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

Tellus (2011), 63A, 805–816
Printed in Singapore. All rights reserved

TELLUS

A POD-based ensemble four-dimensional variational assimilation method

By XIANGJUN TIAN[1,2]*, ZHENGHUI XIE[2] and QIN SUN[2], [1]*ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China;* [2]*LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

(Manuscript received 9 November 2010; in final form 18 April 2011)

ABSTRACT

In this paper, a POD-based ensemble four-dimensional variational data assimilation method (referred to as PODEn4DVar) is proposed on the basis of the proper orthogonal decomposition (POD) and ensemble forecasting techniques. The ensemble forecasts are conducted to obtain the model perturbations (MPs) and their corresponding observation perturbations (OPs). Under the assumption of the linear relationship between the MPs and the OPs, the POD transformation is applied to the OP space rather than the MP space directly, which substantially decreases the computational costs. The optimal MP and its corresponding OPs is thus represented by the transformed MP ensemble and their related OP orthogonal base vectors to fit the 4-D observation innovations in the assimilation window. Further, the implementation of the forecast model ensemble update is successfully implemented by replacing the single 4-D observation innovation with the ensemble of innovation vectors. The feasibility and effectiveness of the PODEn4DVar are demonstrated in an idealized model with simulated observations. It is found that the PODEn4DVar is capable of outperforming both 4DVar and the EnKF under both perfect and imperfect-model scenarios with lower computational costs compared with EnKF.

1. Introduction

The four-dimensional variational data assimilation (4DVar) method (e.g. Lewis and Derber, 1985; Courtier et al., 1994) has been a very successful technique used in operational weather prediction (NWP) at many operational numerical weather forecast centres (e.g. Bormann and Thepaut, 2004; Park and Zou, 2004; Caya et al., 2005; Bauer et al., 2006; Rosmond and Xu, 2006; Gauthier et al., 2007), largely thanks to the increment method (Courtier et al., 1994) and adjoint technique (e.g. Lewis and Derber, 1985; Le Dimet and Talagrand, 1986; Courtier and Talagrand, 1987). The 4DVar technique has two attractive features: (1) the physical model can provide a temporal smoothness constraint, and (2) it has the ability to simultaneously assimilate the observational data at multiple times in an assimilation window. However, 4DVar still faces numerous challenges in coding, maintaining and updating the adjoint model of the forecast model and it requires the linearization of the forecast model.

On the other hand, the ensemble Kalman filter (EnKF; e.g. Evensen, 1994, 2004; Houtekamer and Mitchell, 1998; Houtekamer and Mitchell, 2001) has become an increasingly popular method because of its simple conceptual formulation and relative ease of implementation. By forecasting the statistical characteristics, EnKF can provide flow-dependent error estimates of the background errors using the Monte Carlo method, but it lacks the temporal smoothness constraint as that in 4DVar since it is naturally designed to incorporate sequential information only.

Large efforts have been devoted to seek to advance the state-of-the-science in data assimilation by coupling 4DVar with EnKF (e.g. Lorenc, 2003; Tian et al., 2008; Zhang et al., 2009; Cheng et al., 2010) aiming at maximally exploiting the strengths of the two forms of data assimilation while simultaneously offsetting their respective weaknesses. Lorenc (2003) reviewed EnKF in comparison with 4DVar, and suggested that a hybrid method may be attractive for mesoscale NWP systems. A hybrid EnKF-3DVar scheme was proposed by Hamill and Snyder (2000), and a more sophisticated hybrid approach that combines EnKF and 3DVar was elaborated by Zupanski (2005). On the other hand, some practical attempts, extending EnKF from fixed-in-time to four-dimensions, were suggested by Evensen and van Leeuwen (2000), about which the asynchronous observations could be constructed as linear combinations based on the ensemble perturbations (Hunt et al., 2004; Fertig et al., 2007). Vermeulen and Heemink (2006) have attempted to

*Corresponding author.
e-mail: tianxj@mail.iap.ac.cn
DOI: 10.1111/j.1600-0870.2011.00529.x

combine 4DVar with EnKF, however, the tangent linear model is still needed in their method. Recently, Zhang et al. (2009) also examined the performance of coupling the deterministic 4DVar with the EnKF to produce a superior hybrid approach for data assimilation.

A hybrid method, referred to as POD4DVar, was proposed by Tian et al. (2008) based on the Monte Carlo method and the proper orthogonal decomposition (POD) technique. In the POD4DVar, the POD technique is applied to an ensemble of gridded 4-D MPs sampled from perturbed integrations of the forecast model at observational time levels to construct the orthogonal base vectors. After the analysis variables are represented by a truncated expansion of the base vectors in the 4-D model space, the control (state) variables in the cost function appear explicit so that the adjoint model, which is used to derive the gradient of the cost function with respect to the control variables in the traditional 4DVar, is no longer needed. By using an ensemble of historical samples to define a subspace, Wang et al. (2010) also proposed an economical approach to implement 4DVar on the basis of a dimensional-reduced projection (DRP) technique. In Wang et al. (2010), suitable MPs are chosen to ensure the linear independence of the perturbation samples in the observation space. Unfortunately, it should be noted that, unlike EnKF, only one analysed field is obtained in each analysis procedure in both Tian et al. (2008) and Wang et al. (2010), and the initial ensemble should be produced randomly or updated by the historical forecasts at the start time of the assimilation window in each cycle repeatedly. It implies that only a partially flow-dependent background error covariance is employed in these two methods.

The POD-based ensemble four-dimensional variational data assimilation method (PODEn4DVar) method considered in this study is similar to the POD4DVar method (Tian et al., 2008) but the POD transformation is conducted in the OP space rather than in the model perturbation (MP) space directly. Consequently, under the assumption of tangent linear approximation between the MPs and the OPs, the perturbation ensemble in the model space is transformed in accordance with the POD transformation to the ensemble OP subspace. It means that the transformed MPs could guarantee the orthogonality (and thus independence) of their corresponding OPs (i.e. the orthogonal base vectors). The optimal MP and its corresponding OPs can be represented by the transformed MP ensemble and their related OP orthogonal base vectors respectively to fit the 4-D observation innovation in the assimilation window. The application of this strategy significantly simplifies the data assimilation process and retains most advantages of the traditional 4DVar method. It can moderately increase the assimilation precision while reducing the computational costs substantially. The update of the analysis ensemble is further realized by using the ensemble of innovation vectors instead of the single 4-D observation innovation in the explicit 4DVar analysis equation. Therefore, the PODEn4DVar updates its background error covariance by utilizing the evolving ensemble forecasts, which can provide flow-dependent error estimates of the background errors. Since the PODEn4DVar can be easily degenerated to the 3-D case (referred to as PODEn3DVar), we managed to couple the PODEn4DVar with its 3-D format. As a result, the PODEn3DVar can adjust the model solution trajectory gradually throughout the whole assimilation window, which results in a superior performance.

Assimilation experiments by the Lorenz-96 model with simulated observations show that the PODEn4DVar method is capable of outperforming the POD4DVar, the standard 4DVar and the EnKF in terms of lower root mean square (RMS) errors under both perfect and imperfect-model scenarios. By carefully choosing the relaxation coefficient and localization radius, the PODEn4DVar method adjusted the assimilation results to approach the true solution trajectory quickly with sharply reduced RMS errors. Further, owing to its robust performance, the RMS errors of PODEn4DVar are constrained to a very low level with slight fluctuations throughout the whole assimilation process.

2. The POD4DVar

The traditional 4DVar analysis of x_a obtained through the minimization of a cost function J that measures the misfit between the model trajectory $H_k(x_k)$ and the observations y_k at a series of t_k, $k = 1, \ldots, S$

$$J(x) = (x - x_b)^T B^{-1}(x - x_b) + \sum_{k=1}^{S} \left\{H_k[M_{t_0 \to t_k}(x)] - y_k\right\}^T R_k^{-1} \left\{H_k[M_{t_0 \to t_k}(x)] - y_k\right\}, \quad (1)$$

with the forecast model $M_{t_0 \to t_k}(x)$ imposed as strong constraints, defined by

$$x_k = M_{t_0 \to t_k}(x), \quad (2)$$

where the superscript T stands for a transpose, b is the background value, index k denotes the observation time, H_k is the observation operator, and matrices B and R_k are the background and observational error covariances, respectively.

The POD4DVar (Tian et al., 2008) starts from an initial field ensemble $x_{0,n}$ $(n = 1, \ldots, N)$ produced by using the Monte Carlo method or sampling from the historical forecasts (Wang et al., 2010). Subsequently, the forecast model $x_{k,n} = M_{t_0 \to t_k}(x_{0,n})$ with the initial fields $x_{0,n}$ is integrated throughout the assimilation window to obtain the state series $x_{k,n}$ $(k = 1, \ldots, S)$ and then construct the 4-D fields $X_n = (x_{0,n}, \ldots, x_{S,n})^T$. The analysis $x_{k,a}$ over the same assimilation window can be stored into the following vector

$$X_a = (x_{0,a}, \ldots, x_{S,a})^T. \quad (3)$$

We form the ensemble perturbation matrix, denoted by $A(M \times N) = (\delta X_1, \ldots, \delta X_N)$. It is composed of N column vectors, where $M = M_g \times M_v \times (S + 1)$, and M_g, M_v are the number of the model spatial grid points and the number of the model variables, respectively. Each column vector in A is the

following 4-D perturbation field (with respect to the ensemble mean field) as follows:

$$\delta X_n = X_n - \overline{X}, \quad n = 1, \ldots N, \tag{4}$$

where

$$\overline{X} = \frac{1}{N}\sum_{n=1}^{N} X_n. \tag{5}$$

Note $A^T A$ is a $N \times N$ matrix and $N \leq M$ for practical applications, so the eigenvector decomposition of $A^T A = V\Lambda^2 V^T$ is used to obtain Λ and V, then the POD mode matrix is given by $\Phi = AV$. The truncated reconstruction of analysis variable X_a is given by

$$X_a = \overline{X} + \sum_{n=1}^{r} \beta_n \phi_n, \tag{6}$$

where $r(\leq N)$ is the number of the POD modes (see Tian et al., 2008 for more details). Form the POD mode matrix

$$\Phi = (\phi_1, \ldots, \phi_r), \tag{7}$$

where $\phi_n = [\phi_n(t_0), \ldots, \phi_n(t_S)]^T$, $n = 1, \ldots, r$. Transform (7) into the following format

$$\Phi = (\Phi_0, \ldots, \Phi_S)^T, \tag{8}$$

where $\Phi_k = [\phi_1(t_k), \ldots, \phi_r(t_k)]$.

Substituting (5–8) into (1), the control variable becomes $\beta = (\beta_1, \ldots, \beta_r)^T$ instead of x_0, so the control variable is expressed explicitly in the cost function

$$J(\beta) = (\Phi_0\beta)^T B^{-1}(\Phi_0\beta) + \sum_{k=1}^{S} [y_k - H_k(\overline{x}_k + \Phi_k\beta)]^T R_k^{-1} [y_k - H_k(\overline{x}_k + \Phi_k\beta)], \tag{9}$$

and

$$J(\beta) = (\Phi_0\beta)^T B^{-1}(\Phi_0\beta) + \sum_{k=1}^{S} \left[\mathrm{H}'_k\Phi_k\beta - y'_k\right]^T R_k^{-1} \left[\mathrm{H}'_k\Phi_k\beta - y'_k\right], \tag{10}$$

where

$$\mathrm{H}'_k\phi_k = H_k(\overline{x}_k + \phi_k) - H_k(\overline{x}_k) \quad \text{and} \quad y'_k = y_k - H_k(\overline{x}_k).$$

One can solve the optimization problem (10) as follows:

$$\nabla J(\beta) = 0, \tag{11}$$

$$\left[(r-1)I_{r\times r} + \sum_{k=1}^{S} \left(\mathrm{H}'_k\Phi_k\right)^T R_k^{-1} \left(\mathrm{H}'_k\Phi_k\right)\right]\beta = \sum_{k=1}^{S} \left(\mathrm{H}'_k\Phi_k\right)^T R_k^{-1} y'_k, \tag{12}$$

and

$$\beta = \left[(r-1)I_{r\times r} + \sum_{k=1}^{S} \left(\mathrm{H}'_k\Phi_k\right)^T R_k^{-1} \left(\mathrm{H}'_k\Phi_k\right)\right]^{-1} \times \sum_{k=1}^{S} \left(\mathrm{H}'_k\Phi_k\right)^T R_k^{-1} y'_k. \tag{13}$$

Substituting (13) into (5), one can easily obtain the final analysis results.

3. The PODEn4DVar

3.1. Formulation of PODEn4DVar

By minimizing the following incremental format of the 4DVar cost function (1), one can obtain an optimal increment of initial condition (IC), x'_a, at the initial time

$$J(x') = \frac{1}{2}(x')B^{-1}(x') + \frac{1}{2}\left[y'(x') - y'_{\mathrm{obs}}\right]^T R^{-1}\left[y'(x') - y'_{\mathrm{obs}}\right], \tag{14}$$

where $x' = x - x_b$ is the perturbation of the background field x_b at t_0,

$$y'_{\mathrm{obs}} = \begin{bmatrix} y'_{\mathrm{obs},1} \\ y'_{\mathrm{obs},2} \\ \vdots \\ y'_{\mathrm{obs},S} \end{bmatrix}, \tag{15}$$

$$y' = y'(x') = \begin{bmatrix} (y_1)' \\ (y_2)' \\ \vdots \\ (y_S)' \end{bmatrix}, \tag{16}$$

$$(y_k)' = y_k(x_b + x') - y_k(x_b), \tag{17}$$

$$y'_{\mathrm{obs},k} = y_{\mathrm{obs},k} - y_k(x_b), \tag{18}$$

$$y_k = H_k[M_{t_0\to t_k}(x)], \tag{19}$$

and

$$R = \begin{bmatrix} R_1 & 0 & \ldots & 0 \\ 0 & R_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & R_S \end{bmatrix}. \tag{20}$$

We usually assume that the relationship between $(y_k)'$ and x' is approximately linear according to (17)

$$(y_k)' = L_k x', \tag{21}$$

where $L_k = H'_k M'_{t_0 \to t_k}$, H'_k is the tangent linear operator of H_k, and $M'_{t_0 \to t_k}$ is the tangent linear model of $M_{t_0 \to t_k}$. Here, we assume

$$y' = Lx' = \begin{bmatrix} L_1 & 0 & \dots & 0 \\ 0 & L_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & L_S \end{bmatrix} x'$$

$$\approx \begin{bmatrix} y_1(x_b + x') - y_1(x_b) \\ y_2(x_b + x') - y_2(x_b) \\ \vdots \\ y_S(x_b + x') - y_S(x_b) \end{bmatrix} = L_{x_b} x'. \quad (22)$$

Suppose N samples of observation perturbation (OP) y' : $y'_1, y'_2, \dots, y'_N$ are generated by using the observation operator H_k, the forecast model $M_{t_0 \to t_k}$ and the IC samples x' : $x'_1, x'_2, \dots, x'_N$. The OP samples y' and the MP samples x' are related by

$$y'_k \approx L_{x_b} x'_k. \quad (23)$$

As illustrated by Wang et al. (2010), if the observation of y'_{obs} is available, a weighted mean of the multiple simulations y' can be used to approach the observation and determine all the weight coefficients. In this way, the same weight coefficients can be used to obtain the weighted mean of multiple MPs of IC according to the linear relationship between y' and x' in (23), which may provide a better IC for the forecast of y'. Following this strategy, suitable MPs, $x'_1, x'_2, \dots, x'_N$, should be chosen to guarantee the linear independence of the OP samples, $y'_1, y'_2, \dots, y'_N$, as much as possible. To accomplish this task, we first apply the POD transformation to the OP matrix $y' = (y'_1, \dots, y'_N)$ as follows:

$$y'^T y' = V \Lambda^2 V^{\mathrm{T}}, \quad (24)$$

and

$$\Phi_y = y' V. \quad (25)$$

As a result, the POD transformed OP samples $\Phi_y = (y'_1, \dots, y'_N)$ are orthogonal and thus independent. Since y' and x' is approximately linear according to (23), we can obtain

$$\Phi_y = y'V = L(x')V = L(x'V) \approx L_{x_b}(x'V) = L_{x_b}(\Phi_x). \quad (26)$$

Consequently, through the simple calculations (24–26), the perturbation ensemble x' in the model space is transformed to Φ_x in accordance with the POD transformation to the ensemble OP subspace. Apparently, the transformed MPs $\Phi_x = x'V = (x'_1, \dots, x'_N)$ could ensure the orthogonality (and thus independence) of their corresponding OPs (i.e. the orthogonal base vectors $\Phi_y = y'V$). Let's mark $\Phi_{y,r} = (y'_1, \dots, y'_r)$ and $\Phi_{x,r} = (x'_1, \dots, x'_r)$ ($r \le N$ is also the number of the POD modes). A weighted mean of these MP samples $\Phi_{x,r} = (x'_1, \dots, x'_r)$ is used to provide an optimal solution x' as follows:

$$x'_a = \Phi_{x,r} \beta, \quad (27)$$

where $\beta = (\beta_1, \dots, \beta_r)^T$. Its corresponding optimal OPs are determined by

$$y'_a = L(x'_a) = L(\Phi_{x,r}\beta) = L(\Phi_{x,r})\beta \approx L_{x_b}(\Phi_{x,r})\beta = \Phi_{y,r}\beta. \quad (28)$$

Substituting (27–28) into (14), the control variable becomes β instead of x', so the control variable is expressed explicitly in the cost function

$$J(\beta) = \frac{1}{2}(\Phi_{x,r}\beta)^T B^{-1}(\Phi_{x,r}\beta) + \frac{1}{2}\left(\Phi_{y,r}\beta - y'_{\text{obs}}\right)^T R^{-1}\left(\Phi_{y,r}\beta - y'_{\text{obs}}\right). \quad (29)$$

Similar to the EnKF (Evenson, 2004), the background error covariance can be constructed by the MPs as follows:

$$B = \frac{\Phi_{x,r}\Phi_{x,r}^T}{r-1}. \quad (30)$$

Substituting (30) into (29), we can obtain

$$J(\beta) = \frac{1}{2}(r-1)\cdot\beta^T \Phi_{x,r}^T \left(\Phi_{x,r}^T\right)^{-1}(\Phi_{x,r})^{-1}\Phi_{x,r}\beta + \frac{1}{2}(\Phi_{y,r}\beta - y'_{\text{obs}})^T R^{-1}(\Phi_{y,r}\beta - y'_{\text{obs}}), \quad (31)$$

and

$$J(\beta) = \frac{1}{2}(r-1)\beta^T\beta + \frac{1}{2}(\Phi_{y,r}\beta - y'_{\text{obs}})^T R^{-1}(\Phi_{y,r}\beta - y'_{\text{obs}}), \quad (32)$$

where the inverse of $\Phi_{x,r}^T(\Phi_{x,r})$ should be defined as the Moore–Penrose pseudoinverse. Because R is usually symmetrical, the gradient of the cost function is thus obtained

$$\nabla J(\beta) = \left[(r-1)I_{r\times r} + \Phi_{y,r}^T R^{-1}\Phi_{y,r}\right]\beta - \Phi_{y,r}^T R^{-1} y'_{\text{obs}}. \quad (33)$$

One can solve the optimization problem (33) without an iterative procedure

$$\beta = \left[(r-1)I_{r\times r} + \Phi_{y,r}^T R^{-1}\Phi_{y,r}\right]^{-1}\Phi_{y,r}^T R^{-1} y'_{\text{obs}}. \quad (34)$$

Finally, the formula to calculate the increment of analysis is simplified into the following form

$$x'_a = \Phi_{x,r}\left[(r-1)I_{r\times r} + \Phi_{y,r}^T R^{-1}\Phi_{y,r}\right]^{-1}\Phi_{y,r}^T R^{-1} y'_{\text{obs}}. \quad (35)$$

In general, the number of observational data is less than the degrees of freedom of model variables. For this reason, we mark

$$\tilde{\Phi}_{y,r} = \left[(r-1)I_{r\times r} + \Phi_{y,r}^T R^{-1}\Phi_{y,r}\right]^{-1}\Phi_{y,r}^T R^{-1}, \quad (36)$$

and rewrite (35) as follows:

$$x'_a = \Phi_{x,r}\tilde{\Phi}_{y,r} y'_{\text{obs}}. \quad (37)$$

Therefore, the final 4DVar analysis x_a can be calculated as follows:

$$x_a = x_b + \Phi_{x,r}\tilde{\Phi}_{y,r} y'_{\text{obs}}, \quad (38)$$

and

$$x_a = x_b + K_{x,y} y'_{\text{obs}}, \tag{39}$$

where $K_{x,y} = \Phi_{x,r} \tilde{\Phi}_{y,r}$. Notably, eq. (39) shares some similarity with the EnKF formulation in terms of the matrix $K_{x,y}$ acting as a proxy of the Kalman gain matrix as that in the EnKF analysis equation. Inspired by the similarity, we replace the single 4-D observation innovation y'_{obs} by the ensemble of innovation vectors

$$D' = D - H\Phi_x, \tag{40}$$

where

$$D = (d, \ldots, d), \tag{41}$$

$$d = (y_1, \ldots . y_S)^T, \tag{42}$$

and

$$H = \begin{bmatrix} H_1 & 0 & \cdots & 0 \\ 0 & H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_S \end{bmatrix}. \tag{43}$$

In consequence, the update of the analysis ensemble is successfully implemented as follows:

$$A_a = A_b + K_{x,y} D', \tag{44}$$

where $A_a = (x_b, \ldots, x_b)$. Remarkably, the MP ensemble x' of IC can be replaced by the MP ensemble

$$(x_k)' = M_{t_0 \to t_k}(x_b + x') - M_{t_0 \to t_k}(x_b), \tag{45}$$

at any time level t_k in the assimilation window. That means we can obtain the final analysis results at any time level in the assimilation window, which accordingly facilitates the implementation of data assimilation. In addition, the PODEn4DVar can be easily degenerated to the 3-D case if we set $S = 1$.

Both the PODEn4DVar and the POD4DVar are proposed based on the POD technique, ensemble forecasts and linear assumption between the MPs and the OPs. Whereas, they differ from each other significantly in the following two aspects:

(1) In the PODEn4DVar, the POD transformation is first applied to the OP space and then this transformation is transferred to the MP space in terms of the linear assumption between the MPs and the OPs. Accordingly, the weighted mean of the OP orthogonal base vectors can fit the 4-D innovation data in the observation space directly, which has the potential to improve the analysis accuracy. On the contrary, in the original POD4DVar, the POD transformation is conducted directly in the MP ensemble space rather than in the OP space. Since the OP space dimension is usually much smaller than that of the MP space, the strategy adopted in the PODEn4DVar reduces the computational costs substantially with decaying memory.

(2) Only one analysed field is obtained in each analysis procedure in the original POD4DVar and the initial ensemble has to be produced randomly or updated by the historical forecasts at the start time of the assimilation window in each cycle repeatedly. Contrarily, the analysis ensemble update is successfully implemented in the proposed PODEn4DVar method and the initial condition is perturbed only once throughout the whole assimilation. It implies that this method could suffer from sample preparation little.

3.2. Localization and covariance inflation

The PODEn4DVar approach obtains an optimal solution in (38–39) and updates the analysis ensemble in (44). As an ensemble approach, its ensemble is usually composed of far fewer members than both the number of observational data and the degrees of freedom of model variables, which would lead to many spurious correlations between observation locations and model grids. A practical way to address this issue is through localization technique, which ameliorates the spurious long-range correlations (Houtekamer and Mitchell, 2001). Similar to Wang et al. (2010), the Schur product is applied to the matrix $K_{xy}(L_x \times (\sum_{k=1}^{S} L_{y,k})) = \Phi_{x,r}\tilde{\Phi}_y$ (where $L_x = M_g \times M_v$ is the length of vector x_0, $L_{y,k}$ is the length of vector y_k and $L_y = \sum_{k=1}^{S} L_{y,k}$) to filter out the remote correlation between the observation locations and model grids more continuously, and the final increment of analysis is calculated using the formula:

$$x'_a = \rho \circ K_{xy} y'_{\text{obs}}, \tag{46}$$

where the Schur product of two matrices having the same dimension is denoted by $A = B \circ C$ and consists of the element-wise product such that $a_{i,j} = b_{i,j} \cdot c_{i,j}$. For providing the formula of the filtering matrix, ρ, suppose the $K_{x,i}$ $(i = 1, \ldots, L_x)$ and $K_{y,j}$ $(j = 1, \ldots, L_y)$ are the model states and the observational variables, respectively. Making the horizontal and vertical distances between the spatial locations of $K_{x,i}$ and $K_{y,j}$ as $d_{h,i,j}$ and $d_{v,i,j}$, respectively, then the elements of the matrix ρ can be calculated according to

$$\rho_{i,j} = C_0(d_{h,i,j}/d_{h,0}) \cdot C_0(d_{v,i,j}/d_{v,0}), \tag{47}$$

where the filtering function C_0 is defined as

$$C_0(r) = \begin{cases} -\frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1, \\ \qquad\qquad 0 \le r \le 1 \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1}, \\ \qquad\qquad 1 < r \le 2 \\ 0, \qquad\qquad 2 < r \end{cases} \tag{48}$$

and $d_{h,0}$ and $d_{v,0}$ are the horizontal and vertical covariance localization Schur radii, respectively. The matrix ρ filters out the

small (and noisy) correlations associated with remote observations through the Schur product, which is the localization strategy (Wang et al., 2010). In addition, since $\rho_{i,j}$ is smooth and monotonically decreasing, the Schur product tends to reduce the effect of those observations smoothly at intermediate distances.

Covariance inflation for the PODEn4DVar ensemble is achieved through the covariance relaxation method of Zhang et al. (2004)

$$(x_i')_{\text{new}} = \alpha x'_{f,i} + (1-\alpha)x'_{a,i}, \tag{49}$$

where α $(0 < \alpha < 1)$ is the relaxation coefficient and $(x_i')_{\text{new}}$ is the final perturbation of the analysis ensemble used for the next forecast cycle.

3.3. The fully coupling between the PODEn4DVar and PODEn3DVar

The coupled approach aims to link the distributed in time, maximum likelihood approach of PODEn4DVar and its sequential 3D case (PODEn3DVar). An illustration of their coupling proposed in this paper is depicted in the schematic flowchart of Fig. 1: the whole assimilation implementation is based on a framework of ensemble forecasts. In each assimilation window, the PODEn3DVar is implemented to conduct its assimilation procedures with the observational data being incorporated sequentially. The PODEn3DVar implementation is to adjust the model solution trajectory gradually. The 4-D ensemble samples for the PODEn4DVar can be constructed as soon as the PODEn3DVar accomplishes its assimilation procedures in the assimilation window. Consequently, the PODEn4DVar implementation is carried out over the same assimilation window to obtain the 4-D balanced analysis and take place the ensemble mean for next assimilation cycle.

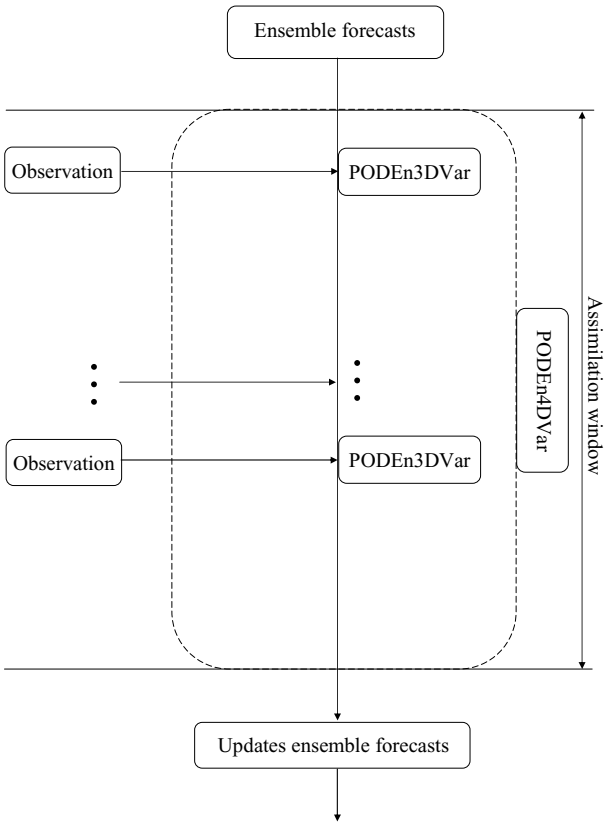

Fig. 1. Schematic diagram of the PODEn4DVar coupling with PODEn3DVar.

4. Evaluations within the Lorenz-96 model

4.1. Experimental design

In this section, the PODEn4DVar approach is evaluated within the model of Lorenz (1996)

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F, \quad i = 1, \ldots n, \tag{50}$$

with cyclic boundary conditions. Although not derived from any known fluid equations, the dynamics of (50) are 'atmosphere-like' in that they consist of non-linear advection-like terms, a damping term, and an external forcing; they can be thought of as some atmospheric quantity distributed on a latitude circle. One can choose any dimension, n, greater than 4 and obtain chaotic behaviour for suitable values of F. In this configuration, we take $n = 40$ and $F = 8$. For computational stability, a time step of 0.05 units (or 6 h in equivalent Lorenz, 1996) is adopted and a fourth-order Runge–Kutta scheme is used for temporal integration in this study.

The performance of the PODEn4DVar approach is examined in comparison to the POD4DVar, the standard 4DVar (Zhang et al., 2009) and the EnSRF-version (Evensen, 2004) of the EnKF. In this study, we employ a limited-memory quasi-Newton method (L-BFGS) (Liu and Nocedal, 1989) for the minimization in the standard 4DVar approach. The L-BFGS method is found to have superb performance in nonlinear minimization problems and has a relatively low computational costs and low storage requirement. The standard 4DVar uses a diagonal background error covariance B, whose values (all equal to 0.01) were determined through long-term statistics of EnKF spread and related to the attractor of the Lorenz-96 model (Zhang et al., 2009). To provide a fair comparison, we also implemented the analysis ensemble update in the POD4DVar same as that in the PODEn4DVar. The default number of observations is $m = 20$ (equally spaced at every observation time). Observations were taken every two steps (or 12 h), which were generated by adding random error perturbations of 0.03 to the time series of the true state. We considered the assimilation window length of 4 (standard 24-h daily assimilation cycle). All experiments were carried out over 365 d. The default parameter setups are the relaxation coefficient $\alpha = 0.9$, the ensemble size $N = 80$, and the

covariance localization Schur radius $r_0 = d_{h,0}(d_{v,0}) = 8$ (only one direction in the Lorez-96 model).

4.2. *Experimental results*

Figure 2a compares the performance of the PODEn4DVar with the original POD4DVar method, the standard 4DVar and the EnKF under the perfect-model assumption ($F = 8$ for all truth, forecast and assimilation runs). It shows that all the four methods behave considerably satisfactory in terms of overall low RMS errors. The PODEn4DVar performs slightly better than the other methods especially during the first ten assimilated days. These days are followed by their almost identical performance until the end of the whole assimilation process, which is apparently attributed to the perfect model assumption. The same conclusion can be drawn from Table 1 that the 365 simulated-days mean of the daily averaged RMS errors for the PODEn4DVar is smallest (only 0.018). The POD4DVar shows almost the same performance as the EnKF (0.026 vs. 0.029). Presumably, only a static background error covariance adopted in the standard 4DVar

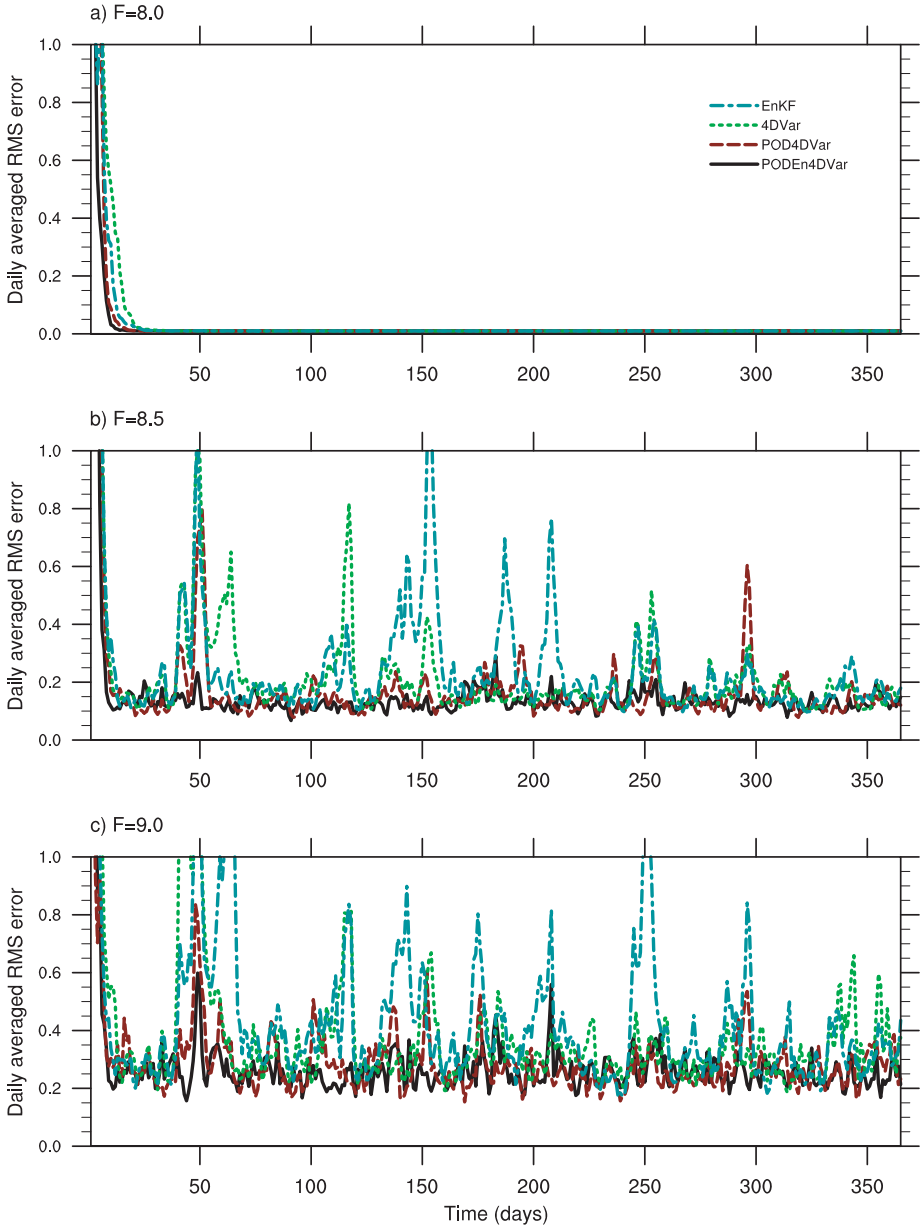

Fig. 2. Time series of the daily averaged root mean square (RMS) error for the four data assimilation techniques (PODEn4DVar, POD4DVar, 4DVar and EnKF) using default parameter setups with the Lorenz-96 model (a) without model error ($F = 8$), (b) with moderate model error ($F = 8.5$) and (c) with severe model error ($F = 9$), respectively.

Table 1. The 365 simulated-days mean of the daily averaged root mean square (RMS) error for the four methods

	$F = 8.0$	$F = 8.5$	$F = 9.0$
PODEn4DVar	0.018	0.16	0.27
POD4DVar	0.026	0.18	0.29
4DVar	0.036	0.24	0.39
EnKF	0.029	0.26	0.44

leads to its inferior performance to the other three ensemble-based approaches under the perfect-model assumption. In this case, the forecast error comes only from the noise of the initial field.

Another group experiments with moderate model error are also conducted using a different (incorrectly specified) forcing coefficient ($F = 8.5$) from that used in the truth simulation ($F = 8$). The truth run is used for verification and for generating observations. Figure 2b compares the performance of the four approaches with the imperfect forecast model ($F = 8.5$). The experiment configurations are exactly the same as those for the perfect model case. Notably, in the presence of moderate model error, the PODEn4DVar performs moderately better than the other three methods and the advantage of the PODEn4DVar over the other three methods becomes quite obvious especially on Day 50 and Day 290 or so. Of course, the POD4DVar can also give a fair good performance if the ensemble size N is increased to a certain degree ($\geq$140, not shown). The experiments with severe model error ($F = 9$, Fig. 2c) produce very similar results with those in the case of $F = 8.5$. The PODEn4DVar approach shows robust performance and works almost same as (slightly worse than) it does under the moderate model error scenario. Different with the above perfect-model scenario, the temporal smoothness constraint in the standard 4DVar results in its general superior performance compared with the EnKF method (Table 1) in the case of imperfect-model assumption ($F = 8.5$ and $F = 9.0$).

For the three groups of experiments, the ratio of the computational costs for the four methods (PODEn4DVar, POD4DVar, 4DVar and EnKF) varies around 1.05:1.0:1.1:1.5. The high computational cost in the EnKF method is mainly due to the fact the analysis process consists of huge matrix and the computation has to be conducted repeatedly when there are observations in the assimilation time window, while in POD4DVar the computation is performed only once in each cycle. As discussed in Tian et al. (2008), the application of the POD technique in the POD4DVar method (also in the PODEn4DVar) greatly lowers their computational costs (especially compared with the EnKF). In the PODEn4DVar, The POD transformation is applied to the OP space rather than the MP space directly, which can also alleviate the computational costs since the dimension of the OP space is usually much lower than the original MP space. Of course, this conclusion is not absolute and case-dependent because the scale of the minimization of cost functional and the iteration times during each assimilation cycle vary greatly within different numerical models.

To examine the sensitivity of the PODEn4DVar assimilation to the variations of the relaxation coefficient α, we designed another group of experiments with different relaxation coefficients. Figure 3a shows that the variations of the relaxation coefficient have slight impacts on the PODEn4DVar performance under the perfect-model assumption. The RMS errors are almost lower than 0.3 with the relaxation coefficients ranging from 0.6 to 1.0 (only partially shown in Fig. 3a). However, under imperfect-model scenarios ($F = 8.5$ and $F = 9.0$), the PODEn4DVar performance is significantly sensitive to the choice of the relaxation coefficient and only those larger than 0.85 or so can result in converged results (Figs. 3b–c).

Further, to explore the sensitivity of the PODEn4DVar assimilation to the variations of the covariance localization radius, we also compared the PODEn4DVar performance with different covariance localization radii. Under the perfect-model assumption, the choice of the covariance localization radius does not seem very critical to the PODEn4DVar performance as long as the ensemble size N is not very small ($\geq$80; Fig. 4a). However, with the imperfect forecast model, only some relatively appropriate covariance localization radii can generate a satisfactory performance (Figs. 4b–c). It shows the choice of an appropriate or inappropriate covariance localization radius is of importance to determine whether the PODEn4DVar could be successfully implemented or not.

Finally, to investigate how much the coupling between the PODEn4DVar and the PODEn3DVar affects the final analysis results, we also designed another group of experiments using the PODEn4DVar with and without coupling the PODEn3DVar respectively. When the assimilation window is 1 d, the coupling performs moderately better that the non-coupling, whose gap is further magnified with the model error becoming severe more and more (Fig. 5). In the case that the assimilation window is 2 d, the coupling superior performance beyond the non-coupling one appears more noticeable (Fig. 5). In the coupling case, the model trajectory ensemble is gradually corrected by the PODEn3DVar incorporating the observational information sequentially during the model ensemble integrations throughout the assimilation window, which contributes substantially to the improvements of the final analysis results. Figures 5a–c show that their RMS error curves for the coupling case almost overlap with the assimilation window being changed from 1 to 2 d. Conversely, the performance of PODEn4DVar without coupling PODEn3DVar becomes significantly poor when the assimilation window is changed from 1 to 2 d (Figures 5b–c). Especially under the sever model error case, even no converged results can be obtained (Fig. 5c). Since the model error would develop without any correction in the non-coupling case during the assimilation window, the assumption of tangent linear approximation

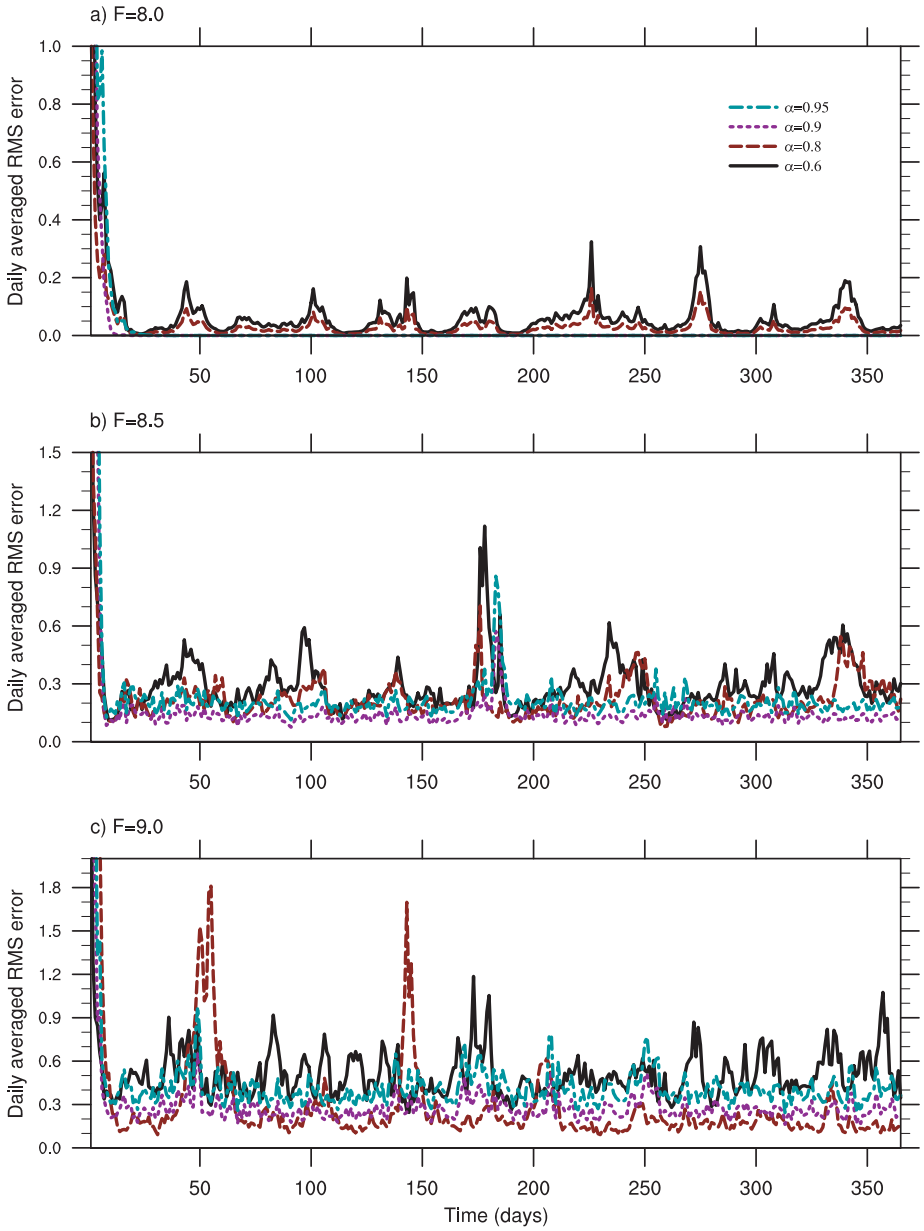

Fig. 3. Time series of the daily averaged root mean square (RMS) error for PODEn4DVar using different relaxation coefficients with the Lorenz-96 model (a) without model error ($F = 8$), (b) with moderate model error ($F = 8.5$) and (c) with severe model error ($F = 9$), respectively.

could likely be destroyed by the non-linearity evolution of the forecast model and the operation operator. Once a longer assimilation window is used, the spread of the ensemble forecasts could collapse rapidly and lead to an inferior performance of the non-coupling case.

5. Summary and concluding remarks

In this paper, a new scheme is developed to improve an ensemble-based explicit POD4DVar proposed by Tian et al. (2008). In this scheme, the POD technique is first applied to the perturbation ensemble in the observation space to construct its orthogonal base vectors. Consequently, under the assumption of tangent linear approximation between the model (state) perturbations and the OPs, the perturbation ensemble in the model space is transformed in accordance with the POD transformation to the ensemble OP subspace. The optimal MP and its corresponding OPs can be represented by the transformed MP ensemble and their related OP orthogonal base vectors respectively to fit the 4-D observation innovations in the assimilation window, which

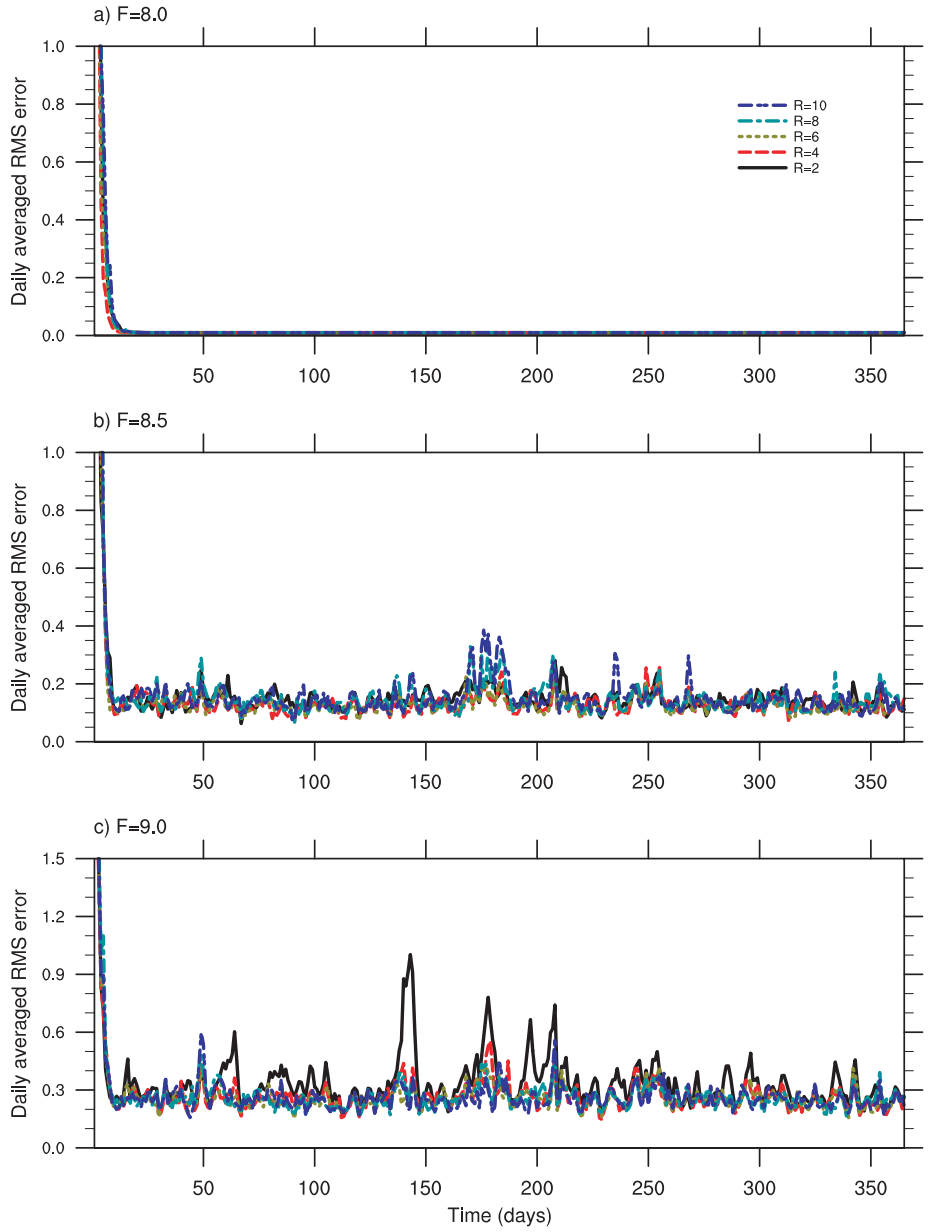

Fig. 4. Time series of the daily averaged root mean square (RMS) error for PODEn4DVar using different covariance localization radii with the Lorenz-96 model (a) without model error ($F = 8$), (b) with moderate model error ($F = 8.5$) and (c) with severe model error ($F = 9$), respectively.

is essentially consistent with the basic function of the 4DVar to optimize the final optimal analysis by making full use of observation information. The application of this strategy can moderately increase the assimilation precision while reducing the computational costs. Further, inspired by the similarity between this scheme and the EnKF, the update of the analysis ensemble is realized by using the ensemble of innovation vectors instead of the single 4-D observation innovation in the explicit 4DVar analysis equation, which finally forms the PODEn4DVar method.

Several numerical experiments performed with the Lorenz-96 model show that the proposed PODEn4DVar method performs better than the original POD4DVar, the standard 4DVar and the EnKF with assimilation errors being reduced to moderately less than those of the latter ones with lower computational costs compared with EnKF, especially under the imperfect model scenarios. The assimilation experiments also indicate that the coupling between the PODEn4DVar and PODEn3DVar can, indeed, help ameliorate the final analysis results significantly, especially when the assimilation window is not short enough. Since the

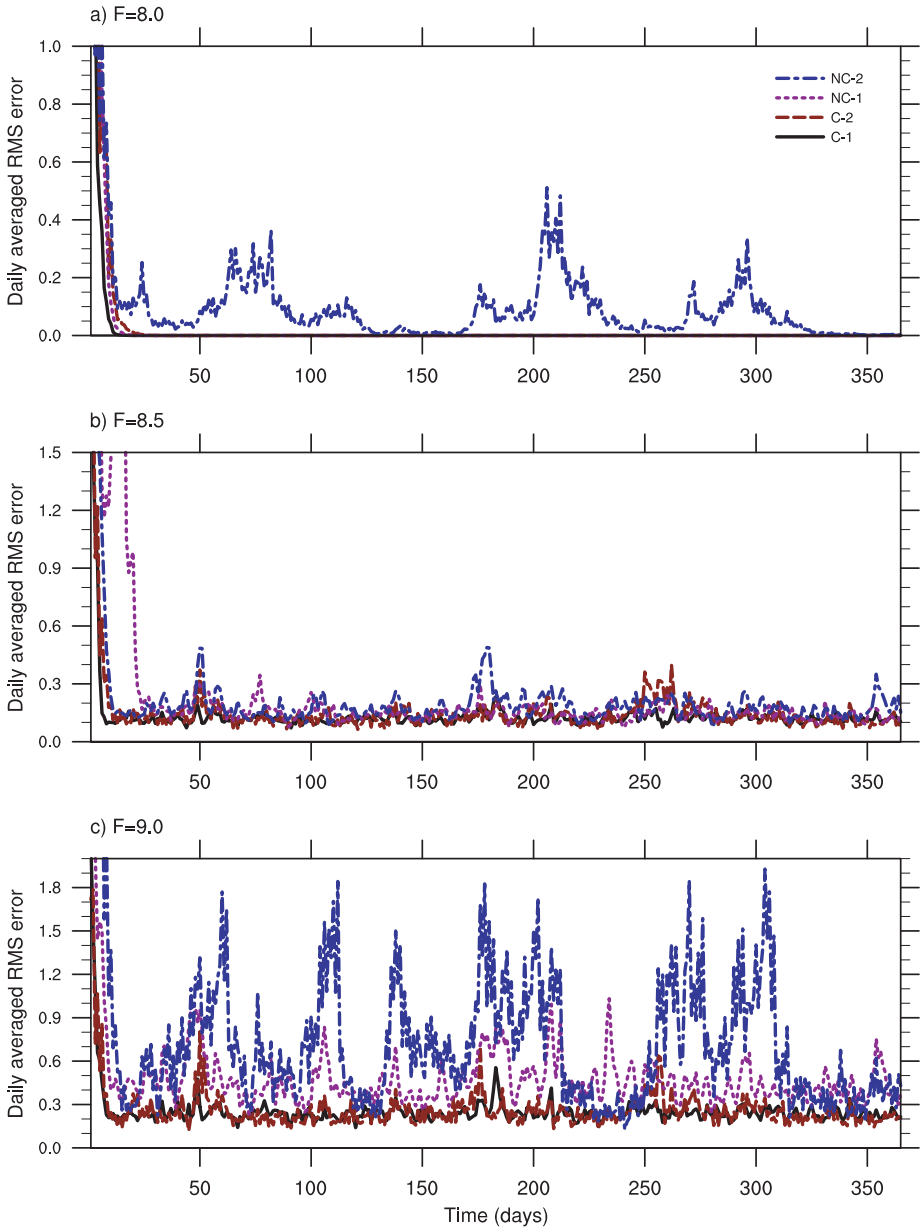

Fig. 5. Time series of the daily averaged root mean square (RMS) error for PODEn4DVar with and without coupling PODEn3DVar using two different assimilation windows (1 and 2 d) with the Lorenz-96 model (a) without model error ($F = 8$), (b) with moderate model error ($F = 8.5$) and (c) with severe model error ($F = 9$), respectively. (C-1 and C-2 denote the two coupling cases with the assimilation windows being 1 and 2 d, respectively; NC-1 and NC-2 stand for the two non-coupling cases with the assimilation windows being 1 and 2 d, respectively).

dynamics of Lorenz-96 equations is "atmosphere-like" and can be thought of as some atmospheric quantity distributed on a latitude circle, which could act as a good proxy of the real atmospheric model. Thus, the pretty good performance of ss shown in the assimilation experiments conducted with the Lorenz-96 implies its potential applications in real numerical weather or climate models. However, several issues, such as how to choose the appropriate localization radius and the relaxation coefficient, still need to be addressed.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 41075076), the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KZCX2-EW-QN207) and the National Basic Research Program (Grant Nos. 2010CB428403 and 2009CB421407). Two anonymous reviewers are thanked for helpful comments and suggestions that improved the presentation.

References

Bauer, P., Lopez, P., Benedetti, A., Salmond, D., Saarinen, S. and co-authors. 2006. Implementation of 1D+4D-Var assimilation of precipitation affected microwave radiances at ECMWF II: 4D-Var. *Q. J. R. Meteorol. Soc.* **132**, 2307–2332.

Bormann, N. and Thepaut, J. N. 2004. Impact of MODIS polar winds in ECMWF's 4DVAR data assimilation system. *Mon. Wea. Rev.* **132**, 929–940.

Caya, A., Sun, J. and Snyder, C. 2005. A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.* **133**, 3081–3094.

Cheng, H., Jardak, M., Alexe, M. and Sandu, A. 2010. A hybrid approach to estimating error covariances in variational data assimilation. *Tellus* **62A**, 288–297.

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation II: numerical results. *Q. J. R. Meteorol. Soc.* **113**, 1329–1347.

Courtier, P., Thepaut, J. N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4DVar using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

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

Evensen, G. 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* **54**, 539–560, doi:10.1007/s10236-004-0099-2.

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

Fertig, E., Harlim, J. and Hunt, B. 2007. A comparative study of 4DVar and 4D ensemble Kalman filter: perfect model simulations with Lorenz-96. *Tellus* **59**, 96–101.

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S. and Morneau, J. 2007. Extension of 3DVAR to 4DVAR: implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Wea. Rev.* **135**(6), 2339–2354.

Hamill, T. M. and Snyder, C. 2000. A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.

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

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

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.

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* **38A**, 97–110.

Lewis, J. M. and Derber, J. C. 1985. The use of the adjoint equation to solve a variational adjustment problem with advective constraints. *Tellus* **37A**, 309–322.

Liu, D. C. and Nocedal J. 1989. On the limited memory BFGS method for large scale optimization. *Math. Prog.* **45**, 503–528.

Lorenc, A. 2003. The potential of the Ensemble Kalman Filter for NWP: a comparison with 4DVar. *Q. J. R. Meteorol. Soc.* **129**, 3183–3203.

Lorenz, E. 1996. Predictability: a problem partly solved. In: *Proc. Seminar on Predictability*. Volume 1, reading, ECMWF, United Kingdom, 1–19.

Park, K. and Zou, X. 2004. Toward developing an objective 4DVAR BDA scheme for hurricane initialization based on TPC observed parameters. *Mon. Wea. Rev.* **132**(8), 2054–2069.

Rosmond, T. and Xu, L. 2006. Development of NAVDAS-AR: nonlinear formulation and outer loop tests. *Tellus A* **58**(1), 45–58.

Tian, X., Xie, Z. and Dai, A. 2008. An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.* **113**, D21124, doi:10.1029/2008JD010358.

Vermeulen, P. T. M. and Heemink, A.W. 2006. Model-reduced variational data assimilation. *Mon. Wea. Rev.* **134**(10), 2888–2899.

Wang, B., Liu, J., Wang, S., Cheng, W., Liu, J. and co-authors. 2010. An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci*. **27**(4), 715–727, doi:10.1007/s00376-009-9122-3.

Zhang, F., Snyder, C. and Sun, J. 2004. Tests of an ensemble Kalman filter for convective-scale data assimilation: impact of initial estimate and observations. *Mon. Wea. Rev.* **132**, 1238–1253.

Zhang, F. Q., Zhang, M. and Hansen, J. A. 2009. Coupling ensemble Kalman filter with four dimensional variational data assimilation. *Adv. Atmos. Sci.* **26**(1), 1–8, doi:10.1007/s00376-009-0001-8.

Zupanski, M. 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Wea. Rev.* **133**, 1710–1726.