Science in China Series D: Earth Sciences

www.scichina.com
earth.scichina.com
www.springerlink.com

An explicit four-dimensional variational data assimilation method based on the proper orthogonal decomposition: Theoretics and evaluation

TIAN XiangJun† & XIE ZhengHui

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

The proper orthogonal decomposition (POD) method is used to construct a set of basis functions for spanning the ensemble of data in a certain least squares optimal sense. Compared with the singular value decomposition (SVD), the POD basis functions can capture more energy in the forecast ensemble space and can represent its spatial structure and temporal evolution more effectively. After the analysis variables are expressed by a truncated expansion of the POD basis vectors in the ensemble space, the control variables appear explicitly in the cost function, so that the adjoint model, which is used to derive the gradient of the cost function with respect to the control variables, is no longer needed. The application of this new technique significantly simplifies the data assimilation process. Several assimilation experiments show that this POD-based explicit four-dimensional variational data assimilation method performs much better than the usual ensemble Kalman filter method on both enhancing the assimilation precision and reducing the computation cost. It is also better than the SVD-based explicit four-dimensional assimilation method, especially when the forecast model is not perfect and the forecast error comes from both the noise of the initial filed and the uncertainty of the forecast model.

POD, data assimilation, 4DVAR, explicit method

The four-dimensional variational data assimilation (4DVAR) method has been a very successful technique used in operational numerical weather prediction (NWP) at many weather forecast centers[1,2]. The 4DVAR technique has two attractive features: (1) the full model is set as a strong dynamical constraint, and (2) it has the ability to assimilate the observation data at multiple times, but it still faces numerous challenges to code, maintain and update the adjoint or linearity model of the forecast model. Usually, the control variables (initial state) are expressed implicitly in the cost function. In order to compute the gradient of the cost function with respect to the control variables, one has to integrate the adjoint model, which is extremely labor-intensive, especially when the forecast model is highly nonlinear and the model physics contains parameterized discontinuities[3,4], which often occures in land surface models. Many efforts have been devoted to avoiding integrating the adjoint model or reducing the expensive computation cost[5−7]. But the linear or adjont model is still required in the methods above. On the other hand, the ensemble Kalman filter (EnKF)[8] is becoming an increasingly popular method for its high precision and simple operation. By forecasting the statistical characteristics of the background errors, the EnKF can provide flow-dependent error estimates of the background errors with the Monte Carlo method, but it lacks the two advantages of the 4DVAR mentioned above. Some studies have been done to relate 4DVAR to the EnKF[9,10], however, the adjoint model is still needed in these methods. In

Received March 5, 2008; accepted July 15, 2008; published online January 5, 2009
doi: 10.1007/s11430-009-0012-4
†Corresponding author (email: tianxj@mail.iap.ac.cn)
Supported by National Natural Science Foundation of China (Grant No. 40705035), National High Technology Research and Development Program of China (Grant No. 2007AA12Z144), Knowledge Innovation Project of Chinese Academy of Sciences (Grant Nos. KZCX2-YW-217 and KZCX2-YW-126-2), and National Basic Research Program of China (Grant No. 2005CB321704)

order to retain the two primary advantages of the traditional 4DVAR method while avoiding the need of an adjoint or linearity model of the forecast model, Qiu et al.[11,12] proposed a new method for 4DVAR by the singular value decomposition (SVD) technique based on the theory of the attractors of atmosphere, which was shown to perform well. We resort to the proper orthogonal decomposition (POD) (or the Karhunan Loève procedure) technique[13,14] and the idea of EnKF. Cao et al.[15] also applied the POD technique to the 4DVAR to reduce the forecast model orders in order to simplify the computation and save CPU and memory requirements, but the adjoint integration is still necessary in their method. Proper orthogonal decomposition technique has been used to obtain low dimensional dynamical models of many applications in engineering and science. In principle, the idea starts with an ensemble of data, called snapshots, collected from an experiment or a numerical procedure of a physical system. The POD technique is then used to produce a set of basis functions which spans the snapshot collection. The goal of the approach is to represent the ensemble of data in terms of an optimal coordinate system. That is, the snapshots can be generated by a smallest possible set of basis functions. An explicit four-dimensional data assimilation method based on the POD technique is proposed in this paper: it begins with a four-dimensional sample ensemble obtained from the forecast ensembles at all the time levels (or at the observational time steps) in an assimilation window produced by the Monte Carlo method, which is similar to that in the ensemble Kalman filter. If we apply the technique of proper orthogonal decomposition to the four-dimensional forecast ensemble, the orthogonal basis vectors can not only capture the spatial structure of the state but also reflect its temporal evolution. After the model status is expressed by a truncated expansion of the basis vectors obtained by the POD technique, the control variables in the cost function appear explicitly, so that the adjoint model is no longer needed. This method is expected to not only simplify the data assimilation procedure but also retain the two main advantages of the traditional 4DVAR method. Several numerical experiments are conducted with a one-dimensional soil water equation model and synthetical observations to evaluate this method in land data assimilation. Comparison is also made between this newly proposed method, Qiu's method (thereafter called SVD-E4DVAR for simplicity) and the EnKF method.

1 Proper orthogonal decomposition

1.1 Continuous case

Let $U_i(\vec{x})$, $i=1, 2, \cdots, N$ denote the set of N observations or simulations (also called snapshots) of some physical process taken at position $\vec{x}=(x, y)$. The average of the ensemble snapshots is given by

$$\overline{U}=\frac{1}{N}\sum_{i=1}^{N}U_i(\vec{x}). \tag{1}$$

We form new ensemble by focusing on deviation from mean as follows:

$$V_i=U_i-\overline{U}. \tag{2}$$

One wish is to find an optimal compressed description of the sequence of data (eq. (2)). One description of the process is a series expansion in terms of a set of basis functions. Intuitively, the basis functions should be representative of the members of the ensemble in a sense. Such a coordinate system, which possesses several optimality properties, is provided by the Karhunan Loève expansion, where the basis functions Φ are, in fact, admixtures of the snapshots and are given by

$$\Phi=\sum_{i=1}^{N}a_iV_i(\vec{x}), \tag{3}$$

where the coefficients a_i are to be determined so that Φ given by eq. (3) will most resemble the ensemble $\{V_i(\vec{x})\}_{i=1}^{N}$. More specially, we look for a function Φ to maximize

$$\frac{1}{N}\sum_{i=1}^{N}|(V_i,\Phi)|^2, \tag{4}$$

subject to $(\Phi, \Phi)=\|\Phi\|^2=1$, where $(\cdot,\cdot)$ and $\|\cdot\|$ denote the usual L^2 inner product and L^2-norm, respectively.

It follows that the basis functions are the eigenfunctions of the integral equation

$$\int C(\vec{x}, \vec{x}')\Phi(\vec{x}')\mathrm{d}\vec{x}'=\lambda\Phi(\vec{x}). \tag{5}$$

Substituting eq. (3) into eq. (5) yields the eigenvalue:

$$\sum_{j=1}^{N}L_{ij}a_j=\lambda a_i, \quad i=1, 2, \cdots, N, \tag{6}$$

where $L_{ij}=\frac{1}{N}(V_i,V_j)$ is a symmetric and nonnegative

matrix. Thus, we see that our problem amounts to solving the eigenvectors of a $N \times N$ matrix where N is the size of the ensemble of snapshots. Straightforward calculation shows that the cost functional

$$\frac{1}{N}\sum_{i=1}^{N}|(V_i,\Phi)|^2=(\lambda\Phi,\Phi)=\lambda, \tag{7}$$

which is maximized when the coefficients a_i of eq. (3) are the elements of the eigenvector corresponding to the largest eigenvalue of L.

1.2 Discrete case

We can consider the discrete Karhunan Loève expansion to find an optimal representation of the ensemble of snapshots. In the two-dimensional case, each sample of snapshots $U_i(x, y)$, (defined on a set of $n \times n$ nodal points (x, y)) can be expressed as an n^2 dimensional vector $\vec{u}_i$ as follows:

$$\vec{u}_i=\begin{bmatrix}\vec{u}_{i1}\\ \vdots\\ \vec{u}_{ij}\\ \vdots\\ \vec{u}_{in^2}\end{bmatrix}, \tag{8}$$

where $\vec{u}_{ij}$ denotes the jth component of the vector $\vec{u}_i$. Here the discrete covariance matrix of the ensemble $\vec{u}$ is defined as

$$C_{\vec{u}}=E\left\{(\vec{u}-\vec{m}_{\vec{u}})(\vec{u}-\vec{m}_{\vec{u}})^T\right\}, \tag{9}$$

where

$$\vec{m}_{\vec{u}}=E\left\{\vec{u}\right\}, \tag{10}$$

the mean vector, E is the expected value. Eqs. (9) and (10) can be replaced by

$$C_{\vec{u}}=\frac{1}{N}\left[\sum_{i=1}^{N}\vec{u}_i\vec{u}_j^{\,T}\right]-\vec{m}_{\vec{u}}\vec{m}_{\vec{u}}^{\,T},$$

and

$$\vec{m}_{\vec{u}}=\frac{1}{N}\sum_{i=1}^{N}\vec{u}_i,$$

respectively.

2 The explicit 4DVAR method based on the POD

In principle, the 4DVAR analysis, x_a is obtained through the minimization of a cost function J, which measures the misfit between the model trajectory $H_k(x_k)$ and the observation y_k at a series of time t_k, t=1, 2, ⋯, K.

The 4DVAR method is defined as a process minimizing the following cost function

$$J(x_0)=(x_0-x_b)^T B^{-1}(x_0-x_b)+\sum_{k=0}^{K}\left[y_k-H_k(x_k)\right]^T R_k^{-1}\left[y_k-H_k(x_k)\right], \tag{11}$$

with the forecast model M imposed as strong constraints, defined by

$$x_k=M_k(x_0), \tag{12}$$

where the superscript T stands for a transpose, b is background value, the index k denotes the observational time, H_k is the observational operator, and matrices B and R are the background and observational error covariances, respectively. The control variable is the initial conditions x_0 (at the beginning of the assimilation time window) of the model.

In the cost function (eq. (11)), the control variable x_0 is connected with x_k through forwarding the model and expressed implicitly, which makes it difficult to compute the gradient of the cost function with respect to x_0.

For simplicity, the proposed method is referred to as POD-E4DVAR. Like the traditional 4DVAR, the POD-E4DVAR also needs to choose an assimilation time window. The four-dimensional sample ensemble is constructed from the forecast ensembles in the assimilation time window produced by the Monte Carlo method, which is similar to that in the ensemble Kalman filter. Then the basis vectors are generated by applying the proper orthogonal decomposition (POD) technique to the four-dimensional sample ensemble. The model state can be expressed by the linear combination of the leading POD modes then. In this way, the control variables are transformed to the expansion coefficients and are expressed explicitly in the cost function. The details are described are follows: assuming there are S time steps during the assimilation time window (0, T). Generate N random perturbation fields and add each perturbation field to the initial background filed at $t=t_0$ and integrate the model to produce a perturbed 4D field (snapshots) X_n (n=1, ⋯, N) over the whole assimilation time window (S time steps) or at the observational time steps (K).

The average of the ensemble of snapshots is given by

$$\overline{X}=\frac{1}{N}\sum_{n=1}^{N}X_n. \tag{13}$$

We form a new ensemble by focusing on deviations from the mean as follows:

$$\delta X_n = X_n - \overline{X},\ n = 1, \cdots, N, \tag{14}$$

which form the matrix A ($M \times N$), where $M = M_g \times M_v \times S(K)$, M_g and M_v are the number of the model spatial grid points and the number of the model variables respectively, and $S(K)$ is the number of all time levels (observational time levels) over each analysis time window. To compute the POD modes, one must solve an $M \times M$ eigenvalue problem

$$T = (AA^T)_{M \times M} = \lambda V.$$

In practice, the direct solution of this eigenvalue problem is often not feasible if $M >> N$, which is very usual in numerical models. We can transform it into an $N \times N$ eigenvalue. In the method of snapshots, one then solves the $N \times N$ eigenvalue

$$TV_k = \lambda_k V_k, k = 1, \cdots, N, \tag{15}$$

where $T = (A^T A)_{N \times N}$. The nonzero eigenvectors λ_k ($1 \leqslant k \leqslant N$) may be chosen to be orthonormal, and the POD modes are given by $\phi_k = AV_k / \sqrt{\lambda_k}$, ($1 \leqslant k \leqslant N$).

The truncated reconstruction of analysis variable in the four- dimensional space X_a is given by

$$X_a = \overline{X} + \sum_{j=1}^{P} \alpha_j \phi_j, \tag{16}$$

where P (the number of the POD modes) is defined as follows:

$$P = \arg\min \left\{ I(p) = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{N} \lambda_i} : I(p) \geqslant \gamma \right\}, \quad 0 < \gamma < 1.$$

It is well known that the expansion (eq. (16)) is optimal[16,17]. Particularly, among all linear combinations (including the linear combination based on the SVD basis vectors), the POD is the most efficient, in the sense that for a given number of modes P, the POD decomposition will capture the most possible kinetic energy. The forward solution is approximately expressed by a truncated expansion of the POD basis vectors in the four-dimensional space. Substituting eq. (16) into eq. (11), the control variable becomes $\alpha = (\alpha_1 \cdots \alpha_P)^T$ instead of x_0, so the control variable is expressed explicitly in the cost function and the computation of the gradient is simplified greatly .

3 Numerical experiments

In this section, the applicability of this newly proposed method is evaluated by several assimilation experiments with a simple one-dimensional soil water equation model used in the NCAR Community Land Model (CLM)[16]. In addition, comparison is made between the POD-E4DVAR method (POD1-E4DVAR: the perturbed 4D fields are produced only at the observational times (K time steps); POD2-E4DVAR: the perturbed 4D fields are produced over the whole assimilation time window (S time steps)), the SVD-E4DVAR method and the EnKF method.

3.1 Setup of experiments

The conservation of water mass (θ) for one-dimensional vertical water flow in a soil column in the CLM is expressed as

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z} - E - R_{fm}, \tag{17}$$

where q is the vertical soil water flux, E is the evapotranspiration rate, and R_{fm} is the melting (negative) or freezing (positive) rate, (for simplicity, E, R_{fm} are taken as zero in the experiments), and z is the depth from the soil surface. Both q and z are positive downward.

The soil water flux q is described by Darcy's law[17]:

$$q = -k \frac{\partial(\varphi + z)}{\partial z}, \tag{18}$$

where $k = k_s (\theta / \theta_s)^{2b+3}$ is the hydraulic conductivity, and $\varphi = \varphi_s (\theta / \theta_s)^{-b}$ is the soil matric potential, and k_s, φ_s, θ_s and b are real constants (Table 1). The CLM computes soil water content in the 10 soil layers through (eq. (17) and eq. (18)) (see ref. [16] for detail). The upper boundary condition is

$$q_0(t) = -k \left. \frac{\partial(\kappa + z)}{\partial z} \right|_{z=0},$$

where $q_0(t)$ is the water flux of the land surface (infiltration), and the lower boundary condition is q_l=0. The time step Δt is 1800 s.

For experiments, the soil water equation model with the "perfect" infiltration series (Figure 1) acts as the

Table 1 The parameters used in the soil water equation model

θ_s	k_s	b	φ_s
0.46	2.07263×10^{-6} m/s	8.634	−3.6779 m

perfect model, and it is forced by the "imperfect" infiltration (Figure 1) which acts as the imperfect model. The "perfect" (or "true") state is produced by integrating the perfect model with the "perfect" initial soil moisture (Figure 2) for 1092 time steps. The observations are made through adding 3% random error perturbation to the "perfect" state series. Figure 3 shows that time series of the skin volumetric soil moisture is simulated by (1) the "perfect" model with the "perfect" initial soil moisture (P with P), (2) the "perfect" model with the "imperfect" initial field (P with Im) and (3) the "imperfect" model with the "imperfect" soil moisture (Im with Im), respectively. The forecast state is produced by integrating the perfect (imperfect) model with the "imperfect" initial soil moisture: The former means that the forecast error comes only from the noise of the initial filed, and the latter indicates that the forecast error comes not only from the noise of the initial field but also from the uncertainty of the forecast model. The length of an assimilation time window in our experiments is 7. Two group experiments are conducted in our research (Table 2). The perfect model with the "imperfect" initial field is used in Group 1 and the imperfect model with the "imperfect" initial field is applied in Group 2. Three assimilating observation frequencies (7 times/window, 4 times/window and 3 times/window) are adopted in each group experiment. The ensemble size of all the three assimilation methods (POD-E4DVAR, SVD-E4DVAR and EnKF) is 60.

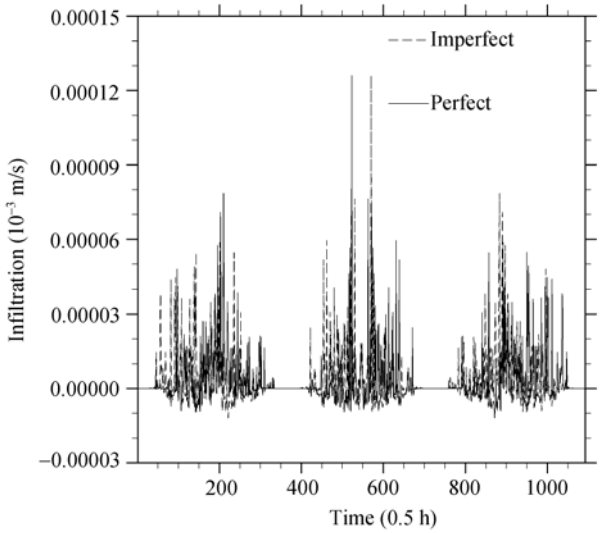

Figure 1 The perfect and imperfect infiltration used in the assimilation experiments.

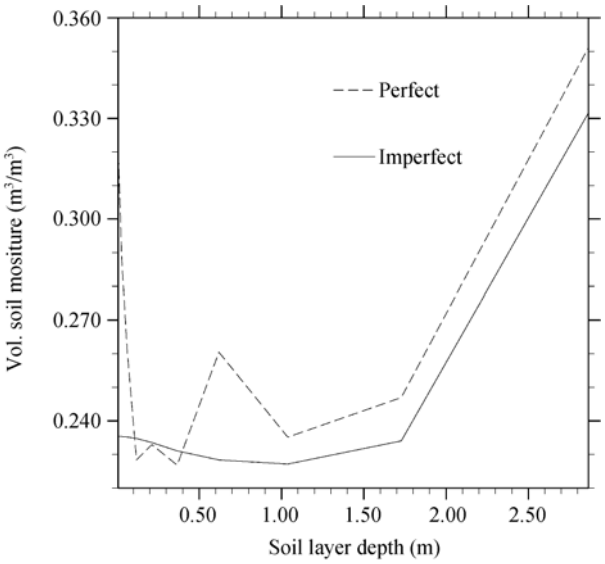

Figure 2 The perfect and imperfect initial soil moisture used in the assimilation experiments.

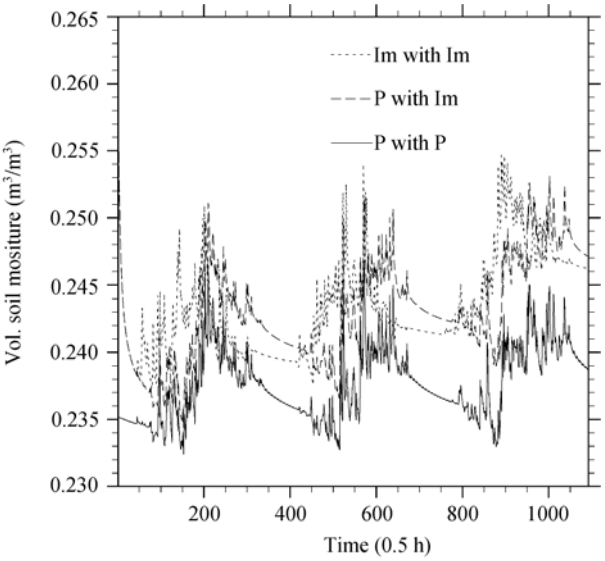

Figure 3 Time series of the simulated skin soil moisture (volumetric soil moisture, $\theta(t)$).

Table 2 Experiments design

Experiments No.	Initial errors	Observation frequency (times/window)	Model errors
1	yes	7	no
2	yes	4	no
3	yes	3	no
4	yes	7	yes
5	yes	4	yes
6	yes	3	yes

3.2 Experimental results

To evaluate the performance of the three algorithms (POD-E4DVAR, SVD-E4DVAR and EnKF), the relative

error is defined as follows:

$$E_n(\theta) = \frac{\sum_{i=1}^{S} (\theta_{n,i}^a - \theta_{n,i}^t)^2}{\sum_{i=1}^{S} (\theta_{n,i}^f - \theta_{n,i}^t)^2}, \quad (19)$$

where the index n denotes the number of the assimilation cycle, S is the length of an assimilation window, f and a denote the forecast state (without assimilation) and the analysis state, respectively, and t represents the "true" ("perfect") state. Figures 4 and 5 show that the POD/SVD-E4DVAR methods perform much better than the EnKF method in the two group assimilation experiments. Their relative errors for soil moisture are very small (most less than 0.01) in the case that the forecast model is perfect (The forecast error comes only from the noise of the initial filed (Figure 4)), however the relative errors of the EnKF method are about 0.03 or so. The two explicit four-dimensional variational methods perform almost the same, but we can still find that the relative errors of the POD-based method are appreciably less than that of the SVD-based method, probably because of the POD basis optimality in the ensemble space. When the forecast model is not perfect, its forecast error comes from both the noise of the initial filed and the uncertainty of the model itself. The relative errors of three methods are becoming bigger as a result (Figure 5). Most of the POD-E4DAVR relative errors are still controlled in the scope between 0.0 and 0.03, only a few are beyond 0.03 but do not exceed 0.06 yet except for those in some windows. Most of the relative errors of the SVD-E4VAR are also less than 0.06, but in several assimilation windows (such as windows 30—40, and windows 15—25), its relative errors of the analyzed soil moisture exceed 0.09, with some being up to 0.15. We

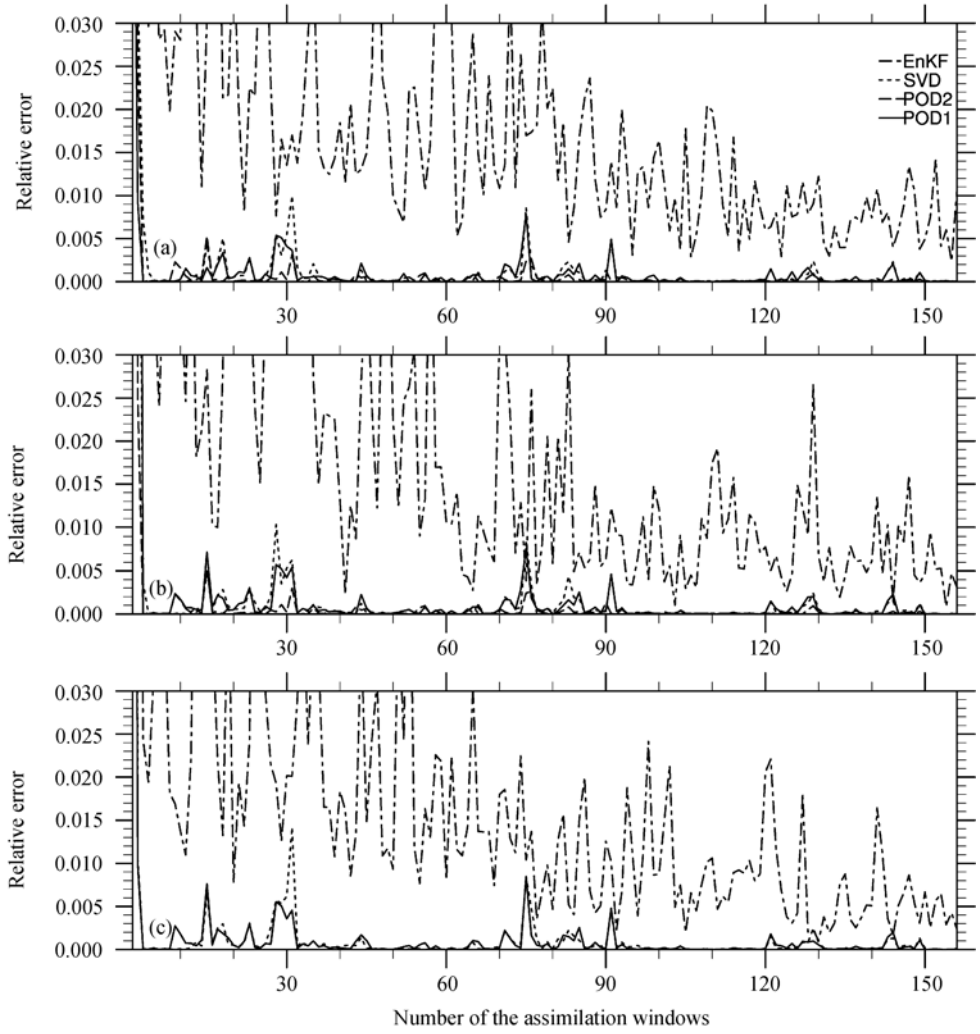

Figure 4 Relative error for analyzed soil moisture in the assimilation experiments by the perfect model with the "imperfect" initial filed. (a), (b) and (c) are the three time, four time and seven time observations in an assimilation window, respectively.

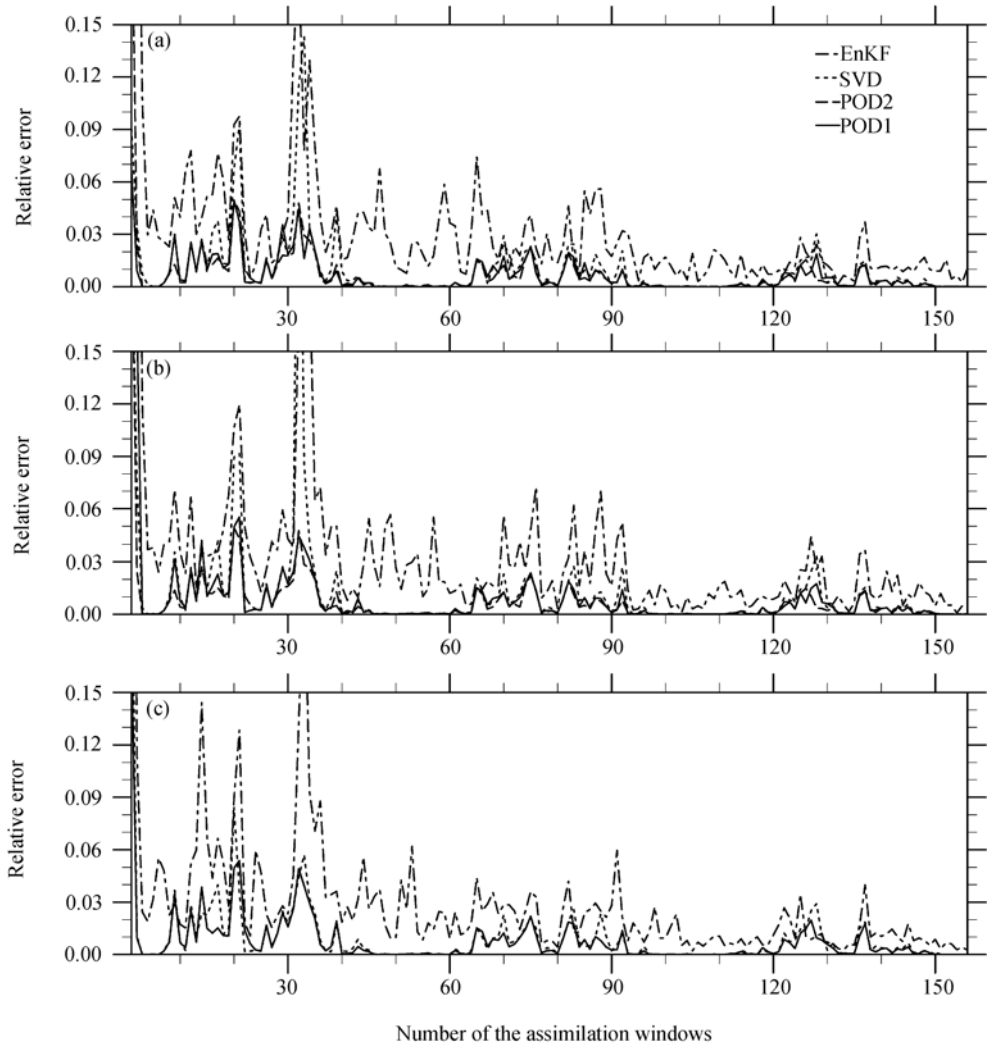

Figure 5 Relative error for analyzed soil moisture in the assimilation experiments by the imperfect model with the "imperfect" initial filed. (a), (b) and (c) are the three time, four time and seven time observations in an assimilation window, respectively.

can also find that the observation frequency has more impacts on the SVD-E4DVAR method compared with the POD-E4DVAR method. Most of the corresponding relative errors of the EnKF method are bigger than 0.03, with some even bigger than 0.15. On the computation efficiency, the ratio of the computation costs for the three methods is about 1∶1∶30 in the two group assimilation experiments.

We can also find that the two different sampling POD-E4DVAR (POD1-E4DVAR and POD2-EDVAR) methods perform almost the same in the two group experiments. Of course, the 4D vector sampled over the whole assimilation time window (S time steps) can sufficiently represent the spatio-temporal evolution of the forecast state, which makes the POD2-E4DAVR method outperform the POD1-E4DVAR to some extent (Figures 4 and 5). More investigations are still needed to explore the impacts of the different sampling on the assimilation results.

4 Summary

We propose an ensemble-based explicit four-dimensional variational data assimilation method in this paper. Several numerical experiments performed with a simple one-dimensional soil water model show that the newly proposed method can perform better than the usual EnKF method not only on the assimilation precision but also on the computation cost, and it is also better than another explicit four-dimensional variational assimilation method, especially when the forecast model is not perfect and the forecast error comes from both the noise

of the initial filed and the uncertainty of the forecast model, which shows that this new method has great potential in data assimilation. Several issues, such as the impacts of the sample size, the initial perturbation fields on this data assimilation method and its actual performance in real numerical forecast models, are under investigation.

The three anonymous reviewers are greatly appreciated for their constructive comments on this paper.

1 Lewis L, Deber J. The use of adjoint equation to solve a variational adjustment problem with advective constraint. Tellus Ser A-Dyn Meteorol Oceanol, 1985, 37: 309－322

2 Le Dimet F X, Talagrand O. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. Tellus Ser A-Dyn Meteorol Oceanol, 1986, 38: 97－110

3 Xu Q. Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basis issues and heuristic examples. J Atmos Sci, 1996, 53(8): 1123－1142

4 Mu M, Wang J. A method for adjoint variational data assimilation with physical "on-off" processes. J Atmos Sci, 2003, 60: 2010－2018

5 Kalnay E, Park S K, Pu Z X, et al. Application of the quasi-inverse method to data assimilation. Mon Weather Rev, 2000,128: 864－875

6 Wang B, Zhao Y. A new data assimilation approach (in Chinese). Acta Meteorol Sin, 2005, 63: 694－700

7 Courtier P, Thepaut J N, Hollingsworth. A strategy for operational implementation of 4D-Var using an incremental approach. Q J R Meteorol Soc, 1994, 120: 1367－1388

8 Evensen G. Sequential data assimilation with a non-linear geostrophic model using Monte Carlo methods to forecast error statistics. J Geophys Res, 1994, 99(C5): 10143－10162

9 Hamill T M, Snyder C. A hybrid ensemble Kalman Filter-3D Variational analysis scheme. Mon Weather Rev, 2000, 128: 2905－2919

10 Lorenc A C. Modelling of error covariances by 4D-Var Variational data assimilation. Q J R Meteorol Soc, 2003, 129: 3167－3182

11 Qiu C J, Chou J F. Four-dimensional data assimilation method based on SVD: Theoretical aspect. Theor Appl Climatol, 2006, 83: 51－57

12 Qiu C J, Zhang L, Shao A M. An explicit four-dimensional variational data assimilation method. Sci China Ser D-Earth Sci, 2007, 50(8): 1232－1240

13 Ly H V, Tran H T. Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor. Q Appl Math, 2002, 60(3): 631－656

14 Ly H V, Tran H T. Modeling and control of physical processes using proper orthogonal decomposition. Math Comput Model, 2001, 33: 223－236

15 Cao Y H, Zhu J, Navon I M. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. Int J Numer Methods Fluids. 2007, 53(10): 1571－1583

16 Oleson K W, Dai Y J, Bonan G, et al. Technical Description of the Community Land Model (CLM). Technical Report, Terrestrial Sciences Section, Climate and Global Dynamics Divison, National Center for Atmospheric Research, 2004. 186

17 Darcy H. Les Fontaines Publiques de la Ville de Dijon, Dalmont, Paris. 1856