumes that observations at different times are noncorrelated, so that the merged observation error covariance matrix $\mathbf{R}$ in (4) is block-diagonal; however, we can see no problem in removing this restriction, that is assuming that the observations within a DA cycle can be time correlated.

4.4. *Ensemble resizing*

One practical implication of the proposed framework for ADA in presence of additive model error is the need for ensemble resizing at the end of a cycle to cut back the increased ensemble size due to (19c). If the augmented ensemble $\mathbf{A}_k$ has the same rank as $\mathbf{A}_0$, then the ensemble size reduction from m_k to m_0 can be performed losslessly (that is, without changing the analysed state error covariance). In large-scale geophysical EnKF systems, the ensemble size is always much smaller than the model size, and realistic implementations of model noise can be expected, generally, to project onto the left null-space of the ensemble. In this case, the resizing will involve removing the principal components with the smallest variance (Verlaan and Heemink, 1997; Raanes et al., 2015; Sakov et al., in press).

4.5. *Observation time binning*

The minimisation problem (9) associates each time step with observations $(\mathbf{y}_i, \mathbf{R}_i)$ and model error $\mathbf{Q}_i$. In practice the observations can be distributed in time continuously. They will need then to be binned by some time intervals, and model error estimates will need to be prescribed for each of these intervals. Because of the overhead associated with augmenting the ensembles of model error anomalies into the ensemble anomalies by (19c) one may have to use a rather coarse time binning for the model error. It is then possible to assume that the forward function $\mathcal{H}_i \circ \mathcal{M}_i$ in (34) propagates ensemble to the time of each particular observation, as the overhead associated with the time binning of ensemble observations can be much smaller. This is the same observation time binning strategy as the one proposed by Trémolet (2006) with his Equation (9).

4.6. *ADA: further extending the KF*

The EnKF yields a large number of advantages over the KF: scalability, derivative-less formulation, ease of implementing the covariance propagation equation, robustness, and so on. Most of these advantages can be viewed as concerning the technical side of the Kalman filtering. By contrast, the ADA aims at extending the very framework of the sequential DA. While in the perfect model case it is possible to modify the square root formulations of the KF to permit the ADA similarly to that in the EnKF, the technique proposed in this note involves using nonsquare covariance factorisations of varying size within a DA cycle. We see it as yet another step in the evolution of the KF-based methods towards greater universality, well beyond the original scope of the KF.

5. Summary

Equations (23), (21) and (19c) provide a framework for the ADA with the EnKF in presence of additive model error. Its derivation in the course of solving the minimisation problem warrants its optimality in the linear case.

The most essential new component of the proposed framework is Equation (19c) for propagation of ensemble anomalies. This equation manifests an effective increase of the ensemble size at each time step by the size of the ensemble of model error anomalies. We argue that this increase is necessary for the optimal ADA in presence of additive model error regardless of whether the model error anomalies project onto the column space

of the ensemble or not; in practice, it deems it necessary to resize the ensemble at the end of the cycle.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. The upper label 'a' refers to analysis as, later on, these initial state and error covariance matrix will coincide with the outcome of a previous analysis.
2. Another reason is that in the linear case functions $\mathcal{M}$ and $\mathcal{H}$ can be affine: $\mathcal{M}(\mathbf{x}) = \mathbf{m} + \mathbf{M}\mathbf{x}$, $\mathcal{H}(\mathbf{x}) = \mathbf{h} + \mathbf{H}\mathbf{x}$.
3. In the EnKF instead of explicit propagation of state, it is common to use the average of the propagated ensemble: $\mathcal{M}(\mathbf{x}) \leftarrow \mathcal{M}(\mathbf{E})\mathbf{1}/m$, where m is the size of ensemble $\mathbf{E}$.
4. The presence of $\mathbf{U}$ here symbolically accounts for the variety of the deterministic EnKF schemes.

References

Bocquet, M. and Sakov, P. 2014. An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.* **140**, 1521–1535.

Carrassi, A. and Vannitsem, S. 2010. Accounting for model error in variational data assimilation: a deterministic formulation. *Mon. Wea. Rev.* **138**, 3369–3386.

Desroziers, G., Camino, J.-T. and Berre, L. 2014. 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Q. J. R. Meteorol. Soc.* **140**, 2097–2110.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162.

Evensen, G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Gu, Y. and Oliver, D. S. 2007. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE J.* **12**, 438–446.

Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J. and co-authors 2004. Four-dimensional ensemble Kalman filtering. *Tellus* **56A**, 273–277.

Hunt, B. R., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D* **230**, 112–126.

Lorenc, A. C. 2003. Modelling of error covariances by 4D-Var data assimilation. *Q. J. R. Meteor. Soc.* **129**, 3167–3182.

Raanes, P. N., Carrassi, A. and Bertino, L. 2015. Extending the square root method to account for additive forecast noise in ensemble methods. *Mon. Wea. Rev.* **143**, 3857–38730.

Sakov, P., Evensen, G. and Bertino, L. 2010. Asynchronous data assimilation with the EnKF. *Tellus A* **62**, 24–29.

Sakov, P., Haussaire, J.-M. and Bocquet, M. In press. An iterative ensemble Kalman filter in presence of additive model error. *Q. J. R. Meteorol. Soc.*. DOI: 10.1002/qj.3213.

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

Verlaan, M. and Heemink, A. W. 1997. Tidal flow forecasting using reduced rank square root filters. *Stoch. Hydrol. Hydraul.* **11**, 349–368.